Port selection from a hinterland perspective

Abstract

The studies investigating the port selection process use to have one thing in common: they analyse the declared preferences of the port agents. However, it is difficult to identify the relevant variables in this process because of the heterogeneity of this group. In this paper we suggest to study the port choice through revealed port selection instead of asking port stakeholders about the main factors in port selection. We propose to analyse the actual inter-port traffic distribution from a holistic view using the *hinterland perspective* and the discrete choice modelling approach in order to answer the question: Does the location of a port still remain important in port selection? As a case study, we use the Spanish inter-port container distribution among the main peninsular ports.

Keywords: Port selection; discrete choice; hinterland perspective; hinterland; port location; inter-port competition.

INTRODUCTION

Transport sector improvements and expansion result in a larger number of ports through which freight can be efficiently transported. Consequently, the modernization of the transport sector leads to increased inter-port competition. Nevertheless, the port selection process is a complex and a rarely analysed issue. Some of the first papers aimed at port selection analysis were published three decades ago (Foster, 1977; Foster, 1978). Several authors have studied this subject since (see, for example, Slack, 1985; Brooks, 1990; D'Este and Meyrick, 1992; Lago *et al.*, 2001; Lirn *et al.*, 2004; or Ugboma *et al.*, 2006) and, even though their conclusions were different, most of them consider that achieving scale economies and reducing the time necessary to offer a *door-to-door* service favours the attraction of traffic to a certain port, more than the physical proximity of the port to clients. In this sense, Robinson (2002) pointed out that port selection depends on a port's inclusion in logistics chains, and Magala and Sammons (2008) have recently published new research in support of this idea. Bergantino (2002), De and Park (2003) or Malchow and Kanafani (2004) went even further by considering that the evolution of the port activity does no longer depend on a port's immediate hinterland, due to the development of intermodal transport. However, Bichou and Gray (2005) reintroduced the idea that each port belongs to a system. They suggest that the evolution of port activity is related to its economic, political and social environment. Also Yap and Lam (2006) associate the evolution of the activity of a port to the economic evolution of its province. It can be seen therefore that the role of the port location on the port selection process is still under discussion.

On the other hand, the reasons why one port is chosen while another is not are usually studied by asking port agents about their preferences. However, there are several economic agents involved in the port selection process, and each of them makes a different evaluation of the main factors of this process depending on its own objectives (see, for example, D'Este and Meyrick, 1992; Matear and Gray, 1993; or Murphy *et al.*, 1997). Consequently, the conclusions about the variables considered fundamental can differ depending on who responds to the survey.

Taking into account this last point, and following Bichou and Gray (2005) and Yap and Lam (2006), we propose to analyse the actual inter-port distribution of traffic to study the port selection process. Since each ton of maritime traffic is channelled through a particular port, i) we consider that each assignment is a selection, and ii) we assume that each selection is made from the province where the flow is generated. With this approach, that we call *hinterland perspective*, we analyse the inter-port traffic distribution by means of a multinomial model to answer the question: Does the location of a port still remain important in port selection? And if this is the case, to what extent?

As a case study, we analyse the Spanish inter-port container traffic distribution among the main peninsular ports for this type of traffic (Algeciras, Barcelona, Bilbao and Valencia). For this purpose, in the next section we formulate a statistical model to explain the inter-port container distribution revealed. We use the database of Foreign Trade from the Spanish Treasury Department as data source, which collects all the movements of cargo derived from the Spanish foreign trade taking into account the provincial origin or destination of each merchandise flow. The main conclusions, shown in the last section, are that the port location (port-province distance) is still an important factor in the port selection process and that, consequently, the hinterland of a port contributes to explain the evolution of its activity.

STATISTICAL MODEL FOR THE CONTAINER TRAFFIC DISTRIBUTION

In order to answer the question about the degree of importance of the distance in the inter-port competition process we use the explicative-stochastic approach (Chasco and Vicens, 1998). This approach uses information revealed by past behaviour, allowing us to understand the dynamics of inter-port competition through the analysis of the port selection made from the provinces where the container traffic is generated. In this paper we use the revealed preference approach, while port selection is understood to be a multiple choice problem with a spatial perspective.

Huff (1963) was the first to use a utility function and introduced the spatial interaction models to explain consumer behaviour. Following this author, the attention is no longer on whether agent *i* maximises his utility/benefit with the alternative selected, but rather on the preferred port to channel the generated flows in province *i*, taking into account all of the possible alternatives. The probability that the chosen option by *i* is the j^{th} alternative can be expressed as a multinomial conditional logit model following the idea in McFadden (1974).

As we are analysing the importance of port distance in the inter-port competition process, the port-province *distance* is the main variable in the proposed model (Garcia-Alonso and Sanchez-Soriano, 2007). Accordingly, we provide a destination-oriented model of inter-port competition, relying on the family of spatial interaction models (for more details on spatial interaction models, see Fotheringham and O'Kelly (1989). A brief review of these models can also be found in Chasco and Vicéns (1998)).

Mathematically, the share of the traffic generated in province *i* and channelled through port *j*, π_{ij}, is given by:

$$\pi_{ij} = \frac{e^{-(a_j + dD_{ij})}}{\sum_{k \in P} e^{-(a_k + dD_{ik})}}, \ \forall i \in R; \forall j \in P \qquad (1)$$

where *P* is the set of ports; *R* is the set of provinces; a_j measures the appeal of port *j* derived from its specific characteristics (volume and type of traffic, efficiency, cost, port services available, infrastructure, size, etc.); *d* gives the attitude of the agents with respect to the land distance that could be called '*aversion to distance*'; and D_{ij} is the distance from province *i* to port *j*. The incorporation of *d* attempts to isolate the importance agents responsible for port selection (considered as a whole) give to port-province distance, from the importance they give to all other aspects taken together (*a*). Also, as denoted in expression (1), we consider that agents do not only take the distance from themselves to each port into account in the same way as each port's specific characteristics but, simultaneously, they compare both aspects with the corresponding ones of the remaining ports (reflected in the denominator of (1)). This way of comparing comes from the axiom proposed by Luce (1959) and it is followed in many classical spatial interaction models (see, for example, Huff, 1963; McFadden, 1974; or Fotheringham, 1983).

In order to evaluate the value of the parameters so that the resulting port market distribution conforms most reliably to the real one, the multidimensional probability distribution function given in (2) is optimised to maximise simultaneously the probability that the port traffic distribution generated in 2004 in each of the 47 provinces is likened to what actually occurred. We consider the inter-port container traffic distribution among the main Spanish peninsular ports of Algeciras, Barcelona, Bilbao and Valencia. These four ports managed in 2004 the 77% of the whole container traffic in the Spanish port system.

$$\Pr(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, \ldots, Y_{ip} = y_{ip}) = n_i! \prod_{j \in P} \frac{\pi_{ij}^{y_{ij}}}{y_{ij}!}, \quad i \in R. \tag{2}$$

where $Y_i = (Y_{i1}, Y_{i2}, ..., Y_{ip})$ is the multinomial variable associated with the distribution among the ports of the traffic generated in province i; y_{ij} is the traffic generated in province i and channelled through port j (such that $\sum_{j \in P} y_{ij} = n_i$); n_i is the marginal row of the traffic distribution matrix (cargo generating provinces are found in the rows, and the ports managing them are in the columns); and π_{ij} represents the probability that the option chosen by province i is port j (or the share of the traffic generated in province i that port j attracts).

Next, we use the maximum likelihood method to estimate the parameters of the model that allows us to fit better the distribution of all actual flows. The likelihood function of the multidimensional random variable $Y = (Y_1, Y_2, ..., Y_r)$ (where r is the number of provinces) is given by (3):

$$L(y;\pi) = \prod_{i \in R} \left(n_i! \prod_{j \in P} \frac{\pi_{ij}^{y_{ij}}}{y_{ij}!} \right) \tag{3}$$

As π_{ij}'s ($\forall\, i \in R$, $\forall\, j \in P$) depend on the parameters d (aversion to distance) and a_j (appeal of each port j), we maximise the likelihood function (given the real traffic distribution) with respect to these five parameters. In this sense, we are interested in maximising the probability (or likelihood) that the inter-port traffic distribution predicted by the proposed multinomial conditional logit model matches the real one. In order to do this, we first transform the likelihood function (3) into the log-likelihood function (4), whose mathematical formulation is less complex (as the logarithmic is a monotonic function, the values that maximize expressions (3) and (4) are the same).

$$l(y;a,d) = K + \sum_{i \in R} \sum_{j \in P} y_{ij} (-(a_j + dD_{ij})) + \sum_{i \in R} n_i \left(-\log \sum_{j \in P} e^{-(a_j + dD_{ij})} \right) \tag{4}$$

where K is a constant and a is the vector of parameters a_j, $j \in P$.

We use the Newton-Raphson iterative algorithm to find the roots of the system of first derivatives of (4) (see Appendix). We also use the Hessian matrix, consisting of the second derivatives of the log-likelihood function, to obtain the *p*-values of the estimated parameters. To avoid over-parametrization problems in our model, we fix the value of $a_{Valencia}$ as zero (i.e. we consider the port of Valencia as a reference point with respect to the appeal of the ports).

Results

The iterations carried out to find the values of the parameters that maximise the log-likelihood function were set out from randomly generated seeds, and concluded when the maximum absolute value of the difference between their last value and the previous one was less than 1/10,000. In order to carry out these iterations, a computer program was developed. The obtained results for the parameters are shown in table 1.

INSERT TABLE 1

The Maximum Likelihood estimators of the parameters a_j show Valencia as the most attractive port, followed by Barcelona, Algeciras and Bilbao, respectively. To check the validity of the results we compare the distribution of container traffic, estimated through our model (incorporating the estimated parameters), with the distribution of container traffic actually observed. The difference between them can be measured by the *Cramer Coeffcient V*, defined by:

$$V = \sqrt{\frac{\sum_{i \in R} \sum_{j \in P} \frac{\left(y_{ij} - e_{ij}\right)^2}{e_{ij}}}{N(r-1)(p-1)}} \tag{5}$$

where the e_{ij}'s are the estimated container traffic from province *i* to port *j*; *r* is the number of provinces; and *p* is the number of ports. This indicator ranges between 0 and 1. The value of 0 means that both sets of data (revealed and estimated) match perfectly; the value of 1 means that both sets of data are very far from each other. In our particular case V = 0.4690 but if we remove one outlier observation (traffic from Barcelona to Algeciras) then V = 0.2458. Therefore, we can conclude that the fitness reached is reasonably satisfactory, and hence, the approach followed in order to analyse the port selection process is adequate.

The above measure of the discrepancy between the revealed and estimated maritime container traffic is quantitative; that is, it takes into account the magnitude of the traffic flows. However we can also use a more qualitative measure, for example an ordinal measure. We considered the following ordinal measure: for each province and each pair of ports we compare whether the revealed and estimated traffic data are in concordance in an ordinal sense; that is, whether the preference of the province is the same with respect to that pair of ports. In our particular case, the percentage of discrepancies over the total number of possible pair comparisons was 10.63%. Therefore, the model correctly predicted almost

90% of all pair-wise comparisons (282 in total). This means the model captures quite well the movement of container traffic in Spain.

On the other hand, in order to analyse the degree of importance of the variable *port-province distance* in our model, and hence in the port selection process, we remove the *distance* from the model. For this second model, V = 0.8184. Therefore, we can conclude that the *distance* variable reduces by 42.69 per cent the difference between the revealed and estimated data of the inter-port container traffic distribution of provinces in Spain. If we remove the outlier (now traffic from the province of Barcelona to the port of Barcelona) in this second model, then we obtain V = 0.7610. Then, without the worst outlier in each case, the *distance* at least reduces by 67.70 per cent the difference between the revealed and estimated data. Even if we only use the *distance* in our model, then V = 0.4138[1] , or V = 0.3297 removing the worst outlier (traffic from Barcelona to Algeciras). All this shows that the port-province distance factor plays a very important role in the port selection process when this is analysed from a hinterland perspective.

CONCLUSIONS

This paper deals with the analysis of port selection from a hinterland perspective. Attending to the suggestions of D'Este (1992) and Mangan *et al*. (2001), the study of port selection is made in a holistic way, focusing on the actual inter-port container traffic distribution. The results show that the port-province distance remains a relevant variable in the port selection process (even for container traffic) despite all the transport sector improvements, and it confirms the conclusion of Sargent (1938): cargo tends to seek the shortest route to access the sea. Hence, the concept of *hinterland* does contribute to explain the evolution of the activity of a port, in spite of the development of the intermodal transport and the increase in inter-port competition. Our hypothesis is that the hinterland distance is still a variable so important that firms, when deciding about their location, take into account the location of the ports offering the services they need, whereas firms already established tend to choose the services offered by the nearest port.

Nevertheless, it is also important to keep in mind that the contribution of their own hinterlands is not enough to explain the success of the ports of Algeciras, Barcelona and Valencia during the last decade. According to Gouvernal *et al.* (2005), the development of new logistics chains using the Mediterranean Sea has been very important for the success of

---

[1] We observe that the value of the Cramer coefficient *V* for the model with only the distance is better than the corresponding for the better model from the point of view of the likelihood. The reason of this is the MLE are efficient and asymptotically unbiased but not robust, so they can be strongly influenced by outlier observations, as it is the case in this paper.

these ports. On the other hand, the growth of the activity of the port of Bilbao has been considerably smaller. This port is located on the Atlantic coast, where the evolution of container traffic (and the development of supply chains) did not favour its activity. Consequently, the evolution of the traffic of the port of Bilbao is much more linked to its own hinterland than in the other ports.

It seems, therefore, that the evolution of the port activity matches the strategic position of the ports according to the terms introduced by Fleming and Hayuth (1994): *centrality* and *intermediacy*. That is, it depends on the dynamism of the hinterland of the ports and on their inclusion in the routes of shipping lines. We can conclude that both perspectives, the maritime (usual in the literature) and the hinterland (proposed in this paper), complement each other. Then, both perspectives are necessary to analyse the evolution of the activity of medium ports: ports included in important maritime lines, but with an important volume of national traffic.

ACKNOWLEGEMENTS

A preliminary version of this paper was presented at the 2007 IAME Conference. We would like to thank all comments from the reviewers and the participants at the Congress. We would like to thank too José Ángel Vallejo-Pinto, professor in the Computer Science Department at the Oviedo University (Spain), for his technical support. Financial support from the Government of Spain and FEDER, under project MTM2008-06778-C02-01, is gratefully acknowledged.

REFERENCES

Bergantino, A. S. (2002): The European Commission approach to Port Policy: some open issues. *International Journal of Transport Economics* **XXIX**, **3**: 337–379.

Bichou, K. and Gray, R. (2005): A critical review of conventional terminology for classifying seaports. *Transportation Research Part A* **39**: 75-92.

Brooks, M.R. (1990): Ocean carrier selection criteria in a new environment. *Logistics and Transportation Review* **26**, 339-355.

Chasco, C. and Vicéns, J. (1998): Spatial interaction models applied to the design of retail trade areas. 38th Congress of the European Regional Science Association.

D'Este, G. M. (1992): Carrier selection in a ro/ro ferry trade. Part 2. Conceptual framework for the decision process. *Maritime Policy and Management* **19**, **2**: 127–138.

D´Este, G.M. and Meyrick, S. (1992): Carrier selection in a Ro/Ro ferry trade. Part 1. Decision factors and attitudes. *Maritime Policy and Management* **19**, **2**: 115-126.

De, P. and Park, R. (2003): Container Port System Concentration. *Transportation Quarterly* **57**, **4**: 69-82.

Fleming, D. K. and Hayuth, Y. (1994): Spatial characteristics of transportation hubs: centrality and intermediacy. *Journal of Transport Geography* **2**, **1**: 3-18.

Foster, T. (1977): Ports: what shippers should look for. *Distribution worldwide* **77**, **1**: 40-43.

Foster, T. (1978): Ports: what shippers do look for. *Distribution worldwide* **77**, **1**: 44-48.

Fotheringham, A.S. (1983): A new set of spatial interaction models: the theory of competing destinations. *Environment and Planning A* **15**: 15-36.

Fotheringham, A.S. and O'Kelly, M.E. (1989): *Spatial Interaction Models: Formulations and Applications*. Studies in Operational Regional Science. Kluwer Academic Publishers.

Garcia-Alonso, L. y Sanchez-Soriano, J. (2007): Structure, location, performance: the paradigm of the Spanish port sector. *Proceedings of IAME,* Athens.

Gouvernal, E., Debrie, J. y Slack, B. (2005): Dynamics of change in the port system of the western Mediterranean. *Maritime Policy and Management* **32**, **2**: 107-121.

Huff, D.L. (1963): *A probabilistic analysis of consumer spatial behaviour*. In William S. Decker (ed.) *Emerging Concepts inmarketing*, American Marketing Association, Chicago: 443-461.

Lago, A. and Malchow, M. and Kanafani, A. (2001): An analysis of carriers´schedules and the impact on port selection. *Presentation at 2001 International Association of Maritime Economists Conference*, Hong Kong.

Lirn, T.C., Thanopoulou, H.A., Beynon, M.J. and Beresford, A. (2004): An Application of AHP on Transhipment Port Selection: A Global Perspective. *Maritime Economics and Logistics* **6**, **1**: 70-91.

Luce, R. (1959): *Individual Choice Behaviour*, John Wiley & Sons, New York.

Magala, M. and Sammons, A. (2008): A new approach to port choice modelling. *Maritime Economics and Logistics* **10**, **1/2**: 9-34.

Malchow, M. and Kanafani, A. (2004): A disaggregate analysis of port selection. *Transportation Research. Part E* **40**, **4**: 317–337.

Mangan, J., Lalwani, C. and Gardner, B. (2001): Identifying relevant variables and modelling the choice process in freight transportation. *International Journal of Maritime Economics* **3**, **3**: 278–297.

Matear, S. and Gray, R. (1993): Factors influencing freight service choice for shippers and freight suppliers. *International Journal of Physical Distribution and Logistics Management* **23**: 25-35.

McFadden, D. (1974): *Conditional logit analysis of qualitative choice bahaviour. Frontiers in Econometrics*, P. Zarembka (ed.), Academic Press, New York.

Murphy, P.R. and Daley, J.M. and Hall, P. (1997): Carrier selection: Do shippers and carriers agree, or not? *Transportation Research. Part E* **33**, **1**: 67-72.

Robinson, R. (2002): Ports as elements in value-driven chain systems: the new paradigm. *Maritime Policy and Management* **29**: 241-255.

Sargent, A. J. (1938): *Seaports and Hinterlands*. Adam and Charles Black. London.

Slack, B. (1985): Containerization, inter-port competition and port selection. *Maritime Policy and Management* **12**, **4**: 293-303.

Ugboma, C., Ugboma, O. and Ogwude, I.C. (2006): An Analytic Hierarchy Process (AHP) Approach to Port Selection Decisions-Empirical Evidence from Nigerian Ports. *Maritime Economics and Logistics* **8**, **3**: 251-266.

Yap, W.Y. and Lam, J. S. L. (2006): Competition dynamics between container ports in East Asia. *Transportation Research Part A* **40**: 35-51.

APPENDIX

The system of first derivatives of the log-likelihood function (6) is given by the following expression:

$$\begin{aligned} &\sum_{i\in R}(n_i\pi_{ij}-y_{ij})=0, \qquad \forall j\in P \\ &\sum_{i\in R}\sum_{j\in P}(n_i\pi_{ij}-y_{ij})D_{ij} \end{aligned} \tag{A.1}$$

The Hessian matrix is given by the following expressions of the second derivatives of the log-likelihood function:

$$\begin{aligned} &\frac{\partial^2 l}{\partial\alpha_j^2}=-\sum_{i\in R}n_i\pi_{ij}(1-\pi_{ij}), && \forall j\in P \\ &\frac{\partial^2 l}{\partial\alpha_j\partial\alpha_k}=\sum_{i\in R}n_i\pi_{ij}\pi_{ik} && \forall j,k\in P, j\neq k \\ &\frac{\partial^2 l}{\partial d^2}=\sum_{i\in R}\sum_{j\in P}n_iD_{ij}\pi_{ij}\left(\sum_{k\in P}D_{ik}\pi_{ik}-D_{ij}\right) \\ &\frac{\partial^2 l}{\partial d\partial\alpha_k}=\sum_{i\in R}\sum_{j\in P}n_iD_{ij}\pi_{ij}(\pi_{ik}-\delta_k(j)) \end{aligned} \tag{A.2}$$

where $\delta_k(j) = \begin{cases} 1 & if\ j = k \\ 0 & if\ j \neq k \end{cases}$.

Some nice mathematical properties of our multinomial logit conditional model are the following:

1. Multiplicative scale changes in the traffic flows, y_{ij} $i \in R$ and $j \in P$, do not modify the solutions of system (A.1). Therefore, the estimation of the paremeters in our model does not depend on the particular way of measuring the traffic flows.
2. Multiplicative scale changes in the distances by a constant k modify the solutions of system (A.1) in the following way: the value of the a_j's are the same and the new value of d is obtained dividing by k.
3. If $(a_1, a_2, ..., a_p, d)$ is a solution for system (A.1), then $(a_1 + k, a_2 + k, ..., a_p + k, d)$ is also a solution for system (A.1).