Analyses of spare parts demand correlations with econophysics approaches

Human and Engineered Environmental Studies
47-126851 Chi Zhang (2014.9 Master graduates)
Name of advisor Ass. Prof. Yu Chen

In this study, we proposed a series effective analysis methods for spare parts demand correlations. First, we applied Random Matrix Theory (RMT) to find the genuine correlations in auto spare parts demand by means of investigating the statistics of correlation coefficients and the eigenvalue distribution of the correlation matrix by comparing with a series of random matrices. We found that the distribution of components for the eigenvectors corresponding to the eigenvalues outside the RMT prediction reflects the stable deviations in time, which indicated the correlated demand in different groups. Second, we applied the network analysis to visualize the topological structure of auto spare parts. We found that the threshold method in network analysis successfully detect the change of clusters of correlated demand, and the structure of demand factors could be built by applying the minimum spanning tree method to auto spare parts kinds network. The branches of MST reflect certain demand factors, the inner layer contains complex demand factors whereas the outer layer indicates simple demand factors.

Key words: Econophysics, Random Matrix Theory, Network Analysis, Spare Parts Demand, Demand Correlations.

1 Introduction

1.1 Background

Croston's method[1] and its modified versions by Boylan[2] and Synteton[3] are the most widely used approaches for intermittent spare parts demand forecasting. Croston assumed inter-arrival times as a Bernoulli process, making the intervals between demand independent and identically distributed (iid) geometric, and he further assumed the demand size to be iid Normal. In the follow-up related researches, some used log transformations of both demands and inter-arrival times to amend the sample space for Croston's model including negative values which is inconsistent with reality that demand is always non-negative, some modified the inter-arrival times to other distributions, and the latest study[4] implied that it is more closer to Negative Binomial Distribution or Stuttering Poisson. To improve the prediction accuracy, some recent studies[5,6] applied the hierarchical processing to change item level times into group level time series by classification methods such as logical regression, and still consider the demand to occur as a non-stationary Bernoulli process.

However all aforementioned models are based on simple exponential smoothing (SES), the forecast variances are all increasing over time.

1.2 Motivation

While analyzing the fluctuations in time series data of auto spare parts demand, we have observed peculiar characteristics such as volatility clustering and heavy-tailed leptokurtic distribution, the so-called stylized facts found in financial time series. These characteristics not only reflect explicit deviations from some previous research assumptions, but also motivate us to use unconventional analysis methods to explain those deviations and to solve the challenges in analyses of auto spare parts demand related to demand forecasting and inventory management.

A latest study of applying RMT to analysis of financial market[7] shows various stylized facts in the distributions of stock return, which are similar to our findings in auto spare parts demand data. Furthermore, these results implied that the demand of auto spare part is not a set of weak-convergence variables, correlations among different parts demand.

Another recent study successfully applied RMT to identify the corresponding business sectors in financial data by means of analyzing the statistics of eigenvalue and eigenvectors calculated from the correlation matrix [8]. This is an empirical proof that RMT is a suitable method to systematically analyze the correlated demand in auto spare parts. To our knowledge, RMT has not been applied for spare parts demand data analysis. Those aforementioned interesting characteristics observed from spare parts data motivate us to use the econophysics approaches to establish a series analysis methods for auto spare parts demand correlations.

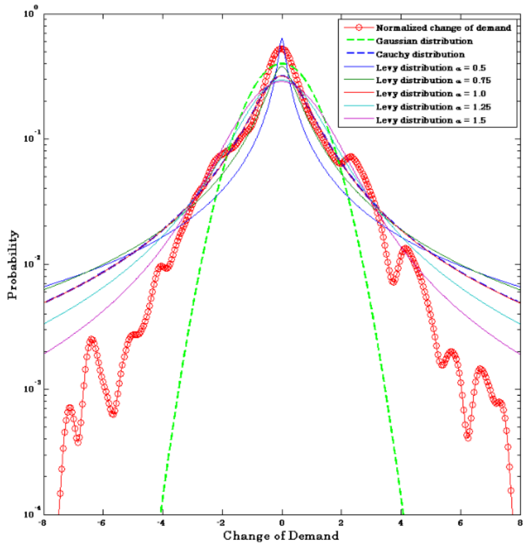

Fig. 1 Heavy-tailed leptokurtic distribution of auto

spare parts demand change

2 Research Goals

Our aim is to propose a series effective analysis methods for spare parts demand correlations. First, to find the genuine correlations of spare parts demand on the basis of full investigation of eigenvalues and eigenvectors calculated from a serious of correlation matrix. Second, to explain the mechanism of the emerged demand correlations by means of applying alternative network approaches.

3 Econophysics approaches

3.1 Statistics of correlation coefficients

We first calculate the demand change of auto spare parts by

$$G_i(t) = S_i(t) - S_i(t-1), \quad (1)$$

where $S_i(t)$ denotes the demand of material i. Since different materials have varying levels of demand scale, we define a normalized change as

$$g_i(t) = \frac{G_i(t) - EG_i(t)}{\sqrt{varG_i(t)}}, \quad (2)$$

where $EG_i(t)$ is the expectation of $G_i(t)$, $\sqrt{varG_i(t)}$ is the standard deviation of $G_i(t)$. In matrix notation, the correlation matrix can be written as

$$\boldsymbol{C} = \frac{1}{T}\boldsymbol{G}\boldsymbol{G}^{\boldsymbol{T}}, \quad (3)$$

where $\boldsymbol{G}$ is an $N \times T$ matrix, and $\boldsymbol{G}^{\boldsymbol{T}}$ denotes the transpose of $\boldsymbol{G}$. Thus, we construct a random correlation matrix $\boldsymbol{R} = (1/T)\boldsymbol{A}\boldsymbol{A}^{\boldsymbol{T}}$ with zero mean and unit variance, where $\boldsymbol{A}$ is a $N \times T$ random normal matrix with zero mean and unit variance. Fig. 2 shows the comparison of correlation matrix elements.

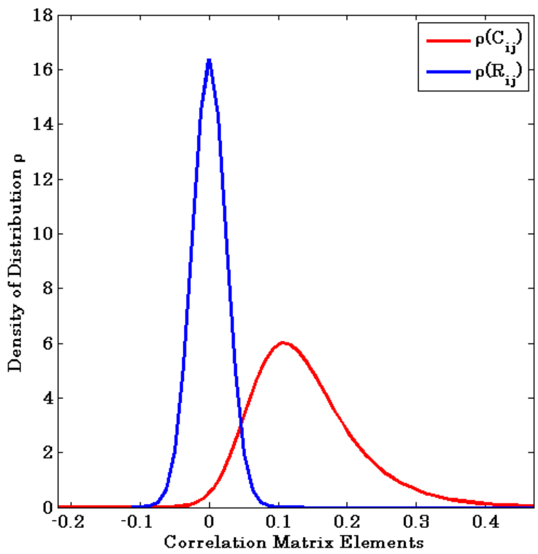

Fig. 2 Statistics of correlation coefficients. N=1516, mean value of correlation matrix elements is 0.1399, variance is 0.0066, skewness is 1.7259 calculated from $\sum_{i,j=1}^{N}(C_{ij} - EC_{ij})^n / [N^2(varC_{ij})^{n/2}]$ when n=3, kurtosis is 12.9280 when n=4.

Fig. 2 initially showed the bulk of correlation matrix elements of auto spare parts demand is not random. Next, we test the eigenvalue distribution of the correlation matrix.

3.2 Eigenvalue distribution of the correlation matrix

According to random matrix theory, For matrix $\boldsymbol{R}$, when $N, T \to \infty$ and $Q \equiv T/N > 1$, the probability density $\rho(\lambda)$ of eigenvalues λ of the random correlation matrix is given by

$$\rho(\lambda) = \frac{Q}{2\pi}\frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}, \quad (4)$$

where the maximum and minimum eigenvalues of $\boldsymbol{R}$, respectively, given by

$$\lambda_\pm = 1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}}. \quad (5)$$

The distribution of eigenvalues λ of the correlation matrix $\boldsymbol{C}$ appeared deviations from RMT scope as shown in Fig. 3.

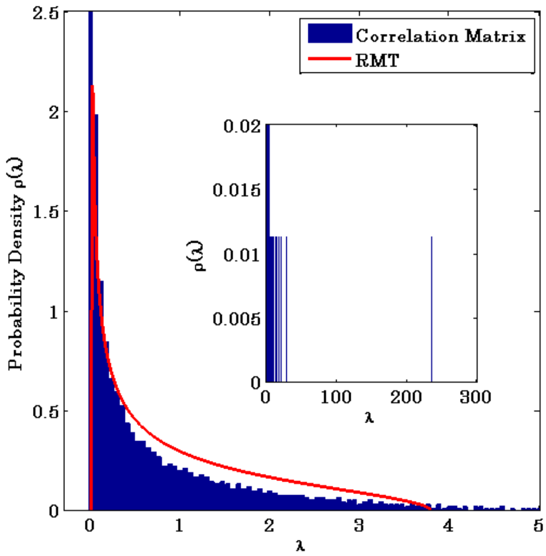

Fig. 3 Eigenvalue distribution of the correlation matrix. The largest eigenvalue λ_{1516}=234.4, $\lambda_{100} < \lambda_- < \lambda_{101}$, $\lambda_{1469} < \lambda_+ < \lambda_{1470}$, λ_-=0.0029, λ_+=3.7858.

We note several eigenvalues outside the RMT upper bound λ_+, this is a clear distinction between correlation matrix $\boldsymbol{C}$ and $\boldsymbol{R}$. Although most eigenvalues of $\boldsymbol{C}$ are included in RMT range $[\lambda_-, \lambda_+]$, we need to further test the statistics of eigenvalue distribution inside RMT range $[\lambda_-, \lambda_+]$ to investigate whether the deviation is stable in time. First, we calculate the unfolding eigenvalue $\xi_i(\lambda_i)$ by

$$\xi_i(\lambda_i) = \tilde{N}\int_{\lambda_-}^{\lambda_i}\frac{Q}{2\pi}\frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}d\lambda, \quad (6)$$

where $\tilde{N}$ is the amount of eigenvalues inside $[\lambda_-, \lambda_+]$. Defining $s_i = \xi_{i+1} - \xi_i$ is the nearest-neighbor spacing, if the correlation matrix have the properties of Gaussian orthogonal ensemble, then the distribution of s_i is given by

$$\rho_{GOE}(s) = \frac{\pi s}{2} e^{-\frac{\pi}{4}s^2}, \quad (7)$$

the distribution of unfolding eigenvalues of correlation matrix $\boldsymbol{C}$ also appeared the some deviation from RMT prediction as shown in Fig. 4. Although the eigenvalue spacing distribution of $\boldsymbol{C}$ deviates from the GOE prediction, general distribution feature like distribution shape remains the same. A small amount of correlation information inside RMT bound cannot reject the hypothesis that the deviations outside the RMT upper bound are the genuine and stable correlation information in auto spare parts demand.

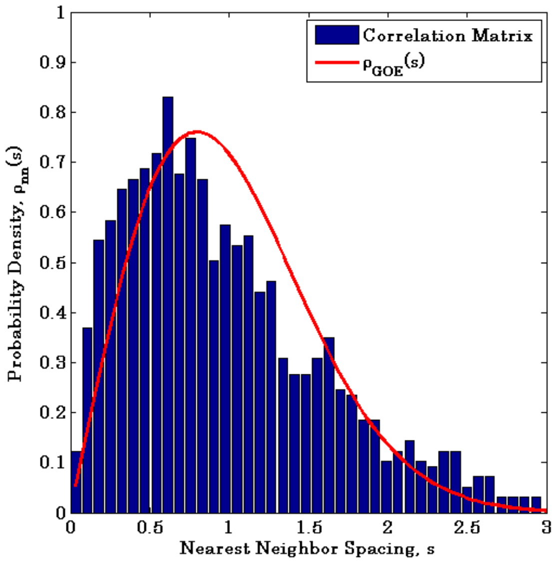

Fig. 4 Nearest-neighbor spacing distribution of the unfolded eigenvalues of the correlation matrix

3.3 Statistics of eigenvectors

The deviations of eigenvalue distribution of correlation matrix from RMT prediction imply that these deviations should also be discovered in the statistics of the corresponding eigenvector components. According to RMT, if a random correlation matrix is constructed by zero mean and unit variance, then the eigenvector components $u_j^{(i)}$ obey Gaussian distribution

$$\rho\left(u_j^{(i)}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u_j^{(i)2}}{2}}, \quad (8)$$

We observed the grouping of spare parts and car series in the eigenvector components corresponding to the eigenvalues outside the upper RMT bound, and there is no grouping occurs in the eigenvector components inside RMT bound, the grouping of spare parts emerges in the eigenvector components corresponding to the eigenvalues outside the lower RMT bound whereas the grouping of car series did not emerge.

Although we achieve the stable demand correlations from the eigenvector components corresponding to the eigenvalues outside the RMT prediction, and do observe the grouping of spare parts in those eigenvector components, the corresponding demand factors and the structure of demand correlations are still vague. We will apply the network analysis to solve the two issues.

4 Network approaches

4.1 Topological structure of spare parts networks

A previous study on resolving the structure of financial markets[9] shows that the network analysis can efficiently and legibly interpret the connection between different data clusters.

We first calculate the demand change of auto spare parts by Eq. (1). Since different materials have varying levels of demand scale, we define a normalized change as Eq. (2).

In matrix notation, the correlation matrix can be written as Eq. (3), this equal-time cross-correlation matrix is computed with elements C_{ij} which are limited to the domain$[-1,1]$. For the spare parts coefficients $C_{ij} = 1$ (absolute value) corresponds to perfect correlation between different spare parts kinds and $C_{ij} = 0$ corresponds to no correlation.

The network of spare parts kinds using the threshold methods is defined as follows. The set of vertices (V) of the network defined by the set of spare parts kinds. A certain threshold θ $(-1 \leq \theta \leq 1)$ will be given and an undirected edge connecting the vertices i and j will be added if C_{ij} is greater than or equal to θ. Therefore different values of threshold generate networks with the same set of vertices but different sets of edges. The edges (E) in the graph $G = (V, E)$ which displayed the network of spare parts kind are given by

$$E = \begin{cases} e_{ij} = 1, i \neq j, and\ C_{ij} \geq \theta, \\ \quad e_{ij} = 0, i = j. \end{cases} \quad (9)$$

We construct spare parts correlation networks of kinds at different thresholds (θ in range 0.1 to 0.9). The Fruchterman-Reingold layout is used to find clusters in all these networks.

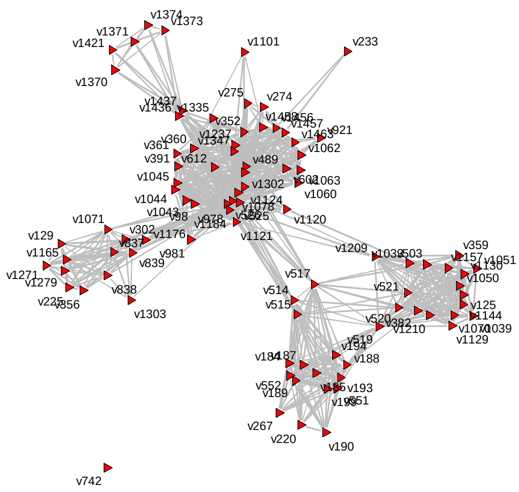

Fig. 5 The network generated by the threshold at 0.3

In Fig. 5, the top right group of spare parts related to the components of u1506, this group mainly contains bumper bar and fender, both of them highly related to the traffic accidents. However, the spare part v1421 in the same cluster is a condenser which is hardly correlated to traffic accident based on intuitive inference. To distinguish between a noise and a hidden demand factor, we then apply the MST method to find the structure of demand factors and the corresponding parts of a certain demand factor.

4.2 Minimum spanning tree

We construct the network of spare parts kinds by using the metric distances $d_{ij} = \sqrt{2(1-C_{ij})}$ forming and $N \times N$ distance matrix with elements vary between 0 and 2.

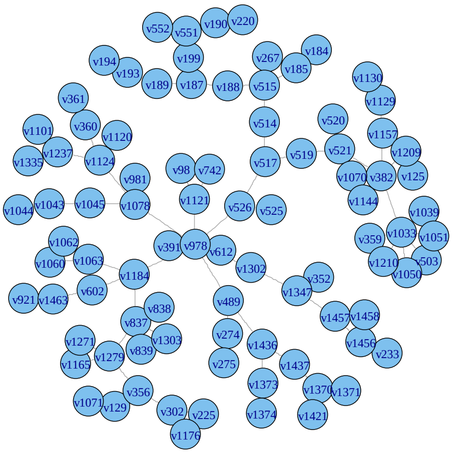

Fig. 6 Minimum spanning tree of auto spare parts

In Fig. 6, we found that those parts near the center of MST are related to the eigenvector components corresponding to the largest eigenvalue. Since the largest eigenvalue contains the most complex demand correlation information, the outer layer branches of MST should reflect certain demand factors. We can locate the v1421 in the most outer layer of MST, this result imply that v1421 have the same demand factor as the other parts on the same branch. Actually, the failure of v1421 often happens, and v1421 could be easily damaged in a traffic accident due to it locates at the front part of engine space, as a consequence, v1421 usually replace with other spare parts related to traffic accidents. A hidden demand factor successfully revealed by MST.

5 Summary

We applied RMT to analyze the auto spare parts demand. We found that RMT successfully met our analysis needs for finding the genuine correlations in auto spare parts demand. In particular, we investigated the statistics of correlation coefficients and the eigenvalue distribution of the correlation matrix by comparing with a series of random matrices. Our main findings is that the distribution of components for the eigenvectors corresponding to the eigenvalues outside the RMT prediction reflects the stable deviations in time, which indicated the correlated demand in different groups.

To further understand and to verify the analysis results of RMT, we applied the network approaches to visualize the topological structure of a simplified network with 91 kinds of auto spare parts. We found that the threshold method in network analysis successfully detect the change of clusters of correlated demand. Our main findings is that the structure of demand factors could be built by applying the minimum spanning tree method to auto spare parts kinds network. The branches of MST reflect certain demand factors, the inner layer contains complex demand factors whereas the outer layer indicates simple demand factors.

Combined with the two steps analyses, we proposed the general analysis methods for spare parts demand correlations.

References

1) T. R. Willemain, C. N. Smart, J. H. Shockor and P. H. DeSautels, "Forecasting intermittent demand in manufacturing: a comparative evaluation of Croston's method," *International Journal of Forecasting,* vol. 10, pp. 529-538, 1994.
2) J. E. Boylan and A. A. Syntetos, "Forecasting for inventory management of service parts," in *Complex System Maintenance Handbook*, vol. 20, 2007, pp. 479-506.
3) R. H. Teunter and B. Sani, "On the bias of Croston's forecasting method," *European Journal of Operational Research,* vol. 194, pp. 177-183, 2009.
4) A. A. Syntetos, M. Z. Babai and N. Altay, "On the demand distribution of spare parts," *International Journal of Production Research,* vol. 50, pp. 2101-2117, 2012.
5) S. Moon, C. Hicks and A. Simpson, "The development of a hierarchical forecasting method for predicting spare parts demand in the Korean Navy—A case study," *Int. J. Production Economics,* vol. 140, pp. 794-802, 2012.
6) S. Moon, A. Simpson and C. Hicks, "The development of a classification model for predicting the performance of forecasting methods for naval spare parts demand," *Int. J. Production Economics,* vol. 143, pp. 449-454, 2013.
7) A. Chakraborti, I. M. Toke, M. Patriarca and F. Abergel, "Econophysics review: I. Empirical facts," Quantitative Finance, vol. 11, pp. 991-1012, 2011.
8) V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral and T. Guhr, "Random matrix approach to cross correlations in financial data," Phys. Rev. E., vol. 65 , p. 066126, 2002.
9) S. Kumar and N. Deo, "Correlation and network analysis of global financial indices," *Physical Review E,* vol. 86, p. 026101, 2012.