SPATIAL DETERMINANTS OF IMPORT TRAFFIC DISTRIBUTION AT PORT HARCOURT (NIGERIA)

by
Soddy I. Ionyang

ABSTRACT

This paper highlights the result of a study carried out to examine the Geographic determinants of import traffic distribution at the Port Harcourt Port. A substantial aspect of the study involved building a regression model to estimate import distribution from the Port. The import function was specified in log-linear form. The adequacy of the model was then tested; this involved statistical exercises to obtain the R-squared, as well as t and f values. Further test on the adequacy of the model was conducted through diagnostic exercises designed to check for multicollinearity and heteroscedasticity, in the data used. Results obtained showed that road distance and manufacturing industries are significant. Geographic determinants of import cargo distribution at the study Port. It is therefore important to take the two variables into account in any policy or planning exercise at Port Harcourt Port.

INTRODUCTION

In the past, several studies, have interested in the task of studying Nigerian Ports (Ogunwada 1966, Ncevo 1970, Bigninski 1979 and Shumwem et al. 1979). Inspite of such attempts very few studies have conducted to examine port traffic from a spatial perspective; that is, few studies have been engaged with examining the role of the hinterland on port traffic. This is inspite of the fact that hinterland correlates (Socio-economic variables) largely determine the volume and direction of flow of imports. Furthermore, effective planning demands that such spatial variables be identified and their significance scientifically tested; it is only within such a framework, that quantitative estimates, can be made, for purposes of policy consideration.

To the best of our knowledge, the (pioneering) work of Onokomaiya and Smith (1972) remains the only work which has scientifically examined import traffic distribution in Nigeria. The study utilized data for 1964 and concerned the distribution of imported Cargo by rail from two Nigerian ports (Apapa and Port

* Soddy I. Ionyang is a lecturer at Department of Geography, Nigerian Defence Academy, Kaduna, Nigeria.
The Gravity model provides a conceptual and theoretical basis for the model to be used. The gravity model has been traditionally applied to migration studies (Gournay 1983). However, variations of the model have been applied to traffic studies (Britton 1967; Bulk 1971; O' Sullivan and Ratliff 1974) and O"Keane (1976).

In its simplest from the Gravity model states that interaction between any two points is a function of some attractive factor (or mass terms) as well as an impedence factor. The attractive variable is usually population, while the impedence factor is distance. Consequently, the larger the mass (suitably defined) the higher the traffic; on the other hand, the larger the operating distance the lower the traffic (all things being equal).

The definition of attractiveness and impedence variables may however take many forms; the mass term may be defined in terms of population, income, retail sales, number of commercial activities as well as industrial development. The fractional effect of distance may be defined in terms of one or more involved in ever-decreasing space.

However, several problems, are associated with the gravity model. A major problem concerns its predictive ability; the model has been criticized for failure to replicate observed trip patterns (Wells 1986; Siddal and Tretkoff 1981). Another problem has to do with the fact that the model is basically misspecified (Fotheringham 1981, 1983 and Baxter 1980); that is, it fails to take into account the spatial structure of origins and destinations.

However, many factors have increased the validity and adoption of the gravity model. For example, Wilson (1967) has proved that the doubly constrained gravity model is equivalent to entropy maximization (Tornquist and Tsonis 1968). Stated that the gravity-entropy model is a generalized form of the linear...
programmed transportation problem. This work therefore takes the gravity model as
given, the model therefore provides a strong theoretical base for our model
specification.

The conceptual model is therefore as follows:

\[ Y = f(X_1, X_2, X_3, \ldots) \]

Where

- \( Y \) = Quantity of imports flowing from the port to inland locations
- \( X_1 \) = Population of inland towns
- \( X_2 \) = Number of manufacturing industries
- \( X_3 \) = Road distance between the port and inland locations
- \( \beta_1, \beta_2, \beta_3 \) are coefficients to be estimated

In this work, the use of income of towns as an additional variable was
dropped because of the following reason:

(i) The difficulty of obtaining adequate data on income at the urban level (that is,
income generated by towns from taxation of individuals, industries and com-
mercial activities).

(ii) Since the study is interested in a function that meets the needs of policy, the
variables to be used must be obtainable on a continuous basis for the past,
present and future.

(iii) The strong possibility of a high correlation between population as well as the
number of manufacturing industries on one hand and income generated on
the other. This is because it is assumed the income generated at the urban
level is a function of the number of taxable adults as well as the number of in-
dustrial and commercial ventures. A high correlation may thus be expected -
thus raising the problem of multicollinearity (see Johnson 1984 pp 246-248).
(We were further discouraged from using income generated from agricultural
exports because of the insignificance of such exports in revenue generation).

With regard to the model earlier specified, we should expect (on theoretical
grounds) the following findings:

(a) The coefficient of population (b1) should have a positive sign given the gravity
model which postulates a direct relationship between volume of Cargo flow
and population.

(b) The coefficient of number of manufacturing industries is also expected to have
a positive sign.

(c) The parameter \( b \) (coefficient of distance) is expected to have a negative sign,
this is because theoretically an inverse relationship is expected between quan-
tity of flow and distance.

If the model is specified.
METHODOLOGY

Methods

Spatial distribution models are usually classified into two groups: these are the multiplicative or logarithmic and linear models. The multiplicative models are non-linear in nature and the parameters of such models are usually estimated using maximum likelihood methods (Batty and Mackie 1977; Wilson 1970 and Mackett 1979).

On the other hand linear models are usually estimated using least squares method (Cesare 1974, 1975). In practice the two methods are found to be insufficient and yield similar results (Ovenden 1976). This study therefore uses the least squares method in estimating the import distribution function for Port Harcourt.

Multiple regression is therefore utilized in this study; the specific functional form used is the log-linear function (This function is used to take into account possible non-linearities in the data).

The log-linear model is specified as follows:

\[ Y = b_0 + b_1 \ln x_1 + b_2 \ln x_2 + b_3 \ln x_3 + U \]

Where the \( b \)’s are coefficients to be estimated.

\( U \) error term

(\( \bar{Y} \) variable in the function have been earlier defined)

The reliability and adequacy of the regression results are tested at various stages. The first stage relates to the theoretical postulates about the sign of the parameters. The second level makes use of first order statistical test; such test involve the use of the \( R^2 \) square, \( t \) and \( F \) values to assess the reliability of the regression results.

Further diagnostic analysis involved testing for multicollinearity and heteroscedasticity in the data used in this study.

Multicollinearity in regression analysis usually occurs when near linear dependencies occur between the independent variables (Koustovias 1977 p.233). This problem leads to inflated sampling variances, imprecise parameter estimates and reduction in the accuracy of significance test (Hair et al and Jackson 1977 pp.86-96; Johnston 1964 p.240). The seriousness of multicollinearity in the data used was tested for using three methods. The first is the intercorrelation between the explanatory variables; a high correlation index normally indicates the presence of multicollinearity. The second method used is the variance inflation factors (VIF). This factor is defined as:

\[ VIF = \frac{1}{1 - R^2} \]

where \( R^2 \) is the squared multiple correlation coefficient when an explanatory variable is regressed on the remaining variables (Thornton 1984 p.247). The VIF index quantifies the magnification of the sampling variance due to collinearity relative to the absence of collinearity \( VIF = 1 \) (Blanda 1991); on the other hand VIF’s in excess of 10 are taken as indicating severe multicollinearity (Margeru 1990). The third method used to test for the severity of multicollinearity is a method due to Koustovias, (see Koustovias 1977 pp.238-241); the method is a revised version.
61

of Fritz's 1934 conference Analysis. The method involves at basic steps. The first
step is to regress the dependent variable on each one of the explanatory variables
separately. The results of the elementary regressions are then checked on the basis of
a priori and statistical criteria. The second step involves adding additional variables
and checking their effect on individual coefficients, standard errors and the R squared
(Koutsoyiannis 1977 p.239). The presence of multicollinearity is indicated by changes
in the signs of coefficients; and a drop in standard errors as more variables are added.

A major assumption in least squares concerns the pattern of the errors; that is
the error is expected to have a constant variance; or the errors should not be
 correlated with the explanatory variables - that is absence of heteroscedasticity. The
presence of heteroscedasticity results in inefficient parameter estimates as well as
invalid standard errors. There are various tests for heteroscedasticity; such tests include
those of Breusch - Pagan (1979); Goldfield - Quandt (1965); Glejser (1969) and the rank
correlation test. In this study, the rank correlation test is used; this is because it is the
simplest; besides Johnson (see Johnston 1972 p.22) seems to accept that the rank
correlation test as well as the Goldfield and Quandt test are preferable to other test.

Data

The port traffic data used in this study was for the year 1980; they were
obtained from the Nigerian Ports Authority. Although, more recent data are available
from the Port Authority, such data are at an aggregated level. That is, current data,
reflect movement of goods from port to state, such Zonal data are associated with the
problems especially when applied in regression analysis (Douglas and Lewis 1970).
The main problem is that Zonal regression can account for only the variation between
zones and not for within zone variabilities (White and Senior 1973 p.155). As a result,
relationships between variables representing aggregated data may be inaccurate
(Dalvi and Martin 1977). It is therefore better for analysis to be undertaken at the most
spatially disaggregated level possible (Dalvi and Martin 1977).

The preceding discussion therefore justifies the use of 1980 data for estimating
port traffic distribution. (The data used in this study are disaggregated and involve
flow of goods from Port Harcourt to twenty Nigerian towns. See appendix D).

Information on population of Nigerian towns were obtained from the National
Population Commission for the year 1963. The data were then projected to obtain
1980 estimates using the formulae Pn = P0e^{nt}

Where

Pn = Future population
P0 = Present population
n = exponential
r = rate
t = time

(The growth rate of Lagos for 1980 is given as 4.7%; other Nigerian towns have
a growth rate 3.2%.) Information on industries were obtained from the Federal
Ministry of Industries; while road distances between Port Harcourt and inland
locations were obtained from the Federal Ministry of Work and Housing. (Cond
distances are used because it is assumed that a large proportion of Port-inland traffic movement now use the highway network rather than the railway.

THE STUDY PORT

Eastern Nigerian rail line and as an outlet for the export of coal from Enugu to Lagos and certain West African Countries. The Port is located 65.6 kilometres up the Benin River and was named after Louis Harcourt the then British Secretary of State. Port Harcourt has witnessed extensive development since its establishment. The first was between 1956 and 1961, the second between 1962 and 1966. The last development took place between 1970 and 1974. The port is capable of berthing eight vessels and has seven stacking areas. The main quay covers an area of 47 hectares and is 1300 metres long. The port remains the most important port in the Eastern part of Nigeria (see Figure 1 for a map of the Port).

EMPIRICAL RESULTS

Table 1 shows the regression results; equation number 1 in the table relates to the regression results for all parameters. The R squared for the equation is 65.3% and the regression is statistically significant as indicated by the F value. Furthermore, all the regression parameters have the correct sign. For example, the population parameter comes out with a positive sign; however, the coefficients of the population parameter is not statistically significant. The industrial parameter also has the correct sign but is not statistically significant; while road distance is statistically significant and also has the expected sign.

Given these, the R squared obtained, was far above average and that only one variable was significant, we decided to examine the importance of individual parameters. In order to do this, we estimated simple functions using each variable one at a time. The result of this exercise is shown on Table 1 from equation Number 2 to 4. Equation Number 2 relates to the population parameter. It shows that population alone explains 12% of the variation in the dependent variable; this variable was found to be insignificant using an F test at the 5% level with 1 and 18 degrees of freedom). Equation Number 2 relates to manufacturing industries; this variable explains 28% of the variation in the import data and was found to be statistically significant. Furthermore, road distance was found to be significant; this variable recorded an R² of 36%.

The preceding paragraph therefore shows that distance and industries are two important variables that explain the distribution of imported goods from Port Harcourt. However, given that in the complete equation (equation number 3) the industrial parameter was found to be insignificant we decided to further check the adequacy of the result obtained by carrying out diagnostic tests. Such tests are important since it is possible to obtain meaningless results despite having high R-squared and significant t and f ratios. Besides a violation of many assumptions governing the application of least squares could lead to wrong conclusions concerning the variables. For example, we could reject a variable as being insignificant while it is indeed very significant.
TABLE 1 Regression Results

In Y = \ln x0 + \ln x1 + \ln x2 + \ln x3 + \ln U

<table>
<thead>
<tr>
<th>Equation number</th>
<th>Constant</th>
<th>Population</th>
<th>Industries</th>
<th>Retail</th>
<th>Services</th>
<th>Adj R-sq</th>
<th>R-sq</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.708</td>
<td>1.544</td>
<td>3.915</td>
<td>-1.4552</td>
<td>-</td>
<td>1</td>
<td>12%</td>
</tr>
<tr>
<td></td>
<td>1.0174</td>
<td>1.0174</td>
<td>1.0174</td>
<td>1.0174</td>
<td>1.0174</td>
<td>1.0174</td>
<td></td>
</tr>
<tr>
<td>VIF</td>
<td>2.038</td>
<td>2.1196</td>
<td>1.9267</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>65.3%</td>
<td>58.9%</td>
<td>10.046</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

|          | b1       | b2         | b3         |        |          |          |      |
| V1(y|x1)  | 5.1629   | 9856       |            |        |          |          |      |
| V1(y|x2)  | 4.4027   | 1.322      |            |        |          |          |      |
| V1(y|x3)  | 15.0311  |            | -1.3186    |        |          |          |      |
|          |          |            |            |        |          |          | 12%  |

DIAGNOSTIC TEST

Table 2 shows the correlation matrix between the explanatory variables. The table shows a high collinearity between population and manufacturing industries. Thus the presence of multicollinearity in the data is confirmed. However, other test were carried out to examine the severity of the collinearity in the data. The variance inflation factor (VIF) was used to assess the severity of the multicollinearity. Results in Table 1 indicate that the VIF for all the parameters were quite low. For instance no variable had a VIF of up to 10, consequently multicollinearity is not a serious problem. However, we note that all the variables have some collinearity since all the VIFs are above 1.

To further examine the severity of multicollinearity in the data we adopted a method suggested by Koutsoyannis. The result of the exercise is presented in Table 3. We used distance in the first elementary regression (Y = b1x1) since it recorded the highest R-square. The introduction of manufacturing industries significantly improves the R-square; the sign of the parameters are correct and both variables are highly significant.
The regression with all three variables shows the effect of multicollinearity. That is, the inclusion of population results in the industrial parameter becoming insignificant. Thus, population has a detrimental effect on the value of the industrial parameter. However, it should be noted that the inclusion of population the R-square is increased and the signs of the parameters are still correct. Consequently, this method also indicates that multicollinearity is not severe, this is because the addition of collinear data did not result in a change of the signs of the parameters.

Table 2. Correlation Matrix of The Explanatory Variables

<table>
<thead>
<tr>
<th></th>
<th>X1</th>
<th>X2</th>
<th>X3</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X2</td>
<td>.773</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>X3</td>
<td>.225</td>
<td>.201</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3. Regression Results

<table>
<thead>
<tr>
<th>Equation number</th>
<th>b0 Uncons</th>
<th>b1 distance</th>
<th>b2 industries</th>
<th>b3 population</th>
<th>R2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Y = \hat{y}</td>
<td>14.2925</td>
<td>-1.0385</td>
<td>-2.0755</td>
<td>.54%</td>
</tr>
<tr>
<td>2</td>
<td>Y = \hat{y}</td>
<td>11.3127</td>
<td>-1.0210</td>
<td>-2.6034</td>
<td>.57%</td>
</tr>
<tr>
<td>3</td>
<td>Y = \hat{y}</td>
<td>1.708</td>
<td>-1.4550</td>
<td>.5255</td>
<td>.60%</td>
</tr>
</tbody>
</table>

Note: An asterisked coefficient is significant at the 5%
Significance level (using a two-tailed test)
From the preceding discussion we can therefore state that road distance and manufacturing industries are important variables that explain the distribution of imports from Port Harcourt. The non-significance of the industrial parameter in the complete equation is due to the collinearity between population and manufacturing industries. The significance of manufacturing industries is not surprising given the emphasis placed on industrial development. Furthermore, the nature of the industries found in Nigeria sheds more light as to the significance of the industrial parameter. Most industries in Nigerian assemble components which to large extent are imported; consequently it is expected that there should be a positive correlation between imports and manufacturing industries.

The spearman rank correlation test was used to test for heteroscedasticity in the data. The rank correlation coefficient between the absolute values of the errors and each of the explanatory variables in turn was then computed. A t test was then used to examine the significance of the rank correlation coefficient. The test used is in the form:

\[ r = \frac{n-2}{\sqrt{n^2 - \sum x^2}} \]

where \( r \) = rank correlation coefficient
\( n \) = sample size
\( v \) = refers to the degrees of freedom used in the test

The results obtained show that the rank correlation coefficient for population against the errors was 7%. The t value obtained was 2.76. This value was compared to the tabular value at the 5% significance level with 18 degrees of freedom; the calculated t value was found to be insignificant. We therefore conclude that heteroscedasticity did not exist in the population data.

In the case of industries the rank correlation coefficient obtained was 51%. The computed t value was found to be insignificant when compared to the tabular value. Consequently heteroscedasticity did not constitute a problem in the industrial data. The test carried out on the data concerning road distance also indicated a lack of heteroscedasticity in the data; (The rank correlation coefficient obtained was 4.76%; at t test indicated that this value was not significant).

CONCLUDING REMARKS

The objectives of this paper have been to provide evidence on the spatial determinants of import traffic distribution from Port Harcourt. We had therefore set out to examine the statistical significance of variables such as population, manufacturing industries and road distance in explaining the distribution of imports. The main conclusion of the paper are therefore as follows:

In the first place road distance was found to be very significant in explaining the distribution of imports. This finding is plausible given the extent of road development in Nigeria. Most of the major highways have been extensively reconstructed or expanded into Express-ways. This fact has helped to develop the
road haulage industry to the extent that importers prefer to patronise haulage companies rather than the railways' co-operation. Consequently, any policy on import distribution in Nigeria must take into account road distance.

Secondly, number of manufacturing industries significantly influences the volume of imports that flows into an inland location. The importance of this variable a plausible given attempt at developing the industrial land-use of Nigeria. It is also worth noting that most of such industries depend almost entirely on imported raw materials. Consequently, it is necessary for policy makers to take into account level of industrialization when examining import distribution. Population was found to be insignificant in explaining the variation in the import data. The variable is indeed detrimental - since its in-operation into the import function seriously reduces the significance of the industrial parameter.

Finally, the diagnostic exercise revealed the presence of multicollinearity in the data; the severity of collinearity was found to be slight. However the presence of collinearity especially between population and industries resulted in the non-significance of the industrial parameter. Further diagnostic test showed that heteroscedasticity was not a problem in the data used.

The results obtained in this study, to a large extent, meet theoretical expectations; besides they appear to be reasonably reliable (given the results of the diagnostic exercise). We may therefore conclude that the methodology adopted as well as the results obtained may prove useful for planning a Port Harcourt. (See Figure 1)

REFERENCES


Cesario F.J. (1975) 'Linear and non-linear regression model of spatial interaction' Geographical Analysis Vol.5 No.1 pp.69-77


Frisch B. (1934) 'Statistical Conjecture Analysis by means of complete Regression Systems'.


Hutchinson B.J. and Smith D.F (1979) 'Empirical studies of journey to work in urban areas' Canadian Journal of Civil Engineering 6, pp.303-318.


## APPENDIX

### DATA USED IN THE REGRESSION ANALYSIS

<table>
<thead>
<tr>
<th>No.</th>
<th>Towns</th>
<th>Quantity of imports from Port Harcourt to towns</th>
<th>Population (1960)</th>
<th>Number of manufacturers (1960)</th>
<th>Road distance from Port Harcourt (1960)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Port Harcourt</td>
<td>774554</td>
<td>319406</td>
<td>336</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>Abe</td>
<td>318532</td>
<td>233628</td>
<td>180</td>
<td>64</td>
</tr>
<tr>
<td>3</td>
<td>Umuahia</td>
<td>3572</td>
<td>53207</td>
<td>26</td>
<td>116</td>
</tr>
<tr>
<td>4</td>
<td>Oriahia</td>
<td>244735</td>
<td>260901</td>
<td>60</td>
<td>208</td>
</tr>
<tr>
<td>5</td>
<td>Enugu</td>
<td>106597</td>
<td>249297</td>
<td>105</td>
<td>253</td>
</tr>
<tr>
<td>6</td>
<td>Lokoja</td>
<td>312565</td>
<td>1555969</td>
<td>624</td>
<td>699</td>
</tr>
<tr>
<td>7</td>
<td>Calabar</td>
<td>6127</td>
<td>139902</td>
<td>47</td>
<td>198</td>
</tr>
<tr>
<td>8</td>
<td>Uyo</td>
<td>18092</td>
<td>58498</td>
<td>40</td>
<td>122</td>
</tr>
<tr>
<td>9</td>
<td>Sokoto</td>
<td>341</td>
<td>158766</td>
<td>12</td>
<td>1929</td>
</tr>
<tr>
<td>10</td>
<td>Katsina</td>
<td>2296</td>
<td>284159</td>
<td>59</td>
<td>1807</td>
</tr>
<tr>
<td>11</td>
<td>Zaria</td>
<td>62</td>
<td>205583</td>
<td>21</td>
<td>1993</td>
</tr>
<tr>
<td>12</td>
<td>Kano</td>
<td>3261</td>
<td>528514</td>
<td>215</td>
<td>1274</td>
</tr>
<tr>
<td>13</td>
<td>Kebbi</td>
<td>1909</td>
<td>175114</td>
<td>150</td>
<td>305</td>
</tr>
<tr>
<td>14</td>
<td>Kwarri</td>
<td>1727</td>
<td>88235</td>
<td>51</td>
<td>383</td>
</tr>
<tr>
<td>15</td>
<td>Kaduna</td>
<td>155</td>
<td>46264</td>
<td>6</td>
<td>217</td>
</tr>
<tr>
<td>16</td>
<td>Jos</td>
<td>6527</td>
<td>166507</td>
<td>163</td>
<td>394</td>
</tr>
<tr>
<td>17</td>
<td>Maiduguri</td>
<td>11002</td>
<td>96656</td>
<td>12</td>
<td>907</td>
</tr>
<tr>
<td>18</td>
<td>Bostin</td>
<td>4659</td>
<td>1115961</td>
<td>125</td>
<td>658</td>
</tr>
<tr>
<td>19</td>
<td>Kebbi</td>
<td>3852</td>
<td>359861</td>
<td>15</td>
<td>446</td>
</tr>
<tr>
<td>20</td>
<td>Zaria</td>
<td>5050</td>
<td>302702</td>
<td>3</td>
<td>465</td>
</tr>
</tbody>
</table>

Sources:
1. Nigerian Pensions Authority
2. National Population Commission
3. Federal Ministry of Industries
4. Federal Ministry of Works and Housing