**Addressing the Problem of Simultaneity in Cargo Shippers’ Port Utility Models: An Econometric Analysis**

Donatus .E. Onwuegbuchunam1, Geoffrey .U. Ugwuanyim2 and Kenneth O. Okeke1

1Department of Maritime Management Technology, School of Management Technology,

2Department of Statistics, School of Physical Science. Federal University of Technology, Owerri 1Email: [donafutow@yahoo.com](mailto:donafutow@yahoo.com).

**Abstract**

The traditional method of modelling shipper port utility behaviour has been the Random Utility models. Such models and their variants have consistently ignored the possibility of obtaining biased parameter estimates due to simultaneity. In this paper, we demonstrate how interrelationship among variables that describe shipper port utilization model can lead to the problem of simultaneity. Using port level shipment data from terminal records of Nigeria ports Authority, we show how this problem can be eliminated using Two-Stage Least Squares techniques. Our post estimation test results confirm the robustness of this technique.

**Keywords: ports, shippers, simultaneity, utilization, Two-Stage Least Squares.**

**Introduction**

Port utility refers to the satisfaction a shipper derives from choosing a particular port and using it for shipment making purposes. This satisfaction is a perceived one and can be measured in utils. But in considering the actual choice (i.e. port) and the shipments made, we can extend this theory to include his level of utilization of a particular port. Shipper’s level of port utilization can then be taken as a measure of the total quantity of shipments which he sends or receives from a particular port based on expected utility. This is same as volume of cargo handled in his port of choice. Onwuegbuchunam (2013) and Tongzon (2002) posit that a shipper’s utilization of any port is a function of his perception about attributes related to shipments and ports. These attributes are: the distance of that port to his warehouse or market for his product, crane efficiency and turn-round time at the port. Turn-round time in any port however, is a function of crane efficiency, the volume of cargo discharged or loaded (or both), number of ships berthed at the port, level of cargo handling facilities at the port and berth occupancy. Mathematically, these variables are related in a system of equations. Again, what appears as a regressor in one equation is a regressand in another. This feature gives rise to simultaneous equation bias or simultaneity problem in models describing shipper port utilization behaviour.

In practice, the interrelationship among these variables (attributes) can be better captured and modelled by employing instrumental variable regression technique. The selection and application of suitable instruments eliminate the problem of correlated error terms. In this paper we adopt instrumental variable regression methodology for assessment of port level shippers utility model to account for the problem of simultaneity.

**Objectives**

The objectives of this paper are to:

1. Assess the structural characteristics of cargo shippers’ port utility function
2. Conduct identification of the formulated model.
3. Estimate the structural parameters of the formulated model using Two-Stage Least Squares.
4. Conduct model validation tests (R2, D/W and F-test) on the estimated model in (iii) above.

**Literature Review**

The traditional method for estimating port utility model of port users particularly the shippers has been estimation of quantile choice models especially the Random Utility Models (RUM). The appeal in the widespread use of random utility models stems from its basis on the theory of consumer behaviour. This theory posits that given a set of ports with equally likely attributes; shippers would choose port(s) which maximize their utility or rather reduce their generalized cost of transport (see McFadden, 1974).

The random utility framework has been applied by many researchers in modelling shipper port choice characteristics. Notable among these are: Gonzalez (2000), Malchow (2001), Tuna (2002), Veldman (2003), Lirn (2004), Malchow and Kanafani (2004). Others include: Tongzon (2002, 2009), Song et al (2004), Guy and Urli (2006), De Langen (2007), Magala and Sammons (2008), Panayides, Song (2012) and Onwuegbuchunam (2013). The basic assumption of RUM is that the expected utility which a shipper derives from choosing a particular port for making shipments is dependent on attributes of the shipper, the consignment (shipment) and the port. Thus, we model the probability that a shipper will utilize a particular port given his expectation of the utility regarding the aforementioned attributes. The parameters of the utility model are regarded as fixed. However, recent studies have shown that these parameters vary and the random nature of these parameters has given way to choice modelling based on random coefficient approach. Both random utility and random coefficient modelling frameworks are not efficient in cases of serial correlation. Again, modelling port utilization behaviour of shippers usually involves a system of equations describing variables which are endogenous in one equation and which also appear as exogenous in the other. Thus, we are usually faced with simultaneity problem which cannot be resolved under Random Utility and Random Coefficient modelling framework as these basically involve single equation models.

In contrast to single-equation models, in simultaneous-equation models, more than one dependent or endogenous variable is involved; necessitating as many equations as the number of endogenous variables. A unique feature of simultaneous-equation models is that the endogenous variable in one equation may appear as an explanatory variable in another equation of the system. As a consequence, such an endogenous explanatory variable becomes stochastic and is usually correlated with the disturbance term of the equation in which it appears as an explanatory variable. In this situation, the classical OLS method may not be applied because the estimators thus obtained are not consistent; i.e., they do not converge to their true population values no matter how large the sample size.

In these circumstances, alternative estimating techniques have been developed; such as the Indirect Least Squares (ILS) technique, the Two-Stage Least Squares (2SLS) technique; among others. (Gujurati, 2004). In the subsequent section, we describe the generic simultaneous equation systems and demonstrate how their estimability can be determined by ‘Identification’

**Solving Simultaneous Equations: Derivation of the Reduced Form of the Model:**

The structural form of simultaneous equation model is:



There are ‘M’ equations and ‘M’ endogenous variables denoted Y1, Y2, …, Ym. There are ‘K’ exogenous variables, X1, X2, …,Xk that may include predetermined values of Y1, Y2, …, Ym as well. The first element of Xt will usually be constant 1. Finally, εt1, εt2,…, εtm are the structural disturbances. The subscript‘t’ will be used to index observations, t = 1, 2, …, T.

In matrix terms, the system may be written:

t+

t

Or 

Each column of the parameter matrices is the vector of coefficients in a particular equation, whereas each row applies to a specific variable. The solution of the system of equations determining  in terms of  and  is the reduced form of the model:

t



For this solution to exist, Г must be nonsingular (Greene, 2008).

**Methodology**

Data for this study consists of shipments records from the Traffic Department of Nigeria Ports Authority. It consists of data on port utilization measured as volumes of cargo (in tonnes) shipped/received by shippers at the ports and other port performance measures. The data set covers a period of six years beginning from years 2008 through 2013. Three equations based on a priori relationship and interrelationships existing among the variables were formed. In simultaneous equations with possibly correlated error terms; a dependent variable (endogenous variable) in one equation may turn out to be exogenous variate in the other equation. The equations constitute a simultaneous system of equations describing port utilization by a shipper. Under this circumstance, assumption of independence of error terms in least squares regression modelling is violated hence the application of instrumental variable or 2-stage least squares techniques to address the problem. Shipper’s level of port utilization is equivalent to the total shipment (in tonnes) which he sends or receives from a particular port. This is same as Cargo volume handled. But a shipper’s utilization of any port has been postulated to be a function of his perception of the distance of that port to his warehouse or market for his product (as seeks to minimize costs), crane efficiency (speed of cargo handling) and turn-round time (cargo dwell time) at the port. However, turn-round time of any port is a function of number of ships visits at the port (proxy for level of congestion), level of berthing (terminal) facilities at the port and berth occupancy (a measure of delay). Finally, berth occupancy is dependent on turn round time and draught of the port. The data set consisting of shipments (in tonnes) made in ports, and port characteristics over a period of time are used for analysis of the structural relationship explained above. The data has the following variables as shown below:

Cargo volume (Y1t), Shipper’s Warehouse Distance from the port (X1t);

Turn-round Time at the port (Y2t), Crane Efficiency of the port (X2t);

Frequency of Ship Visits at the port (X3t), Berth facilities in the port (X4t);

Draft of port channel (X5t) and Berth Occupancy (Y3t)

From the foregoing, the complete model for estimation of shipper’s port utilization is as given below:



We need to identify the system of equations (1-3) using the rank and order conditions to establish that solution exists. The process of identification is carried out in tables 1 & 2.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: Model Identification Using the Rank Condition | | | | | | | | | |
|  | **Coefficient of Variables** | | | | | | | | |
| **EQUATION No.** | **1** | **Y1** | **Y2** | **Y3** | **X1** | **X2** | **X3** | **X4** | **X5** |
| 1 |  | 1 |  | 0 |  |  | 0 | 0 | 0 |
| 2 |  |  | 1 |  | 0 | 0 |  |  | 0 |
| 3 |  | 0 |  | 1 | 0 | 0 | 0 | 0 |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Table 2: Equation Identification Using the Order Condition: | | | |
|  | No. of Predetermined Variable Excluded | No. of Endogenous Variables Included Less Than One |  |
|  |
| **EQUATION No.** | **(K – k)** | **(m -1)** | **Identified** |
| 1 | 3 | 1 | Yes |
| 2 | 3 | 2 | Yes |
| 3 | 4 | 1 | Yes |
| Source: Author | | | |

RANK CONDITION:

From table 2, the following (m-1) x (m-1) or 2 x 2 matrices are formed with their determinants calculated as shown below:





Therefore the rank of  thus formed from table 1 = 2

But Rank = m-1, where m = number of equations. But m-1 = 3-1 = 2, i.e. the rank of = 2 (as calculated above). From the foregoing, equation 1 is identified since = m-1 i.e. 2 = 2. Using the same process for identification through rank condition; equation 2 and 3 are also identified. Having established that the system of equations (1-3) is estimable, we now proceed to estimate their structural parameters using the Two-Stage Least Squares technique

**Data Presentation and Estimation**: The descriptive statistics of the sample data are presented in table 3. The parameters of the structural model were estimated using Two-stage least Squares implemented in Stata for Windows statistical software.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Table 3: Descriptive Statistics of Sample Data | | | | | |
| **Variable** | **Obs.** | **Mean** | **Std. Dev.** | **Min** | **Max** |
| Cargo Volume (Y1t) | 74 | 6165.257 | 2949.496 | 3291 | 15906 |
| Turn-round Time (Y2t) | 74 | 21.297 | 5.786 | 12 | 41 |
| Berth Occupancy (Y3t) | 74 | 3.015 | 0.456 | 2.19 | 3.89 |
| Ware House distance (X1t) | 74 | 2.993 | 0.846 | 1.5 | 5 |
| Crane Efficiency (X2t) | 74 | 13.757 | 4.277 | 5 | 23 |
| Ship Visits (X3t) | 74 | 197.297 | 91.837 | 79 | 425 |
| Berth facilities (X4t) | 74 | 7.979 | 0.266 | 7.47 | 8.48 |
| Draught (X5t) | 74 | 187.932 | 22.266 | 142 | 233 |
| Source: Author | | | | | |

In order to get a better result, three major relevant problems must be solved. The first problem is multi-collinearity among these preliminary variables. In this situation of highly correlated variables existing together, the coefficient estimates may change erratically in response to small changes in the model or the data. Table 4 shows the correlation among preliminary variables. As it clearly shows, the correlated coefficient among some variable pairs are higher than 0.8. For examples the variable pair; berth facilities and ship visits are highly correlated with each other (with coefficient above 0.8) while berth facilities and draught of the port also have correlation coefficient above 0.8. Thus when doing regression, it is better to include only one of the two correlated variables so as to avoid multicollinearity.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Table 4: Pairwise Correlation Among All Explanatory Variables | | | | | | |
|  | **Berth\_Occpncy** | **WareHous\_dist** | **Crane\_Effncy** | **ShipVisits** | **Brth\_facil** | **Draught** |
| **Berth\_Occpncy** | 1.000 |  |  |  |  |  |
| **WareHous\_dist** | **-0.378\*** | 1.000 |  |  |  |  |
|  | [-0.001] |  |  |  |  |  |
| **Crane\_Effncy** | **-0.509\*** | **0.662\*** | 1.000 |  |  |  |
|  | [0.000] | [0.000] |  |  |  |  |
| **ShipVisits** | **-0.829\*** | **0.475\*** | **0.609\*** | 1.000 |  |  |
|  | [0.000] | [0.000] | [0.000] |  |  |  |
| **Brth\_facil** | **-0.753\*** | **0.481\*** | **0.655\*** | **0.866\*** | 1.000 |  |
|  | [0.000] | [0.000] | [0.000] | [0.000] |  |  |
| **Draught** | **-0.696\*** | **0.516\*** | **0.727\*** | **0.835\*** | **0.948\*** | 1.000 |
|  | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] |  |
| *Correlation coefficient in bold asterisks, p-values in p-values in parenthesis.* | | | | | | |

The second problem is simultaneity problem as previously demonstrated. This problem will be taken care of by estimation applying the two-stage least squares techniques. The third problem originated from the dataset itself. Since in our dataset, the 74 observations represent 74 successive yearly data concerning each variable, these adjacent observations are too similar than those that would be expected under independence. As a result, autocorrelation may occur which can make independent variables more significant than they may really be through smaller standard errors (s.e) for the beta coefficient. These anomalies can be eliminated following the steps outlined in figure 1.

Figure 1: Regression Anomalies and Methods for Correction

Check Durbin Watson

Autocorrelation

Correlate and VIF

Multicollinearity

Two-stage least sq. R

Endogeneity

In summary, in our case, the shipper port utilization determinants, multicollinearity, endogeneity, as well as autocorrelation may seriously affect our statistical conclusions.

Steps Followed In Conducting The Two-Stage Least Squares Regression:

1. Regress endogenous variables against all the exogenous variables and save the predicted values of the endogenous variables.
2. Regress original equations, replacing endogenous explanatory variables with their predicted values.

**Results:**

In the table 4, the two-stage least squares regression results are presented. Only the turn-round time variable significantly explains the volume of cargo a shipper handles in a particular port. Therefore the structural coefficient of this variable and the intercept term are significantly different from zero at α = 0.05. Although the coefficient of determination (R2) indicates a poor fit of model to the data (21.1%), this may be due omission of some variable that completely determine cargo volume handled. The F-statistic has a significant p-value (see model fitting information) and indicates that the independent variables jointly explain variations in the Cargo volume (the dependent variable). However, changes in exogenous variable, turnround time are explained by level of berth facilities in the port. The variable berth facility has a structural parameter of value -4.090, which is significant. This model has coefficient of variation of 47% which is higher than that obtained for cargo volume equation. The F-statistic is also significant and implies a better fit of the model.We find also from the regression output that turnround time at the port is the significant variable that explains variations in Berth Occupancy. The coefficient of variation for this equation is however very low and may be accounted by the few explanatory variable considered.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table 4: Structural Parameters of the Port Utility Models (2-SLS) | | | | |
| **Exogenous V’s/Coefficient** | **Estimate** | **Std. Err.** | ***t*-stat** | **P>|t|** |
| **CargoVola** | | | |
| TurnRoundT: | -335.213 | 93.146 | **-3.600** | **0.000** |
| WareHous\_dist: | -698.313 | 494.917 | -1.410 | 0.160 |
| Crane\_Effncy: | 44.490 | 119.585 | 0.370 | 0.710 |
| Constant: | 14782.570 | 3262.242 | 4.530 | 0.000 |
|  | **TurnRoundTa** | | | |
| CargoVol: | 0.000 | 0.001 | -0.220 | 0.823 |
| Berth\_Occpncy: | -9.694 | 20.209 | -0.480 | 0.632 |
| ShipVisits: | -0.030 | 0.063 | -0.480 | 0.632 |
| Brth\_facil: | -19.662 | 4.806 | **-4.090** | **0.000** |
| Constant: | 215.224 | 98.929 | 2.180 | 0.031 |
|  | **Berth\_Occpncya** | | | |
| TurnRoundT: | 0.100 | 0.043 | **2.310** | **0.022** |
| Draught: | 0.006 | 0.009 | 0.700 | 0.488 |
| Constant: | -0.335 | 2.652 | -0.130 | 0.900 |
| Model Fitting Information | | | | |
| **Equation** | **RMSE** | **"R-sq"** | **F-Stat** | **p** |
| CargoVol: | 2675.363 | 0.211 | 7.700 | 0.000 |
| TurnRoundT: | 4.343 | 0.467 | 21.850 | 0.000 |
| Berth\_Occpncy: | 0.451 | 0.050 | 20.780 | 0.000 |
| No. of Obs. = 74 |  |  |  |  |
| Source: Author. aEndogenous Variables | | | | |

**Discussion and Conclusion**

We applied the 2SLS method to address the problem associated with correlated error terms in simultaneous equation systems. The results obtained and presented in table 4 agree with our a priori expectation between endogenous and exogenous dependent variables. In table 4, we note that cargo volume achieved in port is inversely related with the turnround time obtainable in that port. That is to say that the lower the numerical value of turnround time, the more shipments will be made there by shippers since they expect to spend little time (or turnround time) and hence save cost. Turnround time on the other hand depends on the facilities present in the port. The higher the number of facilities, the less the totality of time (or turnround time) spent by a shipper or ship in a port, hence the negative coefficient of berth facility variable. Finally, we also note in table 4, that berth occupancy is positively related to the turnround time obtainable in a port. This means that high turnround time may as a result of congestion at the berths and ultimately result in high berth occupancy rate. In terms of Goodness-of-fit of the models, only equation 2 possesses moderate explanatory power with R-squared value of 46.7%. However, equations 1&3 have low explanatory powers given their R-squared values of 21.1% and 5.0% respectively. Although the F-statistic in each model indicates that the coefficients of the explanatory variable are not significantly different from zero, the R-Squared and Root Mean Square Error (RMSE) values indicate that the estimated port utilization model has a poor fit to the data set. One possible explanation for the generally poor fitness of the models is the omission of other relevant variables which affect shipper port utilization model. Thus, more comprehensive models incorporating all the relevant variables should be considered to improve on the fitness of the present estimated models.

**References**

Gonzalez, J.A. and Gualda, N.D.F. (2000) ‘Stated preference framework for analysis of container port competition’, *Proceedings of Third International Meeting for Research in Logistics*, *(IMRL 2000)*, Trois-Rivieres, 9–11 May [online] <http://www.airl_logistique.org/fr/files> (accessed 23 June 2005).

Tuna, O. and Silan, M. (2002) ‘Freight transportation selection criteria: an empirical investigation of Turkish liner shipping’, *Proceedings from International Association of Maritime* *Economists Panama Conference* [online] <http://www.cepal.org/transporte/perfil/iame_papers/> proceedings/Tuna\_et\_al.doc (accessed 1 June 2013).

Veldman, S.J. and Buckmann, E.H. (2003) ‘A model on container port competition: an application for the West European container hub-ports’, *Maritime Economics and Logistics*, Vol. 5, No. 1, pp.3–22.

Lirn, T.C., Ihanopoulou, H.A., Beynon, M.J. and Beresford, A.K.C. (2004) ‘An application of AHP on transshipment port selection: a global perspective’, *Maritime Economics and Logistics*, Vol. 6, No. 1 pp.70–91.

Malchow, M. and Kanafani, A. (2004) ‘A disaggregate analysis of port selection’, *Transportation Research Part E*, Vol. 40, No. 4, pp.317–337.

Tongzon, J. (2002) ‘Port choice determinants in a competitive environment’, *Proceedings of Annual Conference and Meeting of the International Association of Maritime Economists –IAME*, Panama [online] http://www. eclac.cl/transporte/perfil/iame\_papers/proceedings/Tongzon.doc (accessed 30 May 2013).

Tongzon, J. (2009) ‘Port choice and freight forwarders’, *Transportation Research Part E*, Vol. 45, No. 1, pp.186–195.

Song, D.W. and Tae Yeo, K.I. (2004) ‘A competitive analysis of Chinese container ports using the analytical hierarchy process’, *Maritime Economics and Logistics*, Vol. 6, No. 1, pp.34–52.

Guy, E. and Urli, B. (2006) ‘Port selection and multi-criteria analysis: an application to Montreal- New York alternative’, *Maritime Economics & Logistics*, Vol. 8, No. 2, pp.169–186.

De Langen, P.W. (2007) ‘Port competition and selection in contestable hinterlands: the case of Austria’, *European Journal of Transport and Infrastructural Research*, Vol. 7, No. 1, pp.1–14.

McFadden, D. (1974) ‘Conditional logit analysis of qualitative choice behaviour’, in Hess, S. (Ed.): *Advanced Discrete Choice Models with Applications to Transport Demand*, Unpublished PhD thesis, University of London, pp.20–21.

Magala, M. and Sammons, A. (2008) ‘A new approach to port choice modelling’, *Maritime Economics & Logistics*, Vol. 10, Nos. 1–2, pp.9–34.

Onwuegbuchunam, D.E (2013), Port Selection Criteria by Shippers in Nigeria: A Discrete Choice Analysis. *International Journal of Shipping and Transport Logistics*, Vol. 5, Nos. 4/5, pp.532-550.

Panayides, P. and Song, D.W. (2012) ‘Determinants of users’ port choice’, *The Blackwell Companion to Maritime Economics*, 1st ed., pp.599–622, Blackwell Publishing Ltd.

Greene, W.H (2008), Econometric Analysis, 5th ed. India, Dorling Kindersley Pvt. Ltd, Panchsheel Park, New Delhi