SORT 42 (1) January-June 2018, 73-98 DOI: 10.2436/20.8080.02.70

Empirical analysis of daily cash flow time-series and its implications for forecasting

Francisco Salas-Molina[1], Juan A. Rodríguez-Aguilar[2], Joan Serrà[3]
Montserrat Guillen[4] and Francisco J. Martin[5]

Abstract

Usual assumptions on the statistical properties of daily net cash flows include normality, absence of correlation and stationarity. We provide a comprehensive study based on a real-world cash flow data set showing that: (i) the usual assumption of normality, absence of correlation and stationarity hardly appear; (ii) non-linearity is often relevant for forecasting; and (iii) typical data transformations have little impact on linearity and normality. This evidence may lead to consider a more data-driven approach such as time-series forecasting in an attempt to provide cash managers with expert systems in cash management.

MSC: 62J02, 62J05, 62P20.

Keywords: Statistics, forecasting, cash flow, non-linearity, time-series.

1. Introduction

Cash management is concerned with the efficient use of a company's cash and short-term investments such as marketable securities. The focus is placed on maintaining the amount of available cash as low as possible, while still keeping the company operating efficiently. In addition, companies may place idle cash in short-term investments (Ross, Westerfield and Jordan, 2002). Then, the cash management problem can be viewed as a trade-off between holding and transaction costs. If a company tries to keep balances too low, holding cost will be reduced, but undesirable situations of shortage will force to sell available marketable securities, hence increasing transaction costs. In contrast, if the balance is too high, low trading costs will be produced due to unexpected cash flow, but the company will carry high holding costs because no interest is earned on cash. There–

[1] Universitat de València, Av. Tarongers, s/n, 46022 València, Spain. francisco.salas-molina@uv.es
[2] IIIA-CSIC, Campus UAB, 08913 Cerdanyola, Spain. jar@iiia.csic.es
[3] Telefonica Research, Pl. Ernest Lluch, 5, 08019 Barcelona, Spain. joan.serra@telefonica.com
[4] Universitat de Barcelona, Diagonal, 690, 08034 Barcelona, Spain. mguillen@ub.edu
[5] BigML, Inc, 2851 NW 9th Suite, Conifer Plaza, Corvallis, OR 97330, US. martin@bigml.com
Received: March 2018
Accepted: June 2018

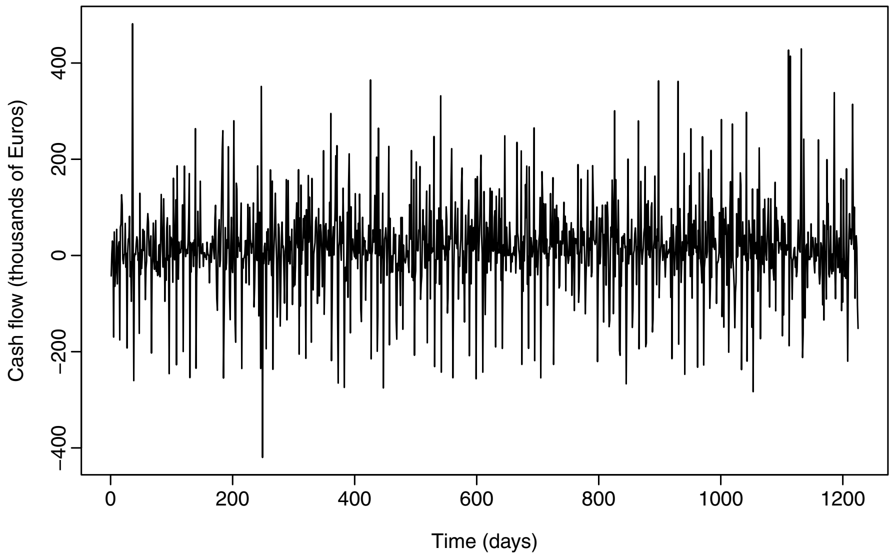

Figure 1: *Example of a cash flow time-series.*

fore, there is a target cash balance which each company must optimize according to the particular characteristics of its cash flows. An example of a raw cash flow time-series is shown in Figure 1, where observations do not apparently follow any seasonal pattern and whose evolution over time seems to be quite stable in terms of mean and variance, similarly to a white noise signal.

Testing the validity of time-series assumptions is an ongoing issue in finance (Marathe and Ryan, 2005; Ewing and Thompson, 2007; Cavaliere and Xu, 2014; Horváth, Kokoszka and Rice, 2014; Arratia, Cabana and Cabana, 2016; Torabi, Montazeri and Grané, 2016). Since Baumol (1952), a number of cash management models have been proposed to control cash balances. These models are based either on the specific statistical properties of cash balances or on cash flow forecasts. A comprehensive review of models, from the first proposals to the most recent contributions, can be found in Gregory (1976), Srinivasan and Kim (1986), and da Costa Moraes, Nagano and Sobreiro (2015). Most of them are based on assuming a given probability distribution for cash flows such as: (i) a random walk in the form of independent Bernouilli trials as in Miller and Orr (1966); (ii) a Wiener process as in Constantinides and Richard (1978), Premachandra (2004), and Baccarin (2009); (iii) a double exponential distribution as in Penttinen (1991). From these and other works, we observe that common assumptions on the statistical properties of cash flow time-series include:

- Normality: cash flows follow a Gaussian distribution with observations symmetrically centered around the mean, and with finite variance.
- Absence of correlation: the occurrence of past cash flows does not affect the probability of occurrence of the next ones.

- Stationarity: the probability distribution of cash flows does not change over time and, consequently, its statistical properties such as the mean and variance remain stable.
- Linearity: cash flows are proportional either to another (external) explanatory variable or to a combination of (external) explanatory variables.

Surprisingly, little and/or contradictory empirical evidence on these assumptions has been provided besides individual cases through time. Early on, negative normality tests were reported in Homonoff and Mullins (1975) for the times series samples of a manufacturing company. Contrastingly, later on, Emery (1981) reported normally distributed cash flow, after data transformation, for two out of three companies, and a small serial dependence for all of them. Pindado and Vico (1996) provided negative normality and independence results on 36 companies, but considering daily cash flow for only a single month. Previous works also reported day-of-week and day-of-month effects on cash flows, in line with the works of Stone and Wood (1977), Miller and Stone (1985), and Stone and Miller (1987). Recently, Gormley and Meade (2007) described the time-series from a multinational company with a non-normal distribution and serial dependence.

We consider that the evidence derived from these works is inconclusive due to: (i) the disagreement between the conclusions of some of the works; (ii) the limited number of companies analysed; and (iii) the short time range of the observations. Moreover, none of the previous works considered the presence of non-linear patterns for forecasting purposes. In this work, we provide an analysis of the statistical properties of 54 real cash flow data sets from small and medium companies in Spain as a representative sample of the most common type of companies in Europe. Indeed, small and medium companies contribute to 99.8% of all enterprises, 57.4% of value added, and 66.8% of employment across the EU28 (Muller et al., 2015). To the best of our knowledge, this is the most comprehensive empirical study on daily cash flow so far. We base this statement on both the length and number of data sets, which amounts to 58005 observations in total, with a minimum, average and maximum time range of 170, 737, 1508 working days, respectively. In addition, we consider a wider range of statistical properties. A further contribution of the present work is to make all the aforementioned data publicly available online[1]. Finally, from a forecasting perspective, we also aim to identify the family of forecasters that best accommodate to cash flow time-series data sets. To this end, we propose a new and simple cross-validated test for non-linearity that provides further knowledge to cash managers in their search for better forecasting models.

Our results show the unlikely occurrence of normality, absence of correlation and stationarity in the data sets under study. These results are consistent with the cited reports of Homonoff and Mullins (1975), based on only one time-series, and Pindado and Vico (1996), based on a very short time range, raising doubts about the claim of indepen-

1. http://www.iiia.csic.es/ jar/54datasets3.csv

dence. We also report that normality could not be achieved through removing outliers, contrary to what was reported by Emery (1981), based on only three time-series. Our analysis also confirms the influence of seasonality as suggested in Miller and Stone (1985) and Stone and Miller (1987). Thus, we consider that our results provide stronger evidence against normality, uncorrelatedness and stationarity than previous works. Note that we do not claim that these results can be extrapolated to all kind of companies. On the contrary, we provide further evidence against standard assumptions in cash management. This evidence may lead to consider a more data-driven approach such as time-series forecasting in order to provide cash managers with expert systems in cash management (Nedović and Devedžić, 2002).

In an attempt to achieve Gaussian and stationary time-seres, practitioners typically use the Box-Cox transformation (Box and Cox, 1964), and time-series differencing (Makridakis, Wheelwright and Hyndman, 2008). Furthermore, some kind of outlier treatment is also a recommended practice. Then, we also study the impact of outlier treatment by replacing them with linear interpolations between two consecutive observations. However, in our study, we find little benefit when these methods are applied to our data sets. As a result, we point out the underlying question about data transformation in relation to the properties of a time-series. Is it always possible to achieve a Gaussian and linear time-series through data transformations? We rely here both on common statistical tests and on our novel non-linearity test to answer this question and we find that: (i) outlier treatment and Box-Cox transformation are not always enough to achieve normality; (ii) outlier treatment produces mixed results in terms of noise reduction and information loss; (iii) outlier treatment and Box-Cox transformations do not produce linearity. These results suggest that non-linear models conform a justifiable alternative for cash flow time-series forecasting, beyond the current conjectures of the literature.

The remaining of the paper is organized as follows. In Section 2, we provide a statistical summary of the contributed 54 real cash flow data sets including normality, correlation and stationarity. In Section 3, we propose a new cross-validated test for non-linearity based on the comparison of a linear model and a non-linear model. Later, we present in Section 4 detailed results on the impact of data transformations on linearity. Finally, we provide some concluding remarks in Section 5.

2. Data summary

The data set contains daily cash flows from 54 different companies from the manufacturing and the service sector in Spain with annual revenue up to €10 million each. No company from the primary sectors is included in the sample. We select only small and medium companies since it is the most common size of companies in both Spain and Europe (Muller et al., 2015). This data set covers a date range of about eight years and is available online. An instance in the data set contains the following fields or columns:

Table 1: *Data sets statistical summary. Mean, standard deviation, minimum, maximum in thousands of €.*

Id	Length	Null %	Mean	Std	Kurtosis	Skewness	Min	Max
1	856	35.7	0.01	3.38	594.81	22.37	−9.07	90.27
2	684	29.8	0.26	5.80	58.98	3.69	−56.51	62.66
3	856	8.5	0.36	35.35	163.62	6.28	−303.20	671.04
4	1201	34.9	−0.12	14.32	78.14	−6.30	−223.38	72.76
5	849	19.4	0.00	1.67	56.10	−0.48	−18.26	16.42
6	799	20.7	0.01	6.63	33.21	−2.42	−68.97	56.27
7	772	38.5	0.07	5.36	86.75	6.74	−24.41	82.91
8	695	21.7	0.05	3.15	14.27	−2.57	−24.21	11.31
9	852	18.8	0.73	56.54	18.92	−0.78	−411.41	473.36
10	744	13.2	0.12	6.95	70.63	0.60	−81.13	78.72
11	639	62.6	−0.05	8.56	391.86	−17.65	−191.53	30.74
12	503	2.6	0.48	35.30	449.38	20.70	−47.27	771.38
13	697	24.7	0.52	24.24	18.81	2.06	−99.39	227.45
14	604	4.6	0.10	13.23	8.51	1.05	−63.23	92.71
15	605	4.1	0.68	11.67	4.43	0.33	−54.75	55.61
16	596	6.4	0.01	1.46	107.82	6.68	−8.48	22.61
17	1102	25.1	0.58	13.31	215.97	11.96	−118.01	250.13
18	552	3.1	0.16	2.16	70.23	5.10	−16.14	26.36
19	503	2.4	−0.31	2.58	6.43	0.50	−15.06	15.28
20	848	27.8	0.02	1.07	96.19	3.86	−12.07	16.04
21	829	18.7	−0.06	5.99	33.36	−1.62	−70.00	53.17
22	494	1.6	−0.46	27.28	22.64	−1.96	−244.29	138.87
23	604	9.1	1.63	20.85	79.99	5.41	−124.19	269.27
24	1097	8.4	0.96	20.36	95.45	6.48	−73.33	317.85
25	587	10.9	0.49	13.94	119.60	6.93	−116.01	201.13
26	751	11.6	−0.02	1.77	15.73	0.15	−10.73	15.56
27	332	8.1	0.29	1.64	10.60	2.14	−4.36	11.84
28	855	5.1	0.00	4.64	13.83	1.77	−18.10	39.01
29	609	13.6	0.04	6.07	108.66	−6.35	−90.04	55.89
30	554	8.1	0.03	1.47	68.26	5.47	−4.81	19.82
31	372	29.6	0.37	8.05	31.46	−2.41	−80.44	34.95
32	1103	24.8	0.28	4.03	11.07	0.54	−25.76	24.50
33	854	31.0	−0.19	6.81	115.63	−1.74	−94.33	95.59
34	1508	11.5	−0.06	10.13	19.89	−2.32	−96.82	49.65
35	501	7.4	0.20	5.40	11.41	−0.58	−31.42	29.19
36	359	11.4	0.42	1.85	12.24	2.44	−7.87	11.84
37	361	3.0	−0.69	17.82	139.06	−1.38	−228.88	218.42
38	170	9.4	−1.20	7.10	43.34	−5.73	−61.93	19.66
39	1104	29.0	0.02	0.95	7.95	−0.07	−5.67	6.57
40	198	0.0	0.78	12.38	0.58	1.02	−25.63	36.91
41	341	17.6	−0.25	8.34	15.80	1.22	−44.29	64.34
42	566	11.0	0.01	1.82	308.62	−15.80	−37.02	7.48
43	750	3.2	0.34	13.10	7.66	−0.04	−65.84	73.40
44	287	4.2	0.52	11.46	81.19	−0.05	−118.74	120.34
45	1465	49.8	0.04	9.12	43.51	−2.89	−107.20	75.47
46	565	44.8	0.54	5.58	75.41	2.91	−51.16	73.83
47	503	4.4	1.98	46.81	46.03	1.37	−338.39	478.26
48	605	13.1	0.21	22.71	34.31	−1.68	−207.04	203.09
49	993	50.5	−0.08	1.36	27.18	−2.18	−10.78	12.73
50	605	45.0	−0.01	27.37	43.79	−2.01	−262.52	221.96
51	1225	0.2	15.09	96.96	2.77	0.12	−419.88	481.66
52	1225	0.4	8.94	49.39	36.23	2.81	−325.46	700.66
53	1223	39.7	0.47	9.13	203.12	−10.25	−196.88	38.48
54	1225	52.3	0.46	77.91	151.93	4.28	−1021.36	1532.10

- Date: standardized YYYY-MM-DD dates from 2009-01-01 to 2016-28-08.
- Company: company identifier from 1 to 54.
- NetCF: daily net cash flow in thousands of €.
- DayMonth: categorical variable with the day of the month from 1 to 31.
- DayWeek: categorical variable with the day of the week from 1 (Monday) to 7 (Sunday).

Table 1 shows the statistical summary of daily net cash flow on non-holidays, grouped by company. Small and medium companies are likely to experiment daily null cash flows, meaning that no monetary movement is observed at a particular working day even under regular activity. As a result, the occurrence of null cash flows is an important characteristic of small and medium companies due to the size of companies. Indeed, almost 30% of the companies in our data set present more than 25% of null cash flow observations even at working days. This fact implies that a null cash flow prediction will be right at least 25% of the times for this group of data sets. Therefore, two good baseline forecasting models for comparative purposes would be an *always-predict-null* or an *always-predict-mean* forecaster (Makridakis et al., 2008).

In addition, the average net cash flow shows that a high percentage of companies present either positive or negative mean with the exception of companies 5 and 28. High positive kurtosis indicates a peaked data distribution in comparison to the normal distribution that has zero kurtosis. The skewness is a measure of the symmetry of the data distribution. Negative skewness indicates that the left tail is longer, and positive skewness indicates that the right tail is longer.

2.1. Normality

First, we study if our cash flows follow a Gaussian distribution. In fact, the observed kurtosis and skewness can be used as a first normality test of the data distribution for each company. Table 1 shows that no company presents zero kurtosis and skewness. Only company 40, with kurtosis 0.58 and skewness 1.02, could be considered close to normality. The proportion of null cash flows is also a strong evidence against normality. Since this situation is likely to be common for SMEs and, due to the high proportion of this type of companies in Europe, we believe that cash managers should test normality before applying cash management models based on this assumption.

Two additional tests can be used to either verify or reject the hypothesis of normality: the Shapiro-Wilk test for normality (Royston, 1982) and the Lilliefors (Kolmogorov-Smirnov) test for normality (Lilliefors, 1967). The results from these two tests applied to the original time-series (summarized in Table 2) allow us to reject the hypothesis of normally distributed cash flows for all the companies in our data set (no exception). However, the presence of correlation and possible changes in the mean of data sets may limit the reliability of these tests. We overcome this problem by performing an additional normality test. More precisely, we check the normality of the residuals of fitting an ARMA model to each of the time-series as suggested by Ducharme and Lafaye de Micheaux (2004). To obtain ARMA models, we follow the automatic fitting procedure described in Hyndman and Khandakar (2008). Finally, we test the normality of the residuals by means of Neyman (1937) smooth tests as recently proposed by Ducharme and Lafaye de Micheaux (2004) and Duchesne, Lafaye de Micheaux and Tagne Tatsinkou (2016). The results from Table 2 before any data transformation suggest the rejection of the normality hypothesis.

As pointed out elsewhere (Emery, 1981; Pindado and Vico, 1996), a possible explanation for non-normality could be the presence of abnormally high values or heavy tails. Thus, we repeated the Shapiro-Wilk, the Lilliefors (Kolmogorov-Simirnov), and the Neyman tests for normality, but now using a trimmed version of the net cash flow time-series by deleting observations greater or lower than three times the sample standard deviation. No difference in the results of the tests was observed, confirming the non-normality hypothesis beyond the conjectures of Emery (1981) and Pindado and Vico (1996).

Non-normal residuals may be problematic in the estimation process when using linear models. Data transformations such as the Box and Cox (1964) transformation to normality represent a possible solution. Forecasts are then calculated on the transformed data, but we must reverse the transformation to obtain forecasts on the original data, resulting in two additional steps. However, these transformations are not always the solution to the non-normality problem. Using both the original observations and the trimmed version of our data sets, we proceeded to transform the data using a Box-Cox transformation of the type:

$$y^{(\lambda)} = \begin{cases} \frac{(y+\lambda_2)^{\lambda_1}-1}{\lambda_1} & \text{if} \quad \lambda_1 \neq 0, \\ \log(y+\lambda_2) & \text{if} \quad \lambda_1 = 0, \end{cases} \tag{1}$$

where y is the original time-series, and λ_1 and λ_2 are parameters. In these experiments, we first set λ_2 to minus two times the minimum value of the time-series to avoid problems with negative and zero observations. Box and Cox (1964) provided the profile likelihood function for λ_1 and suggested to use this function as a way to tune this parameter. Then, we follow the recommendations in Venables and Ripley (2013) to compute the profile likelihood function for λ_1, and we later select the value that maximizes the log-likelihood function when applying a linear regression model of the time-series based on day-of-month and day-of-week dummy variables. After a Box-Cox transformation on the trimmed time-series, we repeated the Shapiro-Wilk, the Lilliefors (Kolmogorov-Smirnov), and the Neyman smooth tests for normality obtaining again negative results as shown in Table 2. A possible explanation of these results is that the correlational structure of a transformed time-series closely depends on the original. A special case of this feature for a logarithmic transformation can be found in Moriña, Puig and Valero (2015). As a result, we must conclude that, even after Box-Cox transformation, the normality hypothesis does not hold.

Table 2: *Normality tests. SW: Shapiro-Wilk; LKS: Lilliefors (Kolmogorov-Smirnov).*

	Before data transformation						After trimming and Box-Cox transformation					
Id	SW	p-value	LKS	p-value	Neyman	p-value	SW	p-value	LKS	p-value	Neyman	p-value
1	0.20	<0.05	0.31	<0.05	13165.72	<0.05	0.76	<0.05	0.23	<0.05	2125.70	<0.05
2	0.36	<0.05	0.35	<0.05	6539.22	<0.05	0.55	<0.05	0.35	<0.05	3884.23	<0.05
3	0.42	<0.05	0.26	<0.05	6452.81	<0.05	0.76	<0.05	0.26	<0.05	839.58	<0.05
4	0.45	<0.05	0.32	<0.05	10249.73	<0.05	0.69	<0.05	0.28	<0.05	4463.36	<0.05
5	0.54	<0.05	0.24	<0.05	5103.79	<0.05	0.83	<0.05	0.23	<0.05	949.19	<0.05
6	0.59	<0.05	0.27	<0.05	3694.65	<0.05	0.78	<0.05	0.26	<0.05	768.83	<0.05
7	0.48	<0.05	0.32	<0.05	4768.18	<0.05	0.74	<0.05	0.29	<0.05	1223.53	<0.05
8	0.64	<0.05	0.33	<0.05	1902.93	<0.05	0.75	<0.05	0.28	<0.05	1724.23	<0.05
9	0.54	<0.05	0.39	<0.05	2558.89	<0.05	0.63	<0.05	0.38	<0.05	1563.74	<0.05
10	0.41	<0.05	0.24	<0.05	6437.00	<0.05	0.78	<0.05	0.25	<0.05	1715.64	<0.05
11	0.17	<0.05	0.39	<0.05	10234.39	<0.05	0.49	<0.05	0.35	<0.05	3164.46	<0.05
12	0.11	<0.05	0.37	<0.05	9828.61	<0.05	0.76	<0.05	0.28	<0.05	1171.58	<0.05
13	0.59	<0.05	0.33	<0.05	2229.85	<0.05	0.65	<0.05	0.31	<0.05	1111.39	<0.05
14	0.84	<0.05	0.17	<0.05	353.45	<0.05	0.89	<0.05	0.16	<0.05	127.69	<0.05
15	0.89	<0.05	0.16	<0.05	216.99	<0.05	0.91	<0.05	0.16	<0.05	148.40	<0.05
16	0.53	<0.05	0.22	<0.05	3145.88	<0.05	0.88	<0.05	0.20	<0.05	175.85	<0.05
17	0.24	<0.05	0.34	<0.05	14724.84	<0.05	0.69	<0.05	0.31	<0.05	2295.82	<0.05
18	0.55	<0.05	0.19	<0.05	2660.82	<0.05	0.89	<0.05	0.19	<0.05	242.65	<0.05
19	0.92	<0.05	0.10	<0.05	456.82	<0.05	0.97	<0.05	0.10	<0.05	54.93	<0.05
20	0.53	<0.05	0.24	<0.05	4530.45	<0.05	0.84	<0.05	0.24	<0.05	578.91	<0.05
21	0.69	<0.05	0.23	<0.05	2015.07	<0.05	0.84	<0.05	0.22	<0.05	588.95	<0.05
22	0.67	<0.05	0.22	<0.05	1263.64	<0.05	0.82	<0.05	0.21	<0.05	285.96	<0.05
23	0.48	<0.05	0.27	<0.05	3365.92	<0.05	0.77	<0.05	0.27	<0.05	500.78	<0.05
24	0.49	<0.05	0.29	<0.05	5986.28	<0.05	0.71	<0.05	0.31	<0.05	1345.56	<0.05
25	0.36	<0.05	0.28	<0.05	5084.02	<0.05	0.74	<0.05	0.28	<0.05	805.87	<0.05
26	0.75	<0.05	0.21	<0.05	1271.00	<0.05	0.84	<0.05	0.21	<0.05	476.22	<0.05
27	0.81	<0.05	0.17	<0.05	3832.19	<0.05	0.92	<0.05	0.14	<0.05	1034.84	<0.05
28	0.77	<0.05	0.21	<0.05	1622.42	<0.05	0.85	<0.05	0.19	<0.05	486.48	<0.05
29	0.31	<0.05	0.39	<0.05	6782.20	<0.05	0.55	<0.05	0.37	<0.05	3712.34	<0.05
30	0.62	<0.05	0.20	<0.05	1988.27	<0.05	0.86	<0.05	0.21	<0.05	196.55	<0.05
31	0.69	<0.05	0.18	<0.05	1707.62	<0.05	0.86	<0.05	0.19	<0.05	580.67	<0.05
32	0.73	<0.05	0.22	<0.05	1838.10	<0.05	0.80	<0.05	0.22	<0.05	741.53	<0.05
33	0.24	<0.05	0.37	<0.05	10442.42	<0.05	0.44	<0.05	0.38	<0.05	6581.93	<0.05
34	0.68	<0.05	0.27	<0.05	2529.16	<0.05	0.77	<0.05	0.23	<0.05	1048.52	<0.05
35	0.74	<0.05	0.24	<0.05	1027.85	<0.05	0.83	<0.05	0.23	<0.05	356.87	<0.05
36	0.67	<0.05	0.26	<0.05	4331.31	<0.05	0.80	<0.05	0.23	<0.05	1949.37	<0.05
37	0.10	<0.05	0.46	<0.05	6768.47	<0.05	0.18	<0.05	0.46	<0.05	5398.45	<0.05
38	0.41	<0.05	0.30	<0.05	1695.03	<0.05	0.68	<0.05	0.21	<0.05	371.92	<0.05
39	0.82	<0.05	0.24	<0.05	497.26	<0.05	0.87	<0.05	0.23	<0.05	191.56	<0.05
40	0.89	<0.05	0.21	<0.05	50.00	<0.05	0.95	<0.05	0.15	<0.05	2155.90	<0.05
41	0.66	<0.05	0.28	<0.05	1117.07	<0.05	0.73	<0.05	0.27	<0.05	507.15	<0.05
42	0.21	<0.05	0.36	<0.05	8189.26	<0.05	0.68	<0.05	0.24	<0.05	725.85	<0.05
43	0.84	<0.05	0.16	<0.05	686.05	<0.05	0.90	<0.05	0.17	<0.05	281.96	<0.05
44	0.37	<0.05	0.30	<0.05	2611.31	<0.05	0.71	<0.05	0.30	<0.05	427.38	<0.05
45	0.42	<0.05	0.36	<0.05	11122.67	<0.05	0.57	<0.05	0.35	<0.05	6084.71	<0.05
46	0.34	<0.05	0.37	<0.05	5194.50	<0.05	0.51	<0.05	0.37	<0.05	2620.56	<0.05
47	0.40	<0.05	0.30	<0.05	3979.90	<0.05	0.64	<0.05	0.30	<0.05	1869.51	<0.05
48	0.52	<0.05	0.34	<0.05	2851.85	<0.05	0.66	<0.05	0.33	<0.05	1247.03	<0.05
49	0.38	<0.05	0.46	<0.05	6891.41	<0.05	0.51	<0.05	0.45	<0.05	2739.27	<0.05
50	0.30	<0.05	0.39	<0.05	6871.79	<0.05	0.40	<0.05	0.38	<0.05	4855.75	<0.05
51	0.93	<0.05	0.11	<0.05	320.22	<0.05	0.94	<0.05	0.12	<0.05	293.46	<0.05
52	0.80	<0.05	0.15	<0.05	1538.44	<0.05	0.93	<0.05	0.13	<0.05	293.37	<0.05
53	0.35	<0.05	0.33	<0.05	9770.54	<0.05	0.62	<0.05	0.29	<0.05	3792.57	<0.05
54	0.30	<0.05	0.37	<0.05	11389.52	<0.05	0.45	<0.05	0.37	<0.05	7568.90	<0.05

2.2. Correlation and seasonality

In what follows, we test the correlation of cash flows and we also explore if seasonality is present. Autoregressive Integrated Moving Average (ARIMA) models by Box and Jenkins (1976), have been extensively used for time-series analysis and forecasting. When dealing with time-series, the autocorrelation coefficient, r_k, describes the relationship between observations that are lagged k time periods (Makridakis et al., 2008). We say that a time-series is not autocorrelated when the r_k values for different lags are close to zero. An example of an independent time-series is the so-called white-noise model where each observation is made by adding a random component to a certain level.

An intuitive plot to assess correlation is the Poincaré map (Kantz and Schreiber, 2004), which is a scatter plot of the original time-series and a k-periods lagged time-series as in Figure 2, which shows a lag of 1 day for time-series 1 and 2 from Table 1. As a reference, we also include the Poincaré map for a white-noise and for a sinusoidal time-series. A cloud of points suggests lack of correlation, as for time-series 1 and white-noise, and the presence of any form suggests a more complex relationship, as for time-series 2 and the sinusoidal. For comparative purposes, we present in Figure 3 the classical plots showing autocorrelation and partial autocorrelation functions for different lags within the range 1-20 with dashed horizontal lines representing 95% confidence intervals. From the analysis of Figure 3, we note correlation for time-seres 1 and 2 at lags 1 and 15, respectively.

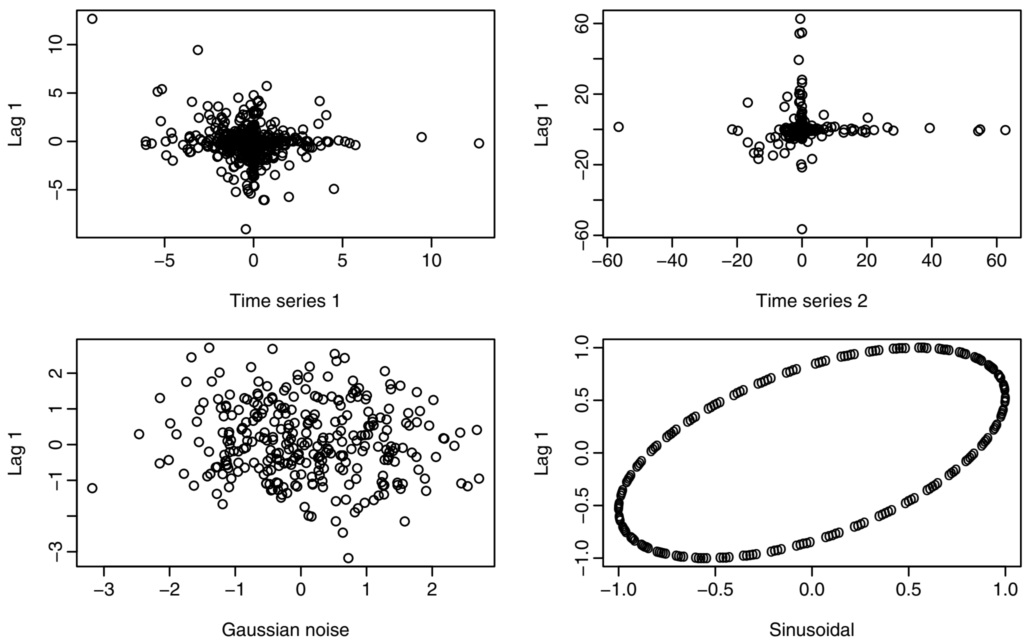

Figure 2: *Poincaré map with lag 1 for time-series 1 and 2.*

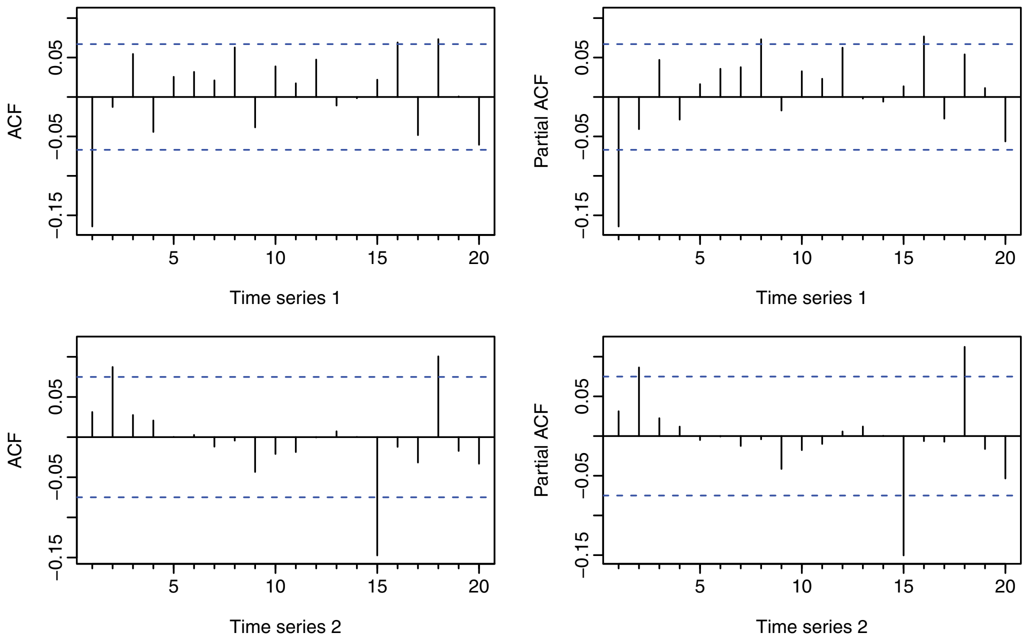

Figure 3: *Autocorrelation plots for time-series 1 and 2.*

A more general approach is to consider a set of the first r_k values as a whole as in the Ljung and Box (1978) test, which we applied to the original time-series and produced mixed results. More precisely, we found that the null hypothesis of independence could not be rejected in 24 out of 54 companies as summarized in Table 3. These results imply that some kind of serial correlation is likely to be present in the case of companies presenting a certain degree of autocorrelation in the sample. A plausible type of serial correlation is seasonality, that is, the existence of a pattern that repeats itself over fixed time intervals in the data (Makridakis et al., 2008). It can be identified by significant autocorrelation coefficients. Seasonal trend decomposition methods Cleveland et al., 1990), seasonal ARIMA models (Box and Jenkins, 1976; Franses and Van Dijk, 2005) or linear (and non-linear) regression models based on seasonal variables are available options to deal with seasonality. In cash flow forecasting, the distribution approach by Miller and Stone (1985) also deserves to be mentioned.

As mentioned in the introduction, previous works by Emery (1981), Miller and Stone (1985), Stone and Miller (1987), and Pindado and Vico (1996), reported the influence of day-of-month and day-of-week effects on cash flow patterns. Here, we test the presence of seasonality by fitting a regression model on raw daily cash flows using day-of-month and day-of-week dummy variables. To avoid co-linearity issues in regression, we use thirty day-of-month dummy variables from the 2nd to the 31st day of the month and four day-of-week variables from Monday to Thursday up to a total of 34 regression variables. At each time step t, predictor x_{ti} is set to one if the corresponding day-of-month is i, zero otherwise, and x_{tj} is set to one if the corresponding day-of-week is j, with j ranging from 1 for Monday to 4 for Thursday. Mathematically, the linear regression model used to test seasonality is expressed as follows:

$$y_t = \sum_{i=2}^{31} \beta_i x_{ti} + \sum_{j=1}^{4} \beta_j x_{tj} + \epsilon. \tag{2}$$

Table 3 reports, on the one hand, the Ljung-Box correlation test applied to raw data and, on the other hand, the F-statistic, the p-value and the coefficient of determination R^2, derived from the regression model. One may expect that the rejection of the correlation null hypothesis results in better regressions. Our results, however, show a different behavior. Non-linear patterns, non-periodical temporal correlations, and the effect of outliers become possible explanations as we will see below.

2.3. Stationarity

In this section, we analyse if cash flows from our data set can be labelled as stationary. More precisely, we focus on weak stationarity that considers the change over time of the first (mean) and second moment (variance) of a random process. We can visually assess stationarity by inspecting a time-series plot as the one shown in Figure 1. Virtually, every process we find in nature is non-stationary, since its parameters depend on time (Kantz and Schreiber, 2004). However, a minimum requirement is that basic statistical properties of a distribution, such as mean and variance, remain constant over time, when measured through appropriately long time windows. It is important to highlight that seasonality is a particular case of non-stationarity, at least, within each periodic fluctuation when we focus on short-term changes in parameters. In what follows, we pay attention to long-term changes (periods longer than a month) as a way to assess stationarity.

Following the recommendations in Kantz and Schreiber (2004), we perform a stationarity test based on the fluctuations of a sample mean and variance. More precisely, we compute the sample mean and variance of each original time-series by months and obtain the standard errors for both. If the observed fluctuations of the running mean and variance are within these errors, then we consider the time-series stationary. The results from this test shows that none of the time-series is stationary. These results are consistent with the fact that most of the p-values of the regression models used for checking seasonality are below 0.05 as summarized in Table 3.

One way of removing non-stationarity is time-series differencing, which is defined as the change between two consecutive observations. Similarly, seasonal differencing is the change between corresponding observations from two consecutive seasonal periods. Since the presence of seasonality is likely (see Table 3), we next explore three alternative

Table 3: *Correlation and seasonality test results.*

Id	Ljung-Box Test	Statistic	p-value	F-statistic	p-value	R^2
1	Non-rejected	11.05	1.00	1.99	< 0.05	0.08
2	Rejected	65.99	< 0.05	1.05	0.39	0.05
3	Non-rejected	34.47	0.72	1.87	< 0.05	0.07
4	Rejected	120.15	< 0.05	1.51	< 0.05	0.04
5	Rejected	120.91	< 0.05	1.85	< 0.05	0.07
6	Non-rejected	46.96	0.21	1.12	0.29	0.05
7	Rejected	166.97	< 0.05	5.47	< 0.05	0.20
8	Rejected	67.15	< 0.05	0.79	0.80	0.04
9	Rejected	97.32	< 0.05	5.30	< 0.05	0.18
10	Rejected	145.00	< 0.05	2.04	< 0.05	0.09
11	Non-rejected	10.57	1.00	0.97	0.51	0.05
12	Non-rejected	3.25	1.00	0.98	0.51	0.07
13	Rejected	139.26	< 0.05	5.21	< 0.05	0.21
14	Rejected	74.58	< 0.05	7.13	< 0.05	0.30
15	Rejected	87.67	< 0.05	1.92	< 0.05	0.10
16	Non-rejected	38.12	0.56	4.31	< 0.05	0.21
17	Non-rejected	14.49	1.00	4.91	< 0.05	0.14
18	Rejected	57.25	< 0.05	2.99	< 0.05	0.16
19	Rejected	75.16	< 0.05	2.58	< 0.05	0.16
20	Non-rejected	43.37	0.33	2.71	< 0.05	0.10
21	Non-rejected	46.65	0.22	1.37	0.08	0.06
22	Non-rejected	33.35	0.76	1.49	< 0.05	0.10
23	Rejected	68.36	< 0.05	5.60	< 0.05	0.25
24	Non-rejected	41.30	0.41	15.41	< 0.05	0.33
25	Non-rejected	33.35	0.76	4.23	< 0.05	0.21
26	Rejected	95.79	< 0.05	1.22	0.18	0.05
27	Non-rejected	44.66	0.28	1.24	0.18	0.12
28	Rejected	112.21	< 0.05	5.64	< 0.05	0.19
29	Non-rejected	42.55	0.36	1.37	0.08	0.08
30	Rejected	107.46	< 0.05	6.18	< 0.05	0.29
31	Non-rejected	47.51	0.19	1.25	0.16	0.11
32	Rejected	105.26	< 0.05	4.81	< 0.05	0.13
33	Rejected	201.50	< 0.05	1.57	< 0.05	0.06
34	Rejected	130.53	< 0.05	11.61	< 0.05	0.21
35	Rejected	66.04	< 0.05	0.99	0.49	0.07
36	Non-rejected	44.66	0.28	1.82	< 0.05	0.16
37	Rejected	96.75	< 0.05	1.58	< 0.05	0.14
38	Non-rejected	45.37	0.26	1.06	0.39	0.21
39	Rejected	192.30	< 0.05	6.11	< 0.05	0.16
40	Rejected	78.81	< 0.05	0.86	0.68	0.15
41	Non-rejected	39.05	0.51	1.72	< 0.05	0.16
42	Non-rejected	22.85	0.99	3.90	< 0.05	0.20
43	Rejected	80.56	< 0.05	2.96	< 0.05	0.12
44	Non-rejected	19.56	1.00	1.89	< 0.05	0.20
45	Rejected	82.69	< 0.05	1.26	0.15	0.03
46	Non-rejected	32.23	0.80	1.32	0.11	0.08
47	Non-rejected	35.67	0.67	0.90	0.63	0.06
48	Non-rejected	42.53	0.36	1.71	< 0.05	0.09
49	Rejected	105.02	< 0.05	26.15	< 0.05	0.48
50	Rejected	135.48	< 0.05	1.24	0.17	0.07
51	Rejected	131.27	< 0.05	16.66	< 0.05	0.32
52	Rejected	66.68	< 0.05	5.01	< 0.05	0.13
53	Non-rejected	18.62	1.00	1.59	< 0.05	0.04
54	Rejected	129.11	< 0.05	0.88	0.67	0.02

seasons (or periods) to apply differencing: 1) one day, equivalent to no seasonality; 2) five days, to account for day-of-week seasonality; and 3) twenty days, to account for day-of-month seasonality. Finally, differencing can be applied only once to data, twice or a number n of times defining the order of differencing. In Table 4, we summarize stationarity results for our data set in terms of the number of time-series that are labelled as stationary in mean and variance according to the test described above. Only a small fraction of time-series can be considered stationary in mean (but not in variance) after first and second-order differencing. From this analysis, we conclude that our cash flow time-series are non-stationary, even after differencing.

Table 4: *Percentage of time-series labelled as stationary in mean and variance.*

Differencing	Zero-order		First-Order		Second-Order	
Seasonality	Mean	Var	Mean	Var	Mean	Var
1	0	0	18.5	0	18.5	0
5	0	0	3.7	0	5.6	0
20	0	0	0	0	0	0

2.4. Discussion

Our results show that the widely extended hypothesis of cash flow normality is not present in our data sets. The presence of high abnormal values does not explain this behavior since non-normality persisted after removing these abnormal values. Non-linearity could be a possible explanation as we will see below. We also reported mixed results on autocorrelation and the influence of day-of-month and day-of-week effects on cash flow along the lines of the literature. We additionally report that common solutions to non-normality and non-stationarity such as data transformation and differencing produced little benefit when applied to our time-series. Since seasonality and serial correlation are also present in our data set, we further explore the usefulness of alternative forecasting models. More precisely, we next study linearity and data transformation as an additional part of our empirical analysis for cash flow forecasting.

3. A simple cross-validated test for non-linearity

Most forecasting models are linear for computational convenience. However, non-linear patterns are likely to be present in finance and business time-series. A time-series linear model is defined as a variable y_t that depends on the additive contribution of a number of explanatory variables in vector $\boldsymbol{x_t}$ for any time t as follows:

$$y_t = \beta^{\mathsf{T}} \boldsymbol{x_t} + e_t \tag{3}$$

where $\boldsymbol{\beta}^{\mathsf{T}}$ is a transposed vector of coefficients, and e_t is the error or the residual component. An alternative and more general model can also be considered:

$$y_t = g(\boldsymbol{x_t}) + \epsilon_t \tag{4}$$

where $g(\boldsymbol{x_t})$ is any function that aims to describe the underlying time-series. By considering non-linear relationships between the set of predictors and the cash flow dependent variable, more complex patterns such as interactions between the day-of-week and the day-of-month may be captured.

Different tests of linearity can be found in Ramsey (1969), Keenan (1985), Lee, White and Granger (1993), and Castle and Hendry (2010). Basically, all of them follow a common approach: first, they choose a function $g(\boldsymbol{x_t})$ in equation (4) including linear and non-linear terms and, second, they test for the significance of the non-linear terms. However, these approaches are not suitable for forecasting purposes owing to the following reasons: (i) the assumption of a specific form $g(\boldsymbol{x_t})$ for the regression equation such as quadratic, cubic or exponential forms; (ii) cross-validation is neglected.

If we relax the assumption of linearity, different non-linear models such as random forests (Breiman, 2001), neural networks (Hornik, Stinchcombe and White, 1989; Zhang, Patuwo and Hu, 1998), or radial basis functions (Broomhead and Lowe, 1988), could also be considered. However, the consideration of non-linear functions may lead to overfitting to the original time-series. To prevent this problem, we propose the use of time-series cross-validation. Cross-validation is a method to assess the predictive performance of a forecasting model that circumvents the problem of overfitting the data by testing the accuracy of the model on subset of data not used in the estimation (Hyndman and Athanasopoulos, 2013). As a result, we here propose a simple cross-validated test for non-linearity based on the following steps:

1. Estimate two alternative forecasting models, one linear and another one non-linear.
2. Cross-validate the predictive accuracy of both models with respect to a baseline.
3. Label as trivial[2] if both models are significantly worse than the baseline.
4. Label as non-linear if the error of the non-linear model is significantly lower than that of the linear model. Otherwise, label as linear as described in Figure 4.

Since we do not assume any distribution for the forecasting results, we use the two-sided Wilcoxon rank-based for statistically significant differences in performance between models. More precisely, we test the null hypothesis that the distribution of the difference is symmetric about zero with a 95% confidence interval (Wilcoxon, Katti and Wilcox, 1970). Approximate p-values are computed based on the asymptotic distribution of the two-sided Wilcoxon test statistic and used to label data sets as detailed in Algorithm 1.

2. Trivial is here used with the meaning of very little value with respect to a basic standard.

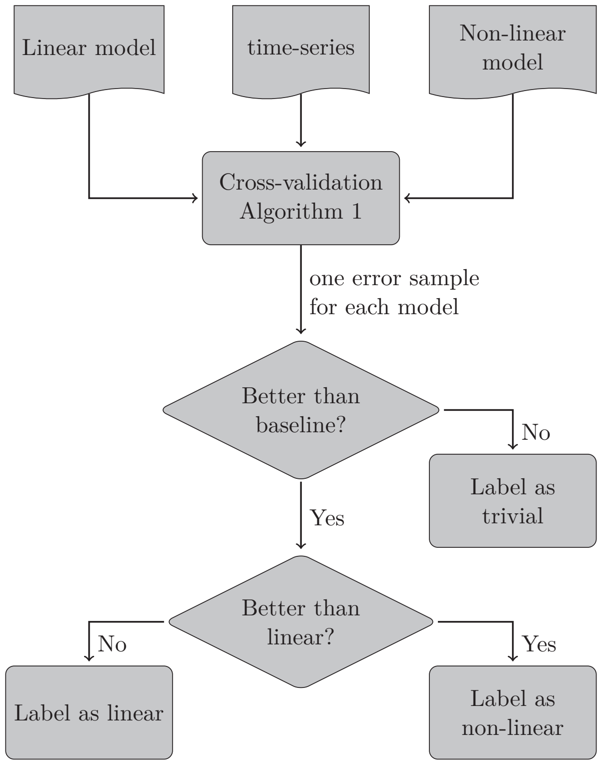

Figure 4: *Simplified flow chart for our cross-validated test for non-linearity.*

Algorithm 1 Algorithm for a simple cross-validation test for non-linearity

---

Input: Cash flow data set of T instances, minimum number k of instances to estimate a model, baseline m_0, linear model m_1, non-linear model m_2, prediction horizon h, level of significance α.
Output: Average prediction error, statistic for the difference in mean errors, confidence interval.
for $i = 1, 2, \ldots, T - k - h + 1$ **do**
    Select the instances from time $k + i$ to $k + h + i - 1$, for the test set;
    Estimate m_0 with instances at times $1, 2, \ldots, k + i - 1$;
    Estimate m_1 with instances at times $1, 2, \ldots, k + i - 1$;
    Estimate m_2 with instances at times $1, 2, \ldots, k + i - 1$;
    Compute test errors ε_0, ε_1, ε_2 from time $k + i$ to $k + h + i - 1$;
Compute average h-step errors $\varepsilon_0(h)$, $\varepsilon_1(h)$, $\varepsilon_2(h)$;
Test for α significant differences between $\varepsilon_0(h)$, $\varepsilon_1(h)$, $\varepsilon_2(h)$;
if $\varepsilon_0(h) < \varepsilon_1(h)$ **and** $\varepsilon_0(h) < \varepsilon_2(h)$ **then**
    Label as trivial;
else if $\varepsilon_2(h) < \varepsilon_1(h)$ **then**
    Label as non-linear;
else
    Label as linear.

A common practice to assess the usefulness of forecasts derived from any model is to compare its accuracy to that of a baseline forecasting model. The use of a baseline model allows us to label our data sets as trivial if neither the linear model nor the non-linear model are able to improve the accuracy of the baseline. We here report accuracy results with respect to a mean forecaster, meaning that forecasts are always the average of all past observations. We also tried with an additional baseline forecaster using the last observed value as a forecast (persistence model) with much worse results in comparison to the mean forecaster.

We consider the minimum length k to estimate a model as the 80% of the oldest instances forming the training set. The remaining 20% of the instances form the test set for cross-validation. Initially, both the linear and the non-linear model are estimated using the first 80% of the instances. Then, forecasts for a prediction horizon up to 20 days are computed using the estimated models and squared errors are recorded. Then, forecasting accuracy is evaluated on a rolling basis, since both the last observation of the training set and the first observation of the test set roll forward in time. As a result, forecasting errors are recorded for each remaining observation in the test set resulting into two paired error samples, one for the linear model and one for non-linear model.

A critical point when using our cross-validated test for non-linearity is the selection of both the linear and the non-linear forecast model. In essence, our test is a comparative tool based on forecasting accuracy as a proxy for non-linearity. Given a set of explanatory variables, a linear label result from our test implies that the non-linear model is not able to capture non-linearity. However, chances are that alternative non-linear models might perform differently. In this sense, if the time-series is not a white-noise process, then the search for a more informative set of features is meant to play a key role. As a result, multiple runs of our test are necessary to discard/assess non-linearity by using alternative linear and non-linear models

For illustrative purposes, we here restrict ourselves to a linear regression model and a non-linear random forest model, both using day-of-month and day-of-week variables as predictors. Salas-Molina et al. (2017) report that these two models perform significantly better than autoregressive models when producing forecasts for usual prediction horizons up to one hundred days. Here, we are interested in comparing forecasting models that perform well for a wide range of planning horizons from the information available at some point in time. Thus, we expect that forecasting models based on seasonal variables capture patterns for common prediction horizons better than time-series models based on previous observations due to lack of relevant information as reported by Salas-Molina et al. (2017).

In the case of the linear regression model, each instance contains 34 dummy predictor variables, 30 for day-of-month and 4 for day-of week, and a cash flow observation. This linear regression model is the same that we used in Section 2.2 to check seasonality. In the case of random forests, each instance contains two categorical variables, one for day-of-month and one for day-of-week. Random forests are ensembles of slightly

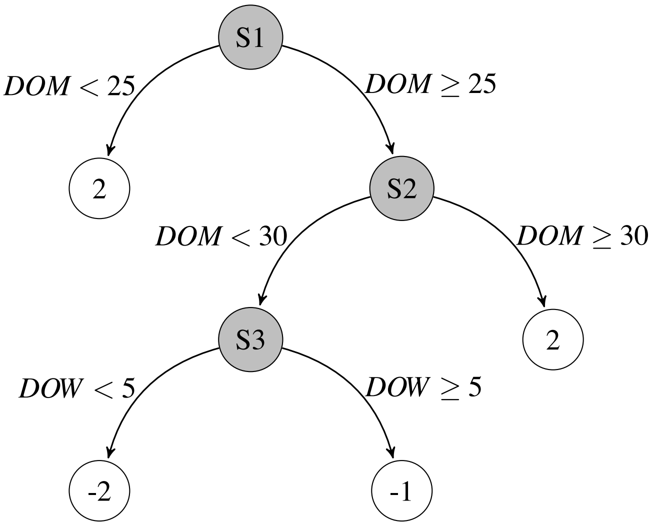

Figure 5: *A basic decision tree. DOM = Day-of-month; DOW = Day-of-week.*

different decision trees (Ho, 1998; Breiman, 2001). An ensemble methodology is able to construct a predictive model by integrating multiple trees in what is called a decision forest (Dietterich, 2000). Decision trees split the input space in subsets based on the value of features such as the day-of-month and day-of-month. In the example in Figure 5, for days comprised between the 25th (node S1) and the 29th of each month (node S2) occurring on Friday (node S3), the predicted cash flow is -1.

Recent examples of time-series forecasting using random forests can be found in Booth, Gerding and Mcgroarty (2014), Zagorecki (2015) and Salas-Molina et al. (2017). Summarizing, random forests are used to forecast variables based on an ensemble of different trees. Unlike linear regression, random forests allow to capture (if any) more complex relationships between predictor variables allowing us to identify possible non-linearities in the underlying cash flow process represented by our sample data sets.

In Table 5, we summarize results only for data sets that can be labelled as trivial because neither the linear model nor the non-linear model were able to significantly beat the baseline forecaster. As described in Algorithm 1, we label time-series as linear when lower normalized squared errors are obtained using the regression model. Similarly, we label time-series as non-linear when lower errors are obtained using the random forest model. In addition, we test the significance of the difference in performance between regression and random forest models. When p-values from these tests are below 0.05, we consider that sample errors for the linear and the non-linear model are significantly different.

From those time-series in which the absence of correlation could no be rejected (see Ljung-Box test at Table 3), 20 out of 24 were labelled as trivial. On the other hand, only 6 of them were labelled as non-linear according to our cross-validated definition. As mentioned above, these results depend on the selected forecasting models. Instead

Table 5: *Results of the test for non-linearity. Reg NSE = Regression normalized squared error; RF NSE = Random forest normalized squared error.*

Id	Reg NSE	RF NSE	Statistic	p-value	Triviality	Linearity
1	0.99	1.00	26	< 0.05	Non-Trivial	Linear
3	0.99	1.01	8	< 0.05	Non-Trivial	Linear
4	1.00	1.01	0	< 0.05	Non-Trivial	Linear
7	0.81	0.83	0	< 0.05	Non-Trivial	Linear
9	0.90	0.93	3	< 0.05	Non-Trivial	Linear
13	0.86	0.88	13	< 0.05	Non-Trivial	Linear
14	0.76	0.77	45	< 0.05	Non-Trivial	Linear
16	0.85	0.86	64	0.13	Non-Trivial	Linear
18	0.86	0.88	63	0.12	Non-Trivial	Linear
19	0.96	0.94	182	< 0.05	Non-Trivial	Non-linear
20	0.99	0.98	209	< 0.05	Non-Trivial	Non-linear
23	0.78	0.79	78	0.33	Non-Trivial	Linear
24	0.73	0.79	0	< 0.05	Non-Trivial	Linear
25	0.77	0.81	21	< 0.05	Non-Trivial	Linear
28	0.84	0.90	0	< 0.05	Non-Trivial	Linear
29	0.99	0.99	30	< 0.05	Non-Trivial	Linear
30	0.73	0.80	5	< 0.05	Non-Trivial	Linear
33	0.94	0.93	166	< 0.05	Non-Trivial	Non-linear
34	0.97	0.95	172	< 0.05	Non-Trivial	Non-linear
39	0.96	0.96	36	< 0.05	Non-Trivial	Linear
42	0.88	0.87	149	0.11	Non-Trivial	Linear
43	0.99	0.96	210	< 0.05	Non-Trivial	Non-linear
48	1.01	0.99	191	< 0.05	Non-Trivial	Non-linear
49	0.63	0.65	7	< 0.05	Non-Trivial	Linear
51	0.77	0.80	0	< 0.05	Non-Trivial	Linear
52	0.94	0.94	116	0.70	Non-Trivial	Linear

of claiming that random forests are able to better capture non-linear patterns than alternative models, we encourage practitioners to consider additional combinations of both linear and non-linear models.

One may assume either linearity or non-linearity from the results of our non-linearity test, but it is important to analyse the robustness of these results to both the presence of outliers and the impact of other data transformations.

4. The impact of data transformations

In this section, we aim to analyse the impact of outlier treatments on noise reduction, as intended, and on information loss, as an undesirable effect. We also study the influence of Box-Cox data transformations on the results of our cross-validated non-linearity test. Detection and treatment of outliers is an ongoing issue in data mining (Rousseeuw and Leroy, 1987; Hodge and Austin, 2004). An outlier is an observation that appears to significantly deviate from other members of the sample in which it occurs (Grubbs, 1969). Outliers arise due to changes in systems, measurement errors or simply due to deviations from average activity. It is also important to note that an outlier may also be the most interesting part of the data.

On the one hand, from the set of cash flow time-series labelled as trivial, some of them may be labelled as non-trivial after removing outliers as a way of noise reduction. On the other hand, from those data sets labelled as non-trivial, some of them may be labelled as trivial due to the information loss produced by the treatment. We here measure the effect of removing outliers on the prediction error using time-series cross validation for different thresholds of outlier replacement. For each data set, we progressively identify as outliers cash flow observations greater than 5, 4, and 3 times the standard deviation in a training set with the 80% oldest observations. We replace outliers with a linear interpolation of the previous and the posterior observation and we proceed as detailed in Algorithm 1 to cross-validate triviality and linearity. The results from this analysis are summarized in Table 6, where global performance after treatments is assessed by averaging noise (error) reduction. Note that some time-series in Table 6 are not present in Table 5 because outlier treatment and Box-Cox transformation produced a improvement in accuracy.

By following this procedure, we identify data sets 5, 10, 17, 32, 44 and 54 (6 out of 28), initially labelled as trivial that, after outlier treatment, can be labelled as non-trivial due to noise reduction. Similarly, data sets 4 and 48 that were initially labelled as non-trivial can be labelled as trivial after outlier treatment due to information loss. If we measure noise reduction by the error reduction and information loss by the error increase, then we can assess the impact of outlier treatment. Following this approach, we obtained mixed results for non-trivial data sets after outlier treatment: an average noise reduction of 22%, and an average information loss of 14%. It is important to recall that unexpected observations are often the most interesting part of the data to predict, e.g., when the goal is to forecast unusual but genuine cash flows.

Non-linearity and outliers are closely linked. Indeed, Castle and Hendry (2012) hypothesized that non-linear functions can align with outliers, causing functions to be considered relevant spuriously, which can be detrimental for generalizing and forecasting. If this hypothesis is correct, the relative forecasting ability of a linear model in comparison to a non-linear model would increase as the presence of outliers in a training set is reduced. From the set of time-series finally labelled as non-trivial, data sets 33, 34 and 54, initially labelled as non-linear changed their labels to linear. Surprisingly, data sets 17, 18, 23, 25, 39, 44 and 49 (7 out of 30), could be labelled as non-linear after outlier treatment. Except for data sets 17 and 44, in all cases there was information loss, i.e., error increase, suggesting that non-linear models can deal better with information loss.

We also considered a Box-Cox transformation to analyse if this kind of data transformation may influence the results from our cross-validated non-linearity test. From the set of non-trivial data sets we compare linearity labels, first, after outlier treatment,

Table 6: *Results of the test for non-linearity after outlier treatment and Box-Cox transformation. Changes in labels are marked with* *.

Id	Triviality	After outliers		After outliers and Box-Cox	
		Linearity	Noise reduction	Linearity	Noise reduction
1	Non-Trivial	Linear	0.00	Non-linear*	−0.01
3	Non-Trivial	Linear	0.02	Non-linear*	0.00
5	Non-Trivial	Non-linear	0.40	Non-linear	0.41
7	Non-Trivial	Linear	−0.10	Linear	−0.13
9	Non-Trivial	Linear	−0.04	Linear	−0.04
10	Non-Trivial	Non-linear	0.46	Non-linear	0.47
13	Non-Trivial	Linear	−0.18	Linear	−0.21
14	Non-Trivial	Linear	−0.05	Linear	−0.07
16	Non-Trivial	Linear	−0.18	Linear	−0.17
17	Non-Trivial	Non-linear*	0.71	Non-linear	0.71
18	Non-Trivial	Non-linear*	−0.20	Non-linear	−0.20
19	Non-Trivial	Non-linear	−0.03	Non-linear	−0.04
20	Non-Trivial	Non-linear	−0.02	Non-linear	−0.02
23	Non-Trivial	Non-linear*	−0.22	Non-linear	−0.22
24	Non-Trivial	Linear	−0.20	Linear	−0.06
25	Non-Trivial	Non-linear*	−0.26	Non-linear	−0.25
28	Non-Trivial	Linear	−0.05	Linear	−0.04
29	Non-Trivial	Linear	0.07	Non-linear*	0.00
30	Non-Trivial	Linear	−0.06	Linear	−0.04
32	Non-Trivial	Non-linear	0.18	Non-linear	0.21
33	Non-Trivial	Linear*	−0.12	Linear	−0.11
34	Non-Trivial	Linear*	0.12	Linear	0.09
39	Non-Trivial	Non-linear*	−0.02	Linear*	−0.01
42	Non-Trivial	Linear	−0.23	Linear	−0.14
43	Non-Trivial	Non-linear	0.04	Non-linear	0.03
44	Non-Trivial	Non-linear*	0.48	Non-linear	0.82
49	Non-Trivial	Non-linear*	−0.56	Non-linear	−0.61
51	Non-Trivial	Linear	−0.03	Linear	−0.03
52	Non-Trivial	Linear	0.01	Linear	0.03
54	Non-Trivial	Linear*	0.17	Linear	0.17
Average performance			0.00		0.02

and second, after outlier treatment and Box-Cox transformation as described in equation (1). In addition, we compare information loss computed as the difference between the sum of errors of the linear and non-linear forecasting models before and after the outlier treatment. A positive value means noise reduction or error reduction while a negative value means information loss or error increase. Results from Table 6 show a similar performance after Box-Cox transformation since the change in labels occurs in data sets with similar linear and non-linear noise reduction.

Table 7 shows the impact of outlier treatment and data transformation on the classification of time-series derived from our cross-validated non-linearity summarized in

Table 7: *Number of time-series data sets and their labels after transformation. OT=Outlier treatment; DT=Data transformation.*

Label	Raw data	After OT	After OT and DT
Trivial	28	24	24
Non-trivial	26	30	30
-Linear	20	17	15
-Non-linear	6	13	15

Table 6. The high number of trivial data sets may be caused by the general inherent randomness of cash flows. In addition, an increase in the number of time-series classified as non-trivial after treatments suggests a positive impact. However, non-linear models seem to obtain a higher benefit from treatments. First, outlier treatment produced a small improvement in non-triviality but also an outstanding increase in non-linearity. Second, after both outlier treatment and Box-Cox data transformation, resulted in similar results but with better performance for non-linear models.

It is worth mentioning that global performance in terms of error reduction remained unchanged after outlier treatment and slightly improved after data transformation (see Table 6). Thus, we conclude that: (i) common data transformations had little impact on our time-series in terms of linearity and accuracy; and (ii) outlier treatment and Box-Cox transformation were unable to transform non-linear into linear cash flows.

5. Concluding remarks

Small and medium companies contribute to a high percentage of all enterprises, value added and employment in Europe. In this paper, we provide a complete empirical study of the statistical properties of daily cash flows based on 54 real-world time-series for small and medium companies. To the best of our knowledge, this work is the most comprehensive empirical study on daily cash flows so far in terms of the range of statistical properties considered, and also in terms of the number and the length of the data sets. Particularly, we focus on the implications of our analysis for forecasting due to its key role in cash management. An additional contribution of this work is to make all data publicly available online for further research.

5.1. Summary of findings

Our results show that the extended hypotheses of normal, stationary and uncorrelated cash flows are hardly present in our cash flow data set. Thus, we conclude that the standard assumptions of normality, stationarity and uncorrelatedness that have been extensively used in the cash management literature must be verified before the deployment of any cash management model based on them. We do not claim that these results can be generalized to all small and medium companies. Indeed, we hypothesize that companies with a larger number of daily cash flows may be closer to satisfy these usual assumptions than small and medium companies. This hypothesis represents an interesting subject of future research and we here set the path to this research by providing the methods to verify such hypothesis. We also highlight that common solutions to non-normality and non-stationarity such as data transformation and differencing produce little benefit when applied to our data sets, with the risk of losing important information on extreme cash flows. Alternative and more complex data transformations are nevertheless an option to consider in further research to achieve Gaussian cash flows.

In an attempt to discover the attributes of actual-world cash flows, we also studied the presence of non-linearity. To this end, we proposed a new simple test for non-linearity with two main advantages in comparison to alternative approaches. First, our test does not assume any non-linear function. Second, it is based on time-series cross validation to increase robustness and to avoid overfitting. It is important to note that our cross-validated definition of non-linearity depends on the alternative models considered, one linear and another one non-linear.

Our cross-validated non-linearity test labelled as either trivial, linear or non-linear our cash flow data set after outlier treatment resulting in an important increase in the number of data sets labelled as non-linear. After both outlier treatment and Box-Cox transformation, linearity could not be achieved and non-linear models showed more robust. However, the overall impact of data transformations on forecasting performance was limited. The application of our test to provide further evidence on these topics when using alternative cash flow data sets represents a natural extension of our work.

5.2. Implications

Our results raise questions about two common assumptions in cash flow time-series since we found that: (i) the usual assumption of normality, absence of correlation and stationarity is hardly present; and (ii) common data transformations such as outlier treatment and Box-Cox transformation have little impact on normality and linearity. Contrary to the rather common assumption in the literature, these results imply that neither it is always possible to achieve a Gaussian, white-noise and linear time-series through data transformation nor it is always desirable due to information loss. In this paper, we are interested in models that produce forecasts for a wide range of planning horizons. Thus, autoregressive and linear models should be considered as an initial step towards more realistic ones which are better adapted to real cash flow situations. The results from our cross-validated test for non-linearity suggest that non-linear models represent a justifiable alternative for time-series forecasting. Moreover, since our test is both model and outlier dependent, a promising line of future work is the integration of outlier treatment in the test itself in an attempt to assess noise reduction or information loss.

We claim that a number of preliminary steps are necessary in cash flow forecasting before model selection: (i) statistical summary including normality, correlation and stationarity; (ii) impact of data transformations such as outlier treatment and Box-Cox transformation; (iii) non-linearity test to determine the type of model which is expected to deliver a better performance. This process is not limited to daily cash flow, since it can also be applied to any other time-series data set when cross-validation is required.

Acknowledgements

Work partially funded by projects Collectiveware TIN2015-66863-C2-1-R (MINECO/ FEDER) and 2014 SGR 118.

References

Arratia, A., Cabana, A. and Cabana, E.M. (2016). A construction of continuous-time arma models by iterations of ornstein-uhlenbeck processes. *SORT-Statistics and Operations Research Transactions*, 1, 267–302.

Baccarin, S. (2009). Optimal impulse control for a multidimensional cash management system with generalized cost functions. *European Journal of Operational Research*, 196, 198–206.

Baumol, W.J. (1952). The transactions demand for cash: An inventory theoretic approach. *The Quarterly Journal of Economics*, 66, 545–556.

Booth, A., Gerding, E. and Mcgroarty, F. (2014). Automated trading with performance weighted random forests and seasonality. *Expert Systems with Applications*, 41, 3651–3661.

Box, G.E. and D.R. Cox (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 211–252.

Box, G.E. and Jenkins, G.M. (1976). *Time Series Analysis: Forecasting and Control*, revised ed. Holden-Day.

Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.

Broomhead, D. and Lowe, D. (1988). Multivariable functional interpolation and adaptive networks. *Complex Systems*, 2, 321–355.

Castle, J.L. and Hendry, D.F. (2010). A low-dimension portmanteau test for non-linearity. *Journal of Econometrics*, 158, 231–245.

Castle, J.L. and Hendry, D.F. (2012). Automatic selection for non-linear models. In *System Identification, Environmental Modelling, and Control System Design*, pp. 229–250. Springer.

Cavaliere, G. and Xu, F. (2014). Testing for unit roots in bounded time series. *Journal of Econometrics*, 178, 259–272.

Cleveland, R.B., Cleveland, W.S., McRae, J.E. and Terpenning, I. (1990). Stl: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6, 3–73.

Constantinides, G.M. and Richard, S.F. (1978). Existence of optimal simple policies for discounted-cost inventory and cash management in continuous time. *Operations Research*, 26, 620–636.

da Costa Moraes, M.B., Nagano, M.S. and Sobreiro, V.A. (2015). Stochastic cash flow management models: A literature review since the 1980s. In *Decision Models in Engineering and Management*, pp. 11–28. Springer International Publishing.

Dietterich, T.G. (2000). Ensemble methods in machine learning. In *Multiple Classifier Systems*, pp. 1–15. Springer.

Ducharme, G.R. and Lafaye de Micheaux, P. (2004). Goodness-of-fit tests of normality for the innovations in arma models. *Journal of Time Series Analysis*, 25, 373–395.

Duchesne, P., Lafaye de Micheaux, P. and Tagne Tatsinkou, J. (2016). Estimating the mean and its effects on neyman smooth tests of normality for arma models. *Canadian Journal of Statistics*, 44, 241–270.

Emery, G.W. (1981). Some empirical evidence on the properties of daily cash flow. *Financial Management*, 10, 21–28.

Ewing, B.T. and Thompson, M.A. (2007). Symmetric and asymmetric stationarity tests of productivity. *The Engineering Economist*, 52, 355–365.

Franses, P.H. and Van Dijk, D. (2005). The forecasting performance of various models for seasonality and nonlinearity for quarterly industrial production. *International Journal of Forecasting*, 21, 87–102.

Gormley, F.M. and Meade, N. (2007). The utility of cash flow forecasts in the management of corporate cash balances. *European Journal of Operational Research*, 182, 923–935.

Gregory, G. (1976). Cash flow models: a review. *Omega*, 4, 643–656.

Grubbs, F.E. (1969). Procedures for detecting outlying observations in samples. *Technometrics*, 11, 1–21.

Ho, T.K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20, 832–844.

Hodge, V.J. and Austin, J. (2004). A survey of outlier detection methodologies. *Artificial Intelligence Review*, 22, 85–126.

Homonoff, R. and Mullins, D.W. (1975). *Cash Management: an Inventory Control Limit Approach*. Lexington Books.

Hornik, K., Stinchcombe, M. and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2, 359–366.

Horváth, L., Kokoszka, P. and Rice, G. (2014). Testing stationarity of functional time series. *Journal of Econometrics*, 179, 66–82.

Hyndman, R. and Athanasopoulos, G. (2013). *Forecasting: Principles and Practice*. http://otexts.org/fpp/. Last accessed: 2017-05-01.

Hyndman, R.J. and Khandakar, Y. (2008). *Automatic Time Series for Forecasting: the Forecast Package for R*.

Kantz, H. and Schreiber, T. (2004). *Nonlinear Time Series Analysis*, Volume 7. Cambridge University Press.

Keenan, D.M. (1985). A tukey nonadditivity-type test for time series nonlinearity. *Biometrika*, 72, 39–44.

Lee, T.-H., White, H. and Granger, C.W. (1993). Testing for neglected non-linearity in time series models: A comparison of neural network methods and alternative tests. *Journal of Econometrics*, 56, 269–290.

Lilliefors, H.W. (1967). On the kolmogorov-smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 62, 399–402.

Ljung, G.M. and Box, G.E. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65, 297–303.

Makridakis, S., Wheelwright, S.C. and Hyndman, R.J. (2008). *Forecasting Methods and Applications*. John Wiley & Sons.

Marathe, R.R. and Ryan, S.M. (2005). On the validity of the geometric brownian motion assumption. *The Engineering Economist*, 50, 159–192.

Miller, M.H. and Orr, D. (1966). A model of the demand for money by firms. *The Quarterly Journal of Economics*, 80, 413–435.

Miller, T.W. and Stone, B.K. (1985). Daily cash forecasting and seasonal resolution: Alternative models and techniques for using the distribution approach. *Journal of Financial and Quantitative Analysis*, 20, 335–351.

Moriña, D., Puig, P. and Valero, J. (2015). A characterization of the innovations of first order autoregressive models. *Metrika*, 78, 219–225.

Muller, P., Caliandro, C., Peycheva, V., Gagliardi, D., Marzocchi, C., Ramlogan, R. and Cox, D. (2015). *Annual Report on European SMEs*.

Nedović, L. and Devedžić, V. (2002). Expert systems in finance: a cross-section of the field. *Expert Systems with Applications*, 23, 49–66.

Neyman, J. (1937). "Smooth test" for goodness of fit. *Scandinavian Actuarial Journal*, 1937, 149–199.

Penttinen, M.J. (1991). Myopic and stationary solutions for stochastic cash balance problems. *European Journal of Operational Research*, 52, 155–166.

Pindado, J. and Vico, J. (1996). Evidencia empírica sobre los flujos de caja. Un nuevo enfoque en su tratamiento. *Revista Española de Financiación y Contabilidad*, 25, 497–517.

Premachandra, I. (2004). A diffusion approximation model for managing cash in firms: An alternative approach to the miller-orr model. *European Journal of Operational Research*, 157, 218–226.

Ramsey, J.B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 31, 350–371.

Ross, S.A., Westerfield, R. and Jordan, B.D. (2002). *Fundamentals of Corporate Finance* (Sixth ed.). McGraw-Hill.

Rousseeuw, P.J. and Leroy, A.M. (1987). *Robust Regression and Outlier Detection*, Volume 589. John Wiley & Sons.

Royston, J. (1982). An extension of Shapiro and Wilk's W test for normality to large samples. *Applied Statistics*, 31, 115–124.

Salas-Molina, F., Martin, F.J., Rodríguez-Aguilar, J.A., Serrà, J. and Arcos, J.L. (2017). Empowering cash managers to achieve cost savings by improving predictive accuracy. *International Journal of Forecasting*, 33, 403–415.

Srinivasan, V. and Kim, Y.H. (1986). Deterministic cash flow management: state of the art and research directions. *Omega*, 14, 145–166.

Stone, B.K. and Miller, T.W. (1987). Daily cash forecasting with multiplicative models of cash flow patterns. *Financial Management*, 16, 45–54.

Stone, B.K. and Wood, R.A. (1977). Daily cash forecasting: a simple method for implementing the distribution approach. *Financial Management*, 6, 40–50.

Torabi, H., Montazeri, N.H. and Grané, A. (2016). A test for normality based on the empirical distribution function. *SORT-Statistics and Operations Research Transactions*, 1, 55–88.

Venables, W.N. and Ripley, B.D. (2013). *Modern Applied Statistics with S-PLUS*. Springer Science & Business Media.

Wilcoxon, F., Katti, S. and Wilcox, R.A. (1970). Critical values and probability levels for the wilcoxon rank sum test and the wilcoxon signed rank test. *Selected Tables in Mathematical Statistics*, 1, 171–259.

Zagorecki, A. (2015). Prediction of methane outbreaks in coal mines from mulivariate time series using random forest. In *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, pp. 494–500. Springer.

Zhang, G., Patuwo,B.E. and Hu, M.Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14, 35–62.