OPERATIONS AND SUPPLY CHAIN MANAGEMENT
Vol. 4, No. 2, May 2011, pp. 85-89
ISSN 1979-3561|EISSN 1979-3871

Examination of the Interrelation Among the Price of the Fuel, the Cost of Transport Freight and the Profit Margin

A. Papana[1], A. Papana[2], M. Dagiasis[3], D. Folinas[4], E. Diamantopoulos[5]
[1,4,5]Department of Logistics, Alexander Technological Educational Institute of Thessaloniki, Branch of Katerini, Greece
[2]Department of Mathematics, Cleveland State University, USA
[3] Phalanx Logistics Solutions, Logistics and Supply Chain, Cleveland, USA
[1]papanagel@yahoo.gr, [2]ariadnipapana@hotmail.com, [3]mdagiasis@hotmail.com, [4]dfolinas@gmail.com,
[5]epdiamantopoulos@yahoo.gr

Abstract

The investigation of the variation of the prices of fuel and their interrelation with the cost of hauling freight and how this effects profit margins for various freight transports (ie. carriers), is of great interest for all players in the supply chain (suppliers, manufacturers and carriers). Although supply chains may vary from country to country, the end result will help establish trends that can be used to further identify links between various markets both domestically and internationally. This paper aims to examine the interrelation of the daily price of fuel, the cost of hauling freight from various US geographic locations and the profit margin that various freight haulers may have thereof. Data for the three aforementioned variables will be analyzed over a period of 4 years. The conclusions of this work can be helpful in the cost accounting of provided logistic services from a company concerning the prices of oil, aiming to help the company check the costs and the margin of profit. The conclusion of this work will help various players in the supply chain determine how the chosen variables affect net profit.

Keywords: *logistics, correlation, dependence, mutual information.*

1. Introduction

The correlation among the daily price of fuel in USA, the cost of hauling freight from various US geographic locations and the profit margin of freight haulers is examined in this work. The collection of data is from a logistics company based in Cleveland, Ohio, United States, compiled between the dates January of 2007 to the current date. The number of data used for each variable is 1034. The findings are limited to include only East Coast destination points since fuel prices and freight charges vary greatly between geographic regions. In order to collect the data concerning the cost of hauling freight from various US geographic locations and the profit margin of freight haulers, the total distance in miles for the designated shipment and the revenue for the trucking company to haul this shipment are used. The cost per gallon for diesel fuel on the date of the particular shipment is used in order to calculate the gross profit of the truck after paying for fuel and determine the cost per mile. A universal figure was used, 5 miles to the gallon, which is a standard throughout the trucking and logistics industries in North America. On average a semi-truck will be able to attain 5 miles of driving per gallon of fuel.

*Corresponding Author

The aim of this work is to investigate the type of dependence (linear/ nonlinear) among the pairs of the three examined variables, i.e. the daily price of fuel, the cost of hauling freight from various US geographic locations and the profit margin of freight haulers. Further, causal measures are used in order to differentiate among direct and indirect dependence. These findings are specific to the domestic trucking portion of the logistics industry. In order to identify the relationship of the three variables, the linear and partial linear correlation coefficients of Pearson are estimated, and respectively the mutual information and the conditional mutual information of the variables are estimated in order to measure the global (linear and nonlinear) correlation of them. The necessity to use of more complex measures such as mutual information instead of the standard linear ones is also investigated. The findings of the above research could help 3rd Party Logistics and transportation companies that operate to other countries / regions, etc. to study the relationship between price of fuel, cost and profit margin so as to control more effective their operation costs.

2. Methodology

A relationship refers to the correspondence between two variables. The relationship between two variables may be linear or/ and nonlinear, while may be direct or indirect. There are several terms to describe the different kind of patterns one might find in a relationship. If there is no relationship at all, then the knowledge of the values on one variable provides no information about the values on the other variable. A positive relationship indicates that high (low) values on one variable are associated with high (low) values on the other, respectively. On the other hand, a negative relationship implies that high values on one variable are associated with low values on the other.

2.1. Definition of correlation measures

The correlation is one of the most common and most useful statistics (Galton, 1885). A correlation is a single number that describes the degree of relationship between two variables. Pearson derived in the 1880s the analytic product-moment formula of the "population correlation coefficient" between two variables, which is defined as the covariance of the two variables divided by the product of their standard deviations

$$\rho = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y} \tag{1}$$

(Stigler, 1989). Substituting the estimates of the covariance and variances based on a sample gives the Pearson's linear correlation coefficient

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \tag{2}$$

The index *r* takes values in [-1, 1]; if *r* is negative, we have a negative relationship while if it's positive, then the relationship is positive. Although several authors have offered guidelines for the interpretation of a correlation coefficient, e.g. (Cohen, 1988), this depends on the context and purposes of each study. In general, values of *r* close to zero ($|r|<0.1$) indicate the lack of correlation, if $0.1<|r|<0.3$ then there is a week correlation, and if $0.7<|r|<1$ then there is a strong correlation among the variables.

In order to determine whether the non-zero correlations are direct or indirect, causal measures should be used. Partial correlation coefficient measures the linear correlation between two variables after removing the effect of other variables. Its use aims in finding spurious correlations and revealing hidden correlations. For three variables X, Y and Z, the partial correlation coefficient $r_{xy.z}$ between variables X and Y adjusted for the third variable Z may be computed in terms of simple correlation coefficients as

$$r_{XY.Z} = \frac{r_{XY} - r_{XZ} r_{YZ}}{\sqrt{(1 - r_{XZ}^2)(1 - r_{YZ}^2)}} \tag{3}$$

The index $r_{xy.z}$ takes again values between -1 and 1. If X and Y are both uncorrelated with Z, then $r_{XY.Z} = r_{XY}$.

Mutual information is a measure of the dependence between two random variables which measures how much knowing one of these variables

reduces our uncertainty about the other (Shannon, 1948). Mutual information is a very general measure of dependence as it makes no assumptions regarding the nature of the relationship that exists between the variables; it does not assume a linear, nor functional correlation but only a predictable relationship. For its computation uses the joint and marginal distributions. Mutual information of two discrete variables X and Y is defined as

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p_{X,Y}(x,y) \log \frac{p_{X,Y}(x,y)}{p_X(x)\, p_Y(y)} \quad (4)$$

where $p_{X,Y}(x,y), p_X(x), p_X(y)$ are the joint and marginal probability functions. Mutual information always takes non negative values, as it yields the maximum information we can obtain from a variable X if we know Y. If the variables X and Y are statistically independent, then it is zero.

For inferring causal relation, conditional mutual information is used. The conditional mutual information characterizes the net dependence between two variables *X* and *Y* without the possible influence of another variable, *Z*. For discrete random variables *X*, *Y* and *Z*, conditional mutual information is defined as

$$I(X;Y \mid Z) = \sum_{z \in Z} p_Z(z) \sum_{y \in Y} \sum_{x \in X} p_{X,Y|Z}(x,y \mid z) \log \frac{p_{X,Y|Z}(x,y \mid z)}{p_{X|Z}(x \mid z) p_{Y|Z}(y \mid z)} \quad (5)$$

in terms of the marginal and joint conditional probability mass functions. Conditional mutual information takes always non-negative values. For *Z* independent of *X* and *Y*, $I(X;Y \mid Z) = I(X;Y)$.

2.2. Testing the significance of the correlation measures

In order to determine the probability that the observed correlations occurred by chance, a significance test should be conducted. The null hypothesis H_0 examined is that there are no correlations under the alternative hypotheses of existence of correlations (two-tailed test). When the distribution of the statistic/ measure under H_0 is known analytically, the rejection region is at the tails of the distribution according to a given significance level α. In case of the Pearson's correlation coefficient, the statistic $t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$ follows the Student distribution t_{N-2} with *N*-2 degrees of freedom (N is the number of data) and thus critical values for the standard significance level α=0.05 are known. H_0 is rejected if $|t| > t_{N-2;a/2}$. The *p*-value of the test represents the probability of error that is involved in accepting our observed result as valid. If the p-value of the test is small (smaller than α) then the correlation is significantly different from zero.

In case of the nonlinear measures, the null hypothesis H_0 for the statistical test is that the two examined variables are independent. The discriminating statistic is a single number estimate of a characteristic of the data and its variation is such that it allows us to decide whether data are consistent with H_0 or not. Here, the discriminating statistics used are the mutual information and the conditional mutual information. As the distribution of these statistics is generally not known, therefore it is formed through Monte Carlo simulation from the values of the statistics computed on an ensemble of *M* surrogate data consistent with H_0. If the observed values of the measure are not consistent with this null hypothesis, then one may claim that H_0 is wrong, and significant correlations exist. The surrogate data are generated by random permutations of the values of the variables. To decide for the rejection of H_0, the significance *s* for the two discriminating statistics is computed. If q_0 is the statistic from the original data and $q_1, \ldots, q_M$ from the surrogate data sets then the significance is $s = |q_0 - \overline{q}_s| / \sigma_q$, where $\overline{q}_s$ is the mean of and is their standard deviation. Assuming that follows the standard normal distribution, significance of suggests the rejection of the null hypothesis H_0 at α = 0.05. In this case, the *p*-value of the test is given as $p = 2(1 - \Phi(s))$, where Φ is the cumulative distribution function of the standard normal distribution.

3. Results

The values of the linear correlation coefficients among all pairs of variables are presented in Table 1. The Pearson's correlation coefficient of the price of the gas (X) and the cost of hauling freight (Y) is r_{XY} =-0.2934 indicating the weak negative interrelation of the variables. The profit of freight

haulers (Z) has also a negative correlation to the cost of gas (r_{XZ} = -0.4077), as should be expected. Finally, the cost of hauling freight is positively correlated to the profit of freight haulers (r_{YZ}=0.5976). The *p*-values from the two tailed significance test are zero and therefore the estimated linear correlations are significant.

Table 1. Pearson's correlation coefficient and Pearson's partial correlation coefficients among the three variables, i.e. price of gas (*X*), cost of hauling freight (*Y*) and profit of freight haulers (*Z*).

Pearson's correlation	r_{xy} = -0.2934	r_{xz} = -0.4077	r_{yz} = 0.5976
Pearson's partial correlation	$r_{xy,z}$ = -0.0680	$r_{xz,y}$ = -0.3032	$r_{yz,x}$ = 0.5475

The partial correlation coefficient among the price of gas and the cost of hauling freight is slightly negative indicating the weak direct negative dependence of the two measures. The *p*-value of the significance test is smaller than the significance level (α=0.05) indicating that the partial correlation is significantly different from zero. The fact that is much smaller than $r_{XY.Z}$ indicates the existence of both direct and indirect dependence among the variables. The results for $r_{XZ.Y}$ and $r_{YZ.X}$ are in agreement with r_{XZ} and r_{YZ}, as far as the type of the correlation (negative/ positive) and its strength, i.e. giving similar values, indicating the existence of direct dependence among the variables.

Results from the nonlinear measures are in consistent with the linear ones. The price of the gas (X) and the cost of hauling freight (Y) seem to be the least correlated variables, while the cost of hauling freight (Y) and the profit of freight haulers (Z) seem to be the most correlated ones (see Table 2). All estimated correlations are significant as the *p*-values of the significant tests are zero.

Table 2. As in Table 1, but for mutual information and conditional mutual information

Mutual information	I(X;Y)=0.2878	I(X;Z)=0.4061	I(Y;Z)=0.6474
Conditional Mutual information	I(X;Y|Z)=0.3639	I(X;Z|Y)=0.4622 I	(Y;Z|X)=0.6783

4. Conclusions

Pearson's correlation coefficient distinguishes between positive and negative correlations, while the mutual information does not. However, mutual information can detect global correlation and not only linear ones. Further, the Pearson correlation is bounded by the mutual information, but for cases of numerical or statistical errors. For the analyzed dataset, there seems to be almost a one-to-one correspondence between the Pearson correlation and the mutual information (see Tables 1 and 2). Most of all, it means that investigations using Pearson correlation coefficient as a measure of similarity for such data measurements are justified. The correlations between simultaneously measured values of the daily price of fuel, the cost of hauling freight from various US geographic locations and the profit margin of freight haulers, are essentially linear.

The investigation of the correlation between the daily price of the fuel and the cost of the transport freight may not present such interest; however the interdependence between the daily price of the fuel and the profit margin of transport haulers and between the cost of the transport freight and the profit margin of transport haulers present great interest. The type of correlations (negative/ positive) among these pairs of variables were are rational and were expected. The fact the correlations are found to be mainly linear could be used for further investigation to make predictions for these variables. The need for extension of this investigation to include more related variables (e.g. cost of insurance, other expenses of logistics companies and transport haulers) will also be considered in future works.

References

Cohen, J., (1988) *Statistical power analysis for the behavioral sciences* (2nd ed.*)*, New York University, New York.

Galton, F., (1985) Regression towards mediocrity in hereditary stature. *Journal of Anthropologic Institute* 15, pp. 246-260.

Shannon, C.E., (1948) A Mathematical Theory of Communication. *Bell System Technical Journal* 27, pp. 379–423.

Stigler, S.M., (1989) Francis Galton's Account of the Invention of Correlation. *Statistical Science* 4 (2), pp. 73-79.

Dr Angeliki Papana possesses a PhD in nonlinear time series analysis from the Aristotle University of Thessaloniki, Greece. For the last two years she holds teaching posts in Technical Institutions in Macedonia (ATEI Thessalonikis, TEI of Kavala and Serres) teaching Mathematics and Statistics. My main fields of interest are the linear and nonlinear analysis of time series, dynamical systems and chaos, analysis and prediction of physiological time series. The topics of the current research work is the detection of correlations and interdependencies of variables of interacting systems, the quantification of the coupling strength and the detection of the direction of the causal effects.

Dr Ariadni Papana possesses a PhD in Statistics from Case Western Reserve University, Cleveland, Ohio, USA, and a B.S. in Applied Mathematics from Aristotle University of Thessaloniki, Greece. Dr Papana is an assistant professor at Cleveland State University, Cleveland, Ohio (August 2008-present) in the department of Mathematics teaching statistics at the undergraduate and graduate level. She also taught various statistical courses (Fall 2002, Summer 2005& 2006, Spring 2004 & 2007) at Case Western Reserve University, Department of Statistics, Cleveland, Ohio and at the Papanas Institute, Thessaloniki (1996-2001). She was awarded with the "Graduate Dean's Instructional Excellence Award", Case Western Reserve University, 2003-2004. Her research interests are in high dimensional genomic data, data mining, dimension reduction, graphical methods, design and analysis of experiments, functional data, variable selection methods and time series data. Dr Papana is a member of the Institute of Mathematical Statistics and American Statistical Association.

Michael M. Dagiasis has a Bachelors degree in Business Administration from The University of Toledo and an MBA from Baldwin-Wallace College in Berea, OH. He has over 12 years of hands on logistics experience, primarily in domestic (USA) transportation. He is currently the CEO and Owner of Phalanx Logistics Solutions, a full-service Third Party Logistics company that was founded in 2005 and is based in Lakewood, OH.

Dr. Dimitris Folinas is an Assistant Professor at the Department of Logistics in the ATEI of Thessaloniki, Greece. He possesses a Ph.D in e-Logistics from the University of Macedonia, Thessaloniki, Greece and a Master of Information Systems from the same Institution. He is the author and co-author of over 90 research publications, and as a researcher he has prepared, submitted and managed a number of projects funded by National and European Union research entities. His research interests include: Logistics and Supply Chain Management, e-Logistics, Supply Chain Integration, Logistics Information Systems and Enterprise Information Systems.

Dr Epaminondas Diamantopoulos possesses a Ph.D in Mathematical Analysis from the Aristotle University of Thessaloniki, Thessaloniki, Greece. He has taught mathematics and statistics in various organizations. He is a member and a reviewer of American Mathematical Society.