The Josai Journal of Business Administration (2017), Vol. 13, No. 1, 1−8


# Energy Consumption Analysis and Forecasting for Quay Crane Operations in Container Terminals

**Meilong Le**[*1]
**Toshitake Kohmura**[*2]

## Abstract

Taking a typical modern container terminal as an example, we developed a regressive model for energy consumption by using a method based on multivariate regression and sampling data. The correlation coefficient matrix shows that multi-collinearity exists widely between variables. Therefore, using a backward linear regression to delete iteratively all unnecessary variables, we obtain a correlation-weakened regression model, Model (1). Then, path analysis is used to explain the mechanics influencing the model. Finally, we use a ridge regression to obtain a more knowledge-consistent forecasting model, Model (2). After obtaining the two models, we compare Model (2) with Model (1) according to the degree of data match and forecasting precision, and conclude that Model (2) is more suitable for forecasting. This result could be instructive for terminal managers to reduce energy consumption.

**Key Words:** QC energy consumption analysis, QC energy consumption forecasting, Backward linear regression, Path analysis, Ridge regression

## I. INTRODUCTION

As international trade has increased rapidly during the past twenty years, the maritime transportation industry has become a major energy consumer. In container terminals, large numbers of huge erected shore-mounted gantry cranes, or simply speaking quay cranes (QC), consume large quantities of electric energy every day. According to the statistics of a typical modern container terminal, the QC energy cost comprised 80.02% of its total operating cost in 2011, 75.16% in 2012, and 75.68% in 2013. Energy consumption in container terminals has attracted much attention and concern. How is this energy consumed? Is there any way to save energy? These are the problems we want to address in this paper. In order to answer these questions, we have first carried out a thorough, on-the-spot investigation of a container terminal over a period of six months. Then, we collected one and a half years' continuous data and filtered out the abnormalities. Finally, using statistical methods to analyze these data, we developed regressive models. Based on these models, we will suggest some energy saving measures.

*1 Nanjing University of Aeronautics and Astronautics, China
*2 Josai University, Japan

# II. APPROACH AND MODELS

Our approach is based on statistical methods which include multivariate regression, path analysis, backward linear regression and ridge regression. Multivariate regression is used in a preliminary regression and correlation analysis; and path analysis is used to reveal how the respective independent variables influence the dependent variable related to the energy consumption; backward linear regression is used to weaken the correlations of variables, and ridge regression is used to obtain a more knowledge-consistent regressive model. By using this approach, not only a series of models are established to better explain the sampling data, but also a final forecasting model usable in practice is obtained.

## *2.1. Variables and Sampling Data*

The related factors of QC electricity consumption have been obtained through interviews of container terminal operators, and our independent observation and investigation during a six-month period. These factors include times loading and unloading hatch-covers, the number of special containers (pin containers) loaded and unloaded, the respective numbers of different sized (20', 40' or 45') containers handled, total operating times and container weights including both cargo and tare. All these factors are regarded as explanatory independent variables and the energy consumption is regarded as a dependent variable in our preliminary model. Since all the liner services repeat monthly, we use one month as our basic time frame measure. Eighteen data groups were collected from various information systems during the observation period from July, 2011 to December, 2012. Two months' worth of abnormal data groups were excluded after we carefully examined their abnormality. The remaining sixteen data groups listed in Table 1 were then used for analysis.

**Table 1 The Sampling Data of Explanatory Variables $X_i$ and Dependent Variable $Y$.**

| NO. | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ | $Y$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 2791 | 910 | 38375 | 112296 | 2468.25 | 3244.46 | 969714.9 | 458097.7 |
| 2 | 2547 | 894 | 36979 | 91096 | 2272.5 | 2803.28 | 926516.8 | 405337.1 |
| 3 | 2768 | 840 | 36735 | 95708 | 2139.75 | 2945.85 | 964154.9 | 406528.6 |
| 4 | 2941 | 736 | 40471 | 90908 | 3512.25 | 2919.97 | 1018370 | 413755.4 |
| 5 | 3226 | 792 | 43576 | 110384 | 3300.75 | 3435.65 | 1136463 | 481567.6 |
| 6 | 3388 | 779 | 46103 | 112610 | 2391.75 | 3402.04 | 1142831 | 466833.4 |
| 7 | 2857 | 761 | 40450 | 99532 | 2216.25 | 2863.04 | 992353.9 | 420307.9 |
| 8 | 2841 | 685 | 45235 | 98452 | 2580.75 | 2939.61 | 1129595 | 431141.1 |
| 9 | 3044 | 619 | 39468 | 106896 | 1617.75 | 3014.77 | 1118617 | 390694.7 |
| 10 | 2646 | 644 | 33859 | 102638 | 2911.5 | 2736.49 | 847251.6 | 382005.3 |
| 11 | 2281 | 525 | 34513 | 77974 | 2396.25 | 2279.03 | 823580.9 | 348796.8 |
| 12 | 3205 | 640 | 52274 | 100334 | 3854.25 | 3305.44 | 1302928 | 438725.8 |
| 13 | 2973 | 601 | 48284 | 92508 | 1993.5 | 2922.59 | 1215193 | 405108.1 |
| 14 | 3007 | 652 | 47769 | 96892 | 2767.5 | 2889.5 | 1206299 | 423887.2 |
| 15 | 3056 | 672 | 49376 | 101398 | 3318.75 | 3002.27 | 1060567 | 460563.9 |
| 16 | 3949 | 691 | 55201 | 132384 | 4036.5 | 3921.79 | 1374197 | 551062.1 |

The denotation of the variables is as follows.

- $Y$ Electricity consumption in the form of electricity cost in RMB or Yuan
- $X_1$ Loading and unloading hatch-cover times. A hatch-cover should be unloaded first while loading or if containers to be unloaded are in the ship's hold, *and it should be loaded before loading and unloading containers on deck.*
- $X_2$ Number of pin containers loaded or unloaded
- $X_3$ Number of 20' containers (TEU) loaded or unloaded
- $X_4$ Number of 40' containers (TEU) loaded or unloaded
- $X_5$ Number of 45' containers (TEU) loaded or unloaded
- $X_6$ Total operating duration in hours
- $X_7$ Container weight in tons

## *2.2. Multi-collinearity between Variables*

Multivariate linear regression is used in order to obtain a correlation model between the dependent variable ($Y$) and its explanatory variables ($X_i$). For the sampling data in Table 1, we obtain the correlation coefficient matrix in Table 2, which shows that two or more independent variables in the regression model are correlated. Based on the correlation coefficient matrix, we judge that multi-collinearity exists widely among those seven explanatory variables and some are even highly correlated. The existence of multi-collinearity may make the model distorted or inaccurate. For example, the correlation coefficient between $X_1$ and $X_6$ is 0.923. From the values of the correlation coefficient between an explanatory variable $X_i$ and the dependent variable $Y$, we can know that the total operating duration has the heaviest effect (0.932) on the electricity cost, the number of 20' containers loaded or unloaded an effect in the forth (0.715) and container weight a smaller effect (0.675).

The multi-collinearity between the explanatory variables $X_i$ is also proved by other indices. The biggest *VIF* ("Variance Inflation Factor") is 44.330, and *the biggest condition number* is 194.170. Normally, the value of *VIF* >10 or condition number >100 indicates that the independent variables are quite highly correlated.

**Table 2 Correlation Coefficient Matrix**

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ | $Y$ |
|---|---|---|---|---|---|---|---|---|
| $X_1$ | 1.000 | .013 | .814 | .835 | .536 | .923 | .857 | .873 |
| $X_2$ | | 1.000 | –.223 | .265 | –.080 | .310 | –.186 | .324 |
| $X_3$ | | | 1.000 | .494 | .561 | .678 | .928 | .715 |
| $X_4$ | | | | 1.000 | .374 | .909 | .562 | .859 |
| $X_5$ | | | | | 1.000 | .516 | .432 | .600 |
| $X_6$ | | | | | | 1.000 | .747 | .932 |
| $X_7$ | | | | | | | 1.000 | .675 |
| $Y$ | | | | | | | | 1.000 |

The correlation coefficients, *VIF* value and condition number indicate that high multi-collinearity exists between those seven explanatory variables $X_i$ though the multivariate linear regression model fits the dependent variable $Y$ highly. In order to eliminate the multi-collinearity, we will use a backward regression and ridge regression.

### *2.3. Backward Regression*

Backward regression is an iterative regression. It first uses all the independent variables in the model to perform the regression, then excludes the variable whose partial regression coefficient is the least significant, and repeats the regression based on the remaining variables to decide which variable should be excluded next. The iterative regression terminates when all the partial regression coefficients are significant. The final model obtained by the backward regression, Model (1), is as follows;

$$Y = 55229.348 + 5.459X_3 + 127.404X_6 - 0.229X_7. \quad (1)$$

The result of the backward regression yields the multiple correlation coefficient 0.947, the adjusted multiple correlation coefficient 0.933, and F = 70.889 >16, the significance level of the model $p = 0.000 < 0.05$ (0.05 is the set value for F-test and t-test). In the coefficient t-test, the partial regression coefficients of the remaining independent variables are all significant. The significance levels are 0.001, 0.000 and 0.002 respectively. The result of multiple correlation coefficient test, F-test and t-test indicates that the matching ability of Model (1) is quite good. The maximum, minimum and mean percentage of the deviations between regression value and actual value of electricity cost is 5.61%, 0.19% and 2.01% respectively. The biggest *VIF* is 8.817, which is smaller than 10. This means that the level of multi-collinearity has apparently decreased.

## III. Analysis of Model Mechanics

Model (1) indicates that the container weight is negatively correlated with the electricity cost. This is not in accord with our past understanding. Why does this occur? The reason is that multi-collinearity still exists though its level has apparently decreased. The *VIF* values are acceptable because they are all less than 10, but the maximum *condition number* is as much as 55.579, more than 30. Because of multi-collinearity, the partial regression coefficient cannot reflect the degree to which the independent variable influences the dependent variable. Actually, it only reflects the combined influence of all the independent variables. For example, if the container weight increases by one unit, the total operating time $X_6$ and loading or unloading 20 feet container number $X_3$ will change accordingly, and their total effect is that electricity cost will decrease by 0.229 Yuan. To explicitly reveal the interaction between independent variables and how an independent variable influences the dependent variable, we adopted a path analysis. The goal of the path analysis is to determine the path coefficients, which consist of the direct path coefficient $q_{iY}$ and the indirect path coefficient $q_{ij} \cdot q_{iY}$, which is equal to the standard partial regression coefficient. The $q_{iY}$ means the direct influence effect of $X_i$ on $Y$, while $q_{ij}$ stands for the indirect influence effect of $X_i$ on $Y$ through the independent variable $X_j$. The $q_{ij}$ is equal to the correlation coefficient $r_{iY}$ multiplied by the standard partial regression coefficient of the variable $X_j$. The direct and indirect path coefficients are listed in Table 3.

It is apparent that the sum of the direct path coefficient of $X_i$ and all the indirect path coefficients of $X_i$ is the same as the correlation coefficient between $X_i$ and $Y$. The positive values of $q_{3y}$ (0.747) and $q_{6y}$ (0.992) mean that $X_3$ and $X_6$ have a direct and positive influence on electricity cost, and the value less than zero of $q_{7y}$ (−0.759) means that $X_7$ has a direct and negative influence on electricity cost. However, the sum of all the indirect effects of $X_7$ through $X_3$

**Table 3 Direct and Indirect Path Coefficients**

| Independent variables | Correlated coefficient $r_{iY}$ | Direct path coefficient $q_{iY}$ | Indirect path coefficients $q_{ij}$ | | |
|---|---|---|---|---|---|
| | | | $X_3 \to Y$ | $X_6 \to Y$ | $X_7 \to Y$ |
| $X_3$ | 0.715 | 0.747 | – | 0.673 | −0.704 |
| $X_6$ | 0.932 | 0.992 | 0.506 | – | −0.567 |
| $X_7$ | 0.675 | −0.759 | 0.693 | 0.741 | – |

(0.693) and $X_6$ (0.741) is more than $q_{7y}$, so the combined effect of $X_7$ on electricity cost is still positive.

To further explain the practical meaning of the path coefficients, we take $X_3$ as an example, then $q_{3Y} = 0.747$ indicates that when $X_3$ increases by one standard deviation, $Y$ (the electricity cost) increases by 0.747 of a standard deviation or 46966.73 × 0.747 = 35084.15 units (YUAN). In other words, if $X_3$ increases by 1 unit (TEU), $Y$ increases by 35084.15 ÷ 6423.76 = 5.46 YUAN. The $q_{36} = 0.673$ means that when $X_3$ increases by one standard deviation, $X_6$ increases accordingly which then results in $Y$ increasing by 0.673 of a standard deviation or 4.92 YUAN eventually. Similarly, when $X_3$ increases by 1, the electricity cost decreases 5.15 YUAN indirectly via $X_7$. In total, when $X_3$ increases by 1 (TEU), the electricity cost increases 5.46 + 4.92 − 5.15 = 5.23 YUAN.

## IV. Model Adjustment

As mentioned above, the partial regression coefficient of container weight ($X_7$) in Model (1) doesn't accord with our past understanding due to multi-collinearity. To change the model to a more knowledge-consistent model, a ridge regression will be used. The essence of the ridge regression is to improve ***the least square fitting*** by eliminating the unbiased property of the method and reducing the accuracy so as to get a more reasonable and reliable regression model. The ridge regression model has a higher tolerance to pathological data. The ridge regression is based on the above result of the independent variables $X_3$, $X_6$ and $X_7$. Figure 1 shows the ridge trace of these three variables. The ridge trace is the locus of the value of standard partial regression coefficient of independent variables under a different ridge parameter k.

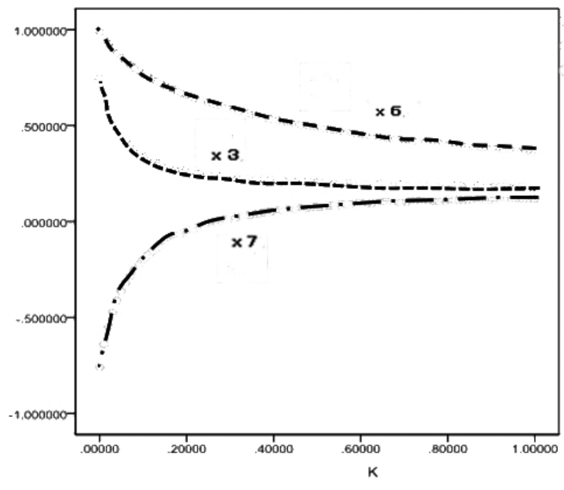


**Figure 1 The Ridge Trace of the Three Independent Variables**

As can be seen from Figure 1, when *k* varies from 0 to 0.5, the standard partial regression coefficient of independent variables $X_3$, $X_6$ and $X_7$ changes quickly, which means that these variables have a significant effect on the electricity cost *Y*. When *k* is bigger than 0.5, each coefficient becomes more stable. Meanwhile, the standard partial regression coefficient of $X_7$ (container weight) moves from negative value to positive value and begins to play a positive role in the electricity cost when *k* is more than 0.28. Therefore, 0.28 is assigned to the parameter *k*, and then we obtain the ridge regression model (2):

$$Y = 120371.2869 + 1.7328X_3 + 76.8957X_6 + 0.00151X_7 \qquad (2)$$

The multiple correlation coefficient of this model is 0.848, which indicates that the matching degree of this model is quite good. The maximum, minimum and mean percentage of deviation of the regression value from the actual value of the electricity cost is 8.08%, 0.70% and 3.25%, respectively. More important is the result that the regression coefficient of $X_7$ (container weight) is $0.0015 > 0$, which shows that the model is more consistent with our knowledge.

# V. Model Assessment

The model is assessed from the two aspects of the matching degree and the forecasting accuracy.

## *5.1. Comparison of the Two Models by Matching Degree*

Good matching means high interpretative ability of sampling data. The multiple correlation coefficients of Model (1) and Model (2) are 0.947 and 0.848, and the mean percentages of matching variation are 2.01% and 3.25%. From these two indices, we see that Model (1) has a higher matching degree. The curves of the regression value are shown in Figure 2.

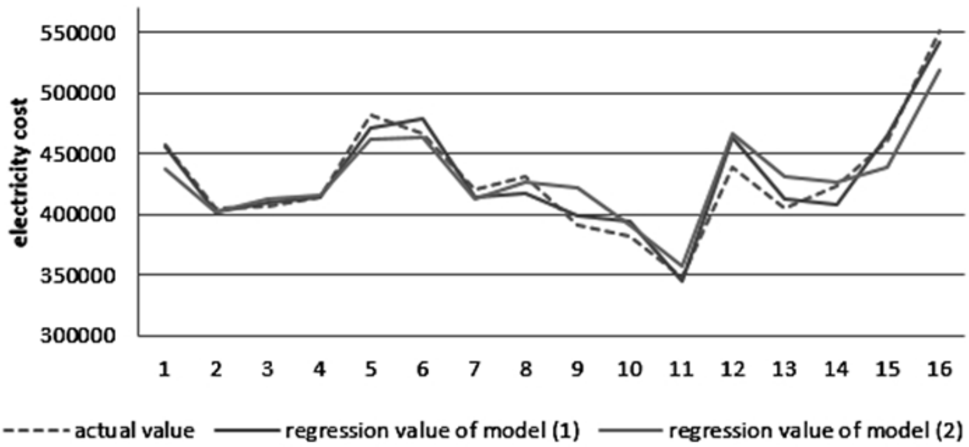


**Figure 2 Matching Curves of the Regression Values**

## *5.2. Comparison of the Two Models by Forecasting Accuracy*

Forecasting accuracy is used to assess the forecasting ability. It can be reflected by the two aspects of forecasting precision and stability of the model. Forecasting precision can be numerated by the mean difference of forecasted value A from actual value P, while stability reflects the volatility of the deviation. In this paper, the forecasting precision is measured by the mean deviation percentage of regression value from actual value, and the stability is evaluated by the standard deviation. The respective mathematical expression is presented in the equations (3) to (5). Here, $i$ denotes the $i$th sampling data and $n$ denotes the total number of sampling data.

Mean deviation $$M_D = \frac{\sum_{i=1}^{n}|A_i - P_i|}{n} \quad (3)$$

Mean deviation percentage $$M_P = \frac{\sum_{i=1}^{n}\frac{|A_i - P_i|}{A_i}}{n} \quad (4)$$

Standard deviation $$D = \sqrt{\frac{\sum_{i=1}^{n}(|A_i - P_i| - M_D)^2}{n}} \quad (5)$$

In order to test the model, we deliberately leave out the data from January to June, which will be used as the actual data for comparison. By calculating the predictive values in Models (1) and (2), we compare them with the actual values for the according month. We obtained the comparative results below.

**Table 4 Comparing Result of Evaluated Values for Forecasting Accuracy**

| Model | $M_D$ | $M_P$ | $D$ |
|---|---|---|---|
| Model (1) | 43866.02 | 9.34% | 76609.27 |
| Model (2) | 23328.49 | 4.92% | 37490.77 |

Apparently, the forecasting precision of Model (2) is higher than that of Model (1). The mean deviation percentage of Model (2) is 4.92%, while that of Model (1) reaches as much as 9.34%. And the standard deviation $D$ also tells us that in terms of volatility of the deviation of forecasting value, Model (2) is much better than Model (1).

In general, the matching degree of Model (1) is quite good and the matching variation (2.01%) is quite small, but when it comes to predictive ability, the performance of Model (1) is quite low. Considering both the matching and forecasting abilities, we conclude that Model (2) is more reasonable.

## VI. Conclusion

Taking a typical modern Chinese container terminal as an example, we examined the QC energy consumption problem. Firstly, by an on-the-spot investigation, we collected data on loading and unloading hatch-cover times, loading and unloading pin container numbers, the TEU number handled, the FEU number handled and the 45 feet container number handled, the total operating times and container weights as influence factors, and collected the relative data as sampling data. Secondly, regarding all the influence factors as explanatory variables and energy cost $Y$ as the dependent variable, we performed multivariate regression and obtained a correlation coefficient matrix that includes correlation coefficients between $Y$ and all explanatory variables. Because of multi-collinearity between variables, we used backward regression to delete iteratively all unnecessary variables and obtained the final regression model. Then we used the path analysis to explain the mechanics influencing the model. Finally, we used the ridge regression to get a more knowledge-consistent forecasting model. After obtaining a new model, we compared the two models by using two indices which stand for matching degree and forecasting precision, and concluded that Model (2) is more suitable for forecasting. This could be instructive for terminal managers to assist them in reducing energy consumption. For example, the terminal could choose more skilled QC operators and follow a more precise schedule to shorten the total operating times. Furthermore, using bigger TEU instead of smaller sized containers is helpful in lowering energy consumption.

Of course, energy consumption is related to the equipment used and the operating mode. The QC system referred here is made by ZPMC, which has greater than a 70% worldwide market share for quaycranes. The operating mode is the QC-truck-YC process mode. If your terminal uses other QC systems or operating modes, the present model should be modified to reflect those differences.

**Acknowledgement**

This work was supported by Jiangsu Province Natural Science Foundation under Grant No. BK20151479 and by Nanjing University of Aeronautics and Astronautics Funds [No. NZ2016109]. Special thanks to Ningbo Daxie Merchants International Terminal Company, Limited for their support of this investigation and collecting data.

**References**

[1] Farrell, S. (2012), "The ownership and management structure of container terminal concessions", Maritime Policy and Management, Vol. 39, Issue 1, pp 7−26.

[2] Hsu, W., Yu, H., Huang, S. (2015), "Evaluating the service requirements of dedicated container terminals: a revised IPA model with fuzzy AHP", Maritime Policy and Management, Vol. 42, Issue 3, pp 1−17.

[3] Abdel-aal, R. E., Al-grani, A.Z. (1997), "Forecasting monthly electric energy consumption in eastern Saudi Arabia using univariate time-series analysis", Energy, Vol. 22, No. 11, pp 1059−1069.

[4] Crompton, P., Wu, Y. (2005), "Energy consumption in china: past trends and future directions", Energy Economics, Vol. 2005, No. 27, pp 195−208.

[5] Dietmair, A., Verl, A. (2009), "Energy consumption forecasting and optimization for tool machines", Energy, Vol. 2009, No. 3, pp 62−67.

[6] Hamzacebi, C. (2007), "Forecasting of Turkey's net electricity energy consumption on sectoral bases", Energy Policy, Vol. 2007, No. 35, pp 2009−2016.

[7] Rudolf, J. F, William, J. W, Ping, S.(2012), "Regression analysis: statistical modeling of a response variable", New York: Academic Press.

[8] Wang, H.W., Huang, H.J. (2008), "The study on measurement by using principle components regression", Journal of Beijing University of Aeronautics and Astronautics, Vol. 36, No. 4, pp 661−664.

[9] Yang, N., (2004), "The unique role of ridge regression analysis in solving multicollinearity", Statistics and Decision, Vol. 2004, No. 3, pp 14−15.