Reputation, Sentiment, Time Series and Prediction

Peter Mitic[1] [a]
[1] *Department of Computer Science, UCL, London, England*
p.mitic@ucl.ac.uk

Keywords: Reputation, Sentiment, Time Series, Prediction, Auto-correlation, ARIMA, Cholesky, Copula, Normal Mixture distribution, Goodness-of-Fit, TNA Test

Abstract: A formal formulation for reputation is presented as a time series of daily sentiment assessments. Projections of reputation time series are made using three methods that replicate the distributional and auto-correlation properties of the data: *ARIMA*, a *Copula* fit, and *Cholesky decomposition*. Each projection is tested for goodness-of-fit with respect to observed data using a bespoke auto-correlation test. Numerical results show that *Cholesky decomposition* provides optimal goodness-of-fit success, but overestimates the projection volatility. Expressing reputation as a time series and deriving predictions from them has significant advantages in corporate risk control and decision making.

1 INTRODUCTION

The title gives the flavour of this study in the order of its words. *Reputation* is derived from *Sentiment* as a *Time Series* which is used for *Prediction*. The sequence starts with wanting to know about product and company performance.

There has been a huge increase since year 2000 in interest in and progress with the analysis of peoples views on products and services, fuelled by technological advances (Liu, 2015). Increased development of the internet, the rise of on-line media (both social and 'traditional' - newspapers and broadcasting), has made it possible for consumers to formulate their own views on products and services in advance of making a decision on purchase or use. Fundamental to such decision making is the concept of *reputation*. Informally, *reputation* is "the opinion that people in general have about someone or something, or how much respect or admiration someone or something receives, based on past behaviour or character" (Cambridge, 2023). The same reference gives an informal definition for *sentiment*: "a thought, opinion, or idea based on a feeling about a situation, or a way of thinking about something". We will give formal definitions for both in Section 3.4. The informal definitions are, however, remarkably close to the ideas we wish to convey formally. We will distinguish between *reputation*, *sentiment* and *opinion*, and link them in a formal way.

[a] https://orcid.org/0000-0002-9845-4435

The purpose of this paper is to predict how the reputation of a corporate body may develop in the future. Reputation is expressed as a time series, to which time series methods apply naturally. However, reputation time series express distinct characteristics which makes it difficult to apply standard methods without some degree of conditioning. In particular, they are highly auto-correlated, are subject to rapid reversals in profile (they look 'spiky'), exhibit high volatility, and are not always stationary. Others have sparse, or almost no sentiment expression. Reputation time series are built using expressions of sentiment, so an initial discussion sets out formal definitions for sentiment and reputation.

We consider predicted reputation because there is some evidence that "reputation means money" (Cole, 2012), (Weber-Shandwick, 2020). On that basis, reputation was quantified in terms of share price in (Mitic, 2024). Specifically, impaired reputation can lead to effects such as loss of profit, share price reduction, and reduced ability to attract and retain staff. These, and similar reports are not quantified in a transparent way, but nevertheless convey the message that a positive reputation matters. Consequently, predicting future reputation also matters.

1.1 Reputation Time Series example

In this section we show an example of a reputation time series. Figure 1 shows Toyota's reputation for the first 6 months of 2023, and a simple exponential smoothed version of it. The plot shows time, measured in days, on the horizontal axis, and numerical expressions of sentiment on the vertical axis on a scale -100 to +100. The trace shows that during that period, Toyota's reputation was entirely negative. To see why, would require detailed analysis of each sentiment value, but a major contributor was the change of Toyota's leadership. That news was widely reported in the financial press at the time. A typical example, which is part of a longer article, appeared in a Reuters report on 26 January 2023. [1]

Reactions to Akio Toyoda stepping down as Toyota CEO. TOKYO, Jan 26 (Reuters) - Toyota Motor Corp (7203.T) said on Thursday that Akio Toyoda will step down as president and chief executive to become chairman from April 1, ...

Figure 1 shows the date 26th January 2023. Interestingly, reputation improved after that date, perhaps indicating that the news was received positively, although that rise did not last long. The reputation trace shows typical features: peaks and troughs in a macro-structure, with a micro-structure of much smaller variations. Toyota's autocorrelation structure is shown in Figure 2. The plot shows typical features of significant autocorrelations at high lags, with some positive and negative regions.

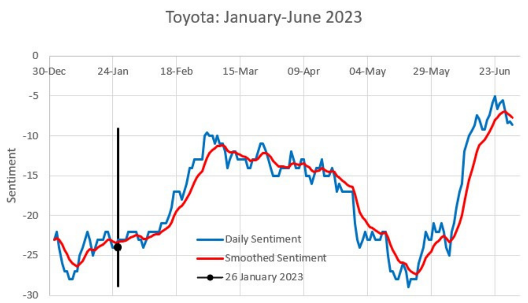

Figure 1: Toyota reputation January-June 2023. Data source: *Penta Group*.

2 RELATED WORK

Reputation time series as described in Section 3.4 are a natural extension of much earlier work on opinion, sourced by survey. The first prominent example of a survey was a correct prediction of

[1]https://www.reuters.com/business/autos-transportation/toyota-leader-akio-toyoda-step-down-president-chief-executive-2023-01-26/

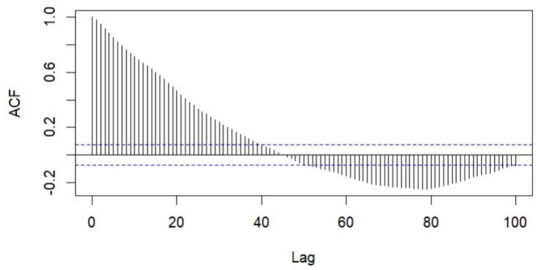

Figure 2: Toyota autocorrelation: 100 lags

the 1936 US Presidential election by the Gallup Company (founded in 1935) (Gallup and Rae, 1968) , although there is a record of an opinion poll from 1824 in the *Harrisburg Pennsylvanian* (https://www.referenceforbusiness.com/history2/84/The-Gallup-Organization.html). Gallup took the view that an opinion poll was simply a reflection of public opinion. There is an interesting counter opinion due to Lippman (Lippman, 1922) that opinion polls manipulate public opinion. The point is discussed in (Jacobs and Shapiro, 1995). In 1995 the internet was relatively young, but since then the means to manipulate opinion have emerged in the form of blogs, social media platforms (such as Facebook, WhatsApp or Twitter ("X")), and product reviews on websites such as Amazon, Google and others. Problems of sample bias are discussed in (Durant, 1954). They centre on location, respondents, and questionnaire design, with additional factors related to administration, cost, and whether or not the results represent a general population.

There is evidence of bias in contemporary opinion procurement. The term 'negative bias' was introduced by (Rozin and Royzman, 2001), and clear numerical illustrations are presented in (Zendesk, 2013).

Early research on sentiment and opinion is summarised in, for example, (Das and Chen, 2007). The emphasis was then on sentiment extraction using lexicons (word lists with tags showing related words or parts of speech), lexical grammar (rules for manipulating a lexicon), and classifiers (Bayes, Voting, Naive, Vector-Distance, Discriminant). Those methods still form the basis of 'traditional' sentiment analysis, and act as a benchmark for assessing later approaches using artificial intelligence.

Prediction of reputation has, to date, been somewhat neglected, largely because of a lack of appropriate data. The problem was tackled, albeit in a difference sense of the word 'prediction' by (Loke and Kachaniuk, 2020), using a bi-directional LSTM.

That study used manual labelling of thousands of product reviews, evaluated on a 3-point scale, aimed at predicting individual review results. Our study aims to produce a forward projection in time, and uses much simpler prediction methods. *Penta Group*, as part of their reputation intelligence website [2] available to subscribers, shows a basic forward (in time) prediction based on exponential smoothing.

2.1 Alternative sources of *Reputation Intelligence*

In this section we summarise the state of online *Reputation Intelligence*. The term *Reputation Intelligence* has been used in the past ten years to refer more general aspects of sentiment and reputation. A reputation time series is one of them. Others include, for example, analysis of sentiment sources (e.g. traditional/social media), analysis of regional sentiment, comparison with peers, and *Environmental, Social and Governance* (*ESG*) issues.

Artiwise, produced by Istanbul Technical University (https://www.artiwise.com) provides (to subscribers) bespoke sentiment analysis services, and calculates a short-term sentiment score based on a limited number of sources to order. The Californian company *Reputation* (https://reputation.com/) provides the same type of service, and makes a *Reputation Experience Management - RXM* platform available to customers. In New York, *Social360* (https://www.social360monitoring.com) provides bespoke analysis of online comments, and tracks influential reporting agents. They specialise in social media checking. *Social360* has recently be acquired by (SignalAI, 2024).

An earlier, and different, approach is typified by the *RepTrak Pulse* metric (Fombrun et al., 2015), published twice yearly by the *Reputation Institute* (https://www.reptrak.com/). *RepTrak* is an updated version of its predecessor, the *Reputation Quotient* (Fombrun et al., 2000). Both are multi-factor snapshot assessments of reputation. *RepTrak Pulse* exports "Good overall reputation", "Good feeling about", "Trust", and "Admire and Respect", all condensed comments amassed throughout the six months prior to publication. In contrast, the *Net Promoter Score - NPS* from Bain and Co. (https://www.bain.com/) is very simple, but limited (Reichheld, 2003). It is based on one question: *On a scale of 0-10, how likely are you to recommend this company to a friend or colleague?*. The *NPS* is then the difference between the percentage of 9-10 (promoter) scores and the percentage of 0-6 scores (detractors). Scores 7 and 8 are regarded as "passive". The imbalance appears to induce negative bias. The study by (Loke and Reitter, 2021) used the same type of multi-factor analysis to measuring reputation, using online review data and 'aspect' extraction by detecting negative sentiment and positive sentiment keywords.

A third strand of reputation measurement is demonstrated by the *Edelman Trust Barometer* (https://www.edelman.com). *Trust* is somewhat distinct from sentiment or reputation, and implies a degree of safety and/or reliable (Cambridge, 2023). The Edelman method of data sourcing is, again, by survey, is targeted at employees, and produces generalised qualitative reports, with some associated data. An example is (Russell, 2023). The argument in (Renner, 2011) is that risk can be minimised by increasing trust, and that corporate reputation is the vehicle to build trust.

A few other attempts to measure reputation have emerged. (Janson, 2014) recommends spending at least 10% of a corporate budget on reputational analysis and sampling, but is otherwise non-specific on methodology. (Carreras et al., 2013) suggests a ranking method in which company executives rank themselves and peers on a multi-factor basis, and produce a score based on those ranks. Overall, these and similar alternatives rely on the subjective opinion of selected individuals.

3 METHODS

We first review data stationarity and a methodology for measuring the appropriateness of a projected time series. Three projection methods are then discussed: *ARIMA*, *Copula* and *Cholesky*.

3.1 Stationarity Test

We cannot assume that distributional properties of reputational time series do not change over time. Therefore we stress that the analyses that follow need to be reviewed periodically. A particular concern is the way changes in the data structure over time affect the effectiveness of a reputation projection. The problem is addressed in Section 3.6. The *Augmented*

[2] https://pentagroup.co/

Dickie-Fuller (*ADF*) test for stationarity is used to test for consistency of mean, variance and autocorrelation structure for the observed data.

The *ADF* test showed that approximately 60% of reputation time series tested were stationary, and 40% were not. That result is more significant for short projections, where auto-correlations may be very different to the observed data. Longer projections are more stable with respect to projection length. In all cases, the general approach is to test whether or not the projection *perturbs* the auto-correlations structure of the observed data unduly.

3.2 Goodness-of-Fit Test

There are indications from histograms of reputation data that Normal distributions might be appropriate for modelling distributions. The established goodness-of-fit for normality is the Shapiro-Wilk test (Shapiro and Wilk, 1965). That test rejected the null hypothesis of normality in all cases that we encountered. The reason appears to be that the Shapiro-Wilk test is weak with respect to distributions with longer tails (Royston, 1992). Isolated outliers can also cause the Shapiro-Wilk test to fail. Many reputation time series have both long tails and/or outliers. As an alternative, we have used the *TNA* test (Mitic, 2015), which is a generalisation of a *Q-Q* plot. The *TNA* test is less powerful than the Shapiro-Wilk test, is insensitive to outliers and long data tails, and is not restricted by data set size. The *TNA* test indicated that the Normal distribution is often not the *best* fit for reputation data, and the null hypothesis was rejected in approximately 8% of cases. The Normal Mixture distribution (Section 3.7) is a better fit in most cases, and is a better model for bimodal distributions and for distributions with long tails. Therefore, we proceed with Normal Mixture distributions, which also subsume Normal distributions.

3.3 Data

Data for this study are sourced from *Penta Group* (https://pentagroup.co). *Penta* can, uniquely, provide time series of daily sentiment scores [3] (i.e. a *reputation* profile) for most organisations that are listed on major world stock exchanges, and a large number of others that are unlisted. We have concentrated on 125 corporate organisations that represent the principal world industrial and service sectors: energy, manufacturing, travel, education, financial, media, mining, food production and retail. The data range was two years: from July 2021 to June 2023. Each recorded data series comprises 730 daily sentiment readings on a scale from -100 (the worst possible) to +100 (the best possible). Zero (or very near to zero) represents neutral sentiment.

[3] Data are available to subscribers only

3.4 Definitions

Following a slightly modified definition from (Liu, 2015) *Opinion* is defined in terms of a numerical value, representing the thought, idea or view that is held or expressed (as defined in, for example (Cambridge, 2023)), Liu's view is slightly different. He represents *Opinion* as an ordered pair: a *polarity* value (+1, 0, or -1) for positive, neutral or negative view respectively, with a positive number representing its intensity. We assume that the view is quantifiable numerically. In principle, the range of permitted values does not matter, but in practice, a meaningful symmetric scale that presents a positive score for positive sentiment and a negative score for negative sentiment (between real numbers -r and +r) is useful. Opinion also incorporates the *holder,* *h* of the view, its *target,* *T*, and a date/time stamp *t*. In addition, Liu labels the opinion value with a *type* flag, used to designate it as either *rational or emotional*. We prefer a wider range *type*, aimed at assessing the influence or importance of the holder, and denote it by u. The definition of *Opinion*, Equation1, incorporates all of those components. The numerical *view* is denoted by x, and the values of h and T are best identified with reference to a set of unique identifiers W (positive integers or guids).

Definition: *Opinion*

$$O_t(x,h,T,u) = \mathcal{F}(x|h,T,u);\ \ x \in [-r,r]; \\ t \in \mathbb{Z}^+;\ \ u \in (0,1);\ \ h,T \in W \qquad (1)$$

At this point, it is acceptable, in principle, to use the terms *Opinion* and *Sentiment* interchangeably. However, to facilitate the ensuing discussion of *Reputation*, it is useful to define *Sentiment* as a function Ψ of a set of holders $H = \{h_1, h_2, ...\} \subset W$, each having expressed corresponding numeric views $X = \{x_1, x_2, ...\}$, and each with having corresponding numeric influences $U = \{u_1, u_2, ...\}$, referred to a single target T on a single day t. The function Ψ acts on the elements of X to produce a single real number in the same range as the x_i, namely $[-r,r]$.

Definition: *Sentiment*

$$S_t(X,H,T,U) = \Psi\Big(\{O_t(x_i,h_i,T,u_i)\}\Big); \\ h_i \in H;\ \ x_i \in X;\ \ u_i \in U \qquad (2)$$

S_t is a single real number representing a set of sentiments at time t. In practice, it is more useful to use a "day" stamp rather than a "time" stamp, so that S_t refers to the sentiment on "day t". It is then easy to define reputation as a sequence of such numbers as t varies. Equation 3 shows a date range from times t_1 to date t_2. No assumption are made about periods within that range that have no sentiment data.

Definition: Reputation

$$y_t(T) = \{S_t(T)\}; \quad t_1 \leq t \leq t_2 \tag{3}$$

The definition of reputation in Equation 3 is hinted at in, for example, (Loke and Vergeer, 2022), in which phrases such as "collective view" and "built over time" are used. Loke and Vergeer make the point that attempts to quantify corporate reputation are limited. We believe that we have made a significant advance in that respect. (Loke and Kisoen, 2022) argue that, essentially, reputation is a summary of internal and external perceptions of an organisation. We argue that reputation should extend much further. Specifically, broadcasting, news reports and trade presentations represent a further strand that provides a more objective view. Reports from the 'popular' press are often not objective. Nevertheless, they are there, and present an opinion. The same applies to reports that contain mistakes or lies.

3.5 Initial Data Preparation

The common basis of the *Copula* and *Cholesky* auto-correlation models used in this analysis is an auto-correlation matrix, A, which contains sequences of lagged data. If a time series of length n has L lags, A takes the form given in Equation 4. The S-values are the daily sentiments in Equation 2.

$$A = \begin{pmatrix} S_1 & S_2 & \dots & S_n \\ S_2 & S_3 & \dots & S_{n+1} \\ \dots & \dots & \dots & \dots \\ S_n & S_{L+1} & \dots & S_{n+L-1} \end{pmatrix} \tag{4}$$

Following construction of A, we calculate a rank correlation matrix (Spearman or Kendall) rather than Pearson's product moment variety, since the latter assumes a linear relation between co-variates.

3.6 Auto-correlation success criterion

Comparing the autocorrelations of any two subsets of the data cannot be expected to give similar correlation structures. Therefore we adopt an alternative strategy, which is to test whether or not a projected simulation does not perturb the correlation structure of the observed data. The test applied is to calculate the auto-correlation function (*ACF*) of the observed data and compare it the observed data augmented by the simulated data. With a fixed number of lags L (typically between 50 and 100), the two applications of an *ACF* function yields parallel sequences of auto-correlation components c_i^O and c_i^{OS} (equation 5.

$$\begin{cases} \{c_1^O, c_2^O, \dots, c_L^O\} & \text{Observed} \\ \{c_1^{OS}, c_2^{OS}, \dots, c_L^{OS}\} & \text{Observed + Simulated} \end{cases} \tag{5}$$

Since the two sequences are paired, a two sample t-test can be used to determine significance of the augmentation of the observed data by the simulation. If the means of the sequences in Equation 5 are denoted by $\mu(c^O)$ and $\mu(c^{OS})$ respectively, the null and alternative hypotheses are $\mu(c^O) = \mu(c^{OS})$ and $\mu(c^O) \neq \mu(c^{OS})$ respectively, and significance is tested at 5% and 1%.

3.7 Normal Mixture distribution

In this section we define a distribution that fits the reputation time series in this study. Although a Normal distribution is a good fit in most cases, a *Normal Mixture* distribution is usually better. We call it *NMix* for short.

NMix is a weighted sum of two Normal distributions, with parameters $\{\mu_1, \sigma_1, \mu_2, \sigma_2, b\}$. Its density function is $\phi_M(t)$ and the corresponding distribution function is denoted by $\Phi_M(t)$ (on day t). The inverse distribution (*quantile*) function takes a probability p as parameter, and is denoted by $\Phi_M^{-1}(p)$. The *quantile* function is needed for the *Copula* algorithm in Section 3.8. In the following equations, $x \in [-r, r]$, $p \in (0,1)$. The parameter ranges are $\mu_1, \mu_2 \in (-r, r)$, $\sigma_1 > 0, \sigma_2 > 0$, and $b \in [0,1]$.

$$\phi_M(t, \mu_1, \sigma_1, \mu_2, \sigma_2, b) = \\ b\phi(t, \mu_1, \sigma_1) + (1-b)\phi(t, \mu_2, \sigma_2) \tag{6}$$

$$\Phi_M(b, \mu_1, \sigma_1, \mu_2, \sigma_2, b) = \\ \int_{-r}^{t} \phi_M(z, \mu_1, \sigma_1, \mu_2, \sigma_2, b)dz \tag{7}$$

$$\Phi_M^{-1}(p, \mu_1, \sigma_1, \mu_2, \sigma_2, b) = \\ t \mid \Phi_M(t, \mu_1, \sigma_1, \mu_2, \sigma_2, b) = p \tag{8}$$

As an example, we return to the Toyota data presented in Figure 1, but plot a density histogram instead. An *NMix* distribution has been fitted and overlaid. The bimodal nature of the data is clear from the histogram, and the fitted *NMix* distribution echoes that. In this case, a Normal distribution is a poorer fit, but nevertheless satisfies the *TNA* goodness-of-fit test described in Section 3.2.

Specifically, the *NMix* parameters were $\mu_1 = -23.16, \sigma_1 = 4.13, \mu_2 = -10.72, \sigma_2 = 4.37, b = 0.56$, and the p-value for the *NMix* fit was 0.011. The Normal distribution parameters were $\mu = -17.69, \sigma = 7.49$, with p-value 0.025.

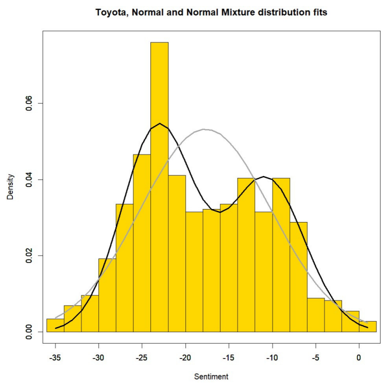

Figure 3: Toyota Normal Mixture and Normal distribution fits (black and grey respectively). Data source: *Penta Group*.

3.8 Copula Model

In Algorithm 1, the symbols used are: Reputation time series R, Lag L, required simulation length n. The internal variables are the auto-correlation matrix A, a multi-variate Normal copula C, uniformly distributed marginal distributions of C $G_i, i = 1...n$, Normal Mixture-distributed marginals $Y_i, i = 1...n$, and their corresponding auto-correlation p-values $\alpha_i, i = 1...n$. The process uses a procedure *FIT(D)* to fit a distribution D (in this case D is a Normal Mixture), a function *MVN* (from the *R* package *mvtnorm*) to initialise a multi-variate normal copula, a function *AC* to test the marginal effect of the simulated data on the autocorrelation of the input data, and a *Loess* smoothing function *LO*.

Data: *R, L, n*
Result: Simulation of length *n*
Calculate best fit parameters $p = FIT(R(D))$;
Derive auto-correlation matrix $A(R)$;
Initialise copula: $C = MVN(A)$;
Generate uniform marginals $G = \Phi(C)$;
for *i in 1:L* **do**
  $Y_i = LO(\Phi_M^{-1}(G_i, p))$ (*NMix* marginals) ;
  Test auto-correlation: $\alpha_i = AC(R, Y_i)$;
end
Select optimal auto-correlation: α_{opt}, Y_{opt};
Return $\{Y_{opt}, \alpha\}$;

Algorithm 1: Copula simulation

3.9 ARIMA Model

The ARIMA modelling incorporates both auto-regressive (AR) and moving average (MA) components, although we suspect that the AR components are much more important. With AR, MA and differencing parameters p, q and d respectively, plus a constant μ and error term ε_t, the ARIMA model used is given in 9. The values of p, q and d are determined using the *auto-ARIMA* method of Hyndman and Khandakar (Hyndman and Khandakar, 2008). Parameter d is determined by carrying out successive unit-root tests (D. Kwiatkowski and Shin, 1992) until a stationary series results. There is a correction for seasonal data, although we would not expect reputation data to exhibit any degree of seasonality since reputation is event-driven. Parameters p and q are determined by a stepwise algorithm in which target values of p and q are tested against for minimal AIC.

$$x_t = \mu + \lambda \sum_{i=1}^{p} p_i x_{t-i} + \sum_{i=1}^{q} q_i \varepsilon_{t-i} + \varepsilon_t \qquad (9)$$

Having determined the parameter values, the ARIMA fit is done using maximum likelihood via a state-space representation of the ARIMA process. The innovations and their variances are found by a Kalman filter (Gardner et al., 1980). In the ARIMA algorithm below, the *auto-ARIMA* function used to determine the ARIMA parameters (Hyndman and Khandakar, 2008) is denoted by $FC(R)$, and the simulation function is denoted by $FSim(R, \ldots)$.

In practice we have never encountered the *White noise* case.

3.10 Cholesky Model

Cholesky decomposition is an established way to derive data that is correlated with a given data set. The autocorrelation matrix, derived from the observed data forms the basis of the *Cholesky* decomposition. As such, the correlation matrix A must be positive definite. That is, it must be symmetric with positive eigenvalues. A proof may be found in, for example, (Golub and van Loan, 1992) (Section 4.2.7). Further details, including points arising from numerical calculations, and supporting literature may be found in (Higham, 1990). Appendix A

Data: *R, L, n*
Result: Simulation with length *n*
Extract ARIMA order $\{p,d,q\} = FC(R)$;
if $(p > 0 \,\&\, d > 0)$ **then**
  ARMA: $Y = FS(R,p,q)$;
end
if $(p > 0 \,\&\, d = 0)$ **then**
  AR: $Y = FS(R,p)$;
end
if $(p = 0 \,\&\, d > 0)$ **then**
  MA: $Y = FS(R,q)$;
end
if $(p = 0 \,\&\, d = 0)$ **then**
  White noise: $Y = FS(R,0,0,0)$;
end
Return(Y)

Algorithm 2: ARIMA simulation

shows how this result applies to auto-correlation matrices. We have found empirically that, in all cases examined, a *Cholesky* decomposition is successful (i.e. all autocorrelation matrices encountered are positive definite). Consequently we have not needed to provide for non-positive definite autocorrelation matrices. There is a work-around for that possibility. (Rebonato and Jaeckel, 2000) describe two methods to cast a non-positive definite matrix into a positive definite state: *hypersphere decomposition* and *spectral decomposition.*

A *Cholesky decomposition* presents problems in the context of autocorrelation. First, the 'base' *Cholesky* result is a matrix that has the same number of columns as the correlation matrix used to derive it. Effectively, in our context where many auto-correlation components are close to 1, each column is an almost carbon copy of the original data. The problem then is to find a reasonable way to derive a single simulation from those columns. To address this problem for an auto-correlation matrix A of dimension $L \times L$, assuming that a simulation of length n is required, L vectors each of length n are generated from a probability distribution D (*NMix* in the case of reputation data). The calculated *Cholesky* matrix is applied to a matrix of the D-distributed vectors, thereby generating L correlated vectors. Each correlated vector is assessed using the autocorrelation test (Section 3.6), and the optimal vector (given by maximum p-value in the auto-correlation t-test) is selected as the simulation.

In Algorithm 2, the symbols used are the same as in Algorithm 1: Reputation time series R, Lag L, required simulation length n. $Chol(A)$ is a function that calculates the Cholesky decomposition of a matrix A. In addition, $G(L,D,n)$ is a function that generates L random samples, each of length n, and each with Normal Mixture distribution D.

Data: *R, L, n*
Result: Simulation, with length *n*
Calculate best fit parameters $p = FIT(R(D))$;
Generate random samples $Z = G(L,D,n)$;
Smooth samples $Z = LO(Z)$;
Derive auto-correlation matrix $A(R)$;
Cholesky decomposition: $C' = Chol(A)$;
Generate correlated samples $Y = XC'$;
for *i in 1:L* **do**
  Test auto-correlation: $\alpha_i = AC(R,Y_i)$;
end
Select optimal autocorrelation
$\alpha_{opt} = \max(\alpha_i(pval))$;
Select optimal sample vector Y_{opt};
Return $\{Y_{opt}, \alpha_{opt}\}$;

Algorithm 3: Cholesky simulation

4 RESULTS

4.1 Prediction Accuracy

The first set of results is a comparison of actual and predicted reputations. The starting point for these results is a partition of the available data into a training set (the first 75%: days 1 to 547) and a test set (the remaining 25%: days 548 to 730). Projections beyond 730 days were not used. Predictions were made using the training data only, and the essential details of the configured models were noted. For the *ARIMA* model, the only necessary component was the *ARIMA fit* object, calculated using the *auto.arima* function in the *R forecast* package. The corresponding *Cholesky* objects were the *Cholesky* decomposition matrix and the fitted Normal Mixture parameters. For the *Copula* model, the *Copula* correlation matrix and the fitted Normal Mixture parameters were needed. Predictions were then made using the test data with the objects derived in the training phase.

Treated in this way, the train/test environments provide a measure of the accuracy of the test prediction compared to the training prediction, via the mean absolute error (*MAE*) for both. To that effect, the proportionate change in *MAE*, $\Delta^{(MAE)}$, was calculated for each target organisation (Equation 10).

$$\Delta^{(MAE)} = \frac{MAE^{(train)} - MAE^{(test)}}{MAE^{(train)}} \quad (10)$$

The distribution of values of $\Delta^{(MAE)}$ then gives an indication of gross deviations of *MAE* between the training and test environments, for every organisation considered . Figure 4 shows a plot of $\Delta^{(MAE)}$ (on the horizontal axis) against quantile (on the vertical axis). The value $\Delta^{(MAE)} = 1$ represents a 100% increase in *MAE* for the test environment relative to the training environment. The corresponding low quantile values shows that in the majority of cases, an order of magnitude difference, which would indicate instability in a model, is absent. Only one value of $\Delta^{(MAE)}$ out of 125 exceeded the nominal order of magnitude limit: 14.19 using the *Cholesky* model. A second instance of the *Cholesky* model had a $\Delta^{(MAE)}$ value of 9.63: just below the limit. The largest $\Delta^{(MAE)}$ values for the *ARIMA* and *Copula* models were 1.17 and 3.82 respectively.

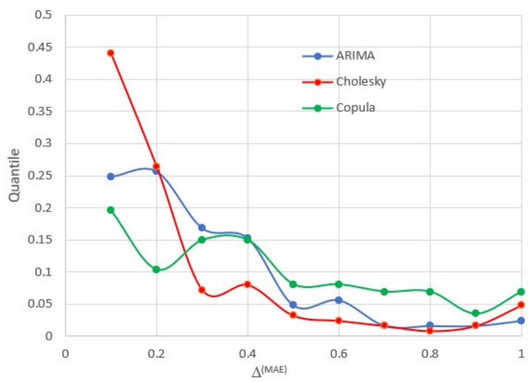

Figure 4: Comparison of *MAE* in training and test environments.

4.2 Auto-correlation results

The principal results of this analysis are presented in this section. The auto-correlation test (Section 3.6) for the three prediction methods (sections 3.8, 3.9 and 3.10) are shown at two significance levels: 5% and 1%. Using five runs in each case, Tables 1, 2 and 3 show the means and standard deviations of the number of organisation that 'passed' the auto-correlation test. A 'pass' is a p-value greater than 0.05 for 5% significance and greater than 0.01 for 1% significance. Column heading 'Simulation length' refers to the percentage augmentation of observed data by simulated data.

The auto-correlation results for the three prediction methods are consistent in that the 'success' rate reduces as the prediction length increases. Of the three, *Cholesky* provides optimal 'success'. There are indications, particularly from the *Cholesky* results, that the 'success' rate levels off for large prediction lengths. It is likely that this effect is due to converging resemblance of the predicted data structure to the observed data structure.

Table 1: Augmentation of observed data by simulated data: Copula method

Simulation	Mean		SD	
length	5%	1%	5%	1%
5%	0.979	1.000	0.004	0.000
10%	0.779	0.906	0.004	0.007
15%	0.672	0.760	0.018	0.009
20%	0.587	0.702	0.012	0.009
25%	0.541	0.603	0.017	0.004
33%	0.448	0.544	0.016	0.006

Table 2: Augmentation of observed data by simulated data: ARIMA method

Simulation	Mean		SD	
length	5%	1%	5%	1%
5%	0.950	0.990	0.009	0.000
10%	0.794	0.896	0.018	0.019
15%	0.623	0.755	0.030	0.013
20%	0.557	0.701	0.036	0.022
25%	0.541	0.663	0.025	0.032
33%	0.475	0.592	0.018	0.033

4.3 Simulation illustrations

This section contains examples of the three simulation modes, to which we add qualitative comments on the characteristics of the simulations. In each case, the observed data is shown in red, the three simulations are shown in green, and the median simulation is shown in blue. The illustrations are for *Microsoft*, which has a typical reputation profile of many large corporates, subject to the general sentiment level (positive, negative or neutral). *Microsoft's* sentiment is mostly positive, and has the characteristic 'jagged' reversing pattern with prolonged upward and downward movements. The two year profile is shown in Figure 5, for which the sentiment mean and standard deviation were 10.76 and 8.10 respectively. The end of the observed data period is marked at day 730. For each simulation type illustrated, the simulation is for 110 days: 15% more than the length of the observed data. Only the latest six months of the observed data are shown, in order to better highlight the profile of each simulation.

Table 3: Augmentation of observed data by simulated data: Cholesky method

Simulation length	Mean 5%	Mean 1%	SD 5%	SD 1%
5%	0.981	1.000	0.007	0.000
10%	0.837	0.933	0.017	0.012
15%	0.722	0.810	0.015	0.019
20%	0.712	0.800	0.032	0.017
25%	0.667	0.739	0.022	0.026
33%	0.662	0.717	0.046	0.040

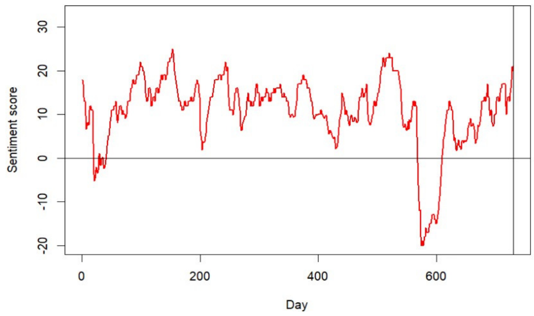

Figure 5: Microsoft: Microsoft observed data

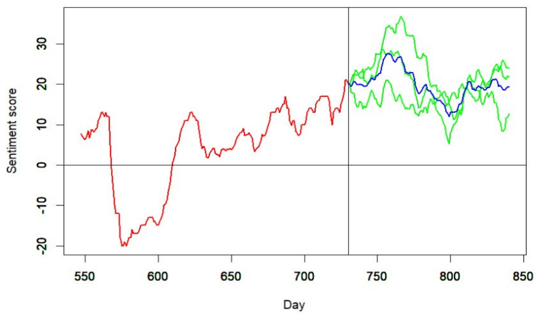

Figure 6: Microsoft: three *ARIMA* simulations.

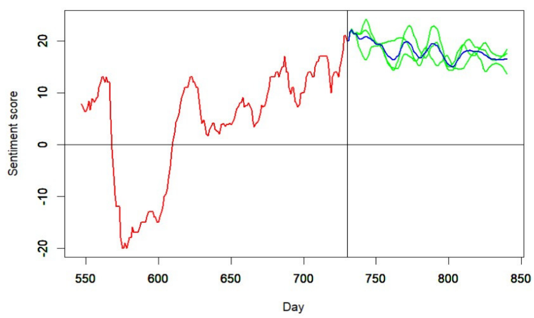

Figure 7: Microsoft: three *Copula* simulations.

5 DISCUSSION

The numerical results in Section 4 invite a choice of which prediction method to use. Table 3 indicates that *Cholesky decomposition* is the optimal method, since

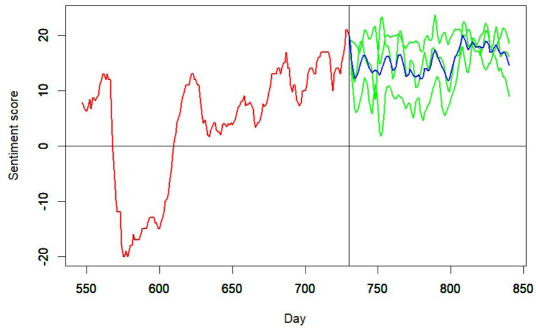

Figure 8: Microsoft: three *Cholesky* simulations.

it provides a higher proportion of auto-correlation 'successes'. The *Cholesky* choice would be clear, were it not for a qualitative examination of the predicted data, and of its microstructure. Figure 8 shows that the day-to-day variation in the prediction is greater than the day-to-day variation for the *ARIMA* and *Copula* methods. Further, the predictions for *ARIMA* and *Copula* appear, subjectively, to be less volatile than the observed data. Examination of similar plots for other organisations confirms that view. We have investigated, albeit briefly, a way to reduce the volatility of the *Cholesky* prediction. A scale factor can be derived as a function of prediction residuals resulting from a piecewise linear fit to the observed data. The same technique can also be used to increase the volatility of the *ARIMA* and *Copula* predictions. Despite some misgivings, we prefer the *Cholesky* method because of its superior conformance to the observed data auto-correlation.

Normally we would not recommend calculating predictions that extend far beyond the bounds of the observed data. A 10-15% extension would be an upper limit. We have extended further in this analysis to illustrate the limitations and capabilities of the overall method. The further extensions have revealed a slow convergence to what appears to be a limiting value for the percentage 'success' metric. Convergence is attributable to convergence of the auto-correlation structures of the observed data and the prediction.

Investigating the predictive nature of reputation is important because it has implications for risk management and corporate decision-making. As part of a generalised risk mitigation process (which nearly always focuses primarily on monetary risk), estimating risk due to reputation can provide insights which balance sheet items cannot. For example, a predicted downturn in reputation could signal future difficulties in selling products or in hiring staff. More generally, tracking reputation following the introduction of new products can indicate whether or not it is worth introducing similar products at a later stage. The question of monetary valuation of reputation was tackled in (Mitic, 2024), in which reputation was valued in terms of share price. Share capitalisations for large corporates are often valued in hundreds of millions of euros, which is not useful for insights into individual products. However, if a company tracks sales with reputation, the possibility of monetising reputation in terms of sales becomes realistic. Thereafter, reputation prediction can be used to predict sales. Further research is required on this topic, but it would probably have to remain in the domain of individual companies who can track their own sales on a daily basis.

5.1 Further work

In addition to monetisation of reputation in terms of product sales (as discussed above), prediction using statistical properties of reputation time series presents possibilities. In particular, neural networks using *Long Short Term Memory* (*LSTM*) is a fruitful area because *LSTM* can mimic the "choppiness" of reputation time series due to its mechanism for selectively retaining or discarding information using *input gates* and *forget gates* respectively. However, this type of neural network is very slow to train. Recent work on this topic in other contexts includes (Yadev and Thakkar, 2024). Adding *attention* layers to a neural network may also be a way forward, provided that the attention can be directed at particular features of the data. A recent study (Wen and Li, 2023) in the contexts of air quality, electricity and share price is encouraging .

ACKNOWLEDGEMENTS

We acknowledge the continuing support and assistance of the staff of *Penta Group*.

REFERENCES

Cambridge (2023). *Cambridge Dictionary online*. CUP https://dictionary.cambridge.org/ dictionary/english/.

Carreras, E., Alloza, A., and Carreras, A. (2013). *Corporate Reputation*. LID Publishing, London, 1st edition.

Cole, S. (2012). The impact of reputation on market value. In *World Economics 13(3), pp. 47-68*. https://www.world-economics-journal.com/Papers/Using-Reputation-to-Grow-Corporate-Value.aspx?ID=563.

D. Kwiatkowski, P.C. Phillips, P. S. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. In *Jnl. Econometrics 54 pp. 159-178*.

Das, S. and Chen, M. (2007). Yahoo! for amazon: Sentiment extraction from small talk on the web. In *Management Science 53(9) pp. 1375-1388*. http://www.icefr.org/icefr2021.html.

Durant, H. (1954). The gallup poll and some of its problems. In *The Incorporated Statistician 5(2) pp. 101–112*. https://www.jstor.org/stable/2986465.

Fombrun, C., Gardberg, N. A., and Sever, J. M. (2000). A multi-stakeholder measure of corporate reputation. In *Journal of Brand Management 7(4) pp. 241–255*. https://link.springer.com/article/10.1057/bm.2000.10.

Fombrun, C., Ponzi, L. J., and Newburry, W. (2015). Stakeholder tracking and analysis: the reptrak system for measuring corporate reputation. In *Corporate Reputation Review 18(1), pp. 3-24*. https://link.springer.com/article/10.1057/crr.2014.21.

Gallup, G. and Rae, S. (1968). *The Pulse of Democracy:the public-opinion poll and how it works*. Simon and Schuster, New York, 1st edition.

Gardner, G., Harvey, A., and Phillips, G. (1980). Algorithm AS 154: An algorithm for exact maximum likelihood estimation of autoregressive-moving average models by means of kalman filtering. In *Applied Statistics 29 pp. 311-322*. 10.2307/2346910.

Golub, G. and van Loan, C. (1992). *Matrix Computations*. Johns Hopkins University Press, Baltimore, 1st edition.

Higham, N. (1990). Analysis of the cholesky decomposition of a semi-definite matrix. In *Reliable Numerical Computation (eds. M. G. Cox and S. J. Hammarling), pp. 161–185*. Oxford University Press, 10.2307/2346910.

Hyndman, R. and Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. In *Jnl. Statistical Software, 26(3) pp. 1-22*. doi:10.18637/jss.v027.i03.

Jacobs, L. and Shapiro, R. (1995). Presidential manipulation of polls and public opinion. In *Political Science Quarterly 110(4) pp. 519-538*. doi:10.2307/21518825.

Janson, J. (2014). *The Reputation Playbook*. Harriman House, Petersfield, UK, 1st edition.

Lippman, W. (1922). *Public opinion*. Harcourt, Brace and Company, available from Project Gutenberg at https://gutenberg.org/ebooks/6456.

Liu, B. (2015). *Sentiment Analysis: Mining Opinions, Sentiments and Emotions*. CUP, New York, 1st edition.

Loke, R. and Kachaniuk, D. (2020). Sentiment polarity classification of corporate review data with a bidirectional long-short term memory (bilstm) neural network architecture. In *Proc, 9th International Conference on Data Science, Technology and Applications (DATA 2020), pp 310-317*. ScitePress doi:10.5220/0009892303100317.

Loke, R. and Kisoen, Z. (2022). The role of fake review detection in managing online corporate reputation. In

Proc, 11th International Conference on Data Science, Technology and Applications (DATA 2022), pp 245-256. ScitePress doi:10.5220/0011144600003269.

Loke, R. and Reitter, W. (2021). Aspect based sentiment analysis on online review data to predict corporate reputation. In *Proc, 10th International Conference on Data Science, Technology and Applications (DATA 2021), pp 343-352*. ScitePress doi:10.5220/0010607203430352.

Loke, R. and Vergeer, J. (2022). Exploring corporate reputation based on sentiment polarities that are related to opinions in dutch online reviews. In *Proc, 11th International Conference on Data Science, Technology and Applications (DATA 2022), pp 423-431*. ScitePress doi:10.5220/0011285500003269.

Mitic, P. (2015). Improved goodness-of-fit tests for operational risk. In *Journal of Operational Risk, 15(1), pp 77-126*. Incisive Media doi:10.21314/JOP.2015.159.

Mitic, P. (2024). What is the value of reputation? In *Proc. ICICT 2024, London*. To appear in Springer LNCS.

Rebonato, R. and Jaeckel, P. (2000). The most general methodology for creating a valid correlation matrix for risk management and option pricing purposes. In *Journal of Risk, 2(2), pp 17-27*. Incisive Media doi:10.21314/JOR.2000.023.

Reichheld, F. F. (2003). The one number you need to grow. In *Harvard Business Review 81(12) pp. 46–54*. https://hbr.org/2003/12/the-one-number-you-need-to-grow.

Renner, M. (2011). *Generating Trust via Corporate Reputation*. Wissenschaftlicher Verlag, Berlin, 1st edition.

Royston, P. (1992). Approximating the shapiro–wilk w-test for non-normality. In *Statistics and Computing 2(3), pp. 117–119*. doi:10.1007/BF01891203.

Rozin, P. and Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. In *Personality and Social Psychology Review. 5(4) pp. 296–320*. doi:10.1207/S15327957PSPR0504_2.

Russell, R. (2023). *The future of Corporate Communications*. Edelman.com, https://www.edelman.com/2023-future-of-corporate-comms.

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). In *Biometrika, 52(3,4), pp. 591-611*. doi:org/10.2307/2333709.

SignalAI (2024). *Signal AI Sharpens Reputation and Risk Capabilities with Acquisition of Social 360*. Press Release, https://signal-ai.com/press release/social-360/.

Weber-Shandwick (2020). *The State of Corporate Reputation in 2020: Everything Matters Now*. Weber Shandwick, https://webershandwick.com/news/the-state-of-corporate-reputation-in-2020-everything-matters-now.

Wen, X. and Li, W. (2023). Time series prediction based on lstm-attention-lstm model. In *IEEE Access 11, pp.48322-48331*. https://doi.org/10.1109/ACCESS.2023.3276628.

Yadev, H. and Thakkar, A. (2024). Noa-lstm: An efficient lstm cell architecture for time series forecasting. In *Expert Systems with Applications 238, pp. 122333*. https://doi.org/10.1016/j.eswa.2023.122333.

Zendesk (2013). *Customer service and Business results: a survey of customer service from mid-size Companies*. Zendesk.com, http://cdn.zendesk.com/resources/whitepapers Zendesk_WP_Customer_Service//_ and_Business_Results.pdf.

APPENDIX A

Proposition

An auto-correlation matrix A is positive definite ($\zeta' A \zeta > 0$ for all vectors ζ) , and therefore admits a *Cholesky* decomposition.

Preliminary result

A positive definite matrix has a *Cholesky* decomposition (Golub and van Loan, 1992) (Section 4.2.7)

Proof

Let A be an $L \times L$ auto-correlation matrix and let its column vectors be $z = \{z_1, z_2, \ldots z_L\}$. Symmetry is assured for (auto-)correlation matrices since for any two vectors z_i and z_j, $cor(z_i, z_j) = cor(z_j, z_i)$; $\quad i, j = 1 \ldots L$.

By definition, $A = \mathbb{E}[(z-\bar{z})(z-\bar{z})']$. Then, for all vectors ζ,

$$\begin{aligned} \zeta' A \zeta &= \zeta' \mathbb{E}[(z-\bar{z})(z-\bar{z})'] \zeta \\ &= \mathbb{E}[\zeta'(z-\bar{z})(z-\bar{z})' \zeta] \\ &= \mathbb{E}[yy'] \quad \text{where } y = \zeta'(z-\bar{z}) \\ &= \mathbb{E}[var(y)] > 0 \quad \forall y > 0 \end{aligned} \tag{11}$$

Also A is symmetric, and therefore A is positive definite.