THE SITING OF MULTI-USER INLAND INTERMODAL CONTAINER TERMINALS IN TRANSPORT NETWORKS

Collins Teye

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Institute of Transport and Logistics Studies (ITLS)
The University of Sydney Business School
The University of Sydney

January 2017

In loving memory of my Dad

To my mother

With love and eternal gratitude

And to all my brothers and sisters

Abstract

Almost without exception, cargo movements by sea have their origins and destinations in the hinterlands and efficient land transport systems are required to support the transport of these cargo to and from the port. Furthermore, not all goods produced are exported or all goods consumed are imported. Those produced and consumed domestically also require efficient transport to move them from their production areas to areas of consumption. The use of trucks for these transport tasks and their disproportionate contribution to urban congestion and harmful emissions has led governments, transport and port authorities and other policy-makers to seek for more efficient and sustainable means of transport.

A promising solution to these problems lies in the implementation of intermodal container terminals (IMTs) that interface with both road and rail (or possibly inland waterway) networks to promote the use of intermodal transport as a more sustainable alternative to road alone transport (e.g. trucks). This raises two important linked questions; where should IMTs be located and what will be their likely usage by individual shippers, each having a choice of whether or not to use the intermodal option. The multi-shipper feature of the problem and the existence of competing alternative modes means that the usage of the IMTs are the outcome of many individual mode choice decisions and the prevailing cargo production and distribution patterns in the study area.

This thesis introduces a novel framework underpinned by the principle of entropy maximisation to link mode choice decisions and variable cargo demand problems with facility location problems. The model allows both decisions on facility location and usage to be driven by shipper preferences. This is the perspective that a rational planning authority would adopt and differs from the perspective of, say, a profit maximising IMT operator. The proposed model takes the form of a non-linear mixed integer programming problem with an entropy objective function subject to a range of constraints. Several properties of the proposed model are presented in the form of propositions, including a general method of dealing with capacity constraints. An important outcome is the demonstration of the link between entropy maximisation and welfare maximisation. In other words, the proposed framework allows IMTs to be strategically placed at locations where shipper welfare is maximised. Exact and heuristic algorithms have been developed

to solve the problem. The computational efficiency, solution quality and properties of the heuristic algorithm are presented along with extensive numerical examples.

Finally, the implementation of the model, illustration of key model features and use in practice are demonstrated through a case study. Specifically, the model was used to determine the best locations in the Greater Sydney Metropolitan Area to locate IMTs for the containerised import market. The full model comprised a linked facility location sub-model, a mode choice sub-model and a cargo distribution sub-model. The model was used for forecasting and testing various policies, like changes in land use patterns, road pricing, subsidies, and strategic transport network expansion or improvements to support more use of intermodal transport in the Sydney region.

Statement of Originality

This is to certify that to the best of my knowledge the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes. I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged

Collins Teye

Acknowledgements

This thesis is the result of many years of continuous research into the siting of multi-user inland intermodal container terminals in transport networks and benefitted from the advice and support of many people who need to be acknowledged.

First, I would like to express my heartfelt gratitude to my cherished supervisor Professor Mike Bell and his lovely wife Dr. Reiko for their love, support, guidance and encouragements throughout my doctoral studies. Mike was the first to see the potential in me and also help to secure funding for me to pursue my PhD in such a world class institution like ITLS. His unique style of supervision and constructive criticisms have brought out the best in me and the confidence to carry out independent and high quality research. Mike also acts as a friend and father figure throughout my candidature. The many coastal walks and dinners help to strengthen this aspect of the relationship. Thanks very much Mike and I am so proud to have you as a supervisor. Special thanks to my second supervisor Professor Michiel Bliemer, who always manage to find the 'hidden holes' in my work. His advice and suggestions have been very valuable in the progression of my candidature.

I also owe a debt of gratitude to Professor David Hensher for believing in me and giving me the opportunity to work on the MetroScan-TI and other projects and also helping to secure the funding for my PhD. David is full of new and innovative ideas and this thesis benefitted from some of his rich ideas. It is a rare privilege working with him and for him and I am most grateful for his strong support of my candidature. I want to also acknowledge all the wonderful people in ITLS, especially Dr. Richard Ellison. I have a special admiration of Richard for always being there for me. He is also one of the finest and most approachable colleagues I have ever worked with. I also benefitted a lot from Dr. Adrian, Dr. Chihn, Dr. Jyotirmoyee and Dr. Claudine. They have provided significant advice and insight, which facilitated my research in one way or the other. Professors Rico and Behnam have also shown interest in my research and gave me numerous advice and encouragements during my candidature. I have also enjoyed the friendship and strong support from fellow PhD students especially Ines, Camilla, Mark and Mahbub. They provided wonderful companionship over the years and I will miss them a lot. Acknowledgements are also due to Ruth

Publications

The following papers were published (either authored or co-authored) during the course of my doctoral candidature. Papers directly related to this thesis topic are marked with an asterisk (∗) at the end.

2017

Journal Article/s

Teye C, Bell MGH and Bliemer MCJ, Locating urban and regional container terminals in competitive environment: An entropy maximising approach, *Transportation Research Part B (2017), http://dx.doi.org/10.1016/j.trb.2017.08.017* (*)

Teye C, Bell MGH and Bliemer MCJ 2017 'Urban intermodal terminals: The entropy maximising facility location problem', *Transportation Research Part B: Methodological*, vol.100, pp. 64-81(*)

Teye C, Bell MGH and Bliemer MCJ 2017 'Entropy maximising facility location model for port city intermodal terminals', *Transportation Research Part E: Logistics and Transportation Review*, vol.100, pp. 1-16(*)

Ellison R, Teye C and Hensher DA 2017 'Modelling Sydney's light commercial service vehicles', *Transportation Research Part A: Policy and Practice*, vol.96, pp. 79-89

Paflioti P, Vitsounis TK, Teye C, Tsamourgelis I and Bell MGH 2017 'Box dynamics: A sectoral approach to analyse containerized port throughput interdependencies', *Transportation Research Part A: Policy and Practice,* vol.106, pp. 396-413

Conference Paper/s

Teye C, Bell MGH and Bliemer MCJ 2017 Forthcoming 'Locating urban and regional container terminals in competitive environment: An entropy maximising approach', *22nd International Symposium on Transportation and Traffic Theory (ISTTT22)*, Evanston, United States, 26th July 2017(*)

Teye C, Davidson P, Bell MGH and Porter H 2017 'Meta-analysis on the value of travel time savings in Africa', *5th International Choice Modelling Conference*, Cape Town, South Africa, 5th April 2017(*)

Ellison R, Teye C and Hensher DA 2017 'Commodity-based heavy vehicle model for Greater Sydney', *5th International Choice Modelling Conference*, Cape Town, South Africa, 5th April 2017

Teye C, Paflioti P, Vitsounis TK and Bell MGH 2017 Forthcoming 'Towards a Diversified Cargo Portfolio for Seaports: A Tool for Forecasting and Policy-Making', *International Association of Maritime Economics (IAME) Conference*, Kyoto, Japan, 30th June 2017

2016

Conference Paper/s

Teye C and Bell MGH 2016 'Entropy maximising facility location model for port city intermodal terminals', *Annual Conference of the International Association of Maritime Economists (IAME)*, Hamburg, Germany, 26th August 2016(*)

Teye C, Bell MGH and Bliemer MCJ 2016 'Urban Intermodal Container Terminals: Entropy Maximization Facility Location Problem', *95th Annual Meeting of the Transportation Research Board TRB*, Washington, D.C., United States, 14th January 2016(*)

Paflioti P, Vitsounis TK, Teye C, Tsamourgelis I and Bell MGH 2016 'Box Dynamics: Leaders, Followers and Transmitting Mechanisms - An intersectoral analysis', *Annual Conference of the International Association of Maritime Economists (IAME)*, Hamburg, Germany, 26th August 2016

Teye C, Bell MGH and Paflioti P 2016 'Investigating alternative approach to the gravity model of international trade in evaluating international trade policies', *International Conference on Logistics and Maritime Systems (LOGMS)*, Sydney, Australia, 23rd June 2016.

Conference Proceeding/s

Teye C, Bell MGH and Bliemer MCJ 2016 'Optimal location of open access urban container terminals under elastic cargo demand', *Australasian Transport Research Forum 2016 Proceedings*, Melbourne, Australia, 18th November 2016 (*)

2015

Conference Paper/s

Teye C, Bell MGH and Bliemer MCJ 2015 'Mixed integer programming formulation of locating optimal Intermodal Terminals in Metropolitan areas', *Conference of Australian Institutes of Transport Research (CAITR)*, Melbourne, Australia, 13th February 2015(*)

Teye C, Bell MGH and Bliemer MCJ 2015 'The efficacy of using genetic algorithm for estimating Cross nested logit models', *Conference of Australian Institutes of Transport Research (CAITR)*, Melbourne, Australia, 13th February 2015

Bell MGH and Teye C 2015 'Dynamic Berth Allocation in Container Terminals', *Global Port Research Alliance GPRA 2015 Conference*, Hong Kong, (China), 22nd May 2015

Conference Proceeding/s

Teye C and Bell MGH 2015 'Lagrangian relaxation technique for solving Intermodal Terminal location problems', *37th Australasian Transport Research Forum ATRF 2015*, Sydney, Australia, 2nd October 2015

Ellison R, Teye C and Hensher DA 2015 'Modelling Sydney's light commercial service vehicles', *37th Australasian Transport Research Forum ATRF 2015*, Sydney, Australia, 2nd October 2015

2014

Journal Article/s

Davidson P, Teye C and Culley R 2014 'Developing a Successful Stated Preference Methodology for Determining Destination Choice Coefficients and Using it to Investigate its Empirical Structural Relationship with Toll Route Choice', *Transportation Research Procedia*, vol.1:1, pp. 36-48

Teye C, Davidson P and Culley R 2014 'Simultaneously Accounting for Inter-alternative Correlation and Taste Heterogeneity among Long Distance Travelers Using Mixed Nested Logit (MXNL) Model so as to Improve Toll Road Traffic and Revenue Forecast', *Transportation Research Procedia*, vol.1:1, pp. 24-35

Table of Contents

List of Figures

List of Tables

Notation

Variable	Meaning
$\mathcal{O}$	set of origin zones indexed by i
$\mathcal{D}$	set of destination zones indexed by j
$\mathcal{T}$	set of candidate IMTs indexed by t
$\mathcal{N}$	set of nodes on a transport network with cardinality n
$\mathcal{K}$	set of located IMTs indexed by t
$\mathcal{S}$	$= \{\{0, s, t\}\ \forall s \in \mathcal{T}, t \in \mathcal{T}\}$ set of elementary modal alternatives
$\mathcal{M}$	$= \{0, \mathcal{T}\}$ set elementary modal alternatives in the metropolitan market
$\mathcal{G}$	Set of cargo generation factors g_{ik} being factor for k for origin zone i
$\mathcal{H}$	Set of cargo attraction factors with a_{jl} being factor for l for destination j
$\mathfrak{M}$	$= \{0, \mathcal{K}\}$ set operational modal alternatives
$\mathcal{R}$	$= \{r = (i, j): i \in \mathcal{O}, j \in \mathcal{D}\}$ set of all origin-destination movements whose cardinality
$\mathcal{W}$	$= \{w = (r, t): r \in \mathcal{R}, t \in \mathcal{T}\}$ set of elemental alternatives
Z_w	Demand assigned to each elemental alternative $w \in \mathcal{W}$
$\mathbb{S}$	Solution space
$\mathcal{U}$	set of all subsets of $\mathcal{T}$ with cardinality p
τ	The cardinality of the set of candidate terminal locations $\mathcal{T}$
Y_t	equals 1 if an IMT is open at location t and 0 otherwise
Y_{ik}	equals 1 if node i is assigned to a hub at k and zero otherwise with $Y_{kk} = Y_k$ equals 1 if node k is a hub and zero otherwise
f_t	fixed cost of locating an IMT at t (\$ per day)
b_t	maximum handling capacity of IMT t (TEUs per day)
p	required number of IMTs to be locate
q_{ij}	Observed quantity of cargo to be transported from origin zone i to destination zone j (TEUs per day)
G_k	Weighted value of variable $k \in \mathcal{G}$
A_l	Weighted value of variable $l \in \mathcal{H}$
Q_{ij}	Estimateed quantity of cargo to be transported from origin zone i to destination zone j (TEUs per day)
q_i	Observed quantity of cargo produced in origin zone i (TEUs per day)
Q_i	Estimateed quantity of cargo produced in origin zone i (TEUs per day)
d_j	Oberved quantity of cargo arriving at destination j (TEUs per day)
W_{istj}	the quantity of q_{ij} transported intermodally through IMTs s and $t \neq s$ (TEUs per day) or regional intermodal transport demand with unit transport cost c_{istj} (in \$ per TEU)
V_{itj}	the quantity of q_{ij} transported intermodally through IMT t (TEUs per day) or metropolitan intermodal transport demand with unit transport cost c_{itj} (in \$ per TEU)
X_{ij}	the quantity of q_{ij} transported by road alone or trucks (TEUs per day) with unit transport cost c_{ij} (in \$ per TEU per day)
$\tilde{V}_{ij}$	Estimated demand for intermodal transport between each origin-destination pair (TEUs per day)

I_{ij}	Demand for intermodal transport between each origin-destination pair
L_{ij}	Measure of access to multiple modes for the movement of cargo between i and j
ℓ_{ij}	Measure of access to intermodal transport for the movement of cargo between i and j
$c^*{}_{ij}$	The minimum unit transport cost of cargo movements between each origin-destination pair ($ per TEU per day)
c_{ij}	The unit transport cost of cargo movements by truck between each origin-destination pair ($ per TEU per day)
Z_{istj}	The quantity of q_{ij} transported through IMT $s, t \in \mathcal{M}$ (TEUs per day)
δ_{ijm}	Equals 1 if mode m ($m = 1,2,3$) is available for that origin-destination pair and 0 otherwise
H	Shannon entropy or amount of missing information for a given probability distribution
S	Entropy
Z	the total quantity of cargo in the study area
E	the number of possible states of modal flows
$\Pr(x)$	the probability density function for random variable x
p_i	Probability of event/alternative i occuring
Λ	objective function
c	given budget of total transport ($ per day)
κ	Budget sensitivity parameter. $\kappa > 0$
Ω	Sample space
$\mathcal{E}$	Event space
Ξ	Set of zones in key cargo markets
$\mathbb{F}$	Set of all possible events of an experiment
γ	Alternative specific constants
λ	Structural parameters
θ, ϕ	Model parameters
T_B	Algorithm running time
vot	Driver's value of travel time savings ($ per min)
voc	Vehicle operating cost ($ per km)
α	Economies of scale factor ($\alpha \geq 0$) from concentration of flows
r_k	Maximum distance or cost between hub k and the demand nodes allocated to it

Glossary

Glossary	Meaning
ABS	Australian Bureau of Statistics
ASGC	Australian Standard Geographical Classification
ATRI	America transport research institute
BITRE	Bureau of Infrastructure, Transport and Regional Economics
CD	Collection Districts
CDM	Cargo demand model
CDP	Cargo demand problem
CFM	Cargo flow model
CFP	Cargo flow problem
DoFD	Department of Finance and Deregulation
EC	European Commission
EM	Entropy maximisation
EMFLP	Entropy maximising facility location problem
EMFL+VDP	Entropy maximising facility location with variable demand problem
FLP	Facility location problem
GMA	Greater metropolitan area
ICTM	Inland containerised transport market
IMT	Intermodal terminal
IPART	Independent Pricing and Regulatory Tribunal of New South Wales 2008, Reforming Port Botany's links with inland transport
IMTLP	(Inland) intermodal terminal location problem
IMTL+VDP	Intermodal terminal location with variable cargo demand problem
MCM	Mode choice model
MCP	Mode choice problem
MCTM	Metropolitan containerised transport markets
MIMTLP	Metropolitan intermodal location problem
NCHRPR	National Cooperative Highway Research Program Report
RCTM	Regional containerised transport market
RIMTLP	Regional intermodal terminal location problem
SFC	Sea freight council
SLA	Statistical local area
TCM	Terminals choice model
TEU	Twenty Foot Equivalent Unit
QMLE	Poisson quasi-maximum likelihood estimator
UNCTAD	United Nations Conference on Trade and Development
VDP	Variable cargo demand problem

Chapter 1 Introduction

1.1 Background

Prior to the advent of containerised shipping in the 1950s, the various modes of transport in the freight system (ships, rail and trucks) were largely used independently, un-integrated and considered as competing modes (Rodrigue and Slack 2013). It was also often the case that through public policy in many countries a company could own and operate only one mode of transport (see Rodrigue and Slack 2013) resulting in fierce competition between transport companies with each exploiting the advantages of the mode under its control regarding modal flexibility, reliability, economies of scale (to drive down costs), environmental friendliness and safety. The biggest barrier to mode integration was, however, not the existence of public policies barring companies from owning more than one mode but rather the difficulties and high costs associated with transferring cargo between modes in a single journey. These difficulties were largely due to the loading units that were employed; sacks, bales, crates, boxes and barrels requiring extensive labour force and taking several hours or days to load and unload the same amount of cargo, which now takes few man-hours to handle at ports (Cudahy 2006). Some studies have estimated that about 20 dock workers were required to load/unload the same amount of cargo that is now loaded/unloaded by one person (Broeze 2012).

Following the container revolution in the 1950s first championed by Malcolm McLean (Cudahy 2006) and the standardisation of container sizes (notably 20-foot and 40-foot container sizes) across the shipping and transport industries in the 1970s, a new and more sustainable means of freight transport emerged – *intermodal transport*. The standardisation also meant that containers could be stacked more efficiently and more importantly, ships, trains, trucks, terminals, cranes and other related equipment could be built to a single size specification. It also removed the potential risks that could have prevented major shipping

companies from investing in the new technology for the fear of it turning obsolete (Cudahy 2006). The basic principle underlying the use of intermodal transport is to combine the strengths of the various modes of transport (e.g., rail, air, inland waterway, sea, and road transport) for the efficient and sustainable shipment of cargo from anywhere to anywhere in the world with minimum complications or disruptions. The European Commission definition of intermodal transport involves the movement of goods in one and the same *loading unit* without handling the goods themselves when changing modes (European Conference of Ministers of Transport 1997). The loading units are often containers, pallets, swap bodies or semi-trailers. This thesis focuses on the container as it is by far the main loading unit for freight and as noted in Levinson (2006) the lifeline of intermodal transport and hence international trade. In this thesis, loading unit and container are used interchangeably.

The intermodal transport concept allows a single journey[1] to be segmented into several connected legs, where the most appropriate and/or cost effective mode is used along each leg. The attractiveness of this method of transport lies in the fact that each mode has its own advantages and disadvantages as each mode may differ in cost, speed, capacity, safety, efficiency and flexibility (Rodrigue and Slack 2013). Comparing rail and truck for example, rail has very high carrying capacity compared to truck and is generally attractive for long distance trips. However, rail is less flexible, usually operates in an inefficient environment and, more crucially, is less accessible to shipper's or customer's facilities. The truck on the other hand is very flexible and can access almost every facility connected to the road network. The truck, due to its limited carrying capacity, has relatively high unit costs, which increases linearly with distance making the truck less attractive for long distance trips and in cases where the volume of goods to be transported is very high (Park et al. 1995). It is also worth noting that almost all commodities (except bulk commodities like coal or iron ore) can be carried for some distance by truck ('first' and 'last' mile logistics).

On a more global scale, consider the movement of cargo, typically expressed in terms of Twenty-foot Equivalent Units (TEUs), by a shipper from, say, warehouses in Shanghai to a consignee's warehouses in Sydney. The shipper can decompose the journey into the sea and landside legs. The creation of a sea leg means that a container vessel can be used on this leg to benefit from economies of scale and other safety and environmental benefits associated with

---

[1] Journey here, refers to the shipment of cargo from the shipper to the consignee

using the sea. The landside legs comprise movements from warehouses to the exporting port in Shanghai and the movements from the importing port in Sydney to various warehouses, where the containers can be finally emptied. The main mode used on these legs is the truck, due to its flexibility and direct access to shippers' or consignee's warehouses. However, if the warehouses can be accessed by rail, then the shipper may opt to use rail to benefit from economies of scale. An alternative option is to seek to avoid congestion problems around the port by first transporting the containers by rail or barge to an inland terminal and then transferring them onto trucks for onward movements to the warehouses. Thus, depending on the transport infrastructure and available freight facilities in the hinterland, the shipper can explore and benefit from several transport solutions. Although, these two modes of transport do still compete, they can complement each other by exploiting the advantages and minimising the disadvantages of each through integration.

The relative ease of integrating the various modes of transport has made it possible to efficiently move products and raw materials over a long distance and transfer them from one mode of transport to another without being opened or having its contents exposed to damage or theft, triggering unprecedented growth in international trade (Bernhofen et al. 2014). This was possible because intermodal transport or mode integration has resulted in a significant reduction in shipping times and the elimination of multiple handling of goods, which usually leads to damage and pilferage. These in turn led to a significant reduction in transport costs of international trade and reduction of congestion within ports. As shown in Figure 1.1, international trade has significantly increased by more than a factor of 10 since the adoption and standardisation of the container in the 1970s.

Intermodal transport has also made it possible to integrate once isolated factories or companies into a global network of international manufacturers (Cudahy 2006). As noted in Cudahy (2006), before Malcolm McLean popularised the idea of container shipping in 1956, the world was full of small manufacturers where almost 100% of the finished product would be consumed in the country of production. Today, purely local markets for goods of any sort have almost disappeared allowing consumers in any part of the world to enjoy unlimited varieties of goods made from other parts of the world (Cudahy 2006). Recent statistics (UNCTAD 2013) even show that containers now carry many more unfinished products than finished ones revealing the deep interdependency among world economies. The success of any intermodal transport system, however, critically depends on the geographical locations of

where the containers are transferred from one mode to another. These locations or nodes are called *intermodal terminals* and for the purposes of this thesis intermodal container terminals (IMTs). The crux of this thesis is to develop models for finding the best locations of these terminals.

The rest of the chapter is organised as follows; Section 1.2 discusses the critical role played by IMTs in promoting intermodal transport. Two types of IMTs were identified; port terminals and inland terminals. The section also contains a brief description of port terminals. Section 1.3 presents a discussion on the role of inland terminals in promoting inland intermodal transport as a viable and more sustainable alternative to road alone transport (e.g., trucks). The potential transport markets for these terminals are discussed in Section 1.4. Section 1.5 discusses the location decisions of these terminals and the factors governing these decisions. Section 1.6 defines the research problems, proposed methods for answering the research questions and presents the main contributions of this research. Finally, Section 1.7 presents the outline of the thesis.
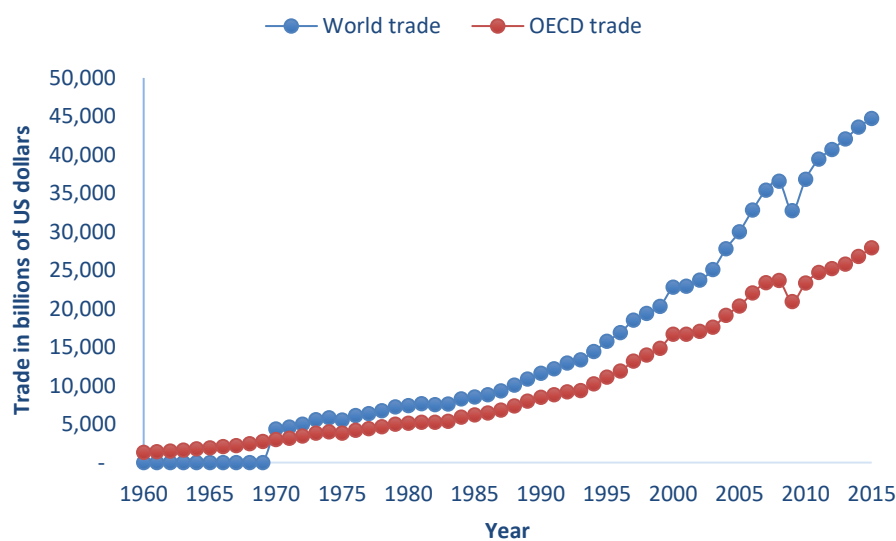
Figure 1.1: World trade since 1960 (Source: OECD Economic Outlook 2016)

1.2 Intermodal terminals

IMTs are an integral part and key promoter of intermodalism. It is the facility where the transfer of containers or loading units from one mode to another takes place. It can also be defined as a 'place equipped for the transshipment and storage of intermodal loading units' (EC 2006). The

Australian government (AHRCR 2007) defined it as any site or facility along the supply chain that contributes to an intermodal movement by providing efficient transfer of goods from one mode of transport to another. The transfers between modes can be done directly or through immediate storage, where the containers arriving at the terminals by one mode (e.g., trucks) are first stored before being loaded onto another mode. A typical IMT must therefore connect at least two modes of transport, which are usually a combination of rail, road and sea. The IMT can also be designed to have interfaces with other modes such as inland waterways or air transport networks.

Although the key feature or primary objective of IMTs is to receive cargo in one mode and transfer it to another mode for an onward journey, there are many auxiliary freight activities that can take place at the terminals. Some of these activities include temporary storage of the goods in their loading units, storage of empty containers, warehousing activities, consolidation activities, and repair and maintenance activities (Meyrick 2006). IMTs of this nature are usually referred to as logistics centres or freight villages. Consolidation/distribution activities could involve consolidating incoming goods of the same type from different sources or separating large volumes of goods into smaller outgoing shipments or direct transfer between modes as suggested by Feldman et al. (1996). In summary, an IMT can simply be a transfer point that provides a limited set of services, to a purpose-built hub, designed for transfers, storage, distribution and a host of associated services (BITRE 2016). IMTs can be classified into two types, depending on their geographical locations on transport networks; the *port terminals* and *inland terminals*, also called dry ports.

Port terminals are terminals located at ports with a direct interface to land transport networks such as rail and/or road networks (Crainic and Kim 2006). They serve as gateways for international trade where, for example, goods to be exported are first transported by rail or trucks to the terminals (of the exporting country) before being transferred onto vessels to be shipped to another country. Similarly, imported goods arriving at the port are first unloaded, temporarily stored at the terminal and then transported by rail and/or truck to consignee warehouses in the hinterland. These terminals sometimes act as *transshipment hubs*, with or without hinterland connections, where containers on small vessels from several ports are consolidated into a large vessel or containers on a large vessel from a given port is divided or distributed into small vessels and then shipped to their final destinations or other ports. A typical example of transshipment hub is the port of Singapore, one of the largest and busiest

ports in the world. However, most of the containers it handles have both their origins and destinations outside Singapore. A more comprehensive treatment of the subject can be found in Crainic and Kim (2006). The focus of this thesis is however, on the development of inland intermodal terminals to support more use of intermodal transport and less use of trucks for the movements of cargo between a given origin-destination pair in the hinterlands.

1.3 Inland terminals

The concept of developing inland container terminals to support intermodal transport is relatively new and largely motivated by the successes chalked up by port terminals in drastically driving down freight rates since the advent of the container. The costs of transporting containers between ports have been driven down further in recent years through the economies of scale of ever increasing container vessel sizes supported by the concept of hub-spoke networks, where as described above containers on small vessels (or large vessel) can be consolidated (de-consolidated) and loaded onto a large vessel (small vessels) for onward movement to another port terminal(s). These benefits are increasingly being offset by inefficiencies in landside container activities, especially the movements of containers between ports and various cargo origins/destinations in the hinterland (Norris 1994; UN 1992).

Land transport is a crucial element along the intermodal transport chain and as a matter of fact, the sea transport cannot operate efficiently if not supported by efficient land transport. This is because goods are generated or consumed in the hinterlands, implying that all goods movements by sea have their origins and destinations in the hinterlands and must be transported to or from the port. Inefficient and lack of sustainable inland transport therefore threatens the growth of international trade and can undo the efficiency gains in ports operations. It is to be expected that port activities can be severely constrained when vehicles arriving to receive or discharge containers are not managed properly. This situation is more acute for city ports like Sydney with increasingly large vessels calling, little room for physical expansion and where the dominant mode of land transport is the truck. Irrespective of the level of investment in port equipment, poor and inefficient hinterland transport can choke port operations, which can, in turn, affect the economy of the host country as the port is the main gateway of trade between the host country and the rest of the world.

Furthermore, not all goods generated are exported or all goods consumed are imported. Goods that are generated and consumed domestically also require efficient transport to move them from the regions of production to the regions of consumption. Increasingly, governments, port authorises and other stakeholders (e.g., carriers and shippers) are turning their attention to land transport, with a view to reducing the dependency on trucks by pursuing policies with the potential to shift freight away from trucks onto more sustainable modes such as rail or barge (EC 2011; NCHRP586 2007; AHRCR 2007).

An important reason why the rail, despite its economies of scale, has a very low mode share in freight transport tasks is largely due to low accessibility (BITRE 2016). Few companies or shippers have direct access to rail yards and in some cases the seaports themselves are not connected to the rail network. Thus, for most shippers, the use of rail is not an option. A promising strategy considered by many policymakers for increasing rail mode share is the development of inland IMTs that interface with both road and rail networks (EC 2011; NCHRP586 2007; AHRCR 2007; Meyrick 2006). For example, imported cargo can now be moved by rail to the inland terminal and then transferred to trucks for onward distribution to their destinations or intermediate warehouses. Similarly, cargo for export can first be consolidated at inland terminals before being transported to the port by rail for export. Additionally, intermodal transport can be used to move cargo from their production areas to their consumption areas in the hinterlands. For example, the movement of cargo from say Sydney to Melbourne can be done intermodally by first moving the cargo from the production areas in Sydney by trucks to a nearby terminal, where they can be transferred onto rail to be moved to a terminal in Melbourne close to the cargo destinations and then transferred onto trucks again for final delivery (see Figure 1.2). The development of inland terminals is therefore critical in making rail accessible to many shippers and thereby promoting more use of rail in the freight tasks.

The development of inland terminals can also be used by governments or local authorities to achieve certain policy objectives and may have to intervene in various ways to create the necessary environment to make the terminals viable and sustainable. One of such policy objective may be to move away from continuous road network expansions to accommodate freight growth to support more use of intermodal transport by investing in intermodal transport systems. It is generally acknowledged that expanding the road network to accommodate growth (both from passengers and freights) is at best a short-term congestion

alleviation strategy as newly created capacity is quickly taken up by induced demand and/or natural growth in demand such as increase in population and/or economic activities (Hills 1996). Additionally, the high cost of land, environmental concerns and physical barriers restricts expansion of existing road networks, especially in urban areas. Furthermore, the disproportionate impacts of truck related congestion on the urban fabric in terms of noise, pollution and safety have also been well documented in the literature (Ellison 2014; Ayres and Kneese 1969; Moseley 1973). Government intervention could take several forms including providing suitable land and related facilities for the development of the terminals (DoFD 2011), subsidising the use of the terminals, investing or improving the rail lines connecting the terminals and between the terminals and other important nodes like the port or giving tax incentives to truck companies willing to make their vehicles available for local pickups and deliveries along the intermodal transport chain (BITRE 2016). Governments can also discourage the use of road alone transport through some form of road pricing, especially around ports, to trigger mode shifts in favour of intermodal transport.

The truck industry also stands to benefit from intermodal transport solutions. The rapidly ageing workforce coupled with growth in freight has resulted in shortage of truck drivers in many advanced economies (Golob and Regan 2000; ATRI 2014). Also, truck drivers generally resent moving freight over long distances since they are often required to spend long hours alone, away from family and colleagues, together with the fact that truck driving is quite a difficult and sometimes dangerous occupation (Golob and Regan 2000). This adds to the daily frustrations experienced by truck drivers when they get stuck in traffic for several minutes or hours (Golob and Regan 2000). As a result, the number of truck drivers is decreasing whilst cargo volumes are increasing, and subsequently, the cost of trucking are rising (ATRI 2014). Intermodal transport solutions open up alternative and more environmentally sustainable solutions requiring fewer drivers.

Although intermodal transport services can be derived from various combinations of modes (road, rail, barge, air), this thesis focuses on rail and road as the main modes of transport for inland intermodal freight movement. The discussion above also reveals two main markets for inland intermodal transport; the regional containerised transport market (RCTM) and metropolitan containerised transport market (MCTM). Detail discussions about these markets are presented below.

1.4 Inland terminal markets

Two main markets for inland IMTs as transfer nodes can be identified in the literature (Arnold et al. 2001; Meyrick 2006); the regional containerised transport market and the metropolitan containerised transport market. The regional market (Arnold et al. 2001; Meyrick 2006) comprises the movement of cargo between their production and consumption areas in the hinterland without the use of the seaport. The cargo is transported over long distances, usually between two urban or metropolitan regions or countries. Two main modes of transport are often available in this market; road alone and intermodal transport. Road alone transport involves the use of only trucks for the transport task. The intermodal transport mode used in this market comprises two terminals as transfer nodes along the intermodal transport chain. The cargo is first consolidated at a terminal close to the cargo origin using trucks and then transported by rail (or other high capacity mode) to another terminal close to cargo destinations where the cargo is finally distributed by truck to their various destinations as shown in Figure 1.2. This type of market is very common in Europe (EC 2011), America (NCHRP586 2007) and Australia (AHRCR 2007; Meyrick 2006) and the development of terminals in this market could be considered as the traditional concept of inland intermodal transport, where both economies of scale and distance are the key drivers in both the location and use of IMTs.

The metropolitan market comprises the export and the import markets and are sometimes referred to as the IMEX (import/export) containerised markets. Here again, two main modes are available; road alone and metropolitan intermodal transport. Metropolitan intermodal transport is a relatively new concept of intermodalism, largely motivated by ports, especially city ports, experiencing high growth in container throughput, increasing port calls by larger vessels, limited physical space for expansion and lack of investment in inland transport infrastructure connecting with the ports, resulting in congestion, safety and environmental problems around ports in addition to choking port operations and increasing the cost and unreliability of container pickups and deliveries. The use of metropolitan intermodal transport in the IMEX market requires the use of only one IMT along the intermodal transport chain, where the seaport is either the cargo origin (import markets) or cargo destination (export markets) and the mode of transport between the port and the terminal is a high capacity mode such as rail. The geographical locations of these terminals do not necessarily have to be within the metropolitan region, they can be located anywhere in the region or outside the region or even at the periphery of the region provided it leads to the use of only one IMT along the

intermodal transport chain. In Australia, for example, the IMEX market accounts for over 99% of all containerised volumes through Australians ports (Piyatrapoom et al. 2006) and over 86% of these volumes are transported to their various destinations by trucks (Shipping Australia, 2011) resulting in road congestion, safety and environmental problems in the vicinity of the ports.

The next section discusses the location decisions of terminals in both markets and their significance in promoting inland intermodal transport use.
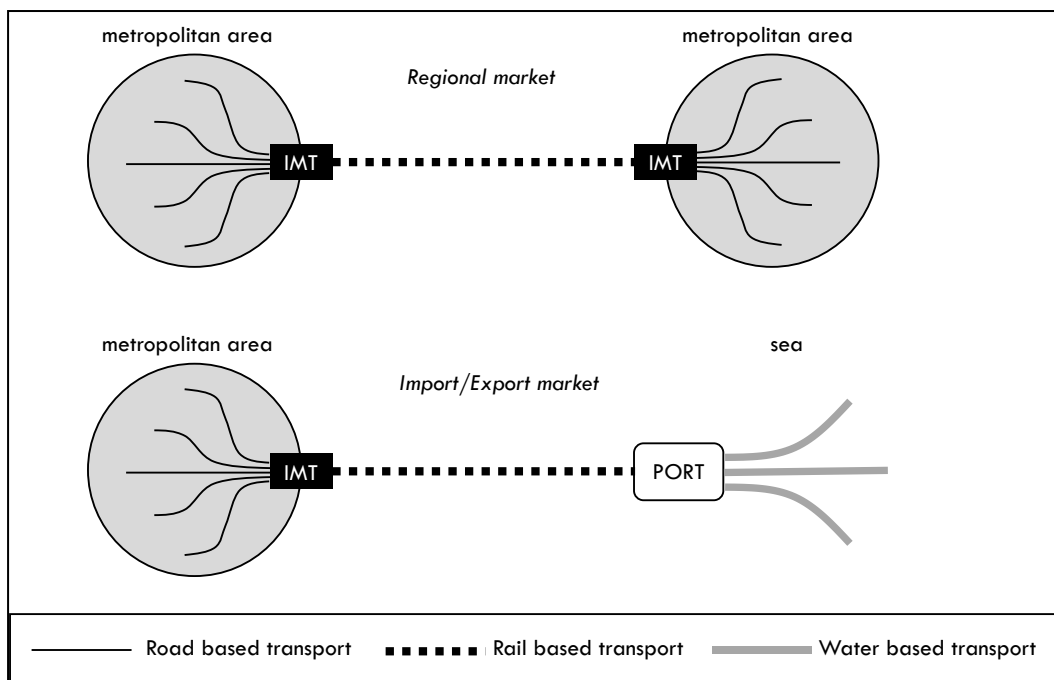
Figure 1.2: Inland intermodal transport markets

1.5 Location decisions

The success or failure of any intermodal terminal depends on it geographical location with respect to cargo origins and destinations. The locations of these terminals are crucial for their survival and are also the key facilitator of intermodalism which in turn drives economic growth. It is hard to imagine any other factor or characteristic of intermodalism or terminal activities that is not influenced by where the terminal is located. If the wrong location is chosen, then whatever facilities or features the terminal possesses may be of little use.

The location also determines users' cost of intermodal transport and whether or not it can compete with road alone transport (e.g., trucks). The cost of using inland intermodal transport comprises three main components; the road leg costs, terminal costs and rail leg costs. Each cost component occupies a certain percentage of the total transport cost with the share depending on the contribution of each mode in the transport task. Crucially, the direct factors governing one cost component are usually related to those governing others. For example, it is to be expected that terminals located in urban areas will have higher user costs than those located in rural areas due to higher installation and operation costs, which are expected to be passed on to the user. However, urban terminals may be close to customers so may reduce the associated truck costs of reaching these customers. The rail and truck cost components form the main part of the total intermodal transport cost.

As noted earlier, the comparative advantage of using intermodal transport hitches largely on the economies of scale and economies of distance of rail. For illustrative purposes, consider intermodal transport use in the MCTM where in this example containers are moved from the port to a warehouse in the metropolitan region as illustrated in Figures 1.3 and 1.4. The only difference between the two figures relates to the location of the terminal with respect to the port and cargo destination in the metropolitan region. In Figure 1.3, the terminal is closer to the cargo destination than in Figure 1.4. The cost of each mode (rail or truck) is assumed to be made up of a fixed cost which is independent of distance or the travel time covered by the mode and a variable cost which is a function of distance and/or journey time. It is to be expected that the fixed cost of rail would be relatively higher than that of the truck (BITRE 2016). However, the economies of scale advantage of rail mean that the overall cost of rail depending on the cargo volumes and distance involved could be lower than that of the truck as illustrated in Figure 1.3 and 1.4.

For intermodal transport to be competitive its total cost (fixed and variable) must be comparable to road alone transport cost and this critically depends on the location of the terminal. For example, in both figures, the total transport cost of intermodal transport is the sum of the fixed (C_3) and variable (C_4) cost of rail and the fixed (C_5) and variable cost (C_6) of truck. As expected, the fixed cost of rail is higher than that of road alone ($C_3 > C_1$). However, due to the economies of scale benefits of rail, the variable cost of rail ($ per TEU) is lower than that of road alone ($C_4 < C_2$) such that there exists a certain distance from the port (the break-

even distance) beyond which the overall cost of rail transport becomes lower than that of road alone transport. In Figure 1.3, the location of the IMT resulted in significant cost savings by rail, which is more than enough to compensate for the additional cost of local distribution by truck (made up of C_5 and C_6), resulting in the total cost of intermodal transport being lower than that of road alone. In Figure 1.4, however, the cost savings by rail is not enough to make up for the extra cost of local truck delivery, making intermodal transport in this instance less competitive. The only difference between the two scenarios presented in Figures 1.3 and 1.4 is the location of the IMT along the intermodal transport chain. This example shows the critical role played by the physical location of the IMT with respect to cargo origins/destinations in making intermodal transport competitive to road alone transport. If the terminal it too close the port, the distance covered by the truck leg will be relatively longer and can undo the benefits that the rail might bring through its economies of scale. In other words, the high cost of one mode can undermine the cost advantage of other modes and may render intermodal transport less competitive. IMTs therefore need to be strategically placed at locations where intermodal transport use will be an attractive modal option to as many shippers as possible.

Additionally, the inland IMT has over the years evolved from a simple transfer node where containers are transferred between two modes (e.g., road and rail) to an extended zone that has taken on additional logistics tasks such warehousing, empty container storage and sometimes port-related activities such as customs and quarantine (BITRE 2016). Revenues from these auxiliary activities are considered vital for the viability and sustainability of many inland terminals especially metropolitan intermodal terminals, partly due to the high setup or rental cost and operation costs associated with these terminals and the need for it operate daily throughout the year (DoFD 2011; Meyrick 2006). It is therefore very important that any procedure used for the selection of the best inland IMT(s) also accounts for these potential revenue generating markets (auxiliary activities).
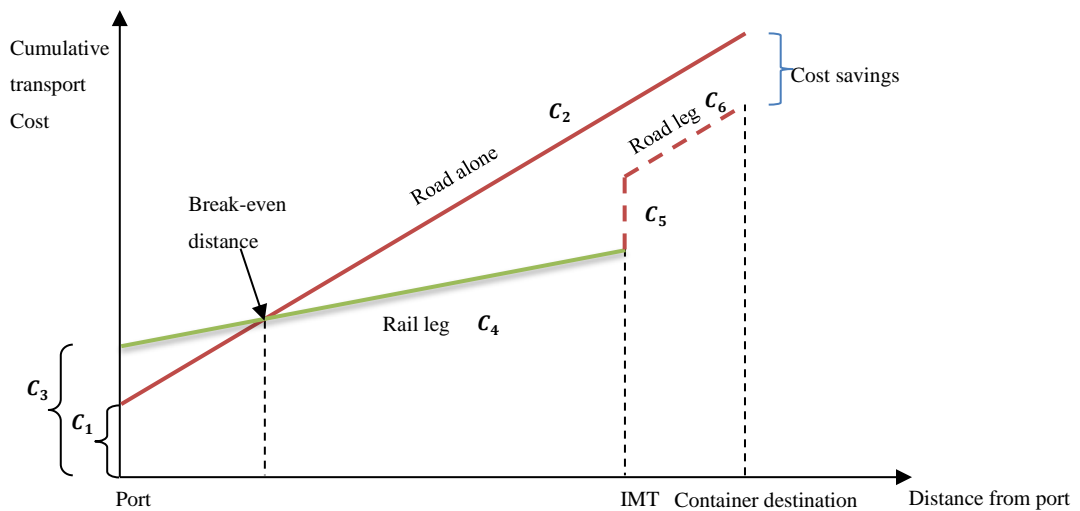
Figure 1.3: Competitive intermodal transport (Source: BITRE 2016 modified)
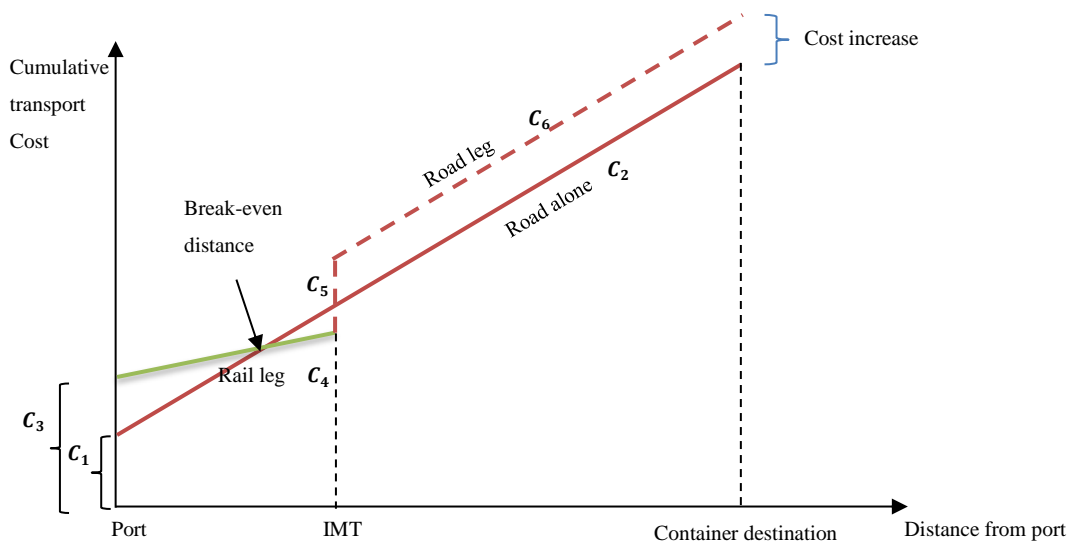
Figure 1.4: Less competitive intermodal transport (Source: BITRE 2016 modified)

1.6 Contribution to literature and practice

The above analysis shows that the geographical distribution of inland IMTs with respect to cargo origins/destinations are key in promoting inland intermodal transport use. An important feature of the IMTs under consideration is that they are open access facilities where a shipper has the choice of using them as part of an intermodal chain or use road alone transport mode (trucks) in the transport task. There are, however, many instances where market forces alone may not be enough to make intermodal transport (especially metropolitan intermodal transport) competitive to road alone transport. In these instances, some form of government intervention

in the form of subsidies or road pricing may be justified to make intermodal transport more competitive. Given the potential environmental benefits of IMTs, it is to be expected that policymakers will be keen on terminal location and demand estimation tools that are responsive to policy variables to help evaluate and select the best policy to promote more use of intermodal transport.

The aim of this research is to develop mathematical models to enable policymakers, both government and private sector, to determine the best locations for current and future IMTs. The model presented here will provide policymakers with a better understanding of the intermodal transport system in general and the means of testing various policy instruments to support more use of intermodal transport. The model can also be connected to an existing transport network model to identify hot spots and bottlenecks on the transport networks that need treating, among others. Additionally, the model can be used to understand factors governing the distribution of freight and the choice of mode and interactively forecast freight volumes by modes and by each IMT. The model can also provide the required inputs to support a business case for an IMT and help identify lands and other resources to reserve for the future development of IMTs.

To achieve the above objective, this thesis introduces a novel framework for locating multi-user facilities and specifically inland IMTs. More importantly, models developed under this framework are suitable for forecasting and testing of various policy instruments. The proposed framework is underpinned by the principle of entropy maximisation or information theory where terminals are located to maximise shippers' or users' welfares. The problem of locating IMTs or intermodal terminal location problem (IMTLP) analogous to the classical facility location problem comprises two linked problems with conflicting objectives; the location problem and the allocation problem. The location problem determines the exact locations of the terminals with the objective of keeping the costs of installations as low as possible, whilst the allocation problem determines the usage of the located terminals with the objective of keeping the transport costs of accessing the terminals as low as possible. These two problems are linked and cannot be solved separately since their objectives are in conflict and therefore require some degree of trade-offs as shown in Figure 1.5.

The multi-user feature of the problem and the existence of a competing alternative mode (road alone) means that the allocation part of the problem can be cast as a mode choice problem (MCP), where potential users of the terminals are assumed to face a choice of choosing the

among the available transport modes (road alone transport versus intermodal transport) the mode they perceive to offer them the highest utility (or least disutility or cost) for the transport task and where the choice of intermodal transport leads to the use of one of the IMTs. Thus, the demand associated with each located IMT is expected to be the outcome of many individual mode choice decisions. As noted by McFadden (1974), in a choice situation not all factors affecting the choice process are known to the analyst or can be quantified and included in the modelling process, making a probabilistic description of modal choices desirable.

Additionally, the choice of mode depends on the cargo origin and where the cargo is destined. Intermodal transport may not be feasible or cost competitive if the cargo destination is sufficiently close to the cargo origin. Conversely, the choice of cargo destination depends on modal accessibility. This implies that cargo origin and destination must be connected to the transport network and must be accessible by at least one available mode of transport. This reveals a link between cargo production and distribution and mode choice, where the choice of mode is conditioned by cargo production and the choice of cargo destination, whilst mode choice influences the production and distribution of cargo as illustrated in Figure 1.6. This leads to three linked problems; facility location problem (FLP), mode choice problem (MCP) and variable cargo demand problem (VDP). The VDP comprises the production and distribution problems and provides a means of quantifying the demand of the located terminals due to auxiliary activities like warehousing or storage, where the terminal can be coded as cargo destination on the transport network. The study refers to the extended problem with VDP as IMT location with variable cargo demand problem (IMTL+VDP). Also for the sake of clarity, the combined MCP and VDP is referred to as the cargo flow problem (CFP) as shown in Figure 1.6. Thus, in applications where the production and distribution of cargo is fixed (not influenced by the choice of mode or changes transport network conditions), the CFP reduces to MCP, and the IMTL+VDP reduces to the basic IMTLP.
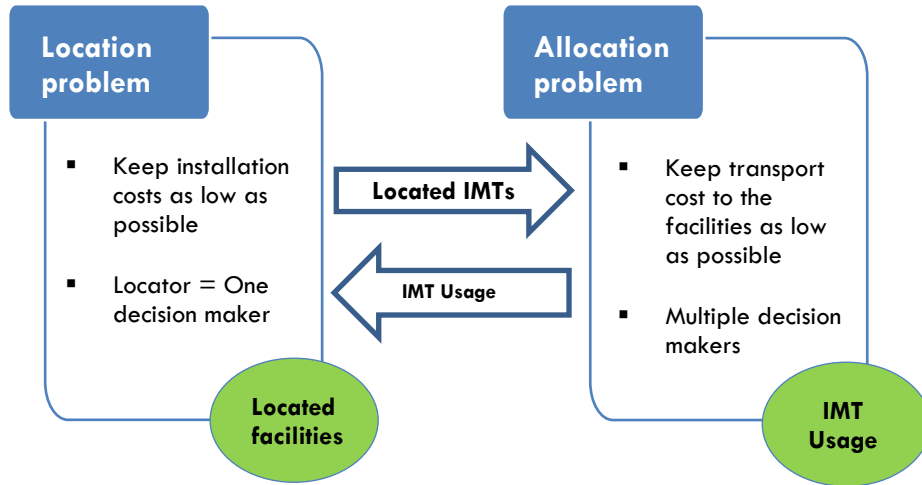
Figure 1.5: Basic intermodal terminal location problem
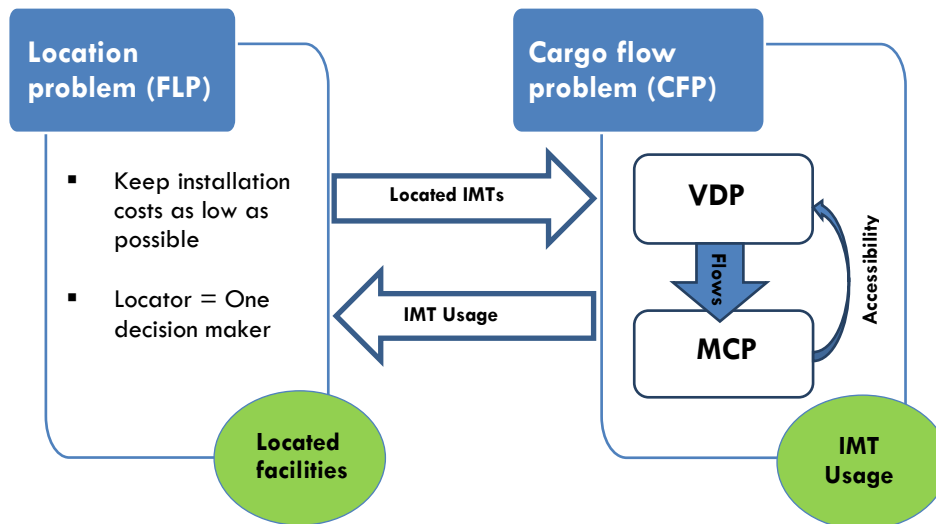
Figure 1.6: Terminal location with variable cargo demand problem

This thesis made several contributions to the literature. Perhaps, the most important contribution is the development of a modelling framework that links behavioural modal decisions and/or variable cargo demand problems with an FLP to determine the best locations of multi-user facilities in general and IMTs in particular and their expected usages. The proposed framework takes the form of a non-linear mixed integer programming problem and involves maximising an objective function subject to a set of constraints. The objective function to optimised is an entropy function describing all possible states of modal decisions and the constraints consist of a linked FLP, MCP and VDP. The framework locates terminal and generates probabilistic models for determining the expected usage of the located terminals. Once the best locations of terminals are determined, the CFP is converted into a nested logit

model suitable for forecasting and testing policies to promote more use of the located IMTs. Models for locating terminals in the metropolitan containerised transport market is first developed, followed by models for locating terminals to serve the regional containerised transport market. The latter model is generalised to also allow for locating terminals to serve the metropolitan or simultaneously serve both markets. Finally, the models developed are extended to incorporate variable cargo demand.

The second important contribution relates to the general formulation of the FLP, MCP, and VDP and how they can be expressed as constraints within the entropy framework. The formulations allow both decisions on facility location and facility allocation or usage to be driven by shippers or users' preferences with one objective function to optimise. Several properties of the entropy model are presented in the form of propositions including a general method of dealing with capacity constraints. An important outcome of this study is the demonstration of the link between entropy maximisation and welfare or consumer surplus maximisation. In other words, the proposed method allows IMTs to be strategically placed at locations where users' welfares are maximised.

The third contribution relates to the development of algorithms for solving the formulated problems. The general solution techniques employed involve decomposition the problem into FLP and the CFP using Lagrangian relaxation technique and developing algorithms to solve each sub-problem. The solution to the CFP (both the MCP and VDP) involves conversion into a behavioural nested logit model to explain the choice behaviour of facility users. To solve the overall model, two main general solutions are proposed; complete enumeration and an entropic greedy heuristic algorithm. The complete enumeration algorithm is proposed to deal with small to medium sized problems and proved to be very useful, especially for locating terminals to serve the metropolitan container market. It also provided a benchmark for gauging the quality of the proposed heuristic for solving large problem instances. The heuristic algorithm was primarily developed for locating terminals to serve the regional market. The geographical region making up the regional intermodal market is usually large and can encompass a whole country or several countries or economic regions and therefore required a more efficient algorithm. The computational efficiency, solution quality and properties of the heuristic algorithm are also presented.

Finally, the implementation of the models in practice, illustration of key model features and use in practice are demonstrated through a case study implementation. Specifically, the model is used to determine the best places in Sydney Greater Metropolitan Area (GMA) to locate terminals to serve the import containerised market. The full model comprises linked FLP, MCP and the VDP formulations. The factors governing the models are discussed followed by the use of the model in forecasting and testing of various policies. Some of the policies tested include changes in land use patterns, road pricing, subsidies, and strategic transport (rail and road) network expansion or improvements to support more use of intermodal transport in the Sydney region.

1.7 Outline of the thesis

The remainder of the thesis is organised in the following way. Chapter 2 discusses the existing literature on IMTLPs, identifies specific gaps in the literature, formulates the research questions and proposed methods for answering the questions. The principle of entropy maximisation and its suitability for developing models for answering the research questions are presented in Chapter 3. In Chapter 4, models suitable for locating terminals to serve the metropolitan containerised market with solution algorithms are developed. The proposed model is underlined by the principle of entropy maximisation and based on the assumption of fixed origin-destination cargo matrix. Chapter 5 addresses the second research question by generalising the models in Chapter 4 to be suitable for locating terminals to serve the regional containerised market and terminals to serve both markets. This chapter also contains a new solution algorithm for solving the generalised problem. The properties of the algorithm including solution quality and computational efficiency together with extensive numerical examples are also presented in this Chapter.

The model proposed in this chapter is also underlined by the fixed origin-destination cargo matrix assumption. Chapter 6, relaxes this assumption by replacing the fixed cargo matrix with cargo production and distribution models to allow changes in cargo production and distribution patterns to affect modal decisions which can in turn influence location decisions. The chapter contains various ways of formulating the cargo production and distribution problem and how they can be expressed as constraints within the entropy framework. Chapter 7 presents a case study implementation of the model developed in the previous chapters. The chapter specifically looks at the location of terminals within the Sydney greater metropolitan

area using import containerised data of the study area. Data sources, estimation of transport costs by each mode, candidate IMTs and general description of the study area including analysis of the results and the testing of various policies are also presented in this chapter. Finally, Chapter 8 provides a summary of how the research questions were addressed, presents the key contributions made in this study to the literature and in practice and finally discusses the limitations of this study and the directions for further research.

Chapter 2 Literature Review

2.1 Background

The previous chapter presents the background and motivation of this study including the need for policy-oriented models to support the development of inland intermodal terminals. The purpose of this review is threefold; first to clearly identify and acknowledge what has been done in the literature in relation to the topic under consideration; identify specific gaps in the literature; formulate the research questions; and finally propose how this thesis address those questions.

Research on intermodal terminal location problems (IMTLP) is relatively young but has been receiving growing attention by both policy makers and academics for more than two decades (Arnold et al. 2001; Bontekoning et al. 2004). The IMTLP can be classified under a more general location problem called the Hub Location Problems (HLP). A hub is characterised by three main features; *Consolidation,* where flows are aggregated from different origins and dispatched to different destinations through other hubs in order to exploit economies of scale; *transfer,* where cargo can be re-directed to different destinations; and *Distribution* where large cargo flows can be decomposed into smaller ones and then distributed to several nodes or destinations (O'Kelly 1987; Alumur and Kara 2008). In addition to these three main features, auxiliary activities such as warehousing, storage and sorting and other freight related activities can be performed in a hub. All the above features also apply to intermodal terminals, making intermodal terminals effectively hubs. The IMTLP also share common features with HLPs, especially in relation to the construction of objective functions, treatments of economies of scale discounts and the fact that both hubs and IMTs act as transfer nodes and can perform value added services.

The review of the literature therefore, starts with work on HLP, followed by review of work on IMTLP and a discussion of the common features between the two problems. These are followed by identifying the gaps in literature on IMTLP that this research aims to fill. The

gaps are synthesised and formulated as research questions followed by the proposed method for answering these questions. The rest of the chapter is, therefore, organised follows; Section 2.2 synthesised and summarised the work on hub location problems in the literature, followed by the literature on intermodal terminals location problems in Section 2.3. The identified gaps in the literature and how this thesis intends to fill them are presented in Section 2.4. Section 2.5 presents the research questions and the proposed methodology for addressing the questions is presented in Section 2.6.

2.2 Hub location problems

HLPs have been investigated by many researchers since the pioneering work by O'Kelly (1987). A recent review by Alumur and Kara (2008) cited over 100 papers related to the problem. The HLP considered by O'Kelly (1987) can be stated as follows;

"*Given a set of demand nodes (flows between origin-destination pairs), locate p-hub facilities at candidate sites to minimise the total transport cost to serve the demands'*.

In formulating the problem, O'Kelly (1987) assumes that freight origins, destinations and candidate hub locations are nodes that interact on the network. Each node is assumed to be assigned to exactly one of the p-hubs (where p is the number of hubs to locate) and all hubs are connected to each other, allowing the movement from one hub to any other hub in the network. For example, if cargo origin node $D_1 \in \mathcal{N}$ is assigned (or connected) to hub $Y_1 \in \mathcal{N}$, and cargo destination node $D_5 \in \mathcal{N}$ is assigned to hub $Y_2 \in \mathcal{N}$, then the flow of cargo from D_1 to D_5 must be first be routed from D_1 to Y_1, then from Y_1 to Y_2, and then finally from Y_2 to D_2 as shown in Figure 2.1a. The mathematical formulation of the HLP due to O'Kelly (1987) can be presented as follows:

$$(O'\text{ Kelly }1987): \text{Min } \Lambda = \sum_i \sum_j q_{ij} \left(\sum_k c_{ik} Y_{ik} + \alpha \sum_k \sum_m c_{km} Y_{ik} Y_{jm} + \sum_m c_{jm} Y_{jm} \right)$$

Subject to:

$$\sum_i Y_{ik} \leq (n - p + 1)Y_{kk}; \quad \forall k \tag{2.1}$$

$$\sum_k Y_{ik} = 1 \; ; \quad \forall i \tag{2.2}$$

$$\sum_k Y_{kk} = p \tag{2.3}$$

$$Y_{ik} \in \{0,1\}; \forall i, k \tag{2.4}$$

In the above formulation, the location variable Y_{ik} equals 1 if node i is assigned to a hub at k and zero otherwise and Y_{kk} equals 1 if node k is a hub and zero otherwise. Here, the flow variables q_{ij} are the quantity of flow between nodes i and j. By construction $q_{ii} = q_{jj} = 0$ and c_{ij} is the transport cost of a unit of flow between node i and j and p is the total number of hubs to be constructed with n being the total number of nodes in the transport network to be interconnected.

The objective function Λ comprises three main terms; the first term (collection costs) captures the weighted transport costs of collecting flows to the assigned hubs; the second term (transfer costs) captures the weighted costs of transfer of flows between hubs. These inter-hub costs are multiplied by a discount factor $0 \leq \alpha \leq 1$ to reflect the economies of scale effects in inter-hubs flows. The third term (distribution costs) reflects the weighted costs of distributing flows to their final destinations. Constraint (2.1) ensures that no node is assigned to a location unless a hub is opened at that site and recognizing that nodes can only be assigned to hubs, and that at most $(n - p + 1)$ nodes can be assigned to any hub (including the hub itself). Constraint (2.2) ensures that each node is assigned to one and only one hub. Constraint (2.3) locates the correct number of hubs. Constraint (2.4) ensures that a hub is either opened or closed. The phrase 'open hub' is used to mean a location where a hub is operating; for this location $Y_{kk} = 1$, conversely a 'closed hub' is a location where $Y_{kk} = 0$.

The second term in the objective function is quadratic making O'Kelly's formulation a quadratic integer program and thus, very difficult to solve exactly for all instances. The benefits of linearizing the quadratic function (Kaufman and Broeckx 1978; Burkard and Stratmann 19780) come at the expense of additional variables and constraints. In addition to the quadratic function, the formulation includes integer variables, making it an NP-hard combinatorial problem (Garey and Johnson 1979) with no known algorithm to efficiently solve every instance of the problem.

Some of the obvious limitations of O'Kelly's formulation include the assigning of each demand node to exactly one hub and the fact that modal decisions are ignored in addition to preventing direct routing of flows between nodes. Subsequent studies on the subject have sought to relax some of these limitations and also expanded the problem to include other aspects of hub activities. The problem considered by O'Kelly (1987) was later classified under a more general class of HLPs by Campbell (1994) called the *p*-hub median problem (*p*-HMP). The other two classes of HLP identified by Campbell (1994) are; the *p*-hub centre problem (*p*-HCP) and the hub covering problems (HCP). The basic formulation of each sub-group together with their extensions and variants are presented below. The discussion starts with the *p*-HMP, followed by *p*-HCP and then finally the HCP.
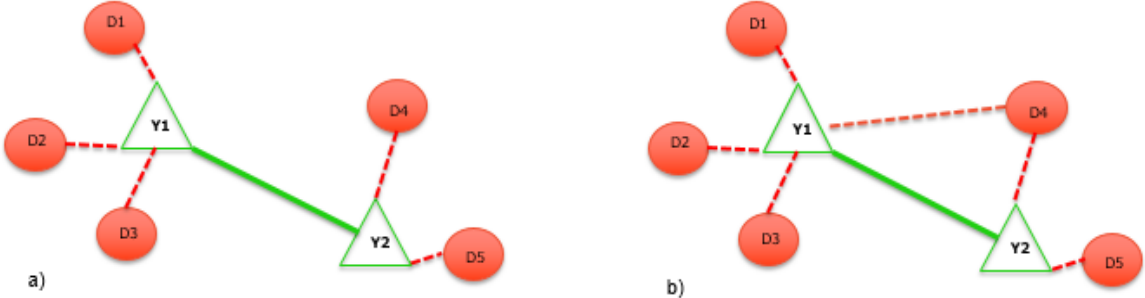
Figure 2.1: a) Single allocation   b) multiple allocations

2.2.1 *p*-Hub median problems

The model proposed by O'Kelly (1987) forms the basis of all problems in this group. All problems under this class are characterised by at least 4 features; 1. Every origin-destination path must visit at least one hub, 2. Exactly *p* number of hubs must be installed on the network, 3. All hubs are assumed to be connected to each other and inter-hub cost per unit flow is discounted by a factor α to reflect the economies of scale benefits from concentration of flows and 4. The objective function is often to minimise the weighted total transport cost of all flow movements. The economies of scale factor α as noted in Campbell (1994) plays an important role in determining the best locations of the *p* hubs and the assignment of demand nodes to the located hubs. It is expected and demonstrated by O'Kelly (1987), O'Kelly et al. (1996) and Campbell (1994) that as α decreases, hubs tend to spread farther apart and the number of spokes (The links or arcs connecting the demand nodes to hubs) decreases, since a lower inter-

hub transport cost favours allocation to the nearest hub. In the extreme case where $\alpha = 0$ the inter-hub cost will reduce to zero and each demand point will be allocated to exactly one hub (least cost hub) and the *p*-HMP collapses to the classical *p*-median problem (Hakimi 1965). For large α ($\alpha > 1$) hub interactions are expensive and hubs are drawn closer together to reduce inter-hub transport costs.

Several variants of the *p*-HMP exist in the literature with the probably the most noticeable one being how demands are allocated to hubs. Two types of allocations are discussed in the literature, the single allocation and the multiple allocations *p*-hub median problems. Single allocation *p*-hub median problems (SApHMP) assign each demand node to exactly one hub. The model by O'Kelly (1987) is a classic example of the SApHMP. Multiple allocation *p*-hub median problems (MApHMP) on the other hand allows the assignment of each demand node to more than one hub. The difference between these two problems is illustrated through Figure 2.1. In Figure 2.1a demand node D_1 is assigned to only hub Y_1 and it implies that flows to and from node D_1 can only go through hub Y_1. Under Figure 2.1b, demand node D_4 is assigned to hubs Y_1 and Y_2 allowing flows to and from this node to go through either hub. This example illustrates the restrictive nature of the SApHMP and may be unrealistic in many real applications especially in transport applications.

Nevertheless, research on SApHMP have been pursued by several authors including the work by Campbell (1994) who presented the linear version of the problem reducing the quadratic integer programming formulation (O'Kelly 1987) to integer linear programming formulation. Skorin-Kapov et al. (1996) show that the linear formulation by Campbell (1994) is not tight enough as it produces highly fractional solutions. They then presented a tighter formulation of the problem and demonstrated using the CAB data set (Ernst and Krishnamoorthy 1996) that their formulation almost always yields integral solutions using the CPLEX software. Ernst and Krishnamoorthy (1996) presented a new mixed integer linear programing (MILP) formulation of the problem also in an attempt to reduce the number of variables and constraints, which are directly linked to the computational time required to solve the problem. In their formulation, they treated the inter-hub transfers as a multicomodity flow problem where each commodity represents the traffic flow from a demand node. They showed the computational benefits of their new formulation using the AP (Australian Post) data set which uses different discount factors for collecting and distributing flows.

Table 2.1 presents a summary of other relevant work on this class of problem and includes the work by O'Kelly and Bryan (1998) who focussed on the effects of the economies of scale factor α on hub locations and potential usage. Sohn and Park (1998) presented a new formulation of the problem using fewer decision variables and constraints. The formulation by Ebery (2001) reduced both the number of decision variables and constraints to the order of $O(n^2)$, making it theoretically the most computationally efficient model on the subject, with n being the number of nodes on the network. However, it was shown that in practice the formulation by Ernst and Krishnamoorthy (1996) is computationally more efficient (Alumur and Kara 2008).

Studies on MApHMP were first conducted by Campbell (1992). He presented an integer linear programming (ILP) formulation of the problem with a total of $O(n^4)$ binary variables and $O(n^4)$ linear constraints. He noted that an optimal solution to the problem exists if the capacity constraints on links are relaxed. Skorin-Kapov et al. (1996) again presented a tighter formulation of the problem with the required number of constraints reducing to the order $O(n^3)$. The model has been reported to return optimal solutions to many instances of the CAB data set and those non-optimal solutions they found were within 1% of the optimal solutions. Extending their work on the SApHMP to the MApHMP, Ernst and Krishnmoorthy (1996) also proposed a new formulation of the problem with significant improvements in computational efficiency. The number of binary variables reduce to the order $O(n^3)$ and required $O(n^2)$ constraints. Other noticeable work on the subject are summarised in Table 2.2, including Sohn and Park in (1998) formulation of the uncapacitated version of the problem and Sasaki et al. (1999) who considered a special case of problem where each route in the network uses only one of the located hubs.

Other important variants of the hub location problem include whether or not the amount of flows through the located hubs are restricted or unrestricted. The restricted versions are called capacitated HLP and work in the area includes Lin and Chen (2012), Stanimirovic (2010) and Ernst and Krishnamoorthy (1999), Aykin (1994). Most of the work described above including O'Kelly (1987) and Campbell (1994) are uncapacitated or unrestricted variants of the problem. Another variant of the HLP is the inclusion of the fixed costs of hub location in the objective function to account for the differential cost of land, labour and other factors in each candidate hub location. HLP with fixed cost of hub locations can be found in

Chen (2007), Topcuoglu et al. (2005), Klincewicz (1996), Aykin (1995) and Ernst and Krishnamoorthy (1998a).

Table 2.1: Summary of work on p-HMP (Single allocation)

Reference	Capacitated (Y/N)	With Fixed Cost (Y/N)	Comments
O' Kelly (1987)	N	N	Quadratic integer program and heuristics
Aykin (1990)	N	N	Exact algorithm
Klincewicz (1991)	N	N	Exchange heuristic
Campbell (1994)	N	N	First integer formulation
O' Kelly et al. (1995)	N	N	Lower bounding technique
Klincewicz (1992)	N	N	Tabu search and GRASP heuristics
Skorin-Kapov and Skorin-Kapov (1994)	N	N	Tabu search heuristics
Campbell (1996)	N	N	Heuristics
Ernst and Krishnamoorthy (1996).	N	N	New formulation, SA and B&B algorithms
O' Kelly et al. (1996)	N	N	New formulation
Smith et al. (1996).	N	N	Heuristics
Sohn and Park (1997)	N	N	Problem complexity
Ernst and Krishnamoorthy (1998a).	N	N	Shortest path based B&B
Pirkul and Schilling (1998).	N	N	Lagrangian relaxation heuristic
Sohn and Park (1998)	N	N	New formulation
Sohn and Park (2000)	N	N	Problem Complexity
Abdinnour-Helm (2001)	N	N	Simulated annealing
Ebery (2001)	N	N	New formulation
Elhedhli and Hu (2005)	N	N	Congestion cost function with Lagrangian heuristic
Stanimirovic (2010)	Y	Y	New formulation with Genetic algorithm
Alumur et al. (2009)	N	Y	Hub location in incomplete hub networks

Table 2.2: Summary of work on p-HMP (Multiple allocation)

Reference	Capacitated (Y/N)	With Fixed Cost (Y/N)	Comments
Campbell (1992)	N	N	First integer program
Campbell (1994)	N	N	New formulations with fixed costs
Campbell (1996)	N	N	Greedy interchange heuristic
Skorin-Kapov and O' Kelly (1996)	N	N	New formulation
Ernst and Krishnamoorthy (1998a)	N	N	New formulation, B&B algorithms and heuristics
Ernst and Krishnamoorthy (1998b)	N	N	Shortest path-based B&B algorithms
Sasaki et al. (1999)	N	N	1-stop problem with B&B algorithm & heuristic
Boland et al. (2004)	N	N	Preprocessing and tightening constraints
O' Kelly (1992)	N	Y	Single allocation hub location with fixed costs
Abdinnour-Helm (1998)	N	Y	Hybrid genetic and Tabu search heuristics

2.2.2 *p*-Hub centre problems

The *p*-hub centre problem (*p*-HCP) was introduced by Campbell (1994) as analogues to the classical *p*-centre problem (Hakimi 1965; Drezner 1984). However, the notion of "centre" in the *p*-HCP is different from that of the classical *p*-centre problem. The objective of the *p*-centre problem is to minimise the maximum weighted distance between each demand node and its closest facility (Hakimi 1965). According to Campbell (1994) *p*-HCP can be characterised by three main objective functions; the first is to minimise the maximum cost or distance for any origin-destination pair; the second is to minimise the maximum cost or distance of any single link in an origin-destination path and finally, to minimise the maximum cost or distance between flow origin or destination and a hub.

Clearly, the first objective function is naturally deduced from the objective function of the *p*-centre problem, where by definition a hub centre represents a set of hubs such that the maximum cost of moving flows between any origin-destination pair is minimised. Campbell (1994) noted that this type of hub centre is important for a hub system involving perishable or time sensitive items. The second objective function is based on the fact that flows between each origin-destination pair goes through a path consisting of three legs; flows from origin-to-hub, hub-to-hub and hub-to-destination and the objective is to minimise the maximum cost or distance of any of these three legs. The third objective was based on the concept of *vertex centre* problem (Hakimi 1965) in which the set of hubs minimises the maximum cost for movement between a hub and an origin/destination.

The *p*-HCP formulation presented here is due to Ernst et al. (2002) which has fewer decision variables and constraints and considered to be computationally more efficient:

$$(p - \text{HCP}): \text{Min } \Lambda$$

Subject to:

$$\sum_k Y_{ik} = 1 ; \quad \forall i \tag{2.5}$$

$$Y_{ik} \leq Y_{kk} ; \quad \forall i, k \tag{2.6}$$

$$\sum_k Y_{kk} = p \tag{2.7}$$

$$r_k \geq c_{ik} Y_{ik} \; ; \forall i, k \tag{2.8}$$

$$\Lambda \geq r_k + r_m + \alpha c_{km} \; ; \forall k, m \tag{2.9}$$

$$Y_{ik} \in \{0,1\}; \forall i, k \tag{2.10}$$

$$r_k \geq 0; \forall k \tag{2.11}$$

The variable r_k is called the radius of hub k and represents the maximum distance or cost between hub k and the demand nodes allocated to it expressed as constraint (2.8). Constraint (2.9) is the objective function to be minimised (the maximum cost or distance for any origin-destination pair). The rest of the constraints have similar interpretations as those under (O'Kelly 1987). Campbell (1994) presented the formulations for the other two objective functions and interested readers are referred to his paper.

The single and multiple allocation versions of the *p*-HCP also exist in the literature. Kara and Tansel (2000) for example show that the single allocation version is NP-complete. The multiple allocation variant was also proposed by Campbell (1994) and was shown to be NP-hard by Ernst et al. (2002).

2.2.3 Hub covering problems

The hub covering problems (HCP) are also analogous to the classical facility covering problems (Hakimi 1965), where demand nodes are considered covered if they are close enough to the facility and satisfy certain pre-defined thresholds. Covering problems are generally classified into two categories (Schilling et al. 1993); Set covering problems (SCP) which deals with the variant of the problem where coverage is required and the maximal covering location problem (MCLP) where coverage is optimised. Analogues to the hub covering problems, the hub set-covering location problem locate hubs to cover all demand nodes such that the cost of opening hub facilities is minimised, whilst for the maximal hub-covering location problems the number of demand nodes covered by a given set of located hubs is maximised (Alumur and Kara 2008). Hub covering problems are also characterised by three objective functions (Campbell 1994); 1. The cost of moving flows between any given origin-destination pair through two hubs does not exceed a specified value; 2. The cost of each leg in the path, origin-to-hub, hub-to-hub and hub-to-destination does not exceed a specified value; and 3. Each of the origin-hub and hub-destination links meets separate specified values. Campbell (1994) presented MILP formulations of the problem under the three objective functions.

Here again, the formulation due to Ernst et al. (2005) is presented as it is computational more efficient with less decision variables and constraints:

$$(\text{HCP}): \text{Min } \Lambda = \sum_k Y_{kk}$$

Subject to: Constraints (2.5), (2.6), (2.8), (2.10), (2.11) and

$$r \geq r_k + r_m + \alpha c_{km} \; ; \forall k, m \qquad\qquad (2.12)$$

where r is the predefined cover radius. The single allocation variant of the problem was studied by Kara and Tansel (2003) and proved it to be NP-hard. The work by Wagner (2004) includes both the single and multiple allocations variants of the problem. His formulations require less decision variables and constraints compared to those of Kara and Tansel (2003).

2.2.4 Other notable variants

2.2.4.1   Hub arc problems

Hub arc problems (HRP) are relatively new class of problems proposed by Campbell et al. (2005a and b) where instead of locating hubs, hub arcs are located. They noted that the assumption of complete graph of hub nodes (inter-connected hubs) underlying p-hub median problems imposes a topological and cost structure that may be undesirable in many applications. Indeed, there are many applications especially in the communication and transport industries where all the hubs do not need to be fully connected (Campbell et al. 2005a and b). They also observed that the economies of scale (discount) factor on hub links may result in some hub arcs having unrealistically lower flows than some hub access links.

The HRP consists of two interrelated network design decisions; hub arcs and access arcs. The *hub arc design* decision is to select the hub arcs and, as a consequence, the hubs that are the end points of these arcs (Campbell et al. 2005a). In comparison with the p-HMP, it is equivalent to locating hub nodes under the assumption of complete graph of hub nodes. The *access arc design* decision is to select the arcs connecting the flow origins and destinations to hubs. The single and multiple allocations variants of the p-HMP also apply to HRP. Campbell et al. (2005a) noted that the HRP can also results in fully connected hubs making the p-HMP

as a special case of the *p*-HRP. Campbell et al. (2005b) proposed integer programming formulations for several variants of the problem including algorithms for solving small instances of the problem with extensive computational analysis using the CAB data set.

2.2.4.2 Deterministic and stochastic hub location problems

The models considered so far are deterministic models as traffic flows, travel times or costs and other variables are assumed to be certain. In these situations, stochastic data are normally averaged and used in the models. The consequences of doing this has long been recognized in the literature as producing sub-optimal solutions (Yang 2009; Savage 2008; Wang 2007). Lium et al. (2009) also noted that uncertainty is a key promoter of hub-spoke-networks since consolidation of flows are done to hedge against uncertainty rather than benefit from economies of scale. The problem with the stochastic hub problem is that they are much harder to solve than their deterministic equivalents (Hult 2011).

Several researchers have extended some of the deterministic models to allow for stochasticity of some of the input variables and in some cases, exact solutions to small instances of the problem are provided. Sim et al. (2009) proposed a stochastic single allocation *p*-hub center problem with stochastic travel times on each link with the objective of ensuring that all path travel times are less than or equal to the maximum time by at least some probability γ. The stochastic uncapacitated multiple allocation hub location problem (UMAHLP) was researched by Yang (2009) where seasonal variations on demand, as well as seasonal variations on the discount factors for hub to hub flights were accounted for. The solution to the problem involves decomposing the problem into two stages, with the first stage determining the number and location of hubs, whilst the second stage accounts for the stochasticity of the seasonal demands and discount factors.

Contreras et al. (2011) investigated the stochastic version of the uncapacitated multiple allocation hub location problem (UMAHLP) where demand is stochastic and show that the stochastic model is equivalent to the deterministic model if demand is considered on an average basis. Their work also involves looking at different stochastic variants of UMAHLP where transport costs on all paths are stochastic. Other relevant works in this area include the work by Hult (2011) on the single allocation *p*-hub centre problem with stochastic travel times; Lium

et al. (2009), who recognized the role play by uncertainty in the evolution of hub-spoke-network structures and the work by Kim and O'Kelly (2009) on the development of both single and multiple allocation models to maximise the expected flow transmitted between origins and destinations, where edges and nodes have reliabilities defined as the probability that the hub or edge transmits flows for a given time period without failing.

2.2.4.3 Competitive hub location problems

Competitive hub location problems involve firms developing hubs to compete with other firms for customers. Several studies (Campbell and O'Kelly 2012; Eiselt and Marianov 2009; Sasaki et al. 2009; Sasaki et al. 2014) have shown that the location of hubs under competitive environments differ greatly in terms of optimal hub locations, network structures, and traffic patterns compared with hub locations under uncompetitive environments.

Work on hub location in a competitive environment include the work by Marianov et al. (1999) where the problem was formulated using a 0-1 integer programming technique and solved by tabu search heuristics. In their model, flows between each origin-destination pair can be routed through one or two hubs and each demand point can be assigned to one or more hubs depending on the traffic destinations. The proportion of each origin-destination demand expected to go through the hubs is determined by a pre-defined set. For example, the set can be defined such that the demand for all origin-destinations pairs is assigned to the new hub if the cost of using it (newly located hub) is lower than that of the existing or competitor hub(s). The set can also be defined such that only a certain percentage of demand is captured by the new hub if the cost difference is not very significant. An improved formulation was provided by Wagner (2008) with associated heuristics for solving it. They have shown that their improved model can be solved to optimality for problem sizes of up to 50 nodes in reasonable computation time. The allocation mechanisms in these two studies were extended by Eiselt and Marianov (2009), where the discrete demand allocation to hubs was replaced with proportional allocation based on relative costs between competing options.

Lüer-Villagra and Marianov (2013) studied the environment where an existing firm operates a hub and spoke network, and where a new entrant is assumed to maximise profit by choosing the best hub location and network topology applying optimal pricing assuming that

the existing company applies mill pricing[2]. The allocation of demand between the firms is logit-based resulting in a non-linear mixed integer programming problem, which was solved using a genetic algorithm. A game theoretic approach of locating competitive hubs was researched by Sasaki and Fukushima (2001), Sasaki (2004) and Lin and Lee (2010). In Sasaki and Fukushima (2001), a 1-stop continuous Stackelberg hub location problem as a bi-level programming problem is formulated and solved using a sequential quadratic programming method. Numerical examples illustrate the significance of rival companies when locating and designing hubs. Their proposed model used a logit function to determine the assignment of customers to available services so as to capture their various choice preferences, compared with the all-or-nothing procedure used in Marianov et al. (1999). To prevent the situations where firms carry an unrealistically low level of flow for any OD pair, Sasaki (2005) introduced flow threshold to ensure that only flows above this threshold are assigned to competing firms. The overall problem was formulated as a bi-level programming problem where the upper level model captures the decisions by the leader firm and the lower level model outputs the decisions of the follower firm.

Lin and Lee (2010) study an oligopolistic market environment. Their approach was motivated by the fact that industries such as freight or airlines can be considered as operating in oligopolistic markets, where prices are controlled by small groups of firms. Thus, a firm's hub network design is largely motivated by the actions or reactions of its competitors in pursuit of profits and market share rather than cost minimisation (Lin and Lee 2010). They showed that the long-term Cournot–Nash equilibrium steady state for the competition game is such that none of the firms may unilaterally change their respective hub networks, demands, or operational plans to increase profit, and also demonstrated that a dense hub network is more likely to earn a higher profit than a sparse hub network under price-elastic demand. The work by Sasaki et al. (2014) considered two firms competing for customers in a Stackelberg framework where the leader firm locates hub arcs to maximise revenue, given that the follower firm will subsequently locate its own hub arcs to maximise its own revenue. They presented an optimal solution algorithm that allocates traffic between the two firms based on the relative utility of travel through the competing hub networks and also demonstrated the importance of competition in locating and designing hub-based transport systems.

---

[2] A mill pricing firm sets a single price at its plant and customers bear the cost of transport.

Similar work in the airline industry includes Adler (2005), where a non-linear mathematical program model is proposed for evaluating the most appropriate hub-and-spoke network for an airline in a competitive environment, which is in turn used to derive a two stage, Nash-type, best-response game. The first stage of the two-stage game determines the choice of network by the airlines, whilst in the second stage, each airline competes for market share, given the other airlines' decisions. Adler and Smilowitz (2007) investigated global alliances and mergers in the airline industry under competition where they proposed a game theoretic framework incorporating a profit-maximising objective and a cost-based network design. Numerical examples show that some mergers may be more successful than others and that optimal international gateway choices change according to the number of competitors remaining in the market. Both studies used logit models to determine the market shares of each competing firm and in both models, global equilibrium cannot be guaranteed due to the non-convexity of the objective function. The location of park and ride facilities using p-Hub approach (where the park and ride locations were treated as hubs) was studied by Aros-Vera et al. (2013). The problem was formulated as a nonlinear programming problem with embedded logit model to determine the demand of the located park and ride facilities. Their solution approach involves a linearizing technique for transforming the nonlinear programming problem to equivalent MILP.

2.2.5 Summary of work on hub location problems

Section 2.2 provides a summary of work on hub location problems and identified three main categories (p-hub median/hub arc problems, hub centre problems, and hub covering problems) analogous to the three classes of classical facility location problems. Several variants of each category were also studied in the literature including single and multiple allocation of demand nodes to hubs; capacitated and uncapacitated hubs, fixed cost hubs and non-fixed cost hubs, deterministic versus stochastic hubs and hubs under competitive and uncompetitive environments. Each of these variants added some insights to the general problem of locating hubs in the literature and applications. Interested readers are also referred to review of hub location problems by Alumur and Kara (2008) and the recent review of facility location problems by Farahani et al. (2014).

Although IMTLP is a problem on its own right, work in the hub location literature seem to provide the basis of the work in the IMTLP literature both in terms of problem formulations

and solution algorithms. Before reviewing the work on IMTLP it is important to recognise the following important features of the HLP considered so far; 1. HLP do not recognise the existence of multiple decision makers when allocating hubs to demand nodes; 2. Direct flows between demand nodes are often ignored; 3. The availability of modes and the competition between the modes in the transport tasks are ignored in the formulations and 4. The demand matrix showing the amount of cargo flows between each origin-destination pair (or demand nodes) is assumed given and fixed. That is, it is assumed to be invariant to changes in transport network conditions, the location of the hubs or any other factors. All these are important features of the IMTLP.

2.3 Intermodal terminal location problems

Research on IMTLP has been receiving growing attention by both policy makers and academics (Bontekoning et al. 2004). The research work began with the mathematical formulation of the problem by Arnold et al. (2001). Arnold et al. (2001) focussed on locating intermodal terminals to serve the regional containerised transport markets, where the use of intermodal transport requires the use of exactly two IMTs along the intermodal transport chain. The chain consists of three leg; the pre-haul leg, the main haul, and the post haul legs as illustrated in Figure 2.2. The pre-leg involves local pickup of cargo by trucks from various cargo origins to the nearest IMT, and the post haul leg consists of local distribution of cargo to various destinations and is also done by trucks due to its flexibility and accessibility to customer facilities. The main haul is done by high capacity mode such as rail.
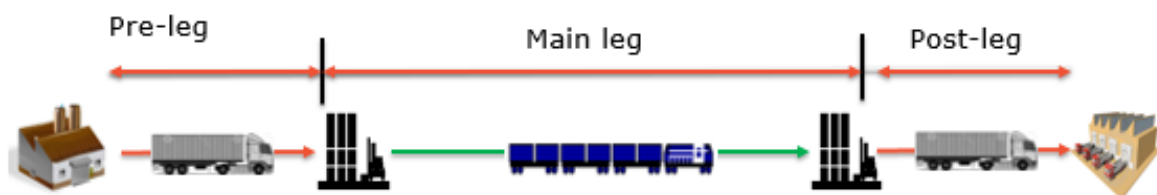
Figure 2.2 : Regional intermodal transport chain

This intermodal transport market as noted in Section 1.4 involves moving freight over long distances and its competitiveness depends on the economies of distance (lower unit cost per kilometer) and the economies of scale (through the use of rail) relative to road alone mode such as trucks. The model formulated by Arnold et al. (2001) can be stated as follows:

(Arnold et al. 2001): Min $\Lambda = \sum\limits_{i \in \mathcal{O}} \sum\limits_{j \in \mathcal{D}} \sum\limits_{s \in \mathcal{T}} \sum\limits_{t \in \mathcal{T}} c_{istj} W_{istj} + \sum\limits_{i \in \mathcal{O}} \sum\limits_{j \in \mathcal{D}} c_{ij} U_{ij} + \sum\limits_{t \in \mathcal{T}} f_t Y_t$

Subject to:

$$W_{istj} \leq q_{ij}\, Y_s \ ; \quad \forall t, s \in \mathcal{T}, i \in \mathcal{O}, j \in \mathcal{D} \tag{2.13}$$

$$W_{istj} \leq q_{ij}\, Y_t \ ; \quad \forall t, s \in \mathcal{T}, i \in \mathcal{O}, j \in \mathcal{D} \tag{2.14}$$

$$\sum\limits_{s \in \mathcal{T}} \sum\limits_{t \in \mathcal{T}} W_{istj} + U_{ij} = q_{ij} \ ; \quad i \in \mathcal{O}, j \in \mathcal{D} \tag{2.15}$$

$$\sum\limits_{i \in \mathcal{O}} \sum\limits_{j \in \mathcal{D}} \sum\limits_{t \in \mathcal{T}} W_{istj} + \sum\limits_{i \in \mathcal{O}} \sum\limits_{j \in \mathcal{D}} \sum\limits_{t \in \mathcal{T}} W_{itsj} \leq b_s \ ; \forall s \in \mathcal{T} \tag{2.16}$$

$$Y_t \in \{0,1\}; \forall t \in \mathcal{T} \tag{2.17}$$

$$W_{istj} \geq 0 \ , \ U_{ij} \geq 0, W_{ittj} = 0 \ ; \forall t, s \in \mathcal{T}, i \in \mathcal{O}, j \in \mathcal{D} \tag{2.18}$$

where W_{istj} the quantity of flow from origin i to destination j routed through IMT s and t with associated unit cost c_{istj}; Y_s equals 1 if IMT s is open and 0 otherwise; $\mathcal{O}$ is the set of all cargo origin nodes and $\mathcal{D}$ the set of destinations nodes; $\mathcal{T}$ is the set of all feasible IMT locations on the network; q_{ij} represent the quantity of cargo to be transported from origin i to destination j; U_{ij} the quantity of q_{ij} to be transported unimodally or by trucks from origin i to destination j with unit cost c_{ij}; b_s the maximum handling capacity of IMT s, and f_s the fixed or setup cost of IMT s. The objective function is composed of three parts; the first part captures the weighted cost of all intermodal transport flows; the second part represents the weighted cost of road alone transport, and the third part comprises the total fixed associated with all opened IMTs. Constraints (2.13) and (2.14) ensure that no cargo can be transported through an IMT, unless it is opened. Constraint (2.15) reveals the existence of competition between road lone and intermodal transport modes and stipulates that, for each origin-destination, the sum of all cargo flows transported by road alone and by regional intermodal transport should equal the demand associated with this origin/destination-pair. Constraint (2.16) enforces capacity limit on each opened IMT. Constraints (2.17) ensure that an IMT should either be opened or closed. Constraint (2.18) ensures that no demand is transported using only one IMT and also ensure that only non-negative amounts are transported.

The model formulated by Arnold et al. (2001) was associated with a large number of decision variables and constraints, making it difficult to efficiently solve for even small problem instances. This limitation motivated a new formulation of the problem by Arnold et

al. (2004). Their reformulated model is similar to the multicommodity fixed-charge network design problem (MCNDP), where IMTs are considered as arcs instead of vertices in a graph. The reformulation resulted in a significant reduction in the number of decision variables and constraints especially in sparse networks (Arnold et al. 2004). The generalisation of the intermodal location problem to include non-linear and concave cost functions to capture economies of scale effects on intermodal usage was developed by Racunica and Wynter (2005). The non-linearity and concavity of the transport cost function means that a linearization procedure is required to solve the problem. The algorithms they proposed includes a linearization algorithm for reducing the non-linear problem to linear and two heuristics for solving large instances of the problem.

Similar work by Rahimi et al. (2008) comprises a location-allocation model for locating hubs to promote the use of rail through the use of hub-and-spoke networks using a concave cost function to capture economies of scale resulting from freight consolidation at hub terminals. Limbourg and Jourquin (2009) cast the intermodal location problem as linked p-hub median problem and multi-modal assignment problem where the demand can be assigned over all the transport modes, with the option of using trans-shipment facilities. The model works by repeatedly solving the *p*-median problem for each update in trans-shipment costs, which in turn is based on previous estimated flow at each terminal until the relative difference in trans-shipment costs between two iterations is smaller than a pre-defined threshold.

Ishfaq and Sox (2011) employed the multiple-allocation *p*-hub median modelling approach to formulate a new mathematical model for IMTLP. Their proposed model includes important transport mode attributes such as transport cost, modal connectivity costs, and fixed location costs under service time requirement. Large instances of the problem were solved using a tabu search metaheuristics algorithm with the quality of the solution measured against lower bounds from a Lagrangian relaxation. Ishfaq and Sox (2012) extended their previous work (Ishfaq and Sox 2011) with the integration of a queuing system to model hub operations and investigated the effects of limited hub resources on the design of intermodal logistics networks under service time requirements. The features of the model were illustrated using a 25-city road-rail intermodal logistics network. Sorensen et al. (2012) proposed meta-heuristic algorithms; the greedy randomised adaptive search procedure (GRASP) and the attribute based hill climber (ABHC) for solving the model proposed in Arnold et al. (2001).

To capture the joint effects of CO_2 emissions and economies of scale on intermodal terminal location, Zhang et al. (2013) proposed as bi-level programming, where a genetic algorithm was used at the upper level to search for the optimal terminal network configurations, while the lower level performs multi-commodity flow assignment over a multimodal network. A similar work by Qu et al. (2016) was conducted but their objective focussed on the effects of greenhouse gas emissions and intermodal transfers on intermodal network design. The resulting non-linear model was linearized and solved for a hypothetical case study of eleven candidate locations in the United Kingdom. Recently, Ghane-Ezabadi and Vergara (2016) proposed a new mathematical formulation and decomposition based solution algorithm for designing intermodal networks. The novelty in their approach was the use of composite variable in representing a complete route for a load from origin to destination, thereby allowing exact solution algorithms to be developed for solving relatively large instances of the problem. The computational efficiency of their proposed decomposition-based algorithm was illustrated through numerical examples, where they show that it could solve for instances of up to 150 nodes in reasonable amount of computational time (few seconds).

The first model for locating city or IMEX intermodal terminals proposed by Teye et al. (2015) was also based on a MILP with Lagrangian heuristics for solving it. They observed that the MILP formulation leads to all-or-nothing (AON) assignment of demand between competing modes for each origin-destination pair, resulting in unintuitive results during forecasting and policy testing. A summary of work on locating intermodal terminals or hubs are presented in Table 2.3. The next section discusses the gaps identified in the literature and how this research intends to fill the gaps.

Table 2.3: Summary of previous research on IMTLP

Reference	Objective	Modelling	Solution	Comments
Arnold et al. (2001)	Total cost minimisation	Mixed integer programming	Heuristic	Here the competition is between modes of transport and intermodal transport involves the use of exactly two terminals
Peeters and Thomas (2004)	Total cost minimisation	Mixed integer programming	Heuristic	Improved formulated of Arnold et al. (2001) as a multicommodity fixed-charge network design problem (MCNDP), to reduce the number of decision variables
Racunica and Wynter (2005)	Total cost minimisation	Mixed integer programming	Linearization and heuristics	The formulation includes non-linear and concave cost functions to capture economies of scale effects on intermodal usage
Rahimi et al. (2008)	Total cost minimisation	Single facility location	Exact (6 nodes)	Location-allocation model for locating hubs terminals to promote the use of rail using a concave cost function to capture economies of scale resulting from freight consolidation at hubs
Limbourg and Jourquin (2009)	Total cost minimisation	Multiple allocation phub median	Exact solution for hubs only locations	Linked p-hub median problem and multi-modal assignment problem for locating intermodal terminals
Ishfaq and Sox (2011)	Total cost minimisation	Single allocation p-hub median	Tabu search	Employed a multiple-allocation p-hub median modelling approach to formulate a new mathematical model for the intermodal terminal location problem
Ishfaq and Sox (2012)	Total cost minimisation	Nonlinear mixed integer programming	Tabu search	Extension of Ishfaq and Sox, (2011) with the integration of a queuing system to model hub operations and investigated the effects of limited hub resources on the design of intermodal logistics networks under service time requirements
Sorensen, Vanovermeire and Busschaert (2012)	Total cost minimisation	Mixed integer programming	GRASP and ABHC	Two heuristic algorithms for solving the intermodal terminal location problem
Zhang et al. (2013)	CO2 emissions cost minimisation	Mixed integer programming	Genetic algorithm	A as bi-level programming, where the upper level determines the optimal terminal network configurations, while the lower level performs multi-commodity flow assignment over a multimodal network
Qu et al. (2014)	Greenhouse gas emissions cost minimisation	Mixed integer programming	Exact (11 nodes)	Focussed on the effects of greenhouse gas emissions and intermodal transfers on intermodal network design
Teye et al. (2015)	Total cost minimisation	Mixed integer programming	Lagrangian heuristic	Intermodal terminals location with and transport mode choice
Ghane-Ezabadi and Vergara (2016)	Total cost minimisation	Integer programming using composite variables	Decomposition approach	Simultaneously determines the location of hubs, routes for loads and their transport modes
Current research	Welfare maximisation	Nonlinear mixed integer Programming with entropy objective function	Exact and heuristic algorithms	The method strategically places IMTs at locations where shippers' or users' welfares are maximised subject to a set of constraints comprising all known information about the freight system

2.4 Research gaps

Almost without exception, MILP techniques have been employed to find the best locations of IMTs in the literature (Arnold et al. 2001; Ishfaq and Sox 2011; Sorensen et al. 2012). The MILP technique has been very successful for locating facilities, especially in uncompetitive environments and can help a single shipper or firm decide on where best to locate a facility as part of the company's supply network. The solution to MILP models are corners to the simplex formed by the linear equality and inequality constraints leading to all-or-nothing (AON) assignment of flows to competing alternatives only constrained by the capacities of the competing alternatives. Solution of this nature may work well in cases where the demand is for a single shipper or decision maker. The solution is less intuitive for the problem under consideration, where the IMTs to be located are open access terminals with multiple users. Under forecasting and testing of various policies, the solution to a MILP model exhibits the so-called 'flip flop' behaviour (disproportionate swings in demands) where a small change to the cost variables leads to large shifts in demand (Teye et al. 2015).

Additionally, the observed costs used in the modelling exercise are the average cost of using each mode for moving cargo between each origin-destination pair and that the actual cost experienced by each shipper varies about the average value. Thus, what we should expect is not the AON assignment of each origin-destination demand to the least cost mode, but the share of each mode in the transport task. The multi-user feature of the problem and the existence of a competing alternative mode (road alone) means that the allocation part of the problem can be cast as an MCP, where potential users of the terminals are assumed to face a choice of mode among those available (road alone transport versus intermodal transport) that offers them the highest utility for the transport task and where the choice of intermodal transport leads to the use of one of the located terminals. Thus, the demand associated with each located IMT is expected to be the outcome of pooled outputs of many individual mode choice decisions. In addition to cost variability, McFadden (1974) noted that in a choice situation not all factors affecting the choice process are known to the analyst or can be quantified and included in the modelling process, making a probabilistic description of modal decisions more desirable.

Another important gap identified in the literature relates to the restrictive assumption of fixed demand of cargo flows underlying existing models on IMTLP. This assumption implies that changes in transport network conditions, the location of terminals or any land use

or economic factors do not alter the distribution pattern of cargo in the study area and are hence invariant to IMT location decisions and usage. It is to be expected that the choice of mode depends on where the cargo is destined and intermodal transport may not be feasible or cost competitive if the cargo destination is sufficiently close to the cargo origin. Conversely, the cargo destination depends on modal accessibility. That is, the cargo destination must be connected to the transport network and must be accessible by at least one available mode of transport. This reveals a link between cargo distribution and mode choice, where the choice of mode is conditioned by the choice of cargo destination, whilst the mode choice influences the choice of cargo destination. Additionally, without a cargo distribution model, there is no means of quantifying the demand of the located terminals due to auxiliary activities like warehousing or storage, where the terminal can be coded as a cargo destination on the transport network. Revenues from these auxiliary activities are considered vital for the viability and sustainability of inland terminals, especially metropolitan terminals (Meyrick 2006).

2.5 Research questions and solution approach

Driven by the gaps identified in the literature, four overarching research questions are raised and answered through the proposed methods, numerical examples and a case study;

Research question I:
Given the distribution of containerised cargo and candidate terminal locations on the transport network, what are the best places to locate p intermodal terminals to best serve the metropolitan containerised market?

The answer to the above question is model based. The research question was cast as a mathematical problem – the metropolitan intermodal terminal location problem (MIMTLP) – with an objective function to optimise and a set of constraints imposed by the metropolitan market. The key assumption underlying this problem is that the distribution of cargo (cargo flow matrix) in the metropolitan region is known and fixed. The cargo flow matrix comprises import cargo and their distributions in the metropolitan region and export cargo from production areas in the region to the port(s) for export. The variable *p* is user defined and represents the number of desired terminals to locate from a candidate set of plausible terminal locations in the region. Once the research question is cast as a mathematical problem, the next stage is solving it. The solution to the formulated problem are then interpreted to answer the

research question. Another important consideration is the type of objective function to optimise. A detailed discussion on the objective to optimise is provided in Chapter 4.

Research question II:

Given the distribution of containerised cargo and candidate terminal locations on the transport network, what are the best places to locate p intermodal terminals to best serve the regional containerised market?

The approach for answering this research question is similar to the approach adopted in answering research question I, except that in answering question II, the demand for the located terminals are derived from the regional containerised market. The associated mathematical problem is called the regional intermodal terminal location problem (RIMTLP). The solution to the RIMTLP is more complex than the solution to the MIMTLP as the study area for the former covers a large geographical area with many potential places to locate the terminals. This calls for an efficient algorithm for solving the RIMTLP.

Research question III:

Given limited demand for intermodal transport in either or both markets, what are the best places to locate p intermodal terminals to serve both the metropolitan and the regional containerised markets?

The problem of insufficient cargo volumes and lack of affordable land with the required scale and features for developing separate terminals for both metropolitan and regional containerised markets is driving the need to develop terminals that can serve both markets. A typical example in practice is the development of the Moorebank intermodal terminal in Sydney with enough cargo handling capacity to serve both markets (DoFD 2011). The proposed model to answer this research question is generalised such that it can also be used to answer research question I and II. The generalised problem is referred to as the intermodal terminal location problem (IMTLP).

The proposed models for answering questions I, II and III are based on the assumption of fixed cargo flow matrix. This assumption as noted earlier limits the application of the models in answering important policy questions such the impacts of changes in network conditions or modal accessibility on cargo distribution and more importantly the demand for the located

terminals associated with auxiliary activities such warehousing and storage of empty containers. This leads to the fourth research question.

Research question IV:

What are the key factors governing the distribution of cargo in the study area and to what extent do changes in these factors alter the distribution pattern and the location and use of intermodal container terminals?

This fourth research question relates directly to the fixed matrix gap identified in the literature and the limitations of the model proposed to answer research questions I, II and III where fixed cargo matrices were used to infer the best locations and usage of terminals. The goal here is to replace the fixed matrix with variable cargo demand models such that changes in cargo production and distribution patterns are allowed to influence terminal locations, whilst terminal location conditions cargo production and distribution and the choice of mode. The model can also be used to gauge the likely revenue that located IMTs can generate from performing auxiliary activities such warehousing and empty container storage, by allowing re-distribution of cargo volumes. This research question is also motivated by the fact that pure transport benefits associated with metropolitan IMTs are less likely to generate sufficient revenue to make them sustainable. This therefore, allows the selection of the best IMTs to also be based on their suitability as transfer nodes (primary purpose) and their attraction for warehousing or other auxiliary activities. Important outcomes of this model are the key factors governing cargo distribution and whose future changes in certain directions may put the sustainability of the located IMTs into question. Knowing these factors and their adverse impacts may help to hedge against them. The associated mathematical problem is called the intermodal terminals location with variable cargo demand problem (IMTL+VDP). To derive the exact factors governing the cargo distribution and hence location decisions, a case study implementation of the model was presented. The implementation of the model also provides the opportunity to test for various policies and quantifying their impacts on intermodal transport usage.

2.6 Proposed methodology

The proposed method to answer the above research questions is model based and departs from the traditional use of mixed integer programming techniques for searching for the best locations

of facilities. An alternative method has been proposed based on information theory or the principle of entropy maximisation, which provides a universal way of constructing probability distributions about a system of interest based on all available information about the system. The notion of entropy maximisation has its root in thermodynamics and dates back to the days of Clausius (1865) and Boltzmann (1872) where it was first proposed, rediscovered in information theory by Shannon (1948) and later enhanced into a general tool for deduction reasoning and statistical inference (Jaynes 1957). It's first application in transport was due to Wilson (196; 1970) who used it to develop trip distribution models for transport planning. However, its suitability for locating facilities of this kind has only recently been explored (Teye et al. 2017a,b). The flexibility of this modelling technique means that probabilities describing the system can be constructed using available information and can readily be updated when new information becomes available. Probabilities constructed under the entropy maximisation principle have been shown to be the least biased distribution possible (Jaynes 1957).

Recall that the classical facility location problem with multiple users is a linked problem and comprises the *location* and the *allocation* subproblems. The location problem is mainly concerned with keeping the cost of establishing the facilities as low as possible. The allocation problem on the other hand, deals with best possible way of allocating demand to the located facilities to reduce transport costs, which has been shown to be mode choice problem and extended to also include cargo production and distribution problems. The application of the entropy maximising principle allows the linking of behavioural choice models with traditional integer programming models in a consistent way. The behavioural models explain shippers' mode choice behaviours and the production and distribution of cargo in the study area, whilst the integer programming part determines the exact location of the IMTs. The problem is formulated in a way to allow shippers or users to drive the decisions on IMT location and usage with one objective function to optimise. This was achieved by including the cost of IMT location in the overall costs confronted by shippers choosing which if any IMT to use. This in turn, enables the demand for new IMTs to be estimated as a function of the fee charged by terminal operators for IMT usage and the transport cost of accessing the terminal. The solution to the model generates a linked facility location sub-model, a mode choice sub-model (Chapter 4 and 5) suitable for explaining shipper's mode choice behaviour and a variable cargo demand model (Chapter 6) for forecasting IMT usage and policy testing. Detailed discussion of the entropy framework and it suitability for answering the research questions are presented in Chapter 3.

A basic entropy model is developed in Chapter 4 to answer research question I. This chapter includes several important properties of the entropy model including showing the link between entropy maximisation and welfare maximisation. The models in Chapter 4 were then generalised to answer research questions II and III in Chapter 5. Different methods of formulating the variable cargo demand problem and its incorporation within the entropy framework is provided in Chapter 6. Finally, a case study for the models developed to address research question IV is presented in Chapter 7. In addition to the several proposed model properties, numerical examples also were used to illustrate the key features of the models and how they can be applied in practice.

2.7 Relationship to practice-based outcome

Successful location of intermodal terminals will promote intermodal transport use hence the use of more sustainable mode such as rail and less trucks on the road on a daily basis. Shippers will benefit from cheaper transport costs and more efficient empty container handling. Port corporations will see a reduction in port congestion, both land side and at the berth. Although these benefits may be reduced by the possible changes in traffic patterns and hence increase in traffic in the surrounding area of the located terminals. Road carriers stand to gain from fast access to containers and an increase in the number of containers they can move in any one day as the result of potentially less congestion at the port. Research work by Meyrick (2006) suggests that properly located IMTs can help address the difficulties that truck operators normally face in coordinating clients' opening hours. For example, a truck operator or driver can pick up the container from port at night or preferred time and temporally store it at an IMT, then deliver it within the customer's operating hours or preferred delivery time. Intermodal transport solutions can also open up alternative and more sustainable services for trucking companies so trucking costs can be kept to a minimum and help solve the problem of truck driver shortage especially for long distance trip services.

The community, as whole, stands to gain as a result of reduction in the cost of the movement of goods, easing of road congestion and reduction in truck related accidents. The community also benefits through newly created jobs and investments. Benefits include jobs from short-term construction and long-term operations of the terminal as well as the potential to attract other businesses near the terminal facility. The terminal can also generate income to the community and the state directly in the form of property tax, corporate income tax, sales tax,

and the various permitting fees from activities at the terminals. For example, an employee at the new warehouse receives wages that otherwise would not have received (AHRCR 2007). Improvements in the environment and safety will result from the reduction in the number of truck trips on the road and the total distance travelled. Shift to a more efficient transport mode like intermodal transport is likely to lead to less damage to the road infrastructure, less congestion, reduction in accidents and other road fatalities and reduction in greenhouse emissions (AHRCR 2007).

Chapter 3 Entropy Maximisation

"As far as the laws of math refer to reality, they are not certain; and as far as they are certain, they do not refer to reality." Albert Einstein (1879 – 1955)

3.1 Background

The various methods used in the literature for locating hubs and inland intermodal terminals have been discussed in chapter 2. The chapter also illustrates why existing techniques are not suitable for addressing the research questions. This chapter discusses the alternative method employed to provide satisfactory answers to the questions. The proposed approach is based on the *principle of entropy maximisation.* This principle provides a general guide on how to make decisions under incomplete information. The lack of complete information about the system of interest raises two important questions; first, how to describe or quantify the current state of partial knowledge of the system and second, how to update from one state of knowledge to another when new information becomes available? The first question can be answered through the construction of probability distributions to describe the system. The probability distribution reflects the fact that not all information about the system is known. For the second question, at least two main strategies can be employed for the updates; the first is based on the conviction that what we learned in the past is important and should only be revised to the extent required by the new information (Caticha 2012). The second strategy is to ignore the knowledge gained in the past and rather combine the old and new information to characterise the current state of knowledge about the system under investigation. It can be shown that the principle of entropy maximisation allows for the implementation of both updating strategies. Thus, both questions (the construction of probability distributions under incomplete information and their subsequent update with new information) can be adequately addressed under the entropy maximisation principle. In the light of this, it is important to first consider the notion of probability, their construction and then how they can be updated with new information.

The rest of the chapter is organised as follows; first, the general concept of probability theory is considered in Section 3.2, followed by the concept of entropy in a thermodynamic sense

and in information theoretic sense in Section 3.3. At this stage, the relationship between probability theory and entropy is established. Section 3.4 presents numerical examples illustrating important features of the entropy framework and Section 3.5 presents a summary of the entropy maximisation concept. Finally, Section 3.6 discusses the suitability of employing the entropy maximisation principle to address the research questions.

3.2 The concept of Probability

3.2.1 Background

Probability of an event is a numeric value constrained between 0 and 1 or 0 and 100% indicating the likelihood of the event happening. A probability of 0 indicates that there is no chance that the event will occur, whereas a probability of 1 indicates that the event is certain to occur. If for example a probability of 0.25 (25%) is assigned to shipper's choice of intermodal transport mode, it indicates that if a shipper makes 100 independent modal decisions over time, 25 of those decisions will favour the use of intermodal transport mode.

Humans have always communicated in probabilistic terms (though not often with numbers), using words like probably, likely or maybe to reveal our complete lack of information about the subject matter. In fact, probabilistic way of thinking can be considered as a process of learning–*inference*. As noted in Giffin (2008) we learn by processing the information available to us which yields answers with uncertainty-*probabilities*. And when we get new information, we add it to what we already know, reprocess and arrive at new probabilities. For example, the analyst who assigned a probability of 0.25 to the shipper's use of intermodal transport may revise this value if new information becomes available to him that the shipper is developing new intermodal terminals or there are restrictions on truck access to the seaport during certain time of the day.

Although the notion of probability has always been with humans, it was not until the 17th century that it was given a mathematical description by Fermat (1601–1665), Pascal (1623–1705), Huyghens (1629–1695) and J. Bernoulli (1654–1705) with the original motivation being to provide answers to various questions regarding the games of chance. Its use is now pervasive in all aspects of science and human learning and to quote Pascal;

"It is remarkable that a science which began with the consideration of games of chance should have become the most important object of human knowledge" (Pascal 1623-1705).

In general, two notions of probabilities, underpinned by two different schools of thoughts can be identified in the literature; the objective probability and the subjective probability. The *objective probability* of an event is defined through random experiments and the computed probability in principle can be verified in every detail. The verification is done by re-computing the probabilities under the same experimental settings or environments. On the other hand, the *subjective probability* of an event is expressed as a measure of our ignorance of the occurrence or non-occurrence of the event based on past knowledge and/or experience. As noted in Jaynes (1957), proponents of subjective probability argue that the purpose of probability theory is to help us in forming plausible conclusions in cases where there is not enough information available to lead to a certain conclusion and that detailed verification is not of interest. The test for a good subjective probability distribution therefore lies in correctly representing the state of knowledge about the events of interest (Jaynes 1957). It turns out that both definitions are useful in most practical problems and as noted in Jaynes (1957) the notion of one being better than the other is not relevant. Before getting into detailed discussion about the two schools of thoughts on probability, the axioms underlying all definitions of probability, also called the Kolmogorov's (1933) axioms, are first discussed.

3.2.2 Kolmogorov's Probability

Before providing a formal definition of probability, two important building blocks are first defined; the *sample space* Ω and *event* $\mathcal{E}$. The sample space (assume the experiment is random) is the set of all possible outcomes of an experiment. For example, if you consider the assignment of 3 containers (c_1, c_2, c_3) to two modes (m_1, m_1) as an experiment called CAM (container assignment to modes) experiment, then the sample space for this experiment is $\Omega = \{I, II, III, IV, V, VI\}$ where the outcome I indicates that c_1, c_2 are assigned to m_1 and c_3 is assigned to m_2 as shown in Table 3.1. An *event* $\mathcal{E}$ is an outcome or a group of outcomes of interest from the experiment. For example, if we are interested in the assignment of two containers to m_1 (mode 1) then the event of interest is $\mathcal{E} = \{I, II\}$ and also for the assignment of no containers to m_2 the event of interest is $\mathcal{E} = \{V\}$. It is worth nothing that the former definition of $\mathcal{E} = \{I, II\}$ involves grouping two elementary outcomes, whilst the later is just one of the 6 possible outcomes. These definitions are useful in the subsequent sections in defining

the possible states of a system of interest and the possible number of ways that each state can occur.

Table 3.1 : Possible outcomes Ω= {I, II, III, IV, V, VI}

Outcomes	m_1	m_2
I	c_1, c_2	c_3
II	c_1, c_3	c_2
III	c_1	c_2, c_3
IV	c_2	c_1, c_3
V	c_1, c_2, c_3	0
VI	0	c_1, c_2, c_3

Once the two basic elements of probability are defined, the next step is to provide a formal definition of probability. All definitions of probability must obey the following three axioms of Kolmogorov's (1933):

1. Axiom 1 (Normalisation): $P(\Omega) = 1$,
2. Axiom 2 (Non-negativity): $0 \leq P(\mathcal{E}) \leq 1$ and
3. Axiom 3 (Finite additivity): $P(\mathcal{E}_1 \cup \mathcal{E}_2) = P(\mathcal{E}_1) + P(\mathcal{E}_2)$, If $\mathcal{E}_1$ and $\mathcal{E}_2$ are two disjoint or independent events.

If we let the set $\mathbb{F}$ represent the collection of all possible events (partial subsets) of the sample space Ω including the empty (or impossible) event $\emptyset$ and the sample space itself Ω, then the collection $(\Omega, \mathbb{F}, P)$ is called *probability space,* the set $\mathbb{F}$ is called *sigma-algebra* or *sigma-field* (σ-algebra or σ-field) and P is the probability function. Axioms 1 and 2 simply define the range of values assigned by the probability function P to each event $\mathcal{E}$ in the set $\mathbb{F}$. Axiom 3 says that if there is nothing in common between two events (disjoint or mutually exclusive) then the occurrence of one should not affect the occurrence of the other. This leads us to the two definitions of probabilities.

3.2.3 Classical and Frequency Probability

The *Classical definition of probability (Equally likely principle):* If each possible outcome of an experiment is equally likely then the probability of an event $\mathcal{E}$ can be computed as:

$$P(\mathcal{E}) = \frac{N(\mathcal{E})}{N(\Omega)} \qquad\qquad (3.1)$$

where $N(\mathcal{E})$ is the number of outcomes in the event $\mathcal{E}$ and $N(\Omega)$ is the number of outcomes in the sample space Ω. With the assumption of equally likely outcomes, the calculation of probabilities reduces to counting the number of outcomes (in $\mathcal{E}$ and Ω) and taking the ratio. For example, in the CAM experiment, the event of assigning two containers to m_1 is $\mathcal{E} = \{\text{I}, \text{II}\}$ and the probability of this event happening is: $P(\mathcal{E}) = N(\mathcal{E})/N(\Omega) = 2/6 = 1/3$. Another way to compute $P(\mathcal{E})$ is to use the assumption that each outcome is equally likely and hence the probability of all the outcomes are the same and equals 1/6 and then apply axiom 3. That is $P(\mathcal{E}) = P(\text{I} \cup \text{II}) = P(\text{I}) + P(\text{II}) = 1/3$. Thus, if we know the probability of each outcome of an experiment, we can compute the probability of any event of interest.

The above definition of probability is very useful but run into many practical problems. One obvious limitation of (3.1) is that one would have to know all of the possible outcomes in Ω in order to be sure that they are equally likely or assign a probability. It is obvious that this is not possible for most practical problems because the set Ω could be unknown or some outcomes may be impossible to identify or the outcomes are simply not equally likely. This practical limitation was actually realised by Pascal who noted that (3.1) is a solution for only games of chance and not reasonable solution for practical matters. He noted and I quote,

'...*what mortal will ever determine, for example, the number of diseases...? It would clearly be mad to want to learn anything this way*' (Pascal 1623-1705).

A natural extension to (3.1) is to define the probability of an event directly by counting the number of times that the event occurred (m) in a given number of repeated trials of an experiment (n):

$$P(\mathcal{E}) \approx \frac{m}{n} \qquad\qquad (3.2)$$

The probability definition in (3.2) is called the *relative frequency* definition of probability. The probabilities are approximated by recording the frequency at which the event occurred m and dividing it by the number of times n that the experiment was repeated. Thus, moving away

from knowing all possible outcomes of an experiment before computing probabilities. Definition (3.2) raises another problem; how many times should the experiment be repeated to achieve accurate estimate for $P(\mathcal{E})$? For example, will the evaluation of $P(\mathcal{E})$ for $n = 10$ be the same for $n = 1000$? It was, however, understood and shown by Bernoulli that the accuracy of computing $P(\mathcal{E})$ using (3.2) increases with increasing n. In other words, as n increases, the ratio $\frac{m}{n}$ gets closer to the true probability $P(\mathcal{E})$

$$P(\mathcal{E}) = \lim_{n\to\infty} \frac{m}{n} \qquad\qquad (3.3)$$

The above outcome led to the law of *large numbers* by Poisson (1835) and its subsequent transformation to weak and strong laws of large numbers by notable mathematicians like Chebyshev, Markov, Borel, Cantelli and Kolmogorov. The practical question that arises from (3.3) is, how large is large enough or what if the experiments are too expensive or impossible to repeat enough number of times? These questions have still not been answered fully. It was, however, suggested by Bernoulli that for practical problems, one could use the probability estimate from (3.2) as the true probability. Pascal also noted regarding (3.1) that if we know the number of possible outcomes of an experiment and have no reason to believe that one outcome should be more likely than another then it is accurate to use (3.1) to compute the probability of events. This is called the principle of *insufficient reasoning*.

These two definitions of probability are generally called *objective* probabilities because anyone repeating the experiments used in producing the probability of an event under the same experimental conditions would be expected to get the same result. Many had argued that these two probabilities have very little practical appeal especially in the sciences and in practical fields like insurance, transport, finance and economics. Yet, these fields were in part founded on probabilistic concepts and tools. As rightly noted in Giffin (2008), some of the best theories of physical, chemical and biological phenomena are all probabilistic in nature as they do not make definite predictions. Rather they predict events, at best, with high objective probability. The question that arises is which concept of probability were they founded on since almost none was based on the above two definitions of probability. For example, the question of 'what is the probability of rain tomorrow?' cannot be answered using the above definitions since there is only one tomorrow and we can't observe tomorrow many times. In reality, we do have

weather forecast almost daily and certainly, these two definitions are not adequate for such forecasts, at least in their classical forms.

This leads us to the third notion of probability called the *belief-type* or *subjective probability*. This approach views probability as a numerical measure of degree of belief that is constrained to certain conditions or axioms (Jaynes 1957; Gilboa et al. 2008). The higher our subjective probability of an event, the higher our expectation that the event will occur. If we consider the weather forecast example, the weather forecaster may compare the weather condition of today with large past dataset and select those days that are in many ways similar to today to form the sample space. The forecaster can then apply the *relative reference* approach to compute the probability of a rain tomorrow, where in this case the number of successes will be the days in the *sample space* where it rained the following day. What is obvious in this approach is that the forecaster could have used a different approach in generating the data with different methods likely to result in different weather forecasts. This makes the subjective way of constructing probabilities partly dependent on the *beliefs* and/or *experience* of the one constructing the probabilities in addition to available evidence on the subject matter.

The subjective approach is by far the most common approach for constructing probabilities and as noted earlier forms the basis of many applied fields. The approach became more attractive and very accessible to many more fields with the development of conditional probability and Bayes' theorem (Bayes 1763). These two techniques allow probabilities to be updated with availability of new information or evidence, making them even more useful in many practical applications. The section below discusses the basic concept of conditional probability and Bayes' theorem.

3.2.4 Conditional Probability and Bayes' Theorem

Conditional probability simply measures or modifies the probability of an event given that another event has occurred. In a broader sense, the probabilities considered so far are conditional probabilities. They all assume that the event Ω (sample space) has already occurred when computing the probability of an event $\mathcal{E}$. Thus, we can effectively deduce the likelihood occurrence of one event by knowing that a similar event has already occurred. This concept can best be explained and then generalised using the dice example in the Venn diagram of

Figure 3.1. The sample space $\Omega = \{1,2,3,4,5,6\}$ is the set of possible outcomes of rolling a die once. Event $\mathcal{E}_1 = \{1,2,3,4\}$ and $\mathcal{E}_2 = \{3,4,5,6\}$ present all outcomes less than 5 and more than 2 respectively. Using the classical definition of probability $P(\mathcal{E}_1) = P(\mathcal{E}_2) = 4/6 = 2/3$. From Figure 3.1, $P(\mathcal{E}_1 \cap \mathcal{E}_2) = 2/6 = 1/3$. The conditional probability of event $\mathcal{E}_1$ occurring given that event $\mathcal{E}_2$ has already occurred (i.e, using $\mathcal{E}_2$ as the sample space) will simply equal the number of elements in $\mathcal{E}_1$ that are also in $\mathcal{E}_2$ divided by the size of $\mathcal{E}_2$: $P(\mathcal{E}_1|\mathcal{E}_2) = 2/4 = 1/2$. It is obvious to deduce a connection among $P(\mathcal{E}_1 \cap \mathcal{E}_2)$, $P(\mathcal{E}_2)$ and $P(\mathcal{E}_1|\mathcal{E}_2)$:

$$P(\mathcal{E}_1|\mathcal{E}_2) = \frac{1}{2} = \frac{P(\mathcal{E}_1 \cap \mathcal{E}_2)}{P(\mathcal{E}_2)} \qquad (3.4)$$

An important consequence of the above result is that the initial probability of event $\mathcal{E}_1$ occurring has decreased from $2/3$ to $1/2$ simply by knowing that event $\mathcal{E}_2$ has already occurred. Re-arranging (3.4) we have $P(\mathcal{E}_1 \cap \mathcal{E}_2) = P(\mathcal{E}_2)\,P(\mathcal{E}_1|\mathcal{E}_2)$. Applying similar logic we have $P(\mathcal{E}_1 \cap \mathcal{E}_2) = P(\mathcal{E}_2|\mathcal{E}_1)P(\mathcal{E}_1)$ and equating the two we have what is called Bayes' theorem:

$$P(\mathcal{E}_1|\mathcal{E}_2) = \frac{P(\mathcal{E}_2|\mathcal{E}_1)P(\mathcal{E}_1)}{P(\mathcal{E}_2)} \qquad (3.5)$$

Equation (3.5) simply implies that the probability of an event is not only a function of current evidence (data) but on prior experience as well. This connection was first discovered by Reverend Thomas Bayes [3](1763) and later developed by other mathematicians such as Laplace (1812), Jeffreys (1939), Cox (1946), Jaynes (1957) among others. Equation (3.5) can have many interpretations depending on the application at hand. One common interpretation is to consider $\mathcal{E}_1$ as *hypothesis* whose truth we want to judge and $\mathcal{E}_2$ as the data. We can then compute the probability that the data would have been obtained if the hypothesis is true $P(\mathcal{E}_2|\mathcal{E}_1)$, followed by the probability (likelihood) that the data would have been obtained whether or not the hypothesis is true, $P(\mathcal{E}_2)$. Equation (3.5) can then be used to update the probability that the hypothesis is true (posterior probability) in light of the observed data $\mathcal{E}_2$.

---

[3] Bayes' work was published after his death by his friend Richard Price who found it among his (Bayes) notes. Unfortunately, the published work was virtually unread by anyone until it was rediscovered by Laplace in 1812.

Thus, the process of *inference* or learning about a subject of interest, usually start with a hypothesis whose truth we want to judge and a belief that the hypothesis is true with a certain probability (prior probability) $P(\mathcal{E}_1)$, and then update the probability in light of new evidence (data) to get posterior probability $P(\mathcal{E}_1|\mathcal{E}_2)$. This process can be repeated several times in light of any new information or data and the final probability (posterior probability) $P(\mathcal{E}_1|\mathcal{E}_2)$ becomes the probability that the hypothesis is true based on all the available information about the hypothesis. This updating rule was first applied by Laplace (1772) to infer the mass of planets such as Jupiter and Saturn using astronomical observations, but the resulting posterior probability distribution was intractable due the choice of the prior probability distribution. He later refined it (Laplace 1812) using the newly developed Gaussian distribution (Gauss 1809).
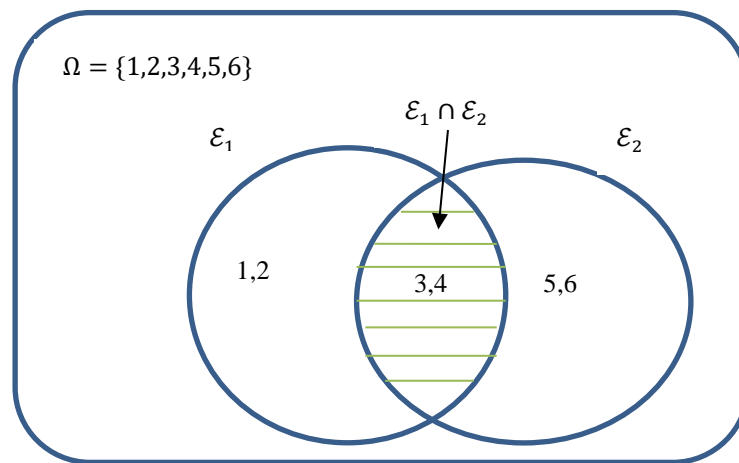
Figure 3.1: Venn diagram

Equation (3.5) is usually referred to as the *product rule* of Bayes'. Another important rule is the *sum rule*. The sum rule deals with the joint occurrence of two or more events. For example, the probability of occurrence of events $\mathcal{E}_1$ or $\mathcal{E}_2$ denoted as $P(\mathcal{E}_1 + \mathcal{E}_2)$ or $P(\mathcal{E}_1 \cup \mathcal{E}_2)$ is expressed as:

$$P(\mathcal{E}_1 + \mathcal{E}_2) = P(\mathcal{E}_1) + P(\mathcal{E}_2) - P(\mathcal{E}_1 \cap \mathcal{E}_2) \tag{3.6}$$

It easy to show that Equation (3.6) is true from the dice example in the Venn diagram given that $\mathcal{E}_1 + \mathcal{E}_2 = \Omega$ and hence $P(\mathcal{E}_1 + \mathcal{E}_2) = 1 = P(\mathcal{E}_1) + P(\mathcal{E}_2) - P(\mathcal{E}_1 \cap \mathcal{E}_2)$. The other probabilities are already computed above. If events are independent (mutually exclusive) then $P(\mathcal{E}_1 \cap \mathcal{E}_2) = 0$ and $P(\mathcal{E}_1 \cup \mathcal{E}_2) = P(\mathcal{E}_1) + P(\mathcal{E}_2)$. The sum rule is found to be particularly

useful in Bayesian parameter estimation as it allows one to investigate some properties of useful parameters whilst removing non-useful ones (Bretthorst 1990). The sum and the product rules together are referred to as the *Bayes' rules* (Bretthorst 1990).

As demonstrated above, Bayesian rules are very useful for updating probabilities when new evidence become available. This flexibility allows the rules of Bayesian probability theory to be applied in parameter estimation or hypothesis selection problems (Bretthorst 1990). The key limitation of Bayesian theory relates to the fact that Bayesian rules only tell us how to manipulate or update probabilities after they have been assigned but not on how to construct the probabilities themselves (Bretthorst 1990). Additionally, the choice of the prior distribution is generally arbitrary and heavily influenced by an individual's subjective belief making its use potentially dangerous in practice. These limitations led to almost the complete neglect and disuse of the Bayesian method by the end of the 19th century (Harney 2003). Although Bayesian and its applications have now been revived (see Jeffreys 1939; Savage 1950; Jaynes 1968; 1980, Bretthorst 1988 and Zellner 1971), none of the methods developed so far for converting new evidence into probabilities and/or choosing of priors has been universally accepted as the best or standard way of dealing with the problems.

An alternative and more appealing theory that possesses the key properties of Bayes' (updating probabilities with new evidence) and also capable of assigning or constructing probabilities is presented. This theory is referred as the *principle of entropy maximisation.*

3.2.5 Summary on Probability

The above section describes the two main schools of thoughts on probabilities. The objective notion of probability which comprised the classical and frequency definitions of probability and the subjective notion, which includes Bayes probability and other belief forms of probabilities. Both descriptions are valid as they both satisfy the three axioms of Kolmogorov's (1933) and the choice of one over the other depends on the application at hand. The probability of an event can therefore, be defined broadly as: (1) a predictor of frequencies of outcomes over repeated trials, or (2) a numerical measure of plausibility that an event will occur. As can be expected sometimes the frequency definition is meaningless, and only the subjective one makes sense. The subjective definition of probability implicitly acknowledges the fact that our beliefs of the occurrence of any event are held on the basis of incomplete information. Thus,

the more information we have the stronger our beliefs. This therefore, calls for a theory for updating probabilities of which the Bayesian theory is one method. The limitations of the Bayesian updating approach have been discussed and the next section presents alternative updating methods and, more importantly, a probability construction method.

3.3 The principle of entropy maximisation

3.3.1 Background

The principle of entropy maximisation could be considered as an extension of Bayes' theorem as it also possesses perhaps the most important property of Bayes' theorem-updating probabilities (one's beliefs) of events occurring with new evidence. In addition and probably more importantly, the entropy principle leads to the construction of probabilities by making use of all available information relating to the event of interest. The mysterious nature of entropy is not limited to these two important properties. Researchers (Burg 1978; Seth and Kapur 1990; Donoso and Grange 2010) have shown the connection between entropy maximisation and the method of maximum likelihood for parameter estimation. The entropy maximisation principle thus allows for the construction of probabilities, estimation of parameters governing the probabilities and updating the probabilities with new information as they become available.

Adding to the mystery are the existence of entropic measures used in many applied fields such as finance, economics, network theory and many other disciplines for selecting between competing alternatives or hypothesis. For example, in Finance, entropy is considered as a measure of *portfolio diversification* where greater entropy is generally considered as a higher degree of portfolio diversification and vice versa (Bera and Park 2009; Usta and Kantar 2011; Jana, Roy and Mazumder 2007). It is also being applied in capital investment and option pricing (Buchen and Kelly 1996; Benth and Groth 2009). In information theory, it is used as a measure of the average amount of information contained in a message (Shannon 1948), in sociology, entropy is considered as a natural decay of structures (Liu, Liu and Wang 2011) and in general statistical inference, it is considered as the amount of missing information in constructed probabilities (Jaynes 1957). Also in economics and transport studies, the similarities between models resulting from utility-maximising and entropy-maximising, have been well documented in the literature, especially with respect to the derivation of the multinomial logit model (Clark and Wilson 1985; Anas 1984; Fisk 1985) and in Physics that

gave birth to the entropy concept, it is often interpreted as the amount of disorder in a system (Boltzmann 1872; Gibbs 1902; Planck 1901).

The pervasive use of entropy in almost all disciplines makes it difficult to give it a precise definition. In fact, from its very conception, it was known to be confusing and mysterious, even among accomplished scientists like Clifford Truesdell and John von Neumann. Clifford Truesdell once noted that;

"Entropy, like force, is an undefined object, and if you try to define it, you will suffer the same fate as the force definers of the seventeenth and eighteenth centuries: Either you will get something too special or you will run around in a circle'' (Truesdell 1966).

And in 1948 when Shannon wanted a measure of the amount of information that was transmitted in signals, he arrived at an equation whose name was suggested by John von Neumann as *entropy*. John gave two reasons for the choice of entropy, his reasons as reported by Shannon and can be found in Tribus (1979) were;

''In the first place, your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more importantly, no one really knows what entropy is, so in a debate you will always have the advantage''.

Caticha (2012) noted that entropy maximisation should be broadly considered as a tool for reasoning without recourse to notions of heat, multiplicity of states, disorder, uncertainty, or even in terms of an amount of missing information. In this way, entropy needs no interpretation, and our focus should be more on how to use it rather than what it means. Any attempt to find a uniquely correct and universally acceptable definition of entropy is likely to fail just like previous attempts and as noted in (Caticha 2012) none can be found.

Indeed, entropy has its origin in statistical mechanics and was first introduced into thermodynamic by Rudolf Clausius (1865) and later given a statistical interpretation by Boltzmann (1872). Since then entropy has played an important role in thermodynamics and a countless number of disciplines. Many researchers have contributed significantly to its development and applications (Planck 1901; Gibbs 1902; Jeffreys 1938; Kullback and Leibler 1951; Wilson 1970; Shore and Johnson 1980). However, this study will focus on the pioneering

contributions of three scholars whose names are often associated with entropy; Boltzmann's entropy, Shannon entropy and Jaynes entropy and the relationship between these entropies.

3.3.2 Boltzmann's Entropy

Ludwig Boltzmann (1844–1906) is generally acknowledged as one of the most important physicists of the nineteenth century and is generally credited with statistical interpretation of entropy and the second law of thermodynamics in general (Boltzmann1872). He was the first to connect entropy with probability through his kinetic theory of gases giving rise to the famous entropy formula:

$$S = k_B \ln E \tag{3.7}$$

where k_B is Boltzmann's constant and is normalised to 1 for non-thermodynamic applications (Jaynes 1957; Shannon 1948; Fisk 1985). 'ln' is the natural logarithm and E is the possible number of ways that a state of the system can occur. The system at any given time can be found in exactly one of the possible states that it can occur. Boltzmann found through experimentation that a system in equilibrium is almost always in the state that has the highest number of ways of occurring and hence has the maximum entropy. The reason is because 'nature' says so. If you consider the CAM (assigning containers to modes) experiment for example. Assume that we want to assign 4 containers to the 2 modes and the state of interest is the number of containers assigned to each mode. Table 3.2 shows that the CAM system can be observed in 5 possible states. Following the classical definition of probability, state $\{M1, M2\} = \{2,2\}$ has the highest probability of occurrence as it has the highest number of ways of occurring and hence the highest entropy according to Boltzmann's Equation (3.7). It is to be expected, at least theoretically, that for large number of containers, the probability of the CAM system being in state $\{2,2\}$ will be closer to one whilst that of the rest of the states gets closer to zero. Thus, all things being equal, without any external influence or information, state $\{2,2\}$ will be the most likely representation of the system as it has the highest number of ways of occurring as shown in Table 3.2. This simple experiment also reveals something very interesting-each mode in $\{M1, M2\} = \{2,2\}$ is assigned the same number of containers. Thus, each alternative or mode has equal 'mode' share, revealing a *uniform probability distribution* at play.

Table 3.2: Possible macrostates $\Omega = \{M1, M2\}$

Distribution of containers to modes (states)		All possible assignments of containers		Number of ways of occurring	Probability of occurrence of each state
M1	M2	M1	M2		
4	0	c_1, c_2, c_3, c_4	0	1	0.063
0	4	0	c_1, c_2, c_3, c_4	1	0.063
3	1	c_1, c_2, c_3	c_4	4	0.250
		c_1, c_2, c_4	c_3		
		c_1, c_3, c_4	c_2		
		c_2, c_3, c_4	c_1		
1	3	c_4	c_1, c_2, c_3	4	0.250
		c_3	c_1, c_2, c_4		
		c_2	c_1, c_3, c_4		
		c_1	c_2, c_3, c_4		
2	2	c_1, c_2	c_3, c_4	6	0.375
		c_1, c_3	c_2, c_4		
		c_1, c_4	c_2, c_3		
		c_2, c_3	c_1, c_4		
		c_2, c_4	c_1, c_3		
		c_3, c_4	c_1, c_2		

We have seen that the most likely state for any isolated system is the one with the highest number of ways of occurring or multiplicity and has the highest entropy according to Equation (3.7). Determining the number of ways and hence the probability that each state can occur can be computed directly using Equation (3.8):

$$E = \frac{n!}{\prod_{i=1}^{m} n_i!}$$

(3.8)

where n the total number of elements is in the system, m is the number of alternatives of interest and n_i is the number of elements assigned to alternative $i = 1, 2, \ldots m$ such that $\sum_{i=1}^{m} n_i = n$. From the CAM experiment, the number of elements (containers) $n = 4$, the number of alternatives (modes) $m = 2$, the number of containers assigned to mode 1 is n_1 and the number assigned to mode 2 is n_2. A state of the system is therefore characterised by $\{n_1, n_2\}$ such that $n_1 + n_2 = n$. The number of ways that each state can occur can be computed directly using

(3.8). For example, the number of ways that the state {4,1} can occur according to (3.8) is 1, and that of {2,2} is 6 and so on. The state with maximum number of ways of occurring can be derived directly by maximising (3.8) subject to the conservation of flow constraint:

$$\max E = \frac{n!}{\prod_{i=1}^{m} n_i!}$$

Subject to the fact that $n_1 + n_2 + \cdots + n_m = n$.

Since the values of $n_i's$ that maximises E, also maximises $\ln E$ but it is easier to work with $\ln E$ so $\ln E$ is maximised instead. Taking the natural logarithm of E and applying the Stirling's[4] approximation we have:

$$\ln E = n\ln(n) - n - \sum_{i=1}^{2} n_i(\ln(n_i) - 1)$$

Forming the Lagrangian equation comprising $\ln E$ above and the constraint and enforcing the first order condition of optimality with respect to n_i we have: $n_i = e^{-\lambda}; \ i = 1,2,$ where λ is the Lagrangian multiplier associated with the conservation constraint. Using the conservation of flow constraint, the parameter λ can be estimated:

$$\sum_{i=1}^{m} n_i = 4 = \sum_{i=1}^{m} e^{-\lambda}$$

resulting in $e^{-\lambda} = \frac{4}{2} = 2$

Hence $n_i = 2; i = 1,2$

and the corresponding entropy $S = 1.792$ where k_B is normalised to 1. Thus, the most likely state is $n_1 = n_2 = 2$ or {2, 2} which agree exactly with the result in Table 3.2. Alternatively, the uniform distribution is the 'most probable state' and it has the maximum number of ways

---

[4] According Stirling $\ln(n!) \approx n\ln(n) - n,$ where 'ln' is the natural logarithm

of occurring. Thus, the most likely state of the system or any system can be found by maximising the entropy of the system.

The entropy in Equation (3.8) can be expressed in terms of probabilities rather than in combinatorial terms. Following the frequentist definition of probability, the probability for alternative i, $p_i = \frac{n_i}{n}$; $i = 1,2,\ldots m$. Applying the Stirling's approximation to (3.8) we have:

$$S = \ln E = n\ln(n) - n - \sum_{i=1}^{m} n_i(\ln(n_i) - 1)$$

Replacing the $n_i's$ with probabilities and simplifying we have:

$$S = -n\sum_{i=1}^{m} p_i\ln(p_i) \quad\quad (3.9)$$

3.3.3 Shannon Entropy

The motivation of Shannon was different from that of Boltzmann and he worked in a completely different subject. Shannon seeks an absolute measure of the amount of missing information in a given probability (p) that an event will occur. He noted that such a measure if it exists must satisfy a set of axioms involving consistency, continuity and additivity (Shannon 1948). He denoted such as measure by the letter H if it exists. Suppose that we have a set of m competing or mutually exclusive and exhaustive discrete alternatives whose probabilities of occurrence are $p_1, \ldots, p_m$ then according to Shannon, H must satisfy the following axioms:

Axiom 1. The measure H should be continuous in p_i; $i = 1,2,\ldots m$

Axiom 2. If $p_i = \frac{1}{m}$; $\forall i = 1,2,\ldots,m$ then H should be a monotonic increasing function of m. Thus, the measure should attain its highest possible value when there is complete lack of information as to which alternative is more likely to occur.

Axiom 3. If a choice can be broken down into two successive choices, the original H should be the weighted sum of the individual values of H. For example, if the probability of using truck, barge and rail are p_{truck}, p_{barge} and p_{rail} respectively and the barge and rail modes are

combined to from one mode called sustainable transport (*st*) with probability $p_{ST} = p_{barge} + p_{rail}$ then the following relation must be satisfied:

$$H(p_{truck}, p_{barge}, p_{rail}) = H(p_{truck}, p_{ST}) + p_{ST}H(p_{barge}, p_{rail})$$ (3.10)

Shannon showed that (3.11) is the only measure that can satisfy the above axioms:

$$H = -K \sum_{i=1}^{m} p_i \log p_i$$ (3.11)

where K is a positive constant and provides a choice of a unit of measure for H since the terms $p_i \log p_i$ are unitless. In information theory, the logarithm (log) is to the base 2 since information is often transmitted in strings of binary digits (0s and 1s) and thus the measure H can be expressed in bits. Shannon called the measure H entropy following the advice of Von Neumann who recognised its physical significance and the similarity in mathematical form to the quantity S in thermodynamics.

A simple approach of deriving (3.11) can presented as follows; assume we are interested in the occurrence of an event $\mathcal{E}$ and all we know is the probability p that the event will occur. By definition, the probability p gives us a measure of information about the likelihood of that event occurring. If the value of p is very high (say 0.99) and the event occurs, then there will be almost no surprise and the amount of information gain in the occurrence of $\mathcal{E}$ or the amount of missing information or uncertainty in p will be low. Conversely, if p is very low (say 0.001), we will be surprised if the event did occur, implying the uncertainty or the amount of missing information in p is high. It is therefore intuitive to assume that the amount of missing information in p is inversely related to p. It is also to be expected that if we have two independent events $\mathcal{E}_1$ and $\mathcal{E}_2$ with p_1 and p_2 as the respective probability of occurrence, then the information gained from the joint occurrence of these two events must equal the sum of the information (I) gained from the occurrence of $\mathcal{E}_1$ (or $\mathcal{E}_2$) followed by the occurrence of $\mathcal{E}_2$ (or $\mathcal{E}_1$). In other words, it is reasonable to assume that the information gained when an event occurs, should be additional to the information already gained, and, thus, one might expect that the information gain on the occurrence of both events to be the sum of each, irrespective of which occurs first: $I(p_1, p_2) = I(p_1) + I(p_2)$. It appears a logarithm function is the only

function that satisfies this condition. Thus, the amount of information gain when an event with probability p occurs can be expressed as:

$$I(p) \propto \ln\left(\frac{1}{p}\right)$$

(3.12)

In general, if there are n independent events with probabilities of occurrence $p_1, \ldots, p_m$, then the expected (average) amount of information gained (entropy) using (3.12) by definition is expressed as:

$$H(p_1, \ldots, p_m) = K \sum_{i=1}^{m} p_i \, I(p_i) = -K \sum_{i=1}^{m} p_i \ln(p_i)$$

(3.13)

where K is constant of proportionality. Clearly, Shannon entropy in Equation (3.13) is equivalent to Boltzmann entropy in (3.9) though both were derived in two completely fields with different motivations and assumptions. This leads to Jaynes entropy, who promoted the Shannon entropy into the method of statistical inference.

3.3.4 Jaynes' principle of entropy maximisation

The work by Jaynes (1957) was key in promoting the entropy concept into a general method of deductive reasoning. Prior to Jaynes, several scholars including John von Neumann recognised the connection between thermodynamics or Boltzmann's entropy and Shannon entropy. However, as noted in Section 3.3.3 the connection took the form of an analogy between the two major disciplines; physics and engineering. Boltzmann's' quantity in (3.9) was meant to be true and reflect the very 'laws of nature' whilst Shannon measure in (3.11) was meant to be true for communications systems. With the work of Jaynes in the 1950s, it became clear that the connection was not an accident.

Jaynes indicated that 'reasoning with incomplete data' is the crucial link between Boltzmann's and Shannon entropies (Caticha 2012). Jaynes argued that what is at play is not a direct 'image' of nature as expounded in thermodynamics but rather a rule for constructing probabilities in the face of incomplete information and that the entropic measure in (3.9) or (3.11) can be interpreted as the uncertainty represented by the probability distribution itself, or

as a measure of the amount of information required to achieve certainty. The impacts on physics is particularly enormous as entropy can no longer be justified as fundamental law of nature, but rather a method of statistical inference about nature (Caticha 2012). Jaynes subsequently announced the principle of entropy maximisation (PEM) as a general technique for constructing probabilities on the basis of partial information. The objective function to maximise is an entropy (either Boltzmann or Shannon) function subject to the available information converted into a set of constraints. Probability distributions constructed using PEM approach are considered the least biased possible on the given information since it maximises the amount of missing or remaining information (Jaynes 1957).

Jaynes' entropy is Shannon entropy with logarithm to base 2 replaced by the natural logarithm and the constant K normalised to 1, making Jaynes' entropy unitless and more relevant to many disciplines:

$$H(p_1, \dots, p_m) = -\sum_{i=1}^{m} p_i \ln(p_i) \tag{3.14}$$

The existence and uniqueness of solutions to the resulting optimisation problem or simply entropy maximising programming problem have been shown by Shore and Johnson (1980).

Although Jaynes' entropy (3.14) is inspired by Shannon's (3.11), the two functions have completely different meaning. Shannon's entropy assumes the probabilities are known and what was needed was a measure that can quantify the amount of information contained in a message and the degradation of that information during transmission, processing and storage (Shannon 1948). For Jaynes, the interest was on finding the least biased ways to construct probability distributions containing all available information about the system of interest and according to Jaynes, these distributions can be constructed using the principle of entropy maximisation.

A further note is also necessary on Boltzmann's and Jaynes' entropies. The motivation behind the work of both scholars is essentially the same—finding the most likely probability distribution based on available information'. Although under Boltzmann, the available information were macroscopic properties like energy or number of particles in the system.

Indeed, it was Boltzmann who first expressed the idea that a system at equilibrium is in a state of highest entropy (Boltzmann 1972). With known macroscopic properties of the thermodynamic system, he constructed a probability distribution of energetic states of particles (called the Boltzmann distribution) by maximising an entropy function subject to what he knows about the system; the total number of particles and the total energy of the system—first application of the principle of maximum entropy. The main difference between the two approaches (Boltzmann verses Jaynes) lies in the functional form of the entropy; Boltzmann entropy is combinatorial in specification whilst Jaynes entropy is probabilistic and as shown through Equation (3.9) one can easily be transformed to the other. Both approaches are valid and depending on the application at hand one may be found more appropriate than the other. In transport applications, both approaches have been used and the choice between the two is usually motivated by the type of data available (Wilson 1967; Evan 1973; 1976; Anas 1983; Bell 1983; Fisk 1985).

In summary, the principle of entropy maximisation provides a universal way of processing information and constructing probability distributions making it suitable for the study of a wide variety of probabilistic systems transcending all disciplinary boundaries. In addition to its practicality and great success in almost every discipline, the principle of entropy maximisation goes to the very heart of the meaning of probability and inductive reasoning (Caticha 2012). The basic idea is that if we want to model or find the most likely state that a system of interest is in, we first define an entropy function of the system, present our knowledge of the system as a set of constraints and then maximise the entropy function subject to these constraints. The constructed probability distribution then represents the current state of knowledge about the system under investigation. Once the probability distribution is constructed, the next important question is, what happens if new information about the system comes along and we want to update our current state of knowledge with the new information? One obvious answer to this question is to employ Bayes's rule as discussed in Section 3.2. It turns out that we do not have to use Bayes' rule for the update. The update can also be done consistently within the entropy framework as discussed in the next section.

3.3.5 Entropic update of probabilities

It has been demonstrated how PEM can be used to construct probability distribution to describe the behaviour of almost any system of interest by construction an entropy function and

converting available information about the system into a set of constraints. It has also been noted that the maximised value of the entropy function captures the amount of missing information in the probabilities constructed. The entropic probability of a system therefore, reflects of our state of knowledge about the system. We therefore, expect that the more information we have about the system under consideration, the better the predictive power of the constructed probability distribution and hence the more we can trust its description of the system. This means that being able to construct probability distribution based on the current state of knowledge about the system is not enough. It is equally important to find the means of updating the distribution as and when new information becomes available. The updating scheme will ensure that the probability distribution at any given time reflects our most current state of knowledge about the system.

One obvious updating scheme is the Bayesian method where the existing probability distribution (called the prior) is replaced by a posterior distribution as described in Section 3.2. Employing Bayesian method required making some assumptions about probability distribution of the new information, which means artificially "adding" information, which may be true or false. The good news is that we do not have to switch to Bayesian updating method when new information becomes available. The framework of entropy maximisation allows for the update of the probability distribution with new information in at least two consistent ways; absolute entropy update (AEU) as proposed by Jaynes (1957) and relative entropy update (REU), also called cross entropy update, first proposed by Kullback and Leibler (1951).

Absolute entropy update (AEU)
The AEU approach simply discards the existing probability distribution and reconstructs a new one using both the old and the new information. This approach is attractive and efficient if the volume of new information is significantly more than the old one and/or if the new information is considered more accurate than certain aspects of the old information. In many practical problems, however, the new information is likely to be significantly smaller in volume and/or reflects some aspect of the system that has not been adequately described by the existing probability distribution. In these situations, the REU method is efficient and more relevant.

Relative update method (REU)

The REU generates new (posterior) probability distribution by allowing the construction of Boltzmann's entropy function to explicitly account for the existing (prior) probability distribution.

Proposition 3.1: Maximising entropy is equivalent to maximising the log-likelihood of the multinomial probability mass function:

$$f(n_1, n_2, \ldots n_m | \mathcal{P}) = \left(\frac{n!}{\prod_{i=1}^{m} n_i!} \right) \prod_{i=1}^{m} \mathcal{p}_i^{n_i} \tag{3.15}$$

with the assumption that the prior probabilities $\mathcal{p}_i$ ($i = 1, 2, \ldots m$) follow a uniform distribution. $\mathcal{p}_i$ is the prior probability that alternative i takes on the value n_i such that $n_1 + n_2 + \cdots + n_m = n$.

Proof 3.1: The term in the bracket of Equation (3.15) is called the multinomial coefficient and corresponds to the entropy function in Equation (3.8). Taking the natural logarithm of (3.15) and applying Stirling's approximation we have:

$$\ln f = n \ln n - \sum_i n_i \ln n_i + \sum_i n_i \ln \mathcal{p}_i$$

If we define the posterior probability as $p_i = \frac{n_i}{n}$, where n is the total size of the system and n_i and is the size occupied by alternative i or the number of times that event i occurred. Converting the $n_i's$ into $p_i's$ and simplifying we have:

$$\ln f = -n \sum_i p_i \ln p_i + n \sum_i p_i \ln \mathcal{p}_i$$

which simplifies to become

$$\ln f = -n \sum_i p_i \ln \frac{p_i}{\mathcal{p}_i} \tag{3.16}$$

If the prior probabilities follow a uniform distribution, $\wp_1 = \wp_2 = \cdots = \wp_m = \wp$ then the above simplifies to become:

$$\ln f = -n \sum_i p_i \ln p_i + n\wp = -n \sum_i p_i \ln p_i + k$$

where $k = n\wp$ is constant and can be ignored in the optimisation process. Clearly, maximising entropy S in (3.8) is equivalent to maximising f with the assumption that the prior probabilities are uniformly distributed. In fact, it has already been shown that the uniform probability distribution is the default distribution when the entropy of a system is maximised with no available information—an outcome that echoes Laplace's principle of insufficient reasoning. Thus, the entropy in (3.8) or (3.11) can be generalised by relaxing the uniform distribution assumption of the prior probabilities:

$$S(p_1, \dots, p_n | \mathcal{P}) = -\sum_i p_i \ln \frac{p_i}{\wp_i} \qquad (3.17)$$

Since maximising $S(p_1, \dots, p_n | \mathcal{P})$ is equivalent to minimising $-S(p_1, \dots, p_n | \mathcal{P})$, the minimisation of $-S(p_1, \dots, p_n | \mathcal{P})$, is referred to in the literature as cross entropy or Kullback-Leibler divergence (Kullback and Leibler 1951):

$$\vec{S}(p_1, \dots, p_n | \mathcal{P}) = -S(p_1, \dots, p_n | \mathcal{P}) = \sum_i p_i \ln \frac{p_i}{\wp_i} \qquad (3.18)$$

The process of minimising (3.18) subject to the new information is generally referred to in the literature as the *principle of minimum information* (Williams 1980; Kullback and Leibler 1951; Caticha and Giffin 2006).

We have demonstrated ways in which existing probabilities can be updated with new information as and when they become available making the entropy framework a truly universal method of deductive inference. It is also worth noting that some scholars such as Willians (1980), Diaconis and Zabell (1982), Jaynes (1988), Caticha and Giffin (2006) have investigated the link between the principle of minimum information and Bayesian updating method and found the later to be a special case of the former. As noted earlier a connection

with maximum likelihood estimation has also been established (Burg 1978; Seth and Kapur 1990). Thus, the principle of entropy maximisation is truly universal; it allows for the construction of probabilities, updating the probabilities with new information as they become available and the estimation of parameters governing the distributions. The next section presents a numerical example to illustrate some important features of the entropy concept.

3.4 Numerical Example

This section provides an illustrative example to demonstrate some features of the entropy concept, especially in relation to probability construction and subsequent update with new data or information.

Problem setup: Consider the system of container movements between a seaport and various warehouses in the metropolitan region. The only information available to the study team are that the only modes of transport available for transporting 120 TEUs of cargo from the port to various warehouses are road, rail and barge. The question posed to the research team was to determine the mode share or the quantity of containers carried by each mode.

Solution approach: Since the researchers wanted to avoid assuming or adding more knowledge than they have, they employed the principle of entropy maximisation to solve the problem. For simplicity, Jaynes' entropy will be used to find the mode shares directly. If $d = 120$ and d_1, d_2, d_3 are quantities of cargo carried by road, rail and barge respectively, with corresponding mode shares p_1, p_2, p_3 then from Equation (3.14) we maximise:

$$H(p_1, p_2, p_3) = -\sum_{i=1}^{3} p_i \ln(p_i)$$

Subject to:

$p_1 + p_2 + p_3 = 1$

$p_i \geq 0; i = 1,2,3$

Forming the Lagrangian equation and enforcing the first order condition for maximum $H(p_1, p_2, p_3)$ with respect to $p_i's$ satisfy the following equation:

$$-\ln(p_i) - 1 - \varphi = 0; \quad i = 1,2,3$$

where φ is the Lagrangian multiplier associated with the equality constraint. Solving for p_i by enforcing the equality constraints we have the maximum of $H(p_1, p_2, p_3)$ occurring at $p_1 = p_2 = p_3 = 1/3$ and the corresponding flows are $d_1 = d_2 = d_3 = 40$. It can also be shown using Equation (3.10) that $d_1 = d_2 = d_3 = 40$ has the highest number of possible ways of occurring (entropy).

Now suppose that new information becomes available in terms average costs of transport by each mode with the average cost of road $c_1 = \$5$, the average cost of rail $c_2 = \$8$, the cost of barge $c_3 = \$10$ and the average cost over all modes $c = \$7$. As discussed in Section 3.3, there are two ways of updating the prior probabilities $\wp_1 = \wp_2 = \wp_3 = 1/3$; the absolute entropy update (AEU) approach and the relative entropy update (REU). Each of these updating methods is applied to the problem, starting with the REU, and then the AEU. Let p_1, p_2, p_3 be the updated (posterior) probabilities.

Relative entropy update (REU): This approach uses the Kullback-Leibler function in (3.18) subject to only the new information converted into a constraint as follows:

$$KL(p_1, p_2, p_3) = \min\left(p_1\log\frac{p_1}{\wp_1} + p_2\log\frac{p_2}{\wp_2} + p_3\log\frac{p_3}{\wp_3}\right)$$

subject to:

$$c_1 p_1 + c_2 p_2 + c_3 p_3 = c$$
$$p_1 + p_2 + p_3 = 1$$
$$p_i \geq 0; i = 1,2,3$$

Solving the above problem, we have the following posterior probability distributions governed by the parameter β:

$$p_i = \frac{\wp_i \, e^{-\beta c_i}}{\sum_{j=1}^{3} \wp_j \, e^{-\beta c_j}} \quad ; \quad i = 1,2,3$$

Since $\wp_1 = \wp_2 = \wp_3 = 1/3$, it implies that:

$$p_i = \frac{e^{-\beta c_i}}{\sum_{j=1}^{3} e^{-\beta c_j}} \quad ; \quad i = 1,2,3$$

With some algebraic manipulation (or use of a root finding algorithm) the parameter value $\beta = 0.1562$ minimises the entropy function and satisfies all the constraints. This produces $p_1 = 0.48$, $p_2 = 0.30$ and $p_3 = 0.22$. Thus $d_1 = 58$, $d_2 = 36$, and $d_3 = 26$ indicating that the new information has increased the share of road by 14.7% (from 33.3% to 48%).

Absolute entropy update (AEU): This approach of update throws away the prior probabilities and re-construct new probabilities using the new information together with the existing information. Thus, we maximise:

$$H(p_1, p_2, p_3) = -\sum_{i=1}^{3} p_i \ln(p_i)$$

subject to:

$$c_1 p_1 + c_2 p_2 + c_3 p_3 = c$$
$$p_1 + p_2 + p_3 = 1$$
$$p_i \geq 0; i = 1,2,3$$

Solving the above problem produces the following probability distributions governed by:

$$p_i = \frac{e^{-\beta c_i}}{\sum_{j=1}^{3} e^{-\beta c_j}} \quad ; \quad i = 1,2,3$$

Again, the value $\beta = 0.1562$ maximises $H(p_1, p_2, p_3)$ and satisfies both old and new constraints and $p_1 = 0.48$, $p_2 = 0.30$ and $p_3 = 0.22$. The probability or the share of each mode can readily be computed for any change in the cost of each mode:

$$p_i = \frac{e^{-0.1562 c_i}}{\sum_{j=1}^{3} e^{-0.1562 c_j}} \quad ; i = 1,2,3 \tag{3.19}$$

In summary, we started with lack of information about the container system, other than the constraint that the sum of the number of containers carried by the three modes should add

up to 120. Clearly, there are many values of d_1, d_2 and d_3 that satisfy this constraint-we could have chosen $d_1 = 120; d_2 = 0; d_3 = 0$ or $d_1 = 20; d_2 = 70; d_3 = 30$ and both will satisfy the constraint. Neither choice of values seems particularly appropriate, because each goes beyond what we know. We are assuming something which can turn out to be true or false. To avoid introducing potential false information, we employed the principle of maximum entropy to select that probability distribution or modal demands which is consistent with the constraint and contains the least added information (maximum entropy). The resulting probability distribution turns out to be the uniform distribution with equal mode share. The researchers were then introduced to new set of information in terms of the average cost of using each mode. The new information required the probability distributions constructed to be updated in light of the new information, which was done using two updating methods; AEU and REU methods. Both methods were shown to yield the same results.

The numerical example also demonstrates the subjective nature of the probabilities constructed by the researchers to cope with their lack of knowledge about the system. It shows that a new research team with new set of information or additional set of information about the system are likely to generate different probability distributions. What is also very clear is that if the two research teams have the same set of information, they will end up constructing the same probability distributions. Additionally, all the information does not have to be available to the two teams at the same time-each can receive different amount of information at a given time and each will produce different probability distributions during the information flow and update them as new information comes along. In the end, both teams will generate the same probability distributions if it turns out that each have access to the same amount of information. It has also been demonstrated that having additional information ought to yield on average better probability distribution in the sense that it produces less entropy or uncertainty.

3.5 Summary

It has been demonstrated that the principle of entropy maximisation provides a universal way of processing information, constructing and updating probability distributions making it suitable for the study of a wide variety of probabilistic systems. The principle can also be seen as a natural extension of Laplace's 'principle of insufficient reasoning', which states that if one wants to assign probabilities to events and there is no reason why one event should occur more than others (absence of prior knowledge), the events must be assigned equal probabilities. This places

entropy maximisation at the very heart of the meaning of probability and inductive reasoning as demonstrated in the previous sections. It was shown in Section 3.2 and by numerical example in Section 3.4 that maximising entropy in the absence of all information yields equal probabilities or uniform distribution and that the distribution becomes non-uniform once we happen to know or learn something about the events. The principle of entropy maximisation thus guides us to the least biased probability distribution that reflects our current state of knowledge and tells us what do to if new information becomes available. This makes entropy maximisation a very powerful framework for model building and arguably the most practical way of constructing unbiased probability distributions to describe the behaviour of a system of interest. Any other distribution would have implied adding information 'artificially', which may be true or false (Jaynes 1975).

It has also been shown that the entropy function possesses desirable properties (many of which will be shown in other sections of this thesis). For example, it was shown that the entropy function attains its maximum possible value when nothing is known about the system and decreases as we know more about the system. This property allowed the optimised value of an entropy function to be interpreted as the amount of missing information about the system or in the probabilities constructed. This interpretation seems intuitive since it implies that the more information we have about the system, uncertainty (entropy) about the resulting probability distribution is less and vice versa and hence the better its predictive power.

Another important property of the entropy function is its additive property. This property underpins Shannon's definition of entropy, where the information gained from the joint occurrence of two independent events should equal the sum of information gained from the occurrence of one followed by the other (irrespective of which occurs first). In other words, the final model is independent of the steps followed in constructing the model. This property is particularly useful when relying on data from several sources to build a model. Data from each source is expected to describe some aspect of the system and the non-availability of some data means that the model can still be constructed to describe the other aspects of the system. As an example, assume that a given data D can be divided into two disjoint groups $D1$ and $D2$, then the additive property requires that $H(D) = H(D1) + H(D2)$. If $D2$ is related to $D1$, then $H(D) = H(D1) + H(D2|D1) = H(D2) + H(D1|D2)$. This property proved useful in developing algorithms for solving entropy problems in Chapter 4 to 6.

3.6 Application of entropy maximisation

This section discusses the suitability of employing the entropy framework to answer the research questions. As noted in Chapter 2, the problem of locating intermodal terminals was decomposed into two linked sub-problems; the location problem and cargo flow problem. The goal of the locater is to strategically place the IMTs at locations that can attract as many users as possible whilst keeping the cost of installations as low as possible. The cargo flow problem comprises many decision makers with each maximising his/her utility in the choice of mode and cargo destination. These two problems are linked and will require some degree of trade-offs between the costs of IMT location and the amount of usage.

This study employs the principle of maximum entropy to solve these problems with one objective to optimise. The entropy approach allows the linking of behaviour models (for describing the cargo flow problem) with facility location problem (integer programming of IMT locations). The use of the entropy principle is also broadly supported by the lack of data about the containerised system in sufficient quantity and quality coupled with the that fact that it is impossible to track every potential user or shipper and extract relevant information about his/her mode and destination choice processes. Additionally, shippers, carriers or other logistics providers are unwilling to provide data at the required level of detail for modelling. They are even sensitive to basic data on cargo volumes and their origins and destinations due to fear of losing competitive advantage of their businesses.

Data limitations coupled with the fact available data (often derived from several sources) are usually in aggregate forms with each data set explaining some aspects of the containerised system. These require the need for a method of combining all these diverse pieces of contextual information or evidence together to explain or describe the system in a consistent and unbiased way. It has been shown in earlier sections that the best and most unbiased way is through entropy maximisation (Jaynes 1957; Shannon 1948; Wilson 1970; Fisk 1985). The principle of entropy maximisation provides the means of combining the various pieces of information as constraints as well as accounting for the fact that not all information about the system can be obtained or quantified (Hensher and Figliozzi 2007; McFadden 1974). The entropy principle was shown to produce probability distributions capturing all known information (aggregate and disaggregate) about the system and can be updated with new information as and when they become available.

Under the entropy maximisation approach, all possible demands of modal decision variables are considered and the most likely demand consistent with all available information on cargo production and distribution and modal decisions is selected. The most likely demand, under the entropy framework is the demand that produces the highest entropy of the system. An equivalence has been shown between entropy and shipper welfare, following from the random utility interpretation of the discrete choice model, which in turn is a consequence of entropy maximisation (see Chapter 4 and 5). In Chapter 4, the principle of entropy maximisation has been applied to develop probabilistic models for locating intermodal terminals to serve the metropolitan containerised market. The model was in turn used to provide suitable answers to research question I. Chapter 5 extends the model in Chapter 4 to answer research questions II and III. The extended model is flexible for locating terminals to serve the metropolitan market, regional market or both markets simultaneously. The incorporation of variable cargo demand as constraints within the entropy framework is presented in Chapter 6. The resulting entropy model comprises linked facility location model, mode choice model and cargo production and/or distribution model suitable for addressing research question IV.

Chapter 4 Metropolitan container terminals

4.1 Background

The previous chapter discussed the methodological framework for the location of inland container terminals. This chapter focuses on addressing research question I by developing models for locating container terminals to serve the import and export markets (IMEX) or the metropolitan containerised transport market (MCTM). The located terminals are expected to promote the use of intermodal transport in the movement of cargo between the port and cargo origin/destinations in the hinterlands, with more use of intermodal transport leading to more use of the located terminals and less visibility of container trucks on the road network.

The motivation for the development of these terminals is driven by the need to find a sustainable solution to the disproportionately negative impacts of container trucks on the urban fabric in terms of road damage, congestion, safety and pollution. These problems are compounded for port, especially city ports faced with continuous growth in trade, increased 'lumpiness' of throughputs produced by larger vessels, limited physical space for expansion to accommodate growth and lack of adequate transport infrastructure connecting the ports and the cargo origins/destinations in the hinterlands or the metropolitan region. Port activities or operations are also adversely affected if large volumes of cargo and trucks within and around the port are not properly managed. Inefficient port systems have direct impact on a nation's economy, its foreign trade and its ability to compete in global markets. This is because the port is the gateway for the greater part of visible trade between countries and links the economy of its host country to the rest of the world.

A promising solution to the above problems is the promotion of inland intermodal transport use through the development of intermodal terminals (IMTs) that interface with both road and rail/barge networks. Import containers can then be transported by rail (a high capacity

mode) to the IMT and then be transferred onto trucks for onward movement to their various destinations in the hinterland as shown in Figure 4.1. Also, export containers can first be consolidated at the IMT before being transported to the port by rail or barge for export. The development of these terminals is expected to promote the use of intermodal transport over road alone transport in the metropolitan (import/export) containerised market. More use of intermodal transport is expected to create significant extra handling and storage capacities at ports, reduce the number of trucks going to deliver and receive container(s) at the port, reduce the number of container trucks on the road network especially around the vicinity of the port and thereby reducing congestion and the number of conflicts with other road users. The positive impact on the environment is realised through the reduction in the number of trucks on the road network or the total kilometres travelled by trucks.

The problem of locating intermodal container terminals to serve the metropolitan market was cast in Chapter 1 and 2 as a linked facility location problem (FLP) and mode choice problem (MCP). The FLP determines the exact locations of the required number of terminals among a candidate set on the transport network and the MCP determines the demand or usage of the located terminals through the demand for intermodal transport. The mathematical formulation of the metropolitan intermodal terminal location problem (MIMTLP) is based on the principle of entropy maximisation. The entropy maximisation principle as discussed in Chapter 3, provides the means of combining all relevant information about the metropolitan containerised market in a consistent and unbiased way to construct probability distributions to describe shippers' mode choice behaviours and hence terminal location decisions in the market. The entropy formulation produced a single level mathematical program where both terminal location decisions and mode choice decisions are driven by shipper preferences. This was achieved by focusing more on the shipper as the decision-maker by including the fixed cost of terminal location in the transport cost confronted by shippers choosing which if any terminal to use. This, in turn, enables the demand for terminals to be estimated as a function of the fee charged by terminal operators for terminal usage and the transport cost of accessing the terminal along the intermodal transport chain.

The rest of the chapter is organised as follows. Section 4.2 presents the mathematical formulation of **MIMTLP** underpinned by the principle of entropy maximisation. The proposed solutions of the formulated problem are presented in Section 4.3. The application of the model

in practice is discussed in Section 4.4. Finally, Section 4.5 presents a summary of how the model addresses the research question.
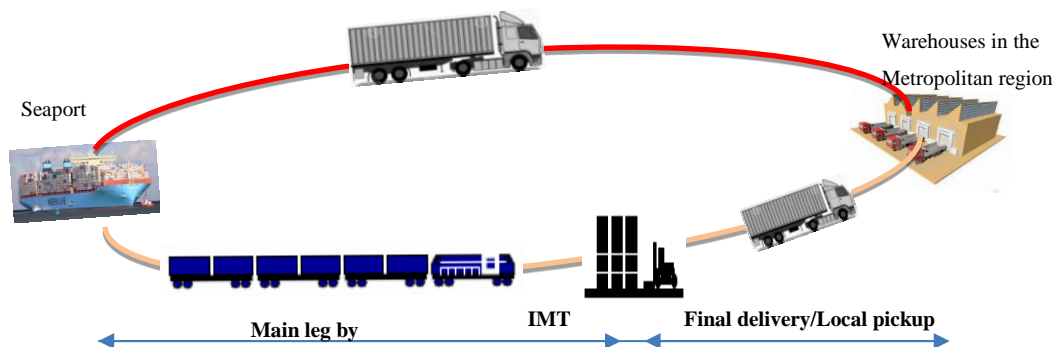
Figure 4.1: Modal options: Import market (Export market is the reverse)

4.2 Methodology

4.2.1 Problem definition and assumptions

The MIMTLP can be stated as follows:

"*Given the distribution of containerised cargo and candidate terminal locations on the transport network, what are the best places to locate p intermodal terminals to best serve the metropolitan containerised market?*"

What is best depends on the objective function and in this study, the objective to be optimised is the entropy, which is a function of the modal decision variables. The motivation for maximising entropy was discussed in Chapter 3, and in this chapter, it will be shown that maximising entropy is equivalent to maximising shippers expected utility or welfare.

In formulating the problem, it was assumed that the study area (e.g., the metropolitan region) is segmented into freight analysis zones where cargo can be seen as originating from one zone and destined to another zone. The zones are connected to both the rail and highway networks so that cargo can be transported from one zone to another using at least one mode of transport. Two main modes of transport are assumed to be available to each user or shipper; road alone transport and intermodal transport. Road alone transport mainly involves the use of trucks to transport containers to and from the port. Intermodal transport, on the other hand, combines the use of trucks and a high carrying capacity mode such as rail for the movement of

containers to and from the port, where the rail is used for the main leg and the trucks for local pickups and/or local deliveries as shown in Figure 4.1. In addition to the above assumptions, following information are assumed to be available or can readily be deduced:

1. Fixed origin-destination movement of cargo (in TEUs) in the study area.
2. The transport budget (known or assumed). Here, the analyst has the opportunity to investigate the implications of different transport budgets on IMT location and usage. Plausible ways of deriving this budget are discussed below.
3. The generalised cost of using each mode of transport between each origin-destination pair. The construction of these costs variables is discussed in below.
4. Candidate IMT sites (plausible places where IMTs can be located), the number of IMTs to locate and handling capacity of each candidate IMT location.

The intermodal transport cost c_{itj} is a very important policy variable and comprises three main components:

$$c_{itj} = c_{it} + c_t + c_{tj} \qquad\qquad (4.1)$$

where for import cargo movements, c_{it} is the unit cost of transporting cargo from the port to IMT t by rail (\$ per TEU); c_{tj} is the unit road cost from IMT $t \in \mathcal{T}$ to cargo destination $j \in \mathcal{D}$ (\$ per TEU). For export cargo movements, c_{it} is the unit truck cost of transporting cargo from origin i to IMT t (\$ per TEU); c_{tj} is the unit rail cost from IMT $t \in \mathcal{T}$ to the port (cargo destination $j \in \mathcal{D}$). The parameter c_t is the terminal usage cost or rental (\$ per TEU) passed on to the shipper, who then decides whether or not to use the terminal and comprised the fixed installation costs and terminal operation costs. The transport network cost (road or rail) between any two zones $i \neq j$ is generally assumed to consist of two cost components:

$$c_{ij} = \tilde{\varphi} + \theta GT_{ij} \qquad\qquad (4.2)$$

where $\tilde{\varphi}$ is the fixed transport cost (\$ per TEU), θ is the cost sensitivity parameter of generalised travel time, GT_{ij} between two locations on the network with the combined term θGT_{ij} representing the variable cost component. The generalised time GT_{ij} can be a function

of distance or time as may be the case for rail or a linear combination of distance and time as expressed below:

$$GT_{ij} = time_{ij} + \frac{voc}{vot} dist_{ij} \qquad (4.3)$$

$time_{tj}$ and $dist_{tj}$ are the travel time (minutes) and distance between location i and location j respectively, voc is the vehicle operating cost (\$ per km) and vot is driver's value of time savings (\$ per minutes). It can further be assumed that the truck travel times and distances will come from an existing suitable transport model of the study area and that the model contains assignment models that adequately capture the non-linearity between flows and travel times on the transport network.

The transport budget is also very important variable in the model and provides the analyst or location planner some degree of flexibility in locating IMTs to achieve certain economic, environmental or social policy targets. The budget c can be derived using $c = \bar{c}Z$, where $\bar{c}$ is the average transport cost within the study area and Z is the total cargo in the system. The average cost can be derived from a sample of origin-destination cargo flows with associated costs. Alternatively, it seems reasonable to assume that shippers individually would not choose to increase their transport costs because a new IMT becomes available, so they would not do so collectively. Thus, collectively shippers can be assumed to behave in such a way that the average transport costs over all origin-destination cargo movements after the addition of IMT(s) are no higher than before. The average transport cost can therefore be computed using equation (4.4), which is the average cost of using road alone transport. Once the average transport is known, the budget can be derived using $c = \bar{c}Z$.

$$\bar{c} = \frac{\sum_{i \in O} \sum_{j \in D} c_{ij} q_{ij}}{\sum_{i \in O} \sum_{j \in D} q_{ij}} \qquad (4.4)$$

where q_{ij} is the quantity of cargo to be transported between the sampled origin-destination cargo movements and c_{ij} is the associated road alone transport cost. The transport budget can more generally be computed using:

$$c = \kappa \bar{c}Z; \quad \kappa > 0 \qquad (4.5)$$

Equation (4.5) allows the budget to be specified such that the location of the terminals can be used to improve the average existing cost of transport in addition to the environmental and other benefits associated with intermodal transport.

Decision variables

The key decision variables are Y_t with $Y_t = 1$ indicating that candidate location $t \in \mathcal{T}$ must be developed as an intermodal terminal. The values of the variable V_{itj} represent the demand for terminal $t \in \mathcal{T}$ in the movement of cargo between zones $i \in \mathcal{O}$ and $j \in \mathcal{D}$ hence determines the demand for intermodal transport, whilst X_{ij} outputs the demand for road alone transport.

4.2.2 The Entropy Maximising Facility Location Problem

The goal is to determine the most likely p IMT locations based on available information on transport budget, cargo distribution patterns, transport costs of cargo distribution, candidate IMT locations with important features such as cargo handling capacities. Here, the entropy maximising principle is used to combine these diverse pieces of information to find the most likely p IMT locations and the least biased probability distribution of located IMT usage. The general entropy maximisation framework comprises an entropy objective function and a set of constraints representing the available information. The entropy function comprises the possible ways that a given state of the system can occur:

$$E = \frac{Z!}{\prod_{i \in \mathcal{O}} \prod_{j \in \mathcal{D}} X_{ij}! \left(\prod_{t \in \mathcal{T}} V_{itj}! \right)} \tag{4.6}$$

where E is the number of possible ways that the state (V_{itj}, X_{ij}) such that $\sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} X_{ij} + \sum_{i \in \mathcal{O}} \sum_{t \in \mathcal{T}} \sum_{j \in \mathcal{D}} V_{itj} = Z$ can occur. The important question is: based on what we know about the system, which of the many states (values of X_{ij} and V_{itj}) is most likely to represent the system? The principle of entropy maximisation is simply asking us to select the states with the maximum number of ways of occurring and consistent with all we know about the system. In general, we seek the values of X_{ij} and V_{itj} that maximises E and also satisfy all the constraints representing the available information about the system. Statistically, the values of X_{ij} and V_{itj} that maximises E also maximises $\ln E$. However, it is easier to maximise $\ln E$ so we maximised $\ln E$ instead. Thus, Equation (4.6) reduces to:

$$\ln E = \ln Z! - \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} \ln(X_{ij}!) - \sum_{i \in \mathcal{O}} \sum_{t \in \mathcal{T}} \sum_{j \in \mathcal{D}} \ln(V_{itj}!) \tag{4.7}$$

It has been shown in Chapter 3 that the term $\ln E$ has special meaning and desirable properties and Boltzmann (1972) referred to it as *entropy*. Thus, maximising (4.7) can equivalently be stated as maximising entropy. One of the desirable properties of $\ln E$ (entropy) is that its corresponds to the amount of missing information (or uncertainty or entropy) in the constructed of the probability distribution and that the maximum amount of missing information is attained when there is no known information about the system under investigation. These properties are expressed in propositions (4.1), (4.2) and (4.3) below:

Proposition 4.1: If the set $\mathcal{M} = \{t : t \in \{0, \mathcal{T}\}\}$ is set of modal alternatives, where $\{0\}$ is the index for road alone and the set $\mathcal{T}$ is the set of indices of IMTs forming the intermodal transport alternatives. Also, let the set of origin-destination movements $\mathcal{R} = \{r = (i,j) : i \in \mathcal{O}, j \in \mathcal{D}\}$ then set the of elemental alternatives $\mathcal{W} = \{w = (r,t) : r \in \mathcal{R}, t \in \mathcal{M}\}$ with cardinality $n = |\mathcal{W}|$. If the probability of each elemental alternative $w \in \mathcal{W}$ is defined by Equation (4.8):

$$P_w = \frac{Z_w}{Z}; \ \forall w \in \mathcal{W} \tag{4.8}$$

Then, equation (4.7) or entropy can be expressed as:

$$\ln E = -Z \sum_{w \in \mathcal{W}} P_w \ln P_w \tag{4.9}$$

Proof 4.1: By definition, equation (4.7) can be simplified as:

$$\ln E = \ln Z! - \sum_{w \in \mathcal{W}} \ln Z_w!$$

Applying Stirling's approximation, the above equation simplifies to:

$$\ln E = Z(\ln Z - 1) - \sum_{w \in \mathcal{W}} Z_w(\ln Z_w - 1) \tag{4.10}$$

Substituting Equation (4.8) into (4.10) and performing some algebraic manipulation we have:

$$\ln E = -Z \sum_{w \in \mathcal{W}} P_w \ln P_w \tag{4.11}$$

Proposition 4.2: In the absence of any other information about the freight system, maximising equation (4.11) produces uniform probability distributions of modal flows:

$$P_w = \frac{1}{n} \; ; \; \forall w \in \mathcal{W} \tag{4.12}$$

with corresponding maximum entropy:

$$H = \ln \tilde{E} = Z \ln(n) \tag{4.13}$$

where n is the cardinality of the set $\mathcal{W}$

Proof 4.2: Now if we assume there is no information available other than obeying the normalisation axiom of probability:

$$\sum_{w \in \mathcal{W}} P_w = 1 \tag{4.14}$$

then from equation (4.11) the first order condition for maximum $\ln E$ with respect to P_w and subject to (4.14) satisfy the following equation:

$$-Z \ln(P_w) - 1 - \varphi = 0; \; \forall w \in \mathcal{W} \tag{4.15}$$

where φ is the Lagrangian multiplier associated with constraint (4.14). Solving for P_w in (4.15) by enforcing constraint (4.14) we have:

$$P_w = \frac{1}{n} \; ; \; \forall w \in \mathcal{W}$$

and substituting the above into Equation (4.11) the maximum entropy can be computed:

$$H = Z\ln(n)$$

Proposition 4.3: The entropy (or amount of missing information (H) in the probability distributions) constructed based on any amount of available information about the system cannot be greater than the entropy in equation (4.13). That is $H \leq Z\ln(n)$

Proof 4.3: Define a convex function $\phi(x) = x\ln(x); \; \forall x \geq 0$. Following Jensen's inequality, the following must hold:

$$-\frac{1}{n} \sum_{w \in \mathcal{W}} \phi(P_w) \leq -\phi\left(\frac{1}{n} \sum_{w \in \mathcal{W}} P_w\right) \tag{4.16}$$

Applying the definition of the convex function ϕ to the term on the left-hand side of Equation (4.16) and using Equations (4.11) we have:

$$-\frac{1}{n} \sum_{w \in \mathcal{W}} \phi(P_w) = -\frac{1}{n} \sum_{w \in \mathcal{W}} P_w \ln(P_w) = (H)\frac{1}{nZ}$$

Also, applying the definition of ϕ to the term on the right-hand side of (4.16) we have:

$$-\phi\left(\frac{1}{n} \sum_{w \in \mathcal{W}} P_w\right) = -\frac{1}{n}\ln\left(\frac{1}{n}\right) = \frac{1}{n}\ln(n)$$

It therefore follows from equation (4.16) that

$$H \leq Z\ln(n) \tag{4.17}$$

The result in equation (4.17) is intuitive since it implies that the more information we have, the less entropy or uncertainty we have about the resulting probability distribution and vice versa. In other words, on average, the amount of missing information (entropy) about the system under investigation is never increased by learning something about it. Equation (4.12) simply states that applying the principle of maximum entropy to the MCTM with no evidence to suggest why a particular modal alternative should be preferred more than the others will result in a uniform probability distribution. The next section presents the information available in the form of constraints to form the entropy maximising facility location problem (EMFLP).

4.2.3 Available evidence as constraints

For the purpose of this exercise, the evidence available are summarised as follows:

1.Budget constraint: It is assumed that the transport budget c is known. This evidence is added as constraint (4.18). The first component captures the weighted cost of using intermodal transport and the second captures the weighted cost of using road alone transport (e.g. truck only), with the sum not exceeding the total allocated transport budget.

$$\sum_{i \in \mathcal{O}} \sum_{t \in \mathcal{T}} \sum_{j \in \mathcal{D}} c_{itj} V_{itj} + \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} c_{ij} X_{ij} \leq c \tag{4.18}$$

2. Conservation of cargo flow constraint. Information on the distribution of origin-destination flows of cargo (q_{ij}) by modes is added as constraint (4.19). It ensures that for each origin-destination pair, the sum of cargo by all available modes equals the total cargo associated with this origin-destination pair.

$$\sum_{t \in \mathcal{T}} V_{itj} + X_{ij} = q_{ij}; \quad \forall \ i \in \mathcal{O}, j \in \mathcal{D} \tag{4.19}$$

3. Definitional constraint. The information on the required number of IMTs (p) to locate is presented by constraint (4.20). It ensures that only the required number of IMTs are located.

$$\sum_{t \in \mathcal{T}} Y_t = p \tag{4.20}$$

4. Capacity constraint. The cargo handling capacity limit of the each IMT is captured in constraints (4.21) and guarantees that no located IMT exceeds its cargo handling capacity. The constraints also ensure that only open IMTs are used.

$$\sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} V_{itj} \leq Y_t b_t ; \quad \forall t \in \mathcal{T} \tag{4.21}$$

Finally, the entropic objection function in (4.7) can be simplified by applying Stirling's approximation to the factorial terms, ignoring the constant term, $\ln Z!$ as it does not influence the optimisation process:

$$\ln E \sim \sum_{i \in \mathcal{O}} \sum_{t \in \mathcal{T}} \sum_{j \in \mathcal{D}} V_{itj}\left(1 - \ln V_{itj}\right) + \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} X_{ij}\left(1 - \ln X_{ij}\right) \tag{4.22}$$

Once the available information are converted to constraints (4.18-4.21), the EMFLP is presented as follows:

$$\text{EMFLP} : \text{Max } \Lambda = \sum_{i \in \mathcal{O}} \sum_{t \in \mathcal{T}} \sum_{j \in \mathcal{D}} V_{itj}\left(1 - \ln V_{itj}\right) + \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} X_{ij}\left(1 - \ln X_{ij}\right)$$

Subject to constraint (4.18) to (4.21) and the following integer and non-negativity constraints:

$$Y_t \in \{0,1\} ; \quad t \in \mathcal{T} \tag{4.23}$$

$$V_{itj} \geq 0; \; X_{ij} \geq 0 ; \forall t \in \mathcal{T}; \; \forall i \in \mathcal{O}; j \in \mathcal{D} \tag{4.24}$$

4.3 Solving the EMFLP

4.3.1 Background

The EMFLP is NP-hard since it includes mixed integer linear programming (Sorensen et al. 2012). This suggests that it is unlikely that efficient algorithms for solving every instance of it (EMFLP) can be found (Garey and Johnson 1979). However, considering the fact that there are few plausible places in a metropolitan region to place IMTs, complete enumeration of IMT locations is feasible for most practical problems. The study therefore explores the structure of the problem and notes that it is possible to separate the location aspect of the problem from the mode choice aspect, allowing for complete enumeration algorithm or other exact algorithms like branch and bound to be used to solve the overall problem. This was achieved by relaxing constraint (4.21) or more generally relaxing all constraints expressed in terms of both location and flow variables:

$$\text{Max}\{\Lambda_R\}$$

Subject to constraints (4.18) to (4.20), (4.23) and (4.24)

where

$$\Lambda_R = \sum_{i \in \mathcal{O}} \sum_{t \in \mathcal{T}} \sum_{j \in \mathcal{D}} V_{itj}\big(1 - \ln V_{itj}\big) + \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} X_{ij}\big(1 - \ln X_{ij}\big)$$

$$+ \sum_{t \in \mathcal{T}} \psi_t \left(Y_t b_t - \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} V_{itj} \right)$$

and where $\psi_t \geq 0; \forall t \in \mathcal{T}$ are Lagrangian multipliers associated with the relaxed constraint (4.21). The relaxed problem can then be decomposed it into two sub-problems; the facility location sub-problem (FLP) and mode choice sub-problem (MCP) with the FLP given as:

$$\text{FLP} : \text{Min } \Lambda_{\text{FLP}} = \sum_{t \in \mathcal{T}} (\psi_t b_t) Y_t$$

Subject to constraints (4.20) and (4.23)

and the MCP with a simplified objective function becomes:

$$\text{MCP}: \text{Max } \Lambda_{\text{MCP}} = \sum_{i \in O} \sum_{j \in D} X_{ij} \left\{ 1 - \ln X_{ij} \right\} + \sum_{i \in O} \sum_{t \in \mathcal{T}} \sum_{j \in D} V_{itj} \left\{ 1 - \ln V_{itj} - \psi_t \right\}$$

Subject to constraints (4.18), (4.19) and (4.24) and $\psi_t \geq 0; \ \forall t \in \mathcal{T}$.

Thus, overall objective function Λ_R reduces to:

$$\Lambda_R = \Lambda_{\text{FLP}} + \Lambda_{\text{MCP}} \tag{4.25}$$

4.3.2 Solution the FLP

Given $\psi_t \geq 0; \ \forall t \in \mathcal{T}$ the FLP can be solved by identifying the p largest elements of $(\psi_t b_t); \ \forall t \in \mathcal{T}$ and setting the corresponding values of Y_t equal to 1. Let $\mathcal{K}$ with cardinality p be the set of located IMTs, which then goes into the MCP. Note that for sufficiently large IMTs, $\psi_t = 0; \ \forall t \in \mathcal{T}$, the objective value ($\Lambda_{\text{FLP}}$) of the FLP will be zero and the selection of the best IMTs will only be based on the value of Λ_{MCP}. Thus, all things being equal IMTs with $\psi_t > 0$ are more likely to be selected as expected. This is because if two competing IMTs have equal demands but one of them has $\psi > 0$, it implies this IMT is more attractive and would have attracted more demand had it not been restricted by its capacity following the definition of $\psi > 0$.

4.3.3 Solution to the MCP

Similarly, the solution to the MCP is based on the assumption that $\mathcal{K}$, the set of IMTs to locate is known. The following propositions investigate the existence and uniqueness of MCP solutions for any given set of located IMTs $\mathcal{K}$.

Proposition 4.4 (Existence): Given that the set $\mathcal{K}$ is known and the set of feasible solutions, $\mathbb{S}$ to the MCP is defined by constraints (4.18), (4.19), (4.21) and (4.24) as:

$$\mathbb{S} = \left\{ Z_{itj} \, \middle| \, (4.18), (4.19), (4.21) \text{ and } (4.24); \ \forall \, i \in O; j \in D; t \in \mathcal{M} \right\}$$

then at least one solution to the MCP exists in $\mathbb{S}$.

Proof 4.4: Set $\mathbb{S}$ is convex, closed and bounded as it is defined by linear equality (4.19) and inequality constraints (4.18), (4.21) and (4.24). The set $\mathbb{S}$ is also not empty in terms of the cargo flow variables V_{itj} and X_{ij}; $\forall\, i \in \mathcal{O}; j \in \mathcal{D}, t \in \mathcal{K}$ since for any $q_{ij} > 0$; $\sum_{t \in \mathcal{K}} V_{itj} + X_{ij} > 0$; $\forall\, i \in \mathcal{O}; j \in \mathcal{D}$. Finally, location variables Y_t; $\forall\, t \in \mathcal{K}$ and the flow variables V_{itj}; $\forall\, i \in \mathcal{O}; j \in \mathcal{D}, t \in \mathcal{K}$ are linearly related through constraint (4.21). Hence the feasible region corresponding to Y_t; $\forall\, t \in \mathcal{K}$ is also convex, compact (closed and bounded) and non-empty. Hence the MCP possesses at least one solution (Boyd and Vandenberghe 2009).

Proposition 4.5 (Uniqueness): If a solution of the MCP model exists in $\mathbb{S}$ then that solution must be unique.

Proof 4.5: To prove uniqueness we first show that the objective function Λ_{MCP} is strictly convex in $\mathbb{S}$. The second-order partial derivatives of Λ_{MCP} at Z_{itj}; $\forall\, i \in O; j \in D; t \in \mathfrak{M} = \{0, \mathcal{K}\}$:

$$\frac{\partial^2 \Lambda_{\text{MCP}}}{\left(\partial Z_{itj}\right)^2} = -\frac{1}{Z_{itj}}; \; \forall\, i \in \mathcal{O}; j \in \mathcal{D}; t \in \mathfrak{M} \tag{4.26}$$

$$\frac{\partial^2 \Lambda_{\text{MCP}}}{\left(\partial Z_{itj}\right)\left(\partial Z_{isj}\right)} = 0; \; \forall\, i \in \mathcal{O}; j \in \mathcal{D}; t \neq s \in \mathfrak{M} \tag{4.27}$$

From (4.26), the second-order derivative or the Hessian of Λ_{MCP} is negative definite since:

$$[\nabla^2 \Lambda_{\text{MCP}}]_{i,t,j} < 0; \; \forall\, Z_{itj} > 0; \; \forall\, i \in \mathcal{O}; j \in \mathcal{D}; t \in \mathfrak{M} \tag{4.28}$$

Hence the objective function Λ_{MCP} of the MCP model is strictly concave for all non-zero $Z_{itj} \in \mathbb{S}$; $\forall\, i \in \mathcal{O}; j \in \mathcal{D}; t \in \mathfrak{M}$ (Boyd and Vandenberghe 2009). Because the feasible region $\mathbb{S}$ is compact, convex and non-empty, and the objective function of the MCP model Λ_{MCP} is strictly concave on $\mathbb{S}$, hence the MCP model has a unique solution (Boyd and Vandenberghe 2009).

The solution to the MCP has been shown to exist and is unique for any given set of IMT locations $\mathcal{K}$. The next step to find this solution, which can be found by constructing a

Lagrangian equation comprising the objective function associated with MCP and the corresponding constraints and enforcing the Karush-Kuhn-Tucker (KKT) condition:

$$\Lambda_{\text{LMCP}} = \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} X_{ij} \{1 - \ln X_{ij}\} + \sum_{i \in \mathcal{O}} \sum_{t \in \mathcal{K}} \sum_{j \in \mathcal{D}} V_{itj} \{1 - \ln V_{itj} - \psi_t\}$$

$$+ \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} \eta_{ij} \left\{ q_{ij} - \sum_{t \in \mathcal{K}} V_{itj} - X_{ij} \right\}$$

$$+ \beta \left\{ c - \sum_{i \in \mathcal{O}} \sum_{t \in \mathcal{K}} \sum_{j \in \mathcal{D}} c_{itj} V_{itj} - \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} c_{ij} X_{ij} \right\}$$

where η_{ij} and $\beta \geq 0$ are Lagrangian multipliers associated with constraints (4.19) and (4.18) respectively.

The KKT conditions for optimality of Λ_{LMCP} with respect to V_{itj} and X_{ij} are:

$$-\ln(V_{itj}) - \psi_t - \eta_{ij} - \beta c_{itj} = 0; \forall \, i \in \mathcal{O}; j \in \mathcal{D}; t \in \mathcal{K} \tag{4.29}$$

$$-\ln(X_{ij}) - \eta_{ij} - \beta \omega_{ij} = 0; \; \forall \, i \in \mathcal{O}; j \in \mathcal{D} \tag{4.30}$$

Solving V_{itj} and X_{ij} in (4.29) and (4.30) we have:

$$V_{itj} = e^{-\psi_t - \eta_{ij} - \beta c_{itj}}; \; \forall \, i \in \mathcal{O}; j \in \mathcal{D}; t \in \mathcal{K} \tag{4.31}$$

$$X_{ij} = e^{-\eta_{ij} - \beta \omega_{ij}} \quad ; \forall \, i \in \mathcal{O}; j \in \mathcal{D} \tag{4.32}$$

Enforcing constraints (4.19), the Lagrangian parameters η_{ij} can be estimated using equations (4.31) to (4.32):

$$e^{\eta_{ij}} = \frac{1}{q_{ij}} \left\{ e^{-\beta c_{ij}} + \sum_{t \in \mathcal{K}} e^{-(\beta c_{itj} + \psi_t)} \right\} \tag{4.33}$$

Also, the Lagrangian multipliers for constraint (4.21) can be estimated as using (4.31):

$$\psi_t = \max\left\{0, \ln\left(\frac{\Omega_t}{b_t}\right)\right\} \ \forall t \in \mathcal{K} \tag{4.34}$$

where

$$\Omega_t = \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} e^{-\eta_{ij}} e^{-\beta c_{itj}} \ ; \ \forall \, t \in \mathcal{K}$$

The Lagrangian parameter β (cost sensitivity) associated with constraint (4.18) can be derived for a given budget c as follows:

$$f(\beta) = \sum_{i \in \mathcal{O}} \sum_{t \in \mathcal{K}} \sum_{j \in \mathcal{D}} c_{itj} X_{itj} + \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} c_{ij} W_{ij} - c \leq 0 \tag{4.35}$$

or from Equations (4.31) and (4.32):

$$f(\beta) = \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} e^{-\eta_{ij}} \left\{ c_{ij} e^{-\beta c_{ij}} + \sum_{t \in \mathcal{K}} c_{itj} e^{-(\beta c_{itj} + \psi_t)} \right\} - c \leq 0 \tag{4.36}$$

The function $f(\beta)$ is continuous and differentiable with respect to β hence the parameter β can be estimated using Newton-Raphson's or Hyman's method (Hyman 1969) with the assumption that η_{ij} and ψ_t are known. Extensive numerical examples show that Hyman's method is computationally more efficient for estimating β, an outcome generally consistent with the conclusion reached by Williams (1976) who compared several estimation methods for calibrating gravity type models with exponential functions. Thus, all estimated values of β in this thesis are based on Hyman's method.

The solution to the MCP can be simplified further by converting the solutions in Equations (4.31) and (4.32) into a two-level nesting structure (see Figure 4.2), where the distribution of demand for each mode of transport (road versus intermodal transport) conditions the distribution of demand for the located IMTs. Conversely, the models are connected in the opposite direction by accessibility measures such that improving accessibility to the located IMTs influences the demand for intermodal transport.
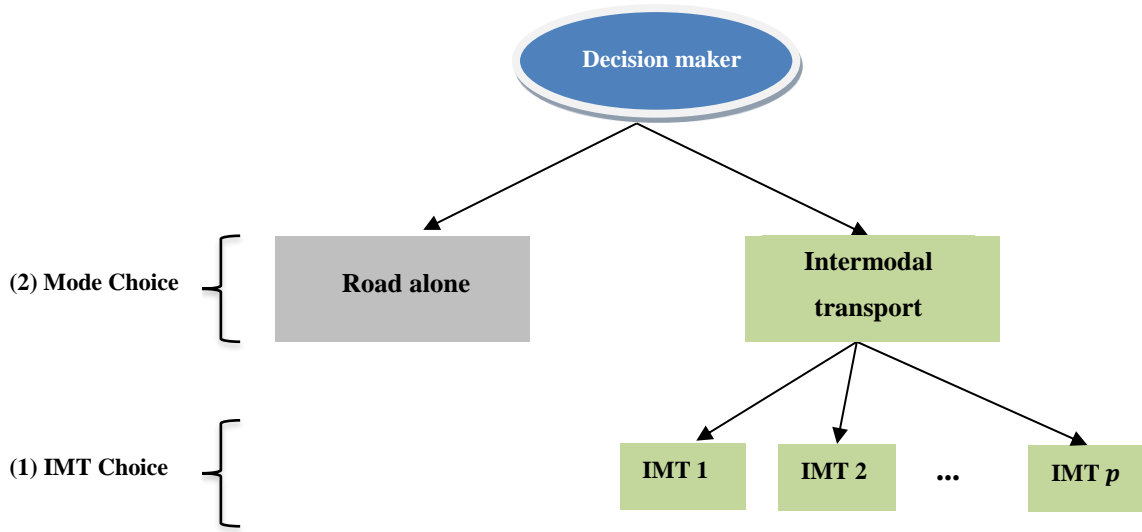
Figure 4.2: Nesting structure of mode choice problem (MCP)

By inserting (4.33) into (4.31) and (4.32) it follows that:

$$V_{itj} = q_{ij} \frac{e^{-\beta c_{itj} - \psi_t}}{\sum_{t \in \mathcal{K}} e^{-\beta c_{itj} - \psi_t} + e^{-\beta c_{ij}}} = q_{ij} \Pr(V_{itj}); \quad \forall \, i \in \mathcal{O}; j \in \mathcal{D}; t \in \mathcal{K} \tag{4.37}$$

$$X_{ij} = q_{ij} \frac{e^{-\beta c_{ij}}}{\sum_{t \in \mathcal{K}} e^{-\beta c_{itj} - \psi_t} + e^{-\beta c_{ij}}} = q_{ij} \Pr(X_{ij}) ; \forall \, i \in \mathcal{O}; j \in \mathcal{D} \tag{4.38}$$

where $\Pr(V_{itj})$ is the probability that demand V_{itj} is realised for IMT $t \in \mathcal{K}$ when transporting cargo between origin zone $i \in \mathcal{O}$ and destination zone $j \in \mathcal{D}$:

$$\Pr(V_{itj}) = \Pr(V_{ij}) \Pr(V_{itj} | \tilde{V}_{ij}); \quad \forall \, i \in \mathcal{O}; j \in \mathcal{D}; t \in \mathcal{K} \tag{4.39}$$

where $\Pr(V_{ij})$ is the probability that demand V_{ij} is realised for intermodal transport between each origin-destination pair:

$$\Pr(V_{ij}) = \frac{e^{\ell_{ij}}}{e^{\ell_{ij}} + e^{-\beta c_{ij}}} ; \forall \, i \in \mathcal{O}; j \in \mathcal{D} \tag{4.40}$$

The variable ℓ_{ij} is the logsum or, with reference to random utility theory, the maximum expected utility over all located IMTs and serves as a measure of access to intermodal transport for each origin-destination pair:

$$\ell_{ij} = \ln \sum_{t \in \mathcal{K}} e^{-\beta c_{itj} - \psi_t} \quad ; \forall i \in \mathcal{O}; j \in \mathcal{D} \tag{4.41}$$

and $\Pr(V_{itj}|\tilde{V}_{ij})$ is the conditional probability for realising the demand V_{itj} for IMT $t \in \mathcal{K}$ given that intermodal transport demand $\tilde{V}_{ij}$ is known for each origin-destination pair:

$$\Pr(V_{itj}|\tilde{V}_{ij}) = \frac{e^{-\beta c_{itj} - \psi_t}}{\sum_{t \in \mathcal{K}} e^{-\beta c_{itj} - \psi_t}} \; ; \forall i \in \mathcal{O}; j \in \mathcal{D}; t \in \mathcal{K} \tag{4.42}$$

Similarly, from (4.38), $\Pr(X_{ij})$ is the probability that the demand X_{ij} is realised for road alone transport between origin zone $i \in \mathcal{O}$ and destination zone $j \in \mathcal{D}$:

$$\Pr(X_{ij}) = \frac{e^{-\beta c_{ij}}}{e^{\ell_{ij}} + e^{-\beta c_{ij}}} \; ; \forall i \in \mathcal{O}; j \in \mathcal{D} \tag{4.43}$$

The maximum expected utility or logsum (Williams 1977) over all modes (road alone and intermodal transport) from the denominator of (4.43) can be expressed as:

$$L_{ij} = \ln\left(e^{-\beta c_{ij}} + e^{\ell_{ij}}\right) = \ln\left(e^{\ell_{ij}} + \sum_{t \in \mathcal{K}} e^{-\beta c_{itj} - \psi_t}\right); \; \forall i \in \mathcal{O}; j \in \mathcal{D} \tag{4.44}$$

The term L_{ij} represent a measure of access to multiple modes of transport for the movement of cargo to destination zone $j \in \mathcal{D}$ from origin zone $i \in \mathcal{O}$. Looking at Equation (4.44) it is clear that the logsum L_{ij} and the Lagrangian multipliers in Equation (4.33) are related as follows:

$$\eta_{ij} = -\ln\left(q_{ij} e^{-L_{ij}}\right) \tag{4.45}$$

An important outcome of the above analysis is that the Lagrangian multipliers ψ_t in (4.34) can be computed more intuitively through proposition 4.6 below:

Proposition 4.6: The Lagrangian parameter $\psi_t; \forall\, t \in \mathcal{K}$ in Equation (4.34) can be evaluated iteratively as a function of the total estimated demand for each IMT $V_t; \forall t \in \mathcal{K}$ and its handling capacity $b_t; \forall t \in \mathcal{K}$ with the kth iterated value evaluated using equation (4.46):

$$\psi_t{}^k = \psi_t{}^{k-1} + \max\left\{0, \ln\left(\frac{V_t{}^{k-1}}{b_t}\right)\right\} \quad ; \forall\, t \in \mathcal{K} \tag{4.46}$$

where $V_t{}^{k-1}$ is the estimated demand for IMT $t \in \mathcal{K}$ from the previous iteration:

$$V_t{}^{k-1} = \sum_{i\in\mathcal{O}}\sum_{j\in\mathcal{D}} V_{itj}{}^{k-1} = \sum_{i\in\mathcal{O}}\sum_{j\in\mathcal{D}} q_{ij}\Pr(V_{itj})^{k-1} \quad ; \forall t \in \mathcal{K} \tag{4.47}$$

Proof 4.6: From Equation (4.34) the computation of Ω_t at iteration k using the logsum in Equation (4.44) can be expressed as:

$$\Omega_t{}^k = \sum_{i\in\mathcal{O}}\sum_{j\in\mathcal{D}} q_{ij}\frac{e^{-\beta c_{itj}}}{e^{L_{ij}{}^{k-1}}}$$

From the probabilities in equations (4.37) we have:

$$\Omega_t{}^k = e^{\psi_t{}^{k-1}}\left\{\sum_{i\in\mathcal{O}}\sum_{j\in\mathcal{D}} q_{ij}\Pr(V_{itj})^{k-1}\right\}$$

The term in the bracket represents the total demand for IMT $t \in \mathcal{K}$ and from (4.47) the above equation reduces to:

$$\Omega_t{}^k = e^{\psi_t{}^{k-1}}V_t{}^{k-1} \tag{4.48}$$

also taking the natural logarithm on both sides of equation (4.48) we have:

$$\ln \Omega_t^{\ k} = \psi_t^{\ k-1} + \ln V_t^{\ k-1}$$

Hence equation (4.34) can be re-expressed as:

$$\psi_t^{\ k} = \max \left\{ 0, \psi_t^{\ k-1} + \ln \left(\frac{V_t^{\ k-1}}{b_t} \right) \right\}; \ \forall t \in \mathcal{K}$$

since by definition $\psi_t \geq 0; \forall t \in \mathcal{K}$ the above equation reduces to:

$$\psi_t^{\ k} = \psi_t^{\ k-1} + \max \left\{ 0, \ln \left(\frac{V_t^{\ k-1}}{b_t} \right) \right\}; \ \forall t \in \mathcal{K}$$

Proposition 4.6 provides simple and intuitive way of computing the Lagrangian parameters associated with the IMT capacity constraint (4.21). It implies that the parameters ψ_t can be updated by simply comparing the estimated demand of an IMT with its handling capacity.

As shown in Equations (4.33) to (4.45) the estimation of the parameters in the MCP are inter-dependent, where the evaluated value of one is required to solve the other. A general technique for solving this kind of problem, called Bregman's balancing method (Lamond and Stewart 1981), has already been developed and been shown to be routed in Brouwer's (1910) fixed point theorem and converge to an acceptable level of accuracy (Wilson 2010). We adapted this method to estimate these parameters in algorithm A₁ below. The convergence criteria can be based on the convergence of both β and ψ parameters.

Algorithm A₁: Modified Bregman's algorithm for solving the MCP

1.Initialisation:
 	
  For a given set of located IMTs $\mathcal{K}$ with size p and starting cost sensitivity parameter $\beta = \frac{1}{\bar{c}}$, where
 	
  $\bar{c}$ can be the average transport budget and $\psi_t = 0; \ \forall \ t \in \mathcal{K}$
 	
2.Logsums Update
 	
  2.1 Update logsums over all located IMTs ℓ_{ij} using Equation (4.41)
 	
  2.2 Update the logsums L_{ij} over all transport modes using Equation (4.44)
 	
3.Flows Update
 	
  3.1 Update the demand V_{itj} for each located IMT using Equation (4.37)
 	
  3.2 Update the demand road alone transport X_{ij} using (4.38)

4.Update β from equation (4.35) using Newton Raphson or Hyman's method (Hyman 1969)

5.Update the Lagrangian multipliers $\psi_t; \forall\ t \in \mathcal{K}$ for IMT capacity constraints using the iterative method in Proposition (4.46).

6. Repeat steps (2) - (5) until convergence is achieved.

Algorithm A₁ can be terminated once the changes in the estimated values of both parameters are smaller than pre-defined thresholds. It has been demonstrated that once the EMFLP is decomposed, the resulting sub-problems can be solved to optimality.

4.4 Solution to the overall EMFLP

As illustrated in Section (4.3), the solution to the FLP relies on the assumption of knowing the evaluated values of the Lagrangian parameters, whilst the solution to the MCP is based on the assumption of knowing the set of located IMTs. Thus, one sub-problem cannot be solved without knowing the solution of the other. However, the decomposition allows Algorithm A₁ for solving the MCP to be embedded in any of the general enumeration algorithms, such as branch and bound (B&B) or complete enumeration (CE), to solve the overall EMFLP to optimality. Considering the fact that the metropolitan region is the main study area for the application of the proposed model, CE based algorithm A₂ with embedded algorithm A₁ is practical for solving the overall EMFLP. Note that the size or cardinality of the set $\mathcal{U}$, which is the set of all subsets of the candidate IMTs $\mathcal{T}$ with cardinality p is polynomially bounded by:

$$|\mathcal{U}| = \binom{\tau}{p} = \frac{\tau!}{p!\,(\tau - p)!} = O(\tau^p) \tag{4.49}$$

where τ is the cardinality of the set $\mathcal{T}$. For example, if the analyst is interested in locating two IMTs, then the number of possible evaluations of the MCP is bounded by $\tau(\tau - 1)$. Also, Bregman's balancing method for solving MCP converges in polynomial time (Lamond and Stewart 1981) to an acceptable level of accuracy, making the use of algorithm A₂ efficient. Once the EMFLP is decomposed, the application of the algorithm A₂ is straight forward. This is because, once the set $\mathcal{K}$ (set of p IMTs) is known, constraints (4.20) and (4.23) are automatically satisfied; the rest of the constraints are satisfied by solving the MCP for the given $\mathcal{K}$. The algorithm A₂ is presented as follows:

Algorithm A₂: Solution by complete enumeration

1. Initialization: $\mathcal{K} = \{0\}, \Lambda^* = -\infty, \mathcal{K}^*$ the set with the optimum IMT sites with associated objective value Λ^*

2. For each subset $\mathcal{K} \in \mathcal{U}$ of size p, with the location variable $Y_t = 1; \forall t \in \mathcal{K}; Y_t = 0 \; \forall t \notin \mathcal{K}$ do:

    2.1. Solve the MCP using algorithm A₁ for the flow variables and the Lagrangian parameters

    2.2. Compute Λ_R using the overall objective function in Equation (4.25)

    2.3. If $\Lambda_R > \Lambda^*$, then $\Lambda^* = \Lambda_R$ and $\mathcal{K}^* = \mathcal{K}$

3. Repeat step (2) for all subsets of $\mathcal{U}$ and stop

4. Set $Y_t = 1, \forall t \in \mathcal{K}^*$ and $Y_t = 0, \forall t \notin \mathcal{K}^*$

Proposition 4.7: For simplicity let $P_{ij} = \Pr(X_{ij})$; and $P_{itj} = \Pr(V_{itj})$. Maximising Λ (the entropy objective function of EMFLP), is equivalent to maximising the weighted expected utility or welfares of all shippers subject to the given transport budget.

Proof 4.7: Using the definitions of probabilities in equations (4.37) and (4.38), the entropy function in (4.22) Λ can be re-expressed as:

$$\Lambda = \sum_{i\in\mathcal{O}}\sum_{j\in\mathcal{D}} q_{ij}P_{ij}\{1 - \ln(q_{ij}P_{ij})\} + \sum_{i\in\mathcal{O}}\sum_{t\in\mathcal{T}}\sum_{j\in\mathcal{D}} q_{ij}P_{itj}\{1 - \ln(q_{ij}P_{itj})\}$$

Expanding, grouping like terms and using the second axiom of probability we have:

$$\Lambda = \sum_{i\in\mathcal{O}}\sum_{j\in\mathcal{D}} q_{ij} - \sum_{i\in\mathcal{O}}\sum_{j\in\mathcal{D}} q_{ij}P_{ij}\ln(q_{ij}P_{ij}) - \sum_{i\in\mathcal{O}}\sum_{t\in\mathcal{T}}\sum_{j\in\mathcal{D}} q_{ij}P_{itj}\ln(q_{ij}P_{itj})$$

Expanding the terms in the logarithm function and grouping like terms we have:

$$\Lambda = \sum_{i\in\mathcal{O}}\sum_{j\in\mathcal{D}} q_{ij} - \sum_{i\in\mathcal{O}}\sum_{j\in\mathcal{D}} q_{ij}\ln(q_{ij}) - \sum_{i\in\mathcal{O}}\sum_{j\in\mathcal{D}} q_{ij}P_{ij}\ln(P_{ij}) - \sum_{i\in\mathcal{O}}\sum_{t\in\mathcal{T}}\sum_{j\in\mathcal{D}} q_{ij}P_{itj}\ln(P_{itj})$$

Using Sterling's approximation, the above can be simplified as:

$$\Lambda = -\sum_{i\in\mathcal{O}}\sum_{j\in\mathcal{D}} q_{ij}! - \sum_{i\in\mathcal{O}}\sum_{j\in\mathcal{D}} q_{ij}\left(P_{ij}\ln P_{ij} + \sum_{t\in\mathcal{T}} P_{itj}\ln P_{itj}\right)$$

The term $-\sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} \ln q_{ij}!$ is constant and can be ignored in the optimisation process, since q_{ij} (input data) are not decision variables. Thus,

$$\Lambda \approx -\sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} q_{ij} P_{ij} \ln P_{ij} - \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} q_{ij} \sum_{t \in \mathcal{T}} P_{itj} \ln P_{itj} \qquad (4.50)$$

From Equation (4.37), $\beta c_{itj} + \psi_t = \beta \tilde{c}_{itj}$, where $\tilde{c}_{itj}$ is the total cost of intermodal transport:

$$\tilde{c}_{itj} = c_{itj} + \frac{\psi_t}{\beta} = c_{it} + \tilde{c}_t + c_{tj} \qquad (4.51)$$

where the term $\tilde{c}_t = c_t + \frac{\psi_t}{\beta}$ is the terminal total user fee passed on to the shipper, who then decides whether or not to use the terminal and comprises the original user cost and a shadow price $\frac{\psi_t}{\beta}$ to dissuade enough users from using IMTs with insufficient handling capacities ($\psi_t > 0$). The shadow price is treated as an out of pocket cost and forms part of the terminal usage cost or rental passed on to the shipper. Replacing the probabilities in (4.50) with those in (4.39) and (4.43) and using (4.37) and (4.38), the entropy function in (4.50) simplifies to become:

$$\Lambda \approx \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} L_{ij} q_{ij} + \beta \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} \left(c_{ij} X_{ij} + \sum_{t \in \mathcal{T}} \tilde{c}_{itj} V_{itj} \right) \qquad (4.52)$$

Thus, maximising entropy Λ is equivalent to maximising the weighted maximum expected utility or weighted consumer surplus subject to the given transport budget:

$$\max \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} L_{ij} q_{ij}$$

Subject to the transport budget constraint (4.18):

$$\sum_{i \in \mathcal{O}} \sum_{t \in \mathcal{T}} \sum_{j \in \mathcal{D}} \tilde{c}_{itj} V_{itj} + \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} c_{ij} X_{ij} \leq c$$

where L_{ij} is the maximum expected utility (see Equation (4.44); Batty 2010; Williams 1977) or consumer surplus (Train 2009; De Jong et al. 2005). The proposition implies that, the maximum entropy yields the maximum consumer surplus or shippers' welfares.

Preposition 4.8: Comparing the solution of MCP and equivalent LP solution

For any given set of located IMTs $\mathcal{K}$, the EMFLP reduces to the MCP and the MILP reduces to equivalent linear programming (LP) solutions or equivalently the solution to EMFLP reduces to the mixed integer linear programming (MILP) solution as $\beta \to \infty$.

Proof 4.8: By using the generalised cost definition $\tilde{c}_{itj}$ in (4.51) in Equations (4.37) and (4.38) the flow variables can be estimated directly using:

$$V_{itj}{}^* = q_{ij} \frac{e^{-\beta \tilde{c}_{itj}}}{\sum_{t \in \mathcal{K}} e^{-\beta \tilde{c}_{itj}} + e^{-\beta c_{ij}}}; \quad \forall i \in \mathcal{O}; j \in \mathcal{D}, t \in \mathcal{K} \tag{4.53}$$

$$X_{ij}{}^* = q_{ij} \frac{e^{-\beta c_{ij}}}{\sum_{t \in \mathcal{K}} e^{-\beta \tilde{c}_{itj}} + e^{-\beta c_{ij}}}; \quad \forall i \in \mathcal{O}; j \in \mathcal{D} \tag{4.54}$$

From constraint (4.18), the total budget used c can be expressed as:

$$c = \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} \left(\sum_{t \in \mathcal{K}} \tilde{c}_{itj} V_{itj}{}^* + c_{ij} X_{ij}{}^* \right) = \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} C_{ij}$$

where

$$C_{ij} = \sum_{t \in \mathcal{K}} \tilde{c}_{itj} V_{itj}{}^* + c_{ij} X_{ij}{}^*$$

Using Equations (4.53) and (4.54), the above cost equation can be expressed as:

$$\frac{1}{q_{ij}} C_{ij} = \frac{1}{\sum_{t \in \mathcal{K}} e^{-\beta \tilde{c}_{itj}} + e^{-\beta c_{ij}}} \left(c_{ij} e^{-\beta c_{ij}} + \sum_{t \in \mathcal{K}} \tilde{c}_{itj} e^{-\beta \tilde{c}_{itj}} \right) \tag{4.55}$$

Suppose that the transport costs $\tilde{c}_{itj}$ and c_{ij} are kept fixed, leaving the origin-destination average budget C_{ij} and β so that if the budget changes β must also change and vice versa. It is clear from Equation (4.55) that as $\beta \to \infty$ the term with the smallest cost (c^*_{ij}) become the biggest term in both the numerator and denominator on the right hand-side of equation (4.55). Thus as $\beta \to \infty$, Equation (4.55) reduces to:

$$\frac{1}{q_{ij}} C_{ij} \to c^*_{ij} \frac{e^{-\beta c^*_{ij}}}{e^{-\beta c^*_{ij}}} = c^*_{ij}$$

which simplified to become:

$$C_{ij} \to q_{ij} c^*_{ij}$$

Hence the total used budget c over all origin-destination pairs as $\beta \to \infty$ becomes:

$$c = \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} C_{ij} \to \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} q_{ij} c^*_{ij} \tag{4.56}$$

which is the optimal solution to equivalent MILP, which assigns all flows to the least cost mode. Thus as $\beta \to \infty$, the solution to the entropy model reduces to the solution of the MILP. A recent study by Teye et al. (2017) has demonstrated the unsuitability of MILP is locating multi-user facilities of this kind (IMTs) as it was shown to produce unrealistic large responses during farecasting and policy testing.

Proposition 4.9: The budget attains its largest possible value as $\beta \to 0$

Proof 4.9: Suppose also that $\beta \to 0$, then all the exponential terms tend to toward unity and equation (4.55) reduces to:

$$\frac{1}{q_{ij}} C_{ij} \to \sum_{t \in \mathcal{K}} \tilde{c}_{itj} \frac{1}{(\sum_{t \in \mathcal{K}} 1) + 1} + c_{ij} \frac{1}{(\sum_{t \in \mathcal{K}} 1) + 1}$$

Or

$$c \to \frac{1}{p+1} \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} q_{ij} \left(\sum_{t \in \mathcal{K}} \tilde{c}_{itj} + c_{ij} \right)$$

<div align="right">(4.57)</div>

The numerator becomes the sum of the cost values of all modes and the denominator is just the total number of modal alternatives. It means that if there is no limit on modal costs, then the budget will tend to infinity. The above analysis gives us some idea of what influence β has on the budget. If β is large, the total budget must be small; if β is small the total budget must be large. In the limit as β approach infinity the budget takes on a minimum value and as β approach zero, the budget takes on the maximum value. This suggests that β is inversely related to the total budget.

4.4 Model Application

It has been demonstrated that once EMFLP is solved to optimality and the optimal location of the IMTs are known, the solution to the MCP can be expressed as a two-level nesting or tree structure model, where the upper model determines the demand of each mode and is influenced by a lower model (through the logsum), which determines the demand for each located IMT. These conversions allow the entropy model to inherit important policy-oriented properties of the logit model (McFadden 1974; Hensher et al. 1996; Hensher and Golob 1998) including the maximum expected utility or logsums, which represent access to intermodal transport modes or access to multiple modes of transport when moving cargo from between each cargo origin-destination pair; an *S*-shaped response curve, which tracks the expected relationship between the attractiveness of intermodal transport and its usage. The curve (see Figure 4.3), which is produced by plotting the probability of intermodal transport against its utility asymptotes to zero (no chance of being used if it is very unattractive) to being the dominant mode if it is very attractive.

Others are direct and cross elasticity properties, which reveal how the probability of say road alone transport changes in response to a given change in the attribute level of the intermodal transport. For example, the analyst or government may want to understand how the demand for intermodal transport (or rail) changes in response to changes in the cost of using road alone and vice versa. This information will be particularly useful for identifying policies that can reduce the usage of trucks or truck km-travelled or promote more use of intermodal transport or rail. This type of information can be derived through the *cross derivative* of

intermodal transport probability $\Pr(V_{ij})$ with respect to the cost (or any policy variable) of road alone:

$$\frac{\partial \Pr(V_{ij})}{\partial c_{ij}} = \beta \Pr(V_{ij})\left(1 - \Pr(V_{ij})\right)$$

(4.58)

Equation (4.58) implies that an increase (decrease) in the cost of road alone transport will increase (decrease) the probability and hence the usage of intermodal transport. The plot of Equation (4.58) implies that the impact of any change is largest when the share of intermodal transport is about 50% and then diminishes as the share moves towards zero or 100% as shown in Figure 4.4. Thus, the average impact of any policy will be greatest if the current share of the intermodal transport is about 50%. It is also of interest to know that the sum of the derivatives over the two modes with respect to the cost of the road alone mode equals zero:

$$\frac{\partial \Pr(V_{ij})}{\partial c_{ij}} + \frac{\partial \Pr(X_{ij})}{\partial c_{ij}} = \beta \Pr(V_{ij})\Pr(X_{ij}) - \beta \Pr(X_{ij})\Pr(V_{ij}) = 0$$

(4.59)

It means that if intermodal transport is improved so that the probability of its being used increases, the additional probability will necessarily be "drawn" from the road alone mode. Thus, to increase the probability of using intermodal transport mode necessitates decreasing the probability of using road alone mode. Finally, the elasticity of $Pr(X_{ij})$ with respect to a change in c_{ij} is:

$$\frac{\partial \Pr(V_{ij})}{\partial c_{ij}} \frac{c_{ij}}{\Pr(V_{ij})} = \beta c_{ij} \Pr(X_{ij})$$

(4.60)

The resulting nested logit model from the MCP can be carried forward for forecasting and policy testing; once the intermodal transport and road alone transport costs are updated or constructed due to changes in transport network conditions or cost parameters, the steps (2), (3) and (5) of algorithm A₁ is used during the forecasting and or policy testing. Steps (5) is only required if an IMT exceeds its handling capacity during the forecasting process.
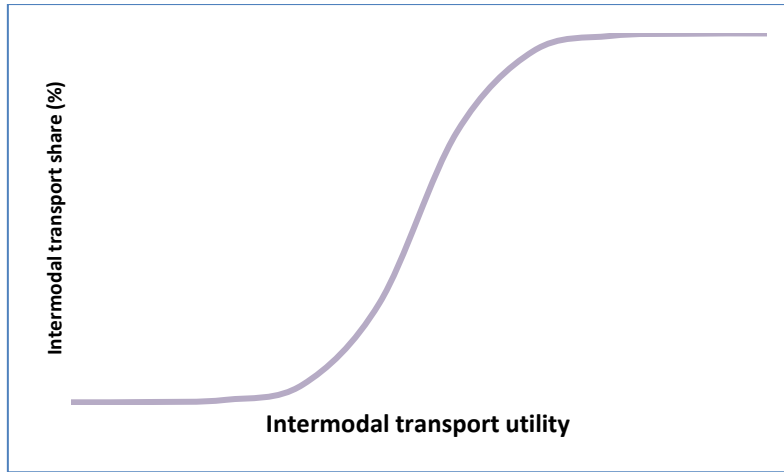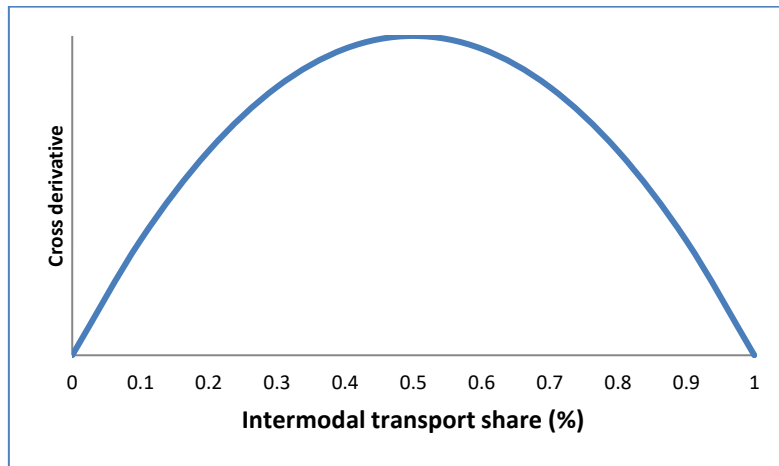
Figure 4.3: S-shaped response function

Figure 4.4: Cross derivative function

4.5 Conclusion

This paper proposes a flexible model based on the principle of entropy maximisation to answer research question I. Research question I was cast a mathematical problem such that solutions to the problem can be inferred to address the research question. The problem was formulated and solved to optimality using Lagrangian relaxation technique so as to make it suitable for applying Bregman and enumeration based algorithms. The Bregman's algorithm was used to solve the mode choice part of the problem and then embeds it in a complete enumeration algorithm to solve the overall problem. Important properties of the solved problem were presented including discussions on how the model can applied in forecasting and testing of various policies to promote intermodal transport use in the metropolitan market.

Chapter 5 Inland container terminals

5.1 Background

Chapter 4 focused on the application of the principle of entropy maximisation to locate intermodal container terminals to primarily serve the import and export containerised markets. The location of these terminals allows the use of high carrying capacity mode such as trains to transport the containers between the seaport and the terminals and the use of trucks for local pickups and deliveries. This type of intermodal transport market was described in Chapter 1 and was referred to as the metropolitan containerised transport market (MCTM). The type of intermodal transport (regional intermodal transport) used in the regional containerised transport market (RCTM) also described in detail in Chapter 1 involves the movements of cargo by combining the strengths of high carrying capacity mode such as rail or barge and trucks in the movements of containerised cargo between cargo origins and destinations in the hinterlands. In this market (see Figure 5.1) the cargo is first consolidated at an IMT close to the cargo origin using trucks and then transported by a high capacity mode (e.g. rail or barge) to another terminal close to the cargo destinations for final distribution by trucks. The cargo movements require the use of exactly two terminals and cargo are usually transported from their production areas in one metropolitan area or economic region to consumption areas in another region or economic areas within the hinterland. This type of intermodal transport is called regional intermodal transport.
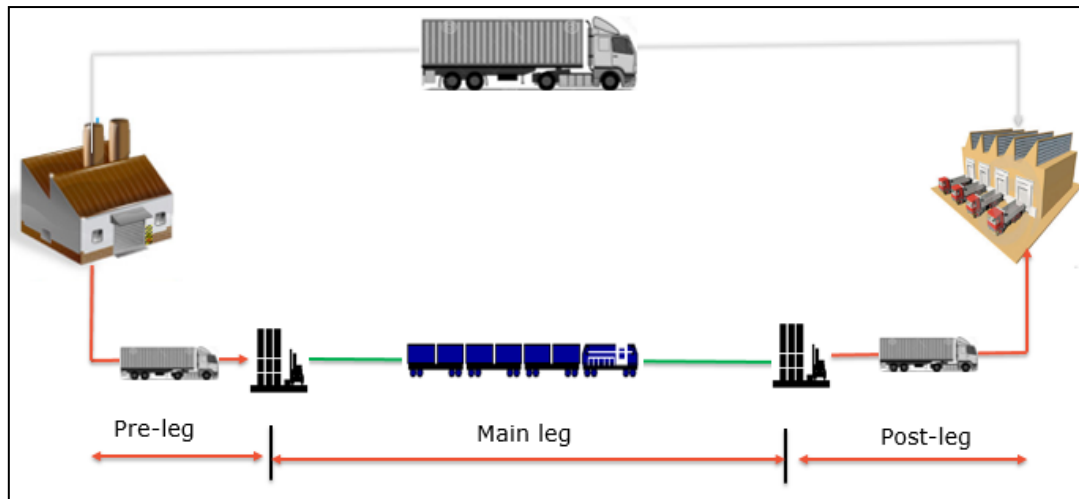
Figure 5.1: Modal options in regional transport market

The competitive advantage of developing these intermodal systems as noted in Chapter 1 is based on exploiting the economies of scale and economies of distance of rail whilst taking advantage of the flexibility and accessibility of trucks in cargo consolidations and final deliveries. In Chapter 4, it was noted that the benefits of using intermodal transport in MCTM lies in the economies of scale benefits (through the use of a high carrying capacity mode) and reduction in congestion related costs, especially around the seaport. For RCTM, the key drivers are economies of scale and economies of distance benefits with respect to cargo origins and destinations (Meyrick 2006; Arnold et al. 2004).

As noted earlier, the key difference between intermodal transport use in the two markets (RCTM and MCTM) is the number of terminals involved in the transport tasks. Additionally, intermodal transport use in both markets benefit from economies for scale, but the economies of distance also play a key role in the choice of intermodal transport in RCTM (Park et al. 1995) with some studies recommending minimum distances (usually between 400-600km) above which regional intermodal transport is considered competitive against road alone transport (NCHRP586 2007; Piyatrapoomi et al. 2006; Klink et al. 1998).

In areas where sufficient cargo volume exit for both markets, separate IMTs can be located to serve each market. This location type of terminal location decisions seems to be the current state of practice (see Meyrick 2006; NCHRP586 2007; Arnold et al. 2004). A classic example is the Australian regional and metropolitan intermodal terminal networks (Meyrick

2006) where separate set of terminals were developed for both markets. Current literature on the subject also treats the location decisions of terminals serving these markets separately. For example, studies on locating IMTs to serve the RCTM can be found in (Park et al. 1995; Arnold et al. 2001; Arnold et al. 2004; Lin et al. 2014) whilst studies on locating terminals to serve the MCTM can be found in (Teye et al. 2015; Piyatrapoomi et al. 2006; Meyrick 2006). The problem of market saturation (several IMTs serving the same markets) and lack of affordable land with the required scale and features for having separate IMTs for both markets is now driving the need to consider the design and location of IMTs that can serve both markets. A typical example in practice is the development of the Moorebank IMT in Sydney to serve both markets (DoFD 2011).

This paper extends the work in the previous Chapter in two main directions; first, the chapter generalised the metropolitan intermodal location problem (MIMTLP) in Chapter 4 to also include regional IMT location problems (RIMTLP), where the former will be a special case of the later. The generalised location problem is referred to as (inland) intermodal terminal location problem (IMTLP). The proposed model for solving IMTLP is suitable for locating IMTs to serve the metropolitan market, regional market or both markets. Second, a new algorithm for solving large instances of the generalised problem efficiently is proposed. The computational efficiency, solution quality and properties of the algorithm are also presented. The computational time of the algorithm can be shown to grow linearly with respect to the number of IMTs to locate. The proposed algorithm is particularly desirable for solving the regional intermodal terminal location problem, which is usually characterised by a large study area such as a whole country or large economic regions.

The rest of the chapter is organised as follows; Section 5.2 presents the assumptions underlying the proposed entropy model for the IMTLP. The IMTLP is also mathematically formulated based on the principle of entropy maximisation in Section 5.3. Section 5.4 discusses solution algorithms for the solving the formulated problem. Numerical example illustrating the key features of the proposed entropy model and algorithms are presented in Section 5.5. Finally, Section 5.6 presents the conclusions drawn from this chapter.

5.2 Assumptions and decision variables

All assumption made during the development of the terminals in the metropolitan market also apply to the generalised case. In addition, it was assumed that the unit cost of regional intermodal transport c_{istj} is made up of four main cost components as follows:

$$c_{istj} = c_{is} + c_s + c_{st} + c_t + c_{tj} \; ; \forall i \in \mathcal{O}, j \in \mathcal{D}, s \neq t, s, t \in \mathcal{T} \tag{5.1}$$

where c_{is} is the road cost (\$/TEU) from cargo origin $i \in \mathcal{O}$ to IMT $s \in \mathcal{T}$; c_{st} is the rail cost (\$/TEU) from IMT $s \in \mathcal{T}$ to IMT $t \neq s \in \mathcal{T}$ and takes into account the economies of scale benefit of using rail; c_{tj} is the road cost (\$/TEU) from IMT $t \in \mathcal{T}$ to cargo destination $j \in \mathcal{D}$ and c_s and c_t are terminal usage cost (\$/TEU) for IMT s and t respectively. These terminal usage costs are assumed to include the fixed cost of IMT location and operation costs (\$ per TEU) passed on to the shipper or user, who then decides whether or not to use the terminal.

Decision Variables

The key outputs of the model are the flow variables $W_{istj}, V_{itj}, X_{ij}$ for determining the demands for regional, metropolitan and road alone transport respectively and the location variables Y_t, which determines the locations for the developments of IMTs.

5.3 Entropy maximising inland IMT location problem

The goal is to find the most likely state of the system based on all we know about it. To do this requires specifying all possible states of the system and selecting the most likely state in accordance with the principle of entropy maximisation. Mathematically, the number of possible ways that a state $(X_{ij}, V_{itj}, W_{istj})$ can occur can be determined using Equation (5.2). It is expected that state with the most ways of occurring are more likely to represent the system.

$$E = \frac{Z!}{\prod_{i \in \mathcal{O}} \prod_{j \in \mathcal{D}} (X_{ij}!) \left(\prod_{\mathcal{T}} V_{itj}! \right) \left(\prod_{s \in \mathcal{T}} \prod_{t \neq s \in \mathcal{T}} W_{istj}! \right)} \tag{5.2}$$

where Z is the total cargo (in TEUs) in the system such that:

$$\sum_{i \in \mathcal{O}} \sum_{s \in \mathcal{T}} \sum_{t \neq s \in \mathcal{T}} \sum_{j \in \mathcal{D}} W_{istj} + \sum_{i \in \mathcal{O}} \sum_{t \in \mathcal{T}} \sum_{j \in \mathcal{D}} V_{itj} + \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} X_{ij} = Z$$

and by construction $W_{issj} = W_{ittj} = 0; \forall s, t \in \mathcal{T}, i \in \mathcal{O}, j \in \mathcal{D}$. Again, $\ln E$ is maximised instead of E so taking the natural logarithm of (5.2) and applying the Stirling's approximation we have:

$$\ln E = \ln Z! + \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} X_{ij} \left\{ 1 - \ln X_{ij} \right\} + \sum_{i \in \mathcal{O}} \sum_{t \in \mathcal{T}} \sum_{j \in \mathcal{D}} V_{itj} \left\{ 1 - \ln V_{itj} \right\} \qquad (5.3)$$
$$+ \sum_{i \in \mathcal{O}} \sum_{s \in \mathcal{T}} \sum_{t \neq s \in \mathcal{T}} \sum_{j \in \mathcal{D}} W_{istj} \left\{ 1 - \ln W_{istj} \right\}$$

It has been shown in Chapter 4 that the maximum value of $\ln E$ corresponds to the amount of missing information (entropy) in the resulting probability distributions and the maximum amount of missing information occurs when we have no information about the system. Under such conditions all states are equally likely to represent the system and the resulting probability distributions are uniform probability distributions. Once information about the system becomes available the system moves way from uniform probability distribution to a distribution consistent with the new information.

5.3.1 Available information as constraints

In developing this model, the following information converted into constraints are assumed to be available or known about the containerised system.

1. *Conservation of cargo flow constraint.* This information is added as constraint (5.4). It ensures that for each origin-destination pair, the sum of cargo transported by all available modes of transport equals the total cargo associated with this origin-destination pair. The first term on the left hand side of (5.4) captures the share of regional transport, the second, the share of metropolitan transport and the third captures the share of road alone transport in the transport task.

$$\sum_{s \in \mathcal{T}} \sum_{t \neq s \in \mathcal{T}} W_{istj} + \sum_{s \in \mathcal{T}} V_{itj} + X_{ij} = q_{ij}; \quad \forall \ i \in \mathcal{O}, j \in \mathcal{D} \qquad (5.4)$$

2. *Transport budget constraint.* This evidence is added as constraint (5.5). The first component captures the weighted cost of using regional transport, the second captures

the weighted cost of using metropolitan intermodal transport and the third represents the weighted cost of road alone transport (e.g. truck only), with the sum not exceeding the total allocated transport budget (c).

$$\sum_{i\in\mathcal{O}}\sum_{s\in\mathcal{T}}\sum_{t\neq s\in\mathcal{T}}\sum_{j\in\mathcal{D}} c_{istj}W_{istj} + \sum_{i\in\mathcal{O}}\sum_{t\in\mathcal{T}}\sum_{j\in\mathcal{D}} c_{itj}V_{itj} + \sum_{i\in\mathcal{O}}\sum_{j\in\mathcal{D}} c_{ij}X_{ij} \leq c \tag{5.5}$$

3. *Budget on capacity constraint.* The maximum quantity of cargo that each IMT can handle is expressed in constraint (5.6). The first and second terms represent the total capacity of each IMT used up by regional intermodal transport, and the third term captures that of metropolitan intermodal transport, the sum of which must not exceed the handling capacity of each IMT.

$$\sum_{i\in\mathcal{O}}\sum_{j\in\mathcal{D}}\sum_{s\neq t\in\mathcal{T}} W_{istj} + \sum_{i\in\mathcal{O}}\sum_{j\in\mathcal{D}}\sum_{s\neq t\in\mathcal{T}} W_{itsj} + \sum_{i\in\mathcal{O}}\sum_{j\in\mathcal{D}} V_{itj} \leq Y_t b_t; \quad \forall t \in \mathcal{T} \tag{5.6}$$

4. *Definitional constraint.* The information on the required number of IMTs to locate is presented by constraint (5.7)

$$\sum_{t\in\mathcal{T}} Y_t = p \tag{5.7}$$

By ignoring the constant term $\ln Z!$, in Equation (5.3) the paper presents the entropy facility location problem (EMFLP) consistent with the above available information:

$$\text{Max } \Lambda = \sum_{i\in\mathcal{O}}\sum_{j\in\mathcal{D}} X_{ij}\left\{1 - \ln X_{ij}\right\} + \sum_{i\in\mathcal{O}}\sum_{t\in\mathcal{T}}\sum_{j\in\mathcal{D}} V_{itj}\left\{1 - \ln V_{itj}\right\}$$

$$+ \sum_{i\in\mathcal{O}}\sum_{s\in\mathcal{T}}\sum_{t\neq s\in\mathcal{T}}\sum_{j\in\mathcal{D}} W_{istj}\left\{1 - \ln W_{istj}\right\}$$

Subject to constraint (5.4) to (5.7) and the following integer and non-negativity constraints:

$$Y_t \in \{0,1\}; \quad t \in \mathcal{T} \tag{5.8}$$

128

$$W_{istj} \geq 0; \ V_{itj} \geq 0; \ X_{ij} \geq 0 \ ; \ W_{issj} \ = \ W_{isttj} \ = \ 0; \forall s, t \in \mathcal{T}; \ \forall \ i \in \mathcal{O}; j \in \mathcal{D} \qquad (5.9)$$

5.4 Solution to the EMFLP

Applying the general solution approach requires the relaxation of constraint (5.6), which links the location problem with the mode choice problem by penalising it in the objective function with Lagrangian multipliers $\psi_t; t \in \mathcal{T}$:

$\text{Max}\{\Lambda_R\}$

Subject to constraints (5.4), (5.5), (5.7) - (5.9)

where

$$\Lambda_R = \Lambda \ + \ \sum_{t \in \mathcal{T}} \psi_t \left(Y_t b_t \ - \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} \sum_{s \neq t \in \mathcal{T}} W_{istj} - \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} \sum_{s \neq t \in \mathcal{T}} W_{itsj} - \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} V_{itj} \right)$$

The above problem then decomposes into two sub-problems, the facility location sub-problem (FLP) consisting of only the location variables as decision variables and the mode choice sub-problem (MCP) consisting of only the flow variables as decision variables. The FLP is expressed as follows:

$$\text{FLP} : \ \text{Max} \ \Lambda_{\text{FLP}} = \sum_{t \in \mathcal{T}} (\psi_t b_t) Y_t$$

Subject to constraints (5.7) and (5.8)

and mode choice mode problem (MCP) simplifies to become:

$$\text{MCP} : \ \text{Max} \ \Lambda_{\text{MCP}}$$

$$= \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} X_{ij} \{1 - \ln X_{ij}\} + \sum_{i \in \mathcal{O}} \sum_{t \in \mathcal{T}} \sum_{j \in \mathcal{D}} V_{itj} \{1 - \ln V_{itj} - \psi_t\}$$

$$+ \sum_{i \in \mathcal{O}} \sum_{s \in \mathcal{T}} \sum_{t \neq s \in \mathcal{T}} \sum_{j \in \mathcal{D}} W_{istj} \{1 - \ln W_{istj} - \psi_s - \psi_t\}$$

Subject to constraints (5.4), (5.5) and (5.9)

Thus, the overall objective function Λ_R can be expressed in terms of Λ_{FLP} and Λ_{MCP}:

$$\Lambda_R = \Lambda_{\text{FLP}} + \Lambda_{\text{MCP}} \tag{5.10}$$

5.4.1 Solution to the FLP

Given $\psi_t \geq 0; \forall t \in \mathcal{T}$. The FLP can be solved by identifying the p largest elements of $(\psi_t b_t); \forall t \in \mathcal{T}$ and setting the corresponding values of Y_t equal to 1. Let the set $\mathcal{K}$ with cardinality p, be the set of located IMTs, which then goes into the MCP as input. And for sufficiently larger IMTs such that $\psi_t = 0; \forall t \in \mathcal{T}$, the objective value ($\Lambda_{\text{FLP}}$) of the FLP will be zero and the selection of the best IMTs will only be based on the value of Λ_{MCP}.

5.4.2 Solution to the MCP

For a given set of located IMTs $\mathcal{K}$, the MCP can be solved by constructing a Lagrangian equation comprising the objective function and the constraints and enforcing the Karush-Kuhn-Tucker (KKT) optimality conditions:

$$
\begin{aligned}
\Lambda_{\text{LMCP}} = &\sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} X_{ij} \left\{ 1 - \ln X_{ij} \right\} + \sum_{i \in \mathcal{O}} \sum_{t \in \mathcal{K}} \sum_{j \in \mathcal{D}} V_{itj} \left\{ 1 - \ln V_{itj} - \psi_t \right\} \\
&+ \sum_{i \in \mathcal{O}} \sum_{s \in \mathcal{K}} \sum_{t \neq s \in \mathcal{K}} \sum_{j \in \mathcal{D}} W_{istj} \left\{ 1 - \ln W_{istj} - \psi_s - \psi_t \right\} \\
&+ \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} \eta_{ij} \left\{ q_{ij} - \sum_{s \in \mathcal{K}} \sum_{t \neq s \in \mathcal{K}} W_{istj} - \sum_{s \in \mathcal{K}} V_{itj} - U_{ij} \right\} \\
&+ \beta \left\{ c - \sum_{i \in \mathcal{O}} \sum_{s \in \mathcal{K}} \sum_{t \neq s \in \mathcal{K}} \sum_{j \in \mathcal{D}} c_{istj} W_{istj} - \sum_{i \in \mathcal{O}} \sum_{t \in \mathcal{K}} \sum_{j \in \mathcal{D}} c_{itj} V_{itj} - \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} c_{ij} X_{ij} \right\}
\end{aligned}
$$

where $\eta_{ij} \geq 0$ are Lagrangian multipliers for the origin-destination cargo flow constraint (5.4). The KKT conditions for a maximum Λ_{LMCP} with respect to the flow variables are:

$$-\ln W_{istj} - \beta c_{istj} - \eta_{ij} - \psi_s - \psi_t = 0; \forall s \in \mathcal{K}, t \neq s \in \mathcal{K} \tag{5.11}$$

$$-\ln V_{itj} - \beta c_{itj} - \eta_{ij} - \psi_t = 0 \; ; \; \forall \, t \in \mathcal{K} \tag{5.12}$$

$$-\ln X_{ij} - \beta c_{ij} - \eta_{ij} = 0 \tag{5.13}$$

Solving for $W_{istj}, V_{itj}, X_{ij}$ in the above equations we have:

$$W_{istj} = e^{-\eta_{ij}} e^{-(\beta c_{istj} + \psi_s + \psi_t)}; \forall s \in \mathcal{K}, t \neq s \in \mathcal{K}; \; i \in \mathcal{O}, j \in \mathcal{D} \tag{5.14}$$

$$V_{itj} = e^{-\eta_{ij}} e^{-(\beta c_{itj} + \psi_t)}; \; \forall \, t \in \mathcal{K}; \; i \in \mathcal{O}, j \in \mathcal{D} \tag{5.15}$$

$$X_{ij} = e^{-\eta_{ij}} e^{-\beta c_{ij}}; \; \forall \, i \in \mathcal{O}, j \in \mathcal{D} \tag{5.16}$$

Equations (5.14) to (5.16) implicitly assume that both metropolitan and regional intermodal transport are available for the movements of containers between origin-destination pairs. In reality, this may not be true in all cases. For example, the metropolitan intermodal transport mode is not available for origin-destination movements where rail is not accessible to either the cargo origin or cargo destination. Conversely, there are cases where the port is the cargo origin or cargo destination and so obviates the need of using two IMTs therefore, making the use of regional intermodal transport mode not feasible. Additionally, the distance between some origin-destination pairs may be too short for regional intermodal transport use to be feasible. To account for some of these limitations, modal choice set variables δ_{ijm} are introduced into the Equations (5.14) to (5.16). The choice set variable variables δ_{ijm} is defined such that $\delta_{ijm} = 1$ means that mode m ($m = 1,2,3$) is available for that origin-destination pair and 0 otherwise. By definition mode 1 ($m = 1$) is road alone transport, mode 2 is metropolitan intermodal transport and mode 3 is the regional intermodal transport. Equations (5.14) to (5.16) are updated with the choices set definitions as follows:

$$W_{istj} = \delta_{ij3} e^{-\eta_{ij}} e^{-(\beta c_{istj} + \psi_s + \psi_t)}; \forall s \in \mathcal{K}, t \neq s \in \mathcal{K}; \; i \in \mathcal{O}, j \in \mathcal{D} \tag{5.17}$$

$$V_{itj} = \delta_{ij2} e^{-(\beta c_{itj} + \psi_t)}; \; \forall \, t \in \mathcal{K}; \; i \in \mathcal{O}, j \in \mathcal{D} \tag{5.18}$$

$$X_{ij} = \delta_{ij1} e^{-\eta_{ij}} e^{-\beta c_{ij}}; \; \forall \, i \in \mathcal{O}, j \in \mathcal{D} \tag{5.19}$$

Depending on the application at hand, more useful choice sets can be developed and applied in a similar fashion. The next stage is to estimate the parameters in the equations. The Lagrangian parameters η_{ij} can be estimated by enforcing constraints (5.4) using Equations (5.17) to (5.19):

$$e^{\eta_{ij}} = \frac{1}{q_{ij}}\left\{ \delta_{ij1}e^{-\beta c_{ij}} + \sum_{t\in\mathcal{K}} \delta_{ij2}e^{-(\beta c_{itj}+\psi_t)} + \sum_{s\in\mathcal{K}}\sum_{t\neq s\in\mathcal{K}} \delta_{ij3}e^{-(\beta\omega_{istj}+\psi_s+\psi_t)} \right\} \tag{5.20}$$

Also by enforcing constraint (5.6) the parameter ψ_t can be also be expressed as:

$$\psi_t = \max\left\{0, \ln\left(\frac{\Omega_t}{b_t}\right)\right\} \quad \forall t \in \mathcal{K} \tag{5.21}$$

where

$$\Omega_t = \sum_{i\in\mathcal{O}}\sum_{j\in\mathcal{D}} e^{\eta_{ij}}\left(\delta_{ij2}e^{-\beta c_{itj}} + \delta_{ij3}\sum_{s\neq t\in\mathcal{K}} e^{-\psi_s}\left(e^{-\beta c_{istj}} + e^{-\beta c_{itsj}}\right)\right); t \in \mathcal{K}$$

Proposition 5.1: The Lagrangian parameters ψ_t can be computed iteratively with the kth iterated values are evaluated using equation:

$$\psi_t^{\ k} = \psi_t^{\ k-1} + \max\left\{0, \ln\left(\frac{F_t^{\ k-1}}{b_t}\right)\right\} \; ; \; \psi_t^{\ 0} = 0; \; k = 1,2,\dots; \forall\, t \in \mathcal{K} \tag{5.22a}$$

where F_t is the total usage of IMT $t \in \mathcal{K}$

$$F_t^{\ k-1} = \sum_{i\in\mathcal{O}}\sum_{j\in\mathcal{D}} V_{itj}^{\ k-1} + \sum_{i\in\mathcal{O}}\sum_{s\neq t\in\mathcal{K}}\sum_{j\in\mathcal{D}} W_{istj}^{\ k-1} + \sum_{i\in\mathcal{O}}\sum_{s\neq t\in\mathcal{K}}\sum_{j\in\mathcal{D}} W_{itsj}^{\ k-1} \; ; \forall\, t \in \mathcal{K} \tag{5.22b}$$

Proof 5.1: The proof follows directly from proposition (4.6) in chapter 4.

Equations (5.17) to (5.19) can be converted into a three-level nesting or tree structure as shown in Figure 5.2 model, where the distribution of demand for each of mode of transport (road verses intermodal transport modes) conditions the distribution of demand for intermodal

transport modes (metropolitan verses regional intermodal transport), which in turn conditions the distribution of demand for the located IMTs. Conversely, the models are connected in the opposite direction by accessibility measures such that improving for example accessibility to the any of the located IMTs influences the demand for intermodal transport.
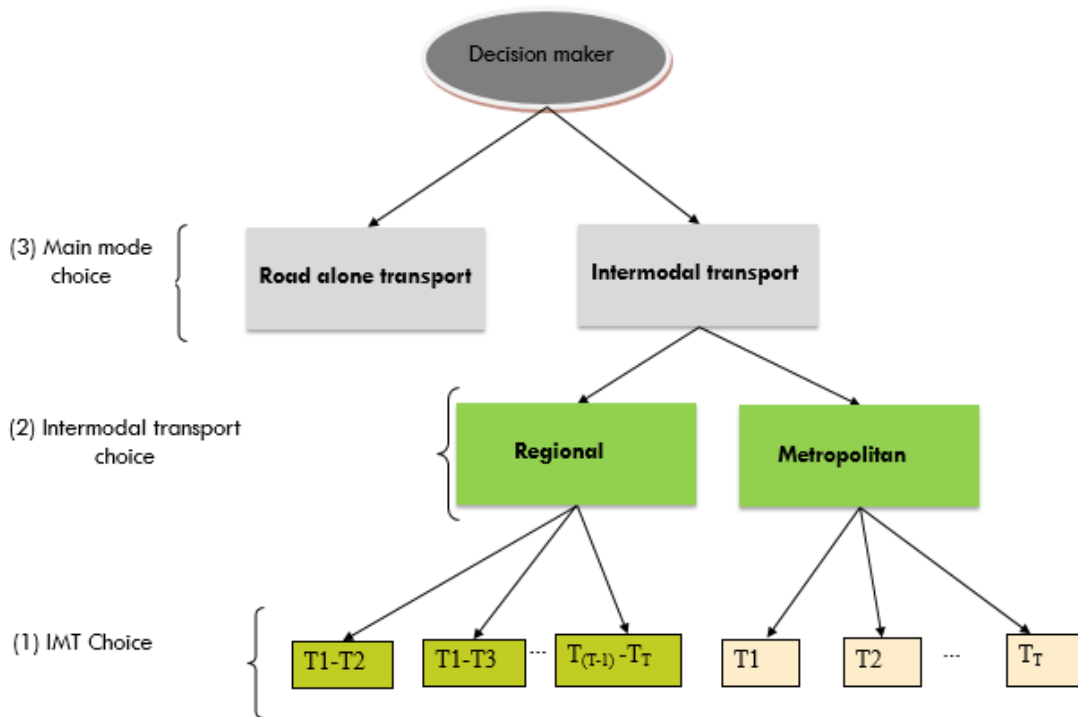

Figure 5.2: Three-level nesting structure for MCP

The expected demand for the three modes can be expressed in probabilistic forms by inserting equation (5.20) into (5.17), (5.18) and (5.19) as follows:

$$W_{istj} = q_{ij}\text{Pr}(W_{istj}); \quad \forall s,t \in \mathcal{K}, s \neq t; \ i \in \mathcal{O}, j \in \mathcal{D} \tag{5.23}$$

$$V_{itj} = q_{ij}\text{Pr}(V_{itj}); \quad \forall \, t \in \mathcal{K}; \ i \in \mathcal{O}, j \in \mathcal{D} \tag{5.24}$$

$$X_{ij} = q_{ij}\text{Pr}(X_{ij}); \quad \forall \, i \in \mathcal{O}, j \in \mathcal{D} \tag{5.25}$$

where $\text{Pr}(W_{istj})$ represents the probability of realising demand W_{istj} for regional intermodal transport and can be evaluated using:

$$\Pr(W_{istj}) = \Pr(W_{istj}|I_{ij}) * \Pr(I_{ij}) \tag{5.26}$$

$\Pr(W_{istj}|I_{ij})$ is the conditional probability of realising W_{istj} given that the demand for intermodal transport I_{ij} is known and is expressed as:

$$\Pr(W_{istj}|I_{ij}) = \frac{\delta_{ij3}e^{-(\beta c_{istj}+\psi_s+\psi_t)}}{\sum_{s\in\mathcal{K}}\sum_{t\neq s\in\mathcal{K}}\delta_{ij3}e^{-(\beta c_{istj}+\psi_s+\psi_t)} + \sum_{t\in\mathcal{K}}\delta_{ij2}e^{-(\beta c_{itj}+\psi_t)}} \tag{5.27}$$

The maximum expected utility (logsum) ℓ_{ij} over all located IMTs is expressed in equation (5.28). ℓ_{ij} serve as a measure of access to intermodal terminals in the movement of cargo between each origin-destination pair. Thus, for any given origin-destination pair, the bigger the value of ℓ_{ij} the better the access to intermodal transport.

$$\ell_{ij} = \ln\left\{\sum_{t\in\mathcal{K}}\delta_{ij2}e^{-(\beta c_{itj}+\psi_t)} + \sum_{s\in\mathcal{K}}\sum_{t\neq s\in\mathcal{K}}\delta_{ij3}e^{-(\beta c_{istj}+\psi_s+\psi_t)}\right\} \tag{5.28}$$

The term $\Pr(I_{ij})$ is the probability of realising the demand for intermodal transport I_{ij} and is expressed as:

$$\Pr(I_{ij}) = \frac{e^{\ell_{ij}}}{e^{L_{ij}}} \tag{5.29}$$

where L_{ij} is the maximum expected utility (logsum) over all modes of transport for each origin-destination pair and serve as a measure of access to multiple transport modes of transport in the movement of cargo between a given origin-destination pair (Williams 1977):

$$L_{ij} = \ln\left\{e^{\ell_{ij}} + \delta_{ij1}e^{-\beta c_{ij}}\right\} \tag{5.30}$$

Knowing $\Pr(I_{ij})$ from (5.29), the realised demand for intermodal transport I_{ij} can be computed for each origin-destination:

$$I_{ij} = q_{ij}\Pr(I_{ij}) \tag{5.31}$$

Similarly, the probability distribution for determining metropolitan intermodal transport demand is expressed as:

$$\Pr(V_{itj}) = \Pr(V_{itj}|I_{ij})\Pr(I_{ij})$$

(5.32)

$\Pr(V_{itj}|I_{ij})$ is the conditional probability of realising the demand for metropolitan intermodal transport V_{itj} given that the demand for intermodal transport I_{ij} is known and are evaluated using:

$$\Pr(V_{itj}|I_{ij}) = \frac{\delta_{ij2}e^{-(\beta c_{itj}+\psi_t)}}{\sum_{s\in\mathcal{K}}\sum_{t\neq s\in\mathcal{K}}\delta_{ij3}e^{-(\beta c_{istj}+\psi_s+\psi_t)} + \sum_{t\in\mathcal{K}}\delta_{ij2}e^{-(\beta c_{itj}+\psi_t)}}$$

(5.33)

Finally, the probability of realising X_{ij} for road alone transport can be expressed as:

$$\Pr(X_{ij}) = \delta_{ij1}\frac{e^{-\beta c_{ij}}}{e^{\ell_{ij}}+\delta_{ij1}e^{-\beta c_{ij}}}$$

(5.34)

A corollary of the above analysis is the direct link between the Lagrangian multipliers η_{ij} in Equation (5.20) and the measure of access to multiple modes of transport parameter L_{ij} in (5.30):

$$e^{-\eta_{ij}} = q_{ij}e^{-L_{ij}}$$

(5.35)

Finally, for a given transport budget c, the cost sensitivity parameter β can be estimated using Hyman's method (Hyman 1966) or Newton-Raphson's method

$$f(\beta) = \sum_{i\in\mathcal{O}}\sum_{s\in\mathcal{K}}\sum_{t\neq s\in\mathcal{K}}\sum_{j\in\mathcal{D}}c_{istj}W_{istj} + \sum_{i\in\mathcal{O}}\sum_{t\in\mathcal{K}}\sum_{j\in\mathcal{D}}c_{itj}V_{itj} + \sum_{i\in\mathcal{O}}\sum_{j\in\mathcal{D}}c_{ij}X_{ij} - c \leq 0$$

(5.36)

Or from equations (5.18) to (5.20):

$$f(\beta) = \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} e^{-\eta_{ij}} \left\{ \delta_{ij1} c_{ij} e^{-\beta c_{ij}} + \sum_{t \in \mathcal{K}} \delta_{ij2} c_{itj} e^{-(\beta c_{itj} + \psi_t)} \right.$$

$$\left. + \sum_{s \in \mathcal{K}} \sum_{t \neq s \in \mathcal{K}} \delta_{ij3} c_{istj} e^{-(\beta c_{istj} + \psi_s + \psi_t)} \right\} - c \leq 0$$

The function $f(\beta)$ is continuous and differentiable with respect to β, and so it can be maximised to obtain the value of β.

Again, the estimation of the Lagrangian multipliers in the MCP are inter-dependent, where the evaluated value of one is required to solve the other. Algorithm (A_1) proposed in Chapter 4 can readily be adapted and used to estimate the MCP parameters. The next section presents the algorithm for solving the overall problem EMFLP by connecting the two sub-problems.

5.4.3 Solving the EMFLP by complete enumeration

The complete enumeration algorithm A_2 proposed in Chapter 4 can also be adapted and used to solve the overall problem. The use of this algorithm to solve the generalized IMT location problem could still be practical for many real world problems. As shown in Equation (4.49) the running time of algorithm A_2 is bounded by $O(\tau^p T_B)$ where T_B is the running time of the modified Bregman's algorithm (A_1) and τ^p gives the maximum running time of selecting p IMTs from a candidate set of $\mathcal{T}$ with cardinality τ.

However, the study area for locating regional IMTs is often very large and may encompass a whole country like Australia, US or China and regions like the European Union. For example, assume there 100 candidate IMT locations in the study area and the planner wants to select the best 10 for the development of IMTs. If it takes about 10^{-3}seconds to solve the MCP, algorithm A_2 will about 560 years to find the best 10 IMT locations. Such large scale applications will benefit from more efficient algorithm than A_2. A new fast and efficient heuristic is proposed in the next section for solving such larger problem instances.

5.4.4 Combined solution of FLP and MCP by heuristics

For large problem instances, a fast algorithm is proposed and its solution quality demonstrated with respect to algorithm A_2 through extensive numerical examples. The heuristic algorithm is

motivated by the convexity of the objective entropy function and the principle of conditional entropy in Propositions 5.2 and 5.3. The key assumption underlying the heuristic is that if $\mathcal{K}^*$ is the set with the optimal IMT locations with IMT $\vartheta \in \mathcal{K}^*$ then IMT $\vartheta 1 \in \mathcal{T}$ must also be in $\mathcal{K}^*$ if:

$$H(Y_\vartheta, Y_{\vartheta 1}) \geq H(Y_\vartheta, Y_t); \forall t \in \mathcal{T}, \vartheta \neq \vartheta 1 \tag{5.37}$$

where $H(Y_\vartheta, Y_{\vartheta 1})$ is the entropy of locating IMTs at locations ϑ and $\vartheta 1$. Thus, following Proposition 5.2 the selection of IMT location $\vartheta 1$ is conditioned on the selection of IMT location $\vartheta \in \mathcal{T}$. The remaining $p - 2$ IMTs locations are selected in a similar way. For example, the selection of the third IMT location $\vartheta 2$ is conditioned on knowing that IMT locations ϑ and $\vartheta 1$ were selected (or in the set $\mathcal{K}^*$):

$$H(Y_\vartheta, Y_{\vartheta 1}, Y_{\vartheta 2}) \geq H(Y_\vartheta, Y_{\vartheta 1}, Y_t); \forall t \in \mathcal{T}, \vartheta 2 \neq \vartheta, \vartheta 2 \neq \vartheta 1 \tag{5.38}$$

The key question that remains is how to select location $\vartheta \in \mathcal{T}$ as it conditions the selection of the remaining $p - 1$ locations. We suggest considering all candidate IMT locations and selecting the one with highest entropy. That is select location $\vartheta^* \in \mathcal{T}$ if

$$H(Y_{\vartheta^*}, \mathcal{K}_1) \geq H(Y_\vartheta, \mathcal{K}_1); \forall \vartheta \in \mathcal{T} \tag{5.39}$$

where $\mathcal{K}_1$ is the set containing the remaining $p - 1$ IMT locations.

Proposition 5.2. By definition each location variable $Y_t; t \in \mathcal{T}$ takes on two values; 0 and 1 and let $\mathcal{I} = \{0,1\}$. If the Shannon entropy (see proposition 5.4) of a location variable Y_1 is defined as $H(Y_1) = \sum_{a \in \mathcal{I}} Pr(Y_1 = a) \ln Pr(Y_1 = a)$, then the joint entropy for locations variables Y_1 and Y_2 can be expressed as:

$$H(Y_1, Y_2) = H(Y_1) + H(Y_2|Y_1) \tag{5.40}$$

Proof 5.2. By definition the joint entropy of Y_1 and Y_2 becomes:

$$H(Y_1, Y_2) = -\sum_{a \in \mathcal{I}} \sum_{b \in \mathcal{I}} Pr(Y_1 = a, Y_2 = b) \ln Pr(Y_1 = a, Y_2 = b)$$

Or for simplicity:

$$H(Y_1, Y_2) = -\sum_{a \in \mathcal{J}} \sum_{b \in \mathcal{J}} Pr(a,b) \ln Pr(a,b) = -\sum_{a \in \mathcal{J}} \sum_{b \in \mathcal{J}} Pr(a,b) \ln(Pr(a)Pr(b|a))$$

Expanding and using the properties of marginal probability distributions we have:

$$H(Y_1, Y_2) = -\sum_{a \in \mathcal{J}} Pr(a) \ln Pr(a) - \sum_{a \in \mathcal{J}} \sum_{b \in \mathcal{J}} Pr(a,b) \ln Pr(b|a)$$

Hence

$$H(Y_1, Y_2) = H(Y_1) + H(Y_2|Y_1)$$

Proposition 5.3. Proposition 5.2 can be generalised for τ number of IMTs, $Y_1, Y_2, \dots Y_\tau$:

$$H(Y_1, Y_2, \dots, Y_\tau) = H(Y_1) + \sum_{t=2}^{\tau} H(Y_t|Y_1, \dots, Y_{t-1}) \qquad (5.41)$$

Proof 5.3 by induction. From proposition 5.2, equation (5.40) is true for $\tau = 2$. Assume proposition 5.3 is also true for any τ then using chain rule:

$$H(Y_1, Y_2, \dots, Y_\tau, Y_{\tau+1}) = H(Y_1, Y_2, \dots, Y_\tau) + H(Y_{\tau+1}|Y_1, Y_2, \dots, Y_\tau)$$

Or

$$H(Y_1, Y_2, \dots, Y_\tau, Y_{\tau+1}) = H(Y_1) + \sum_{t=2}^{\tau} H(Y_t|Y_1, \dots, Y_{t-1}) + H(Y_{\tau+1}|Y_1, Y_2, \dots, Y_\tau)$$

Hence

$$H(Y_1, Y_2, \dots, Y_\tau, Y_{\tau+1}) = H(Y_1) + \sum_{t=2}^{\tau+1} H(Y_t|Y_1, \dots, Y_{t-1})$$

Based on the insight from the above propositions the proposed heuristic is presented as algorithm A$_3$ and labelled as entropic greedy algorithm (EGA).

Algorithm A₃: Entropic greedy algorithm (EGA)

1. Initialization: $\mathcal{K} = \mathcal{K}^* = \{\}, \Lambda^* = -\infty; \mathcal{T}_1 = \mathcal{T}$

2. While the candidate set of IMTs $\mathcal{T}$ is not empty

3.   Choose an IMT $a \in \mathcal{T}$ and delete it from $\mathcal{T}$, i.e., $\mathcal{T} = \mathcal{T} - \{a\}$

4. $\mathcal{T}_3 = \{a\}, \mathcal{K} = \{a\}$

5.   Solve the MCP with $\mathcal{K}$ as the set of located IMTs and return the objective value Λ_R

6.   If $p = |\mathcal{K}|$ then

7.   Go to step 25

8.   Else

9.   For $u \coloneqq 1 \; to \; p - 1$

10. $\Lambda_2 = -\infty;$

11. $\mathcal{T}_2 = \mathcal{T}_1 - \mathcal{T}_3$

12.   While $\mathcal{T}_2$ is not empty:

13.   Choose an IMT $s \in \mathcal{T}_2$

14. $\mathcal{K} \coloneqq \mathcal{T}_3 + \{s\}; \mathcal{T}_2 = \mathcal{T}_2 - \{s\}$

15.   Solve the MCP with $\mathcal{K}$ and return the objective value Λ_R

16.   If $\Lambda_R > \Lambda_2$

17. $\Lambda_2 = \Lambda_R; \vartheta = s;$

18.   Endif

19.   Endwhile

20.   Update the set $\mathcal{T}_3 \coloneqq \mathcal{T}_3 + \{\vartheta\}$

21.   Endfor

22.   Endif

23. $\mathcal{K} = \mathcal{T}_3$

24.   Solve the MCP with $\mathcal{K}$ as the set of located IMTs and return the objective value Λ_R

25.   If $\Lambda_R > \Lambda^*:$ then

26. $\Lambda^* = \Lambda_R; \mathcal{K}^* = \mathcal{K}$

27.   Endif

28.   Endwhile

29.   Return the set $\mathcal{K}^*$ and optimal objective value Λ^*

If T_B be the running time of algorithm A₁ (for solving MCP) for a given set of located IMTs $\mathcal{K}$. It is clear that the running time of the heuristic algorithm A₃ will be dominated by the number executions of algorithm A₁ in line 15 of the inner while loop. Lines 13 to 17 will be executed at most $\tau^2 p$ times each and given the fact that the execution time of Algorithm A₁ will dominate, it suffices that the running time of lines 13 to 17 will be bounded by $O(p\tau^2 T_B)$. Similarly, lines 3 to 7 will be executed at most τ times each with the total running time bounded

by $O(\tau T_B)$. Finally, lines 23 to 26 will also be executed at most τ times each with a total running time also bounded by $O(\tau T_B)$. Therefore, the overall running time of the proposed heuristics algorithm, A₃ is bounded by $O(p\tau^2 T_B)$. Alternatively, the number of executions of Algorithm A₁ for solving MCP is bounded by $O(p\tau^2)$ compared with $O(\tau^p)$ for the enumeration Algorithm A₂. Thus, the running time savings of Algorithm A₃ (heuristic) with respect to A₂ (enumeration algorithm) occurs when $p \geq 3$.

We have shown that the proposed heuristics Algorithm A₃ has a polynomial running time, which increases linearly with increasing p (number of IMTs to locate) and therefore computationally efficient for solving larger problem instances. Algorithm A₃ is optimal for $p = 1$, and $p = 2$ since it reduces to Algorithm A₂, which is a global optimal algorithm. The conjecture is whether it is also optimal for $p \geq 3$. The quality of solutions produced by the heuristics is demonstrated with extensive numerical examples in Section 5.5.

Proposition 5.4. For simplicity let $P_{ij} = \text{Pr}(X_{ij})$; $P_{itj} = \text{Pr}(V_{itj})$ and $P_{istj} = \text{Pr}(W_{istj})$. Maximising Λ (the objective function of EMFLP), is equivalent to maximising the Shannon entropy H:

$$H = -\sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} q_{ij} \left(P_{ij} \ln(P_{ij}) + \sum_{t \in \mathcal{T}} P_{itj} \ln(P_{itj}) + \sum_{s \in \mathcal{T}} \sum_{t \neq s \in \mathcal{T}} P_{istj} \ln(P_{istj}) \right)$$

Proof 5.4. Using the definitions of probabilities in equations (23)-(25), Λ can be re-expressed as:

$$\Lambda = \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} q_{ij} P_{ij} \{1 - \ln(q_{ij} P_{ij})\} + \sum_{i \in \mathcal{O}} \sum_{t \in \mathcal{T}} \sum_{j \in \mathcal{D}} q_{ij} P_{itj} \{1 - \ln(q_{ij} P_{itj})\}$$

$$+ \sum_{i \in \mathcal{O}} \sum_{s \in \mathcal{T}} \sum_{t \neq s \in \mathcal{T}} \sum_{j \in \mathcal{D}} q_{ij} P_{istj} \{1 - \ln(q_{ij} P_{istj})\}$$

Expanding, grouping like terms and using the normalisation axiom of probability we have:

$$\Lambda = \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} q_{ij} - \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} q_{ij} P_{ij} \ln(q_{ij} P_{ij}) - \sum_{i \in \mathcal{O}} \sum_{t \in \mathcal{T}} \sum_{j \in \mathcal{D}} q_{ij} P_{itj} \ln(q_{ij} P_{itj})$$

$$- \sum_{i \in \mathcal{O}} \sum_{s \in \mathcal{T}} \sum_{t \neq s \in \mathcal{T}} \sum_{j \in \mathcal{D}} q_{ij} P_{istj} \ln(q_{ij} P_{istj})$$

Expanding the terms in the logarithm function and grouping like terms we have:

$$\Lambda = \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} q_{ij} - \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} q_{ij} \ln q_{ij}$$

$$- \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} q_{ij} \left(P_{ij} \ln P_{ij} + \sum_{t \in \mathcal{T}} P_{itj} \ln P_{itj} + \sum_{s \in \mathcal{T}} \sum_{t \neq s \in \mathcal{T}} P_{istj} \ln P_{istj} \right)$$

Thus

$$\Lambda = - \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} \ln q_{ij}! + H$$

The term $- \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} \ln q_{ij}!$ is constant and can be ignored in the optimisation process, since q_{ij} (input data) are not decision variables. Hence maximising Λ is equivalent to maximising the entropy H with respect to the decision variables:

$$H = - \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} q_{ij} P_{ij} \ln P_{ij} - \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} q_{ij} \sum_{t \in \mathcal{T}} P_{itj} \ln P_{itj} - \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} q_{ij} \sum_{s \in \mathcal{T}} \sum_{t \neq s \in \mathcal{T}} P_{istj} \ln P_{istj}$$

with the first, second and third terms being the Shannon entropies for road alone, metropolitan and regional intermodal transport decision variables. The above equation can be simplified further by defining the set of elementary modal alternatives; $\mathcal{S} = \{\{0, s, t\}, \forall s \in \mathcal{T}; t \in \mathcal{T}\}$, where $\{0\}$ is the index road alone alternative. The subset $\{\{0, s, t\}, \forall s = t \in \mathcal{T}\}$ represents modal alternatives for the metropolitan transport market whilst $\{\{0, s, t\}, \forall s \neq t \in \mathcal{T}\}$ represent the modal alternatives for the regional transport market:

$$H = - \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} q_{ij} \sum_{m \in \mathcal{S}} P_{imj} \ln P_{imj} \qquad (5.42)$$

Proposition 5.4. Maximising entropy facility location problem (EMFLP) with objective function Λ is equivalent to maximising total welfare:

$$\max \sum_{i\in\mathcal{O}} \sum_{j\in\mathcal{D}} L_{ij} q_{ij}$$

Subject to the transport budget constraint (5.5):

$$\sum_{i\in\mathcal{O}} \sum_{s\in\mathcal{T}} \sum_{t\neq s\in\mathcal{T}} \sum_{j\in\mathcal{D}} \tilde{c}_{istj} W_{istj} + \sum_{i\in\mathcal{O}} \sum_{t\in\mathcal{T}} \sum_{j\in\mathcal{D}} \tilde{c}_{itj} V_{itj} + \sum_{i\in\mathcal{O}} \sum_{j\in\mathcal{D}} c_{ij} X_{ij} \leq c$$

where L_{ij} is the maximum expected utility in Equation (5.30) interpreted as accessibility in Batty (2010) and Williams (1977) or consumer surplus in Train (2009) and De Jong et al. (2005). The cost variable $\tilde{c}_{itj}$ has its usual meaning in Chapter 4. The variable $\tilde{c}_{istj}$ represent the cost of regional intermodal transport: $\tilde{c}_{istj} = c_{is} + \tilde{c}_s + c_{st} + \tilde{c}_t + c_{tj}$ where $\tilde{c}_s = c_s + \frac{\psi_s}{\beta}$ and $\tilde{c}_t = c_t + \frac{\psi_t}{\beta}$ with $\frac{\psi_s}{\beta}$ and $\frac{\psi_t}{\beta}$ been the shadow prices (\$ per TEU) associated with terminal s and t respectively.

Proof 5.4. The proof follows directly from Proposition 4.7 in Chapter 4.

5.5 Numerical Examples

5.5.1 Data Generation

The data used for the numerical examples came from Australian Post (AP) data set (Beasley, 1990) designed for the Capacitated Single Allocation Hub Location Problems (CSAHLP) and has up to 200 nodes with capacity and fixed cost of locating a hub at each node. The data also contains the geographic coordinates of each node, which were used to compute the costs between nodes together with cost parameters. The 20-node version of the dataset was adapted to create both metropolitan and regional intermodal markets with the resulting network structure shown in Figure 5.3.

In creating the market for metropolitan intermodal transport, it was assumed those origin-destination pairs with distances not greater than 50km are too short to justify the use of two IMTs and so a regional intermodal transport mode is not available for such movements. It

is further assumed that for these movements, the longest leg (origin-to-IMT or IMT-to-destination) is the rail leg and benefits from economies of scale. It was also assumed that the average distance between the port and any cargo destination in the metropolitan region is not greater than 50km and so metropolitan intermodal transport modes are not available for movements with distances greater than 50km. Finally, in both cases, intermodal transport movements with an IMT location also acting as cargo origin or destination are removed from the potential markets. With the above assumptions, the metropolitan intermodal transport market accounts for about 45% of the total demand, whilst the regional market accounts for 55%.

Each cost component (see equations 4.11 and 5.1) in both the metropolitan and regional intermodal transport was constructed as functions of distance between the relevant nodes. The metropolitan intermodal transport cost was constructed as: $c_{itj} = \alpha_1 d_{it} + \mu_t + \alpha_0 d_{tj}$ assuming the intermodal leg $(i \rightarrow t)$ is longer than $(t \rightarrow j)$. The cost of regional intermodal transport was constructed using; $c_{istj} = \alpha_0 d_{is} + \mu_s + \alpha_2 d_{st} + \mu_t + \alpha_0 d_{tj}$, where α_1, α_2 are the transfer costs ($ per TEU per km) capturing the economies of scale between IMTs (assuming the port is an IMT) in metropolitan and regional intermodal systems respectively; μ_t is the cost incurred using an IMT ($ per TEU); α_0 is the collection or distribution cost for road alone ($ per TEU per km). Thus $c_{ij} = \alpha_0 d_{ij}$ is the unit road transport cost between two nodes in the AP data set. Unless otherwise stated, the value of α_2 was set at 0.75 and α_0 at 2 (both from the AP data set) and α_1 is the average of α_0 and α_2. The parameter $\mu_t = f_t/b_t$ (fixed cost over capacity).

Finally, the choice set for regional transport alternatives available for each origin-destination pair was expanded to prevent unrealistically small flows or infeasible movements (see Figure 5.4). For example, if IMT $s \in \mathcal{K}$ is sufficiently close to cargo origin zone $i \in \mathcal{O}$ and IMT $t \neq s \in \mathcal{K}$ is sufficiently close to cargo destination $j \in \mathcal{D}$, then it is safe to exclude cargo movement $i \rightarrow t \rightarrow s \rightarrow j$ from the choice set. Although, the model may assign negligible demand to this type of movements, it is practically expedient to exclude them. To control for this, for any pair of movements $i \rightarrow t \rightarrow s \rightarrow j$ and $i \rightarrow s \rightarrow t \rightarrow j$ only the movement with the smallest cost is considered as a feasible alternative for the movements of cargo between origin $i \in \mathcal{O}$ and destination $j \in \mathcal{D}$. The models were implemented in the C/C++ environment, on an Inter(R) Core (TM) i5-3210M CPU@ 2.50GHz, and 8:00 GB RAM of CPU.
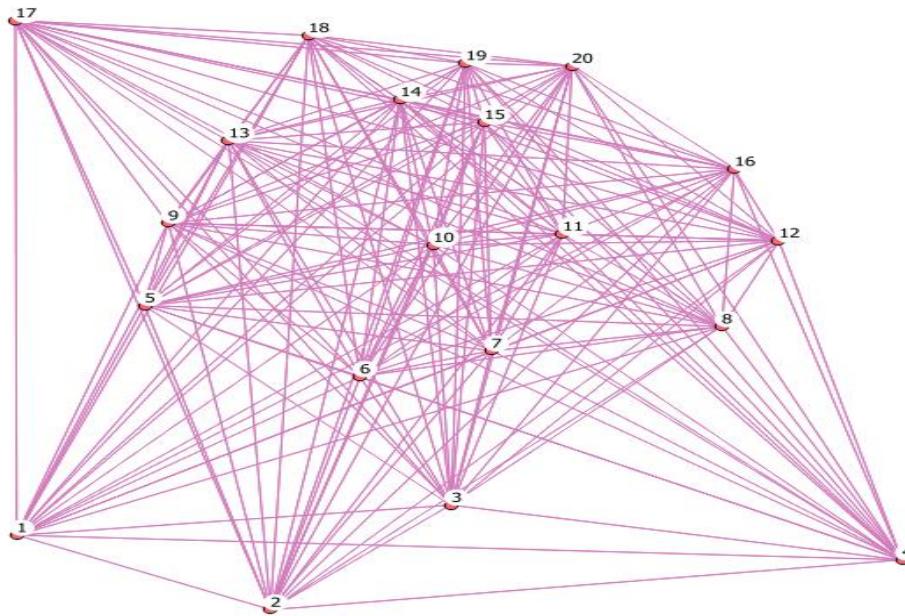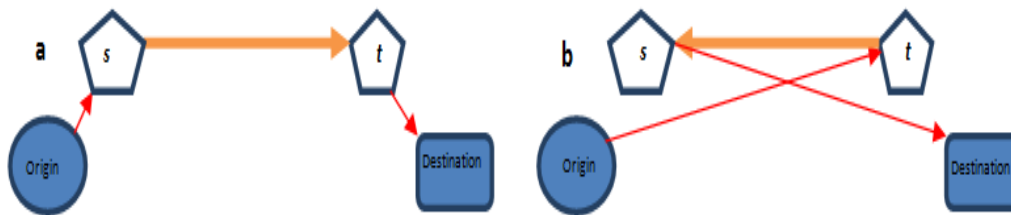
Figure 5.3: Network structure of text data

Figure 5.4: (a) Feasible alternative; (b) infeasible alternative

5.5.2 Analysis of main results

The EMFLP was implemented on the above dataset with the location of $p = 2,3,4,5$ in turn and the results presented in Table 5.1 with estimated cost sensitivity parameters, β, in column 4. The associated Lagrangian multipliers for the capacity constraints ψ_t are all zero since none of the optimal IMT locations reached the handling capacity. From Table 5.1, locations 10 and 15 emerged as the best places for locating IMTs when $p = 2$. The contribution of each located IMT to total intermodal transport demand is shown in Figure 5.5. About 51% of total expected demand for intermodal transport (2056 TEUs) is derived from the metropolitan intermodal transport market. Out of this (metropolitan intermodal transport demand) IMT 10 contributed about 91% with just 9% contribution by IMT 15. The location of two IMTs means that both contributed equally to the demand of regional intermodal transport, although about 60% of the

144

demand first went through IMT 10 before going through IMT 15 and the remaining flow went in the reverse direction.

Table 5.1: Optimal IMTs and computational time (p = 2, 3, 4, 5)

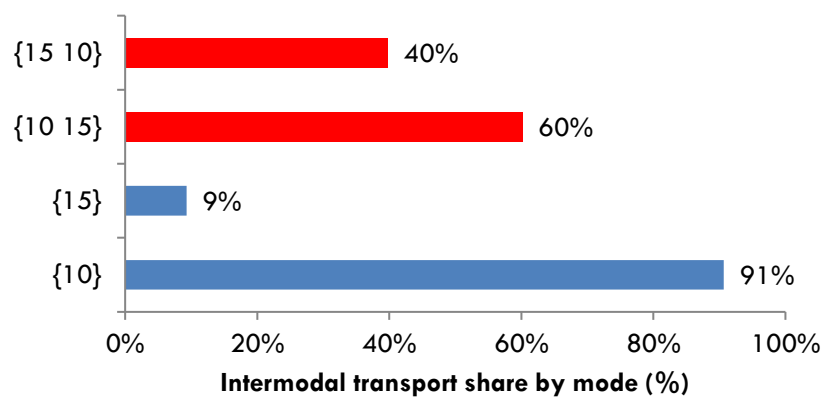| IMTs	CPU(s)	Demand	Beta	Located IMT-------->					IMT demand------->				
2	2	2,056	0.0075	10	15				1311	745			
3	13	3,126	0.0103	10	15	17			1154	841	1131		
4	76	3,676	0.0145	10	15	17	9		952	884	937	903	
5	329	3,936	0.0183	10	15	17	9	5	784	927	765	744	716

Figure 5.5: p=2 located IMTs (metropolitan market is blue)

The results may suggest that the selection of IMT 15 would have been unlikely if the selection process was based on only the metropolitan intermodal market. To support this analysis, we run the model on only the metropolitan intermodal market followed by another run on only the regional market and the results are shown in Figure 5.6. The figure shows that for metropolitan market run, IMTs 7 and 10 are the best locations whilst for the regional market scenario, IMTs 4 and 15 emerged as the best locations. Thus, for the metropolitan market scenario IMT 7 replaced IMT 15 and for the regional market scenario, IMT 4 replaced IMT 10 as the best IMT location. These results show that the wrong IMT location decisions can be made if the target market (metropolitan, regional or both) is not well defined.
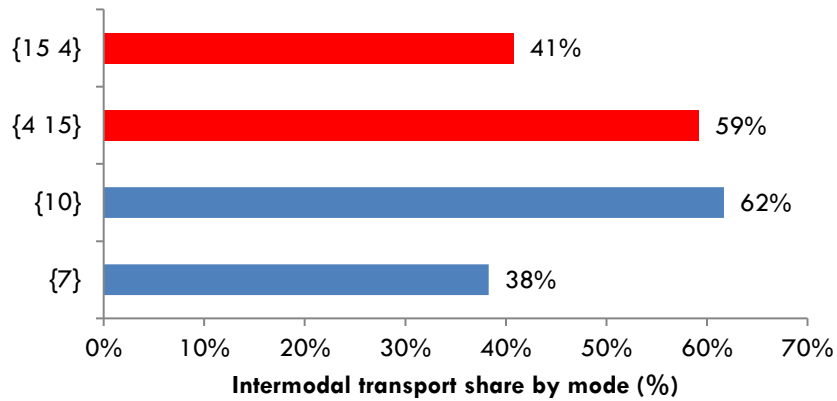
Figure 5.6: Metropolitan and regional markets (metropolitan market in blue)

For $p = 3,$ the best IMT locations were 10, 15 and 17 with the distribution of demand shown in Figure 5.7. IMT 15 only captures 2% of the metropolitan intermodal transport demand, but it is the biggest contributor and a key node in the regional intermodal transport system. Overall, the share of intermodal demand for IMTs 10, 15 and 17 are 37%, 27% and 36% respectively. The overall share of IMT 15 has reduced from 36% for $p = 2$ to 27% for $p = 3$; a reduction due to the increase in the modal alternatives by adding IMT 17, which is a better alternative in the metropolitan intermodal market. Similar conclusions can be drawn from the location of $p = 4$ and $p = 5$ with the contribution of each located IMTs in the overall intermodal transport demand shown in Figure 5.8.
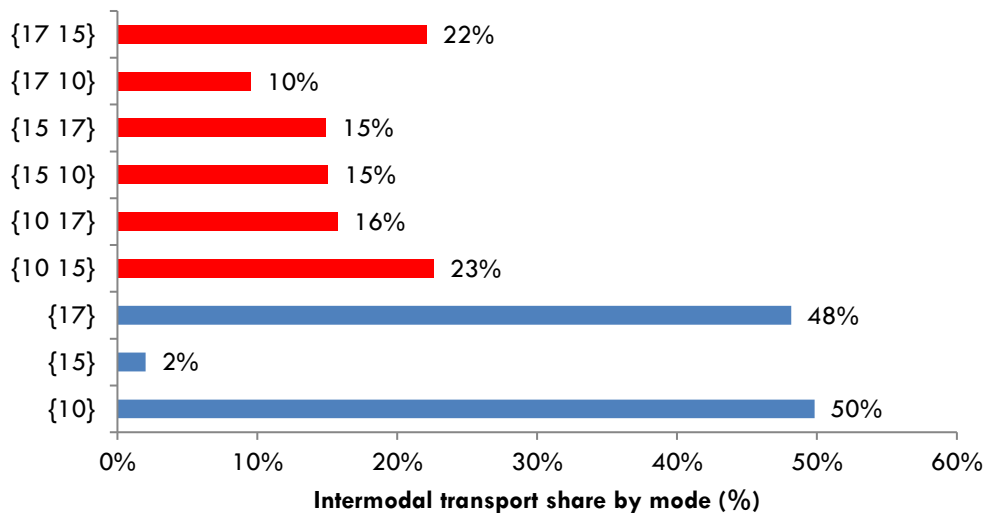

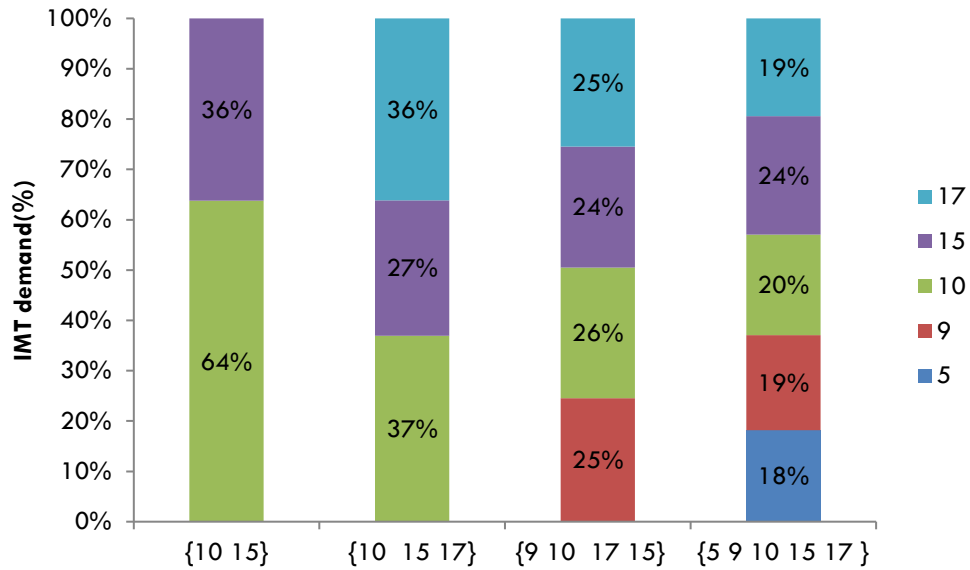**Figure 5.7: p = 3 located IMTs (metropolitan market in blue)**

Figure 5.8: p=2, 3, 4, 5 located IMTs and distribution of flow

5.5.3 Impacts of economies of scale on solutions

This section investigates the impacts of scale economies on IMT location choice and demand. The previous analysis used a fixed distance factor of $\alpha_2 = 0.75$ and $\alpha_1 = 1.4$ (which includes a discount for using rail through economies of scale) in computing the cost for regional and metropolitan intermodal transport modes. These factors were reduced in turn by $10\%, 20\%,$ 30% and 40% followed by $10\%, 20\%, 30\%$ and 40% increase in turn resulting in eight sets of discount factors. Each set was then run in turn for $p = 2,3,4,5$ and the results, presented in Table 5.2. As expected, the results generally show that irrespective of the number of located IMTs and where they are spatially located, the smaller the values of α_1 and α_2 (representing large discount) the higher the demand for intermodal transport and vice versa. For example, for $p = 2$, the demand for intermodal transport with respect the base (see Table 5.2) decreased by 16% for a 40% increase in α_1 and α_2 (small discounts) to about a 10% increase for a 40% reduction in α_1 and α_2 (large discounts). The effects of the discount factors on location decisions (optimal set of IMTs) is very mild and only present for $p = 5$. The mild effects may partly be explained by the sufficient handling capacity at each selected IMT. The results generally indicate that the locations selected for the development of IMTs for each p value are robust at least to variations in economies of scale benefits.

Table 5.2: Impacts of economies of scale on solutions

| scalar	2 IMT system			3 IMT system				4 IMT system					5 IMT system					
	Demand	IMT→		Demand	IMT→			Demand	IMT→				Demand	IMT→				
-0.1	2115	10	15	3215	10	15	17	3796	9	10	15	17		4080	5	9	10	15	17
-0.2	2167	10	15	3287	10	15	17	3886	9	10	15	17		4186	5	9	10	15	17
-0.3	2216	10	15	3350	10	15	17	3959	9	10	15	17		4268	5	9	10	15	17
-0.4	2260	10	15	3405	10	15	17	4021	9	10	15	17		4334	5	9	10	15	17
0.1	1993	10	15	3011	10	15	17	3519	9	10	15	17		3747	5	9	10	15	17
0.2	1922	10	15	2861	10	15	17	3317	9	10	15	17		3507	5	9	10	15	17
0.3	1839	10	15	2667	10	15	17	3067	9	10	15	17		3322	9	10	15	17	19
0.4	1736	10	15	2435	10	15	17	2785	9	10	15	17		3053	9	10	15	17	19

5.5.4 Solution quality of the entropic greedy algorithm

This section compares the solution quality and computation time from the entropy greedy heuristic (Algorithm A_3) with the enumeration algorithm (Algorithm A_2), which guarantees optimal solutions but with high computational costs. The base runs (in Table 5.1) were expanded to include $p = 6,7,8$ and the optimal solutions together with their computational times (CPU in seconds) are presented in Table 5.3. The results show that the computational time of the heuristic increases linearly with increasing value of p whilst the enumeration algorithm appears to increase exponentially with increasing value of p as shown in Figure 5.9. Also, as expected and shown in Section 5.4.4, the enumeration algorithm is faster than the heuristic for $p = 2$, but slower for $p \geq 3$.

In terms of solution quality, the heuristics returns the optimal solution on all the eight instances as shown in Table 5.3. We also repeated the runs for the instances generated in Section 5.5.3 and again the optimal solution was returned for each instance. Finally, we conducted as simple experiment where we used the cost values in the base runs as means and assumed a 50% standard deviation about each mean and generated 50 test instances each for $p = 2,3,4,5$. Again, for all the 200 generated test instances, the heuristic returned the optimal solution for each instance.

Table 5.3: Results for the Heuristics runs

IMTs to locate	CPU(seconds)		Optimal value	
	Enumeration	Heuristic	Enumeration	Heuristic
2	2	4	-4027	-4027
3	13	8	-2551	-2551
4	76	13	-1405	-1405
5	329	21	-547	-547
6	2643	61	137	137
7	8784	82	724	724
8	17757	106	1226	1226
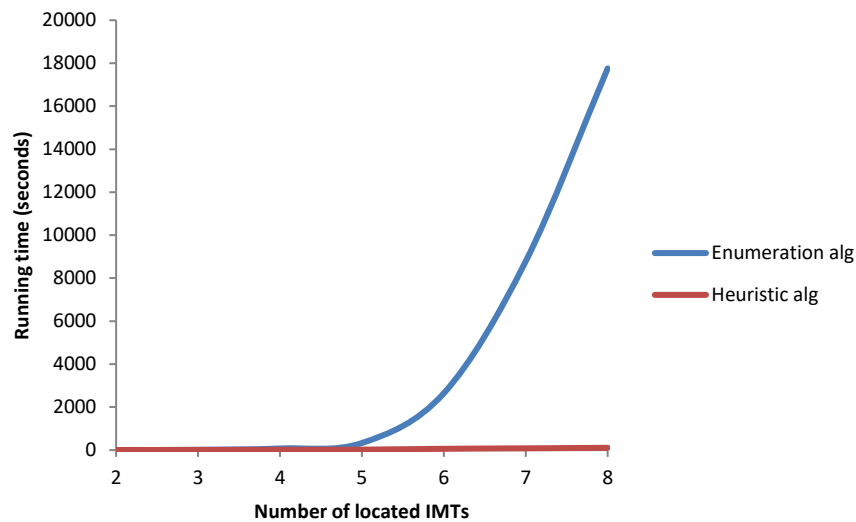
Figure 5.9: Running time comparison of algorithm A3 (heuristic) and A2

5.6 Conclusion

In this chapter, a flexible model based on the principle of entropy maximisation for locating inland multi-shipper intermodal container terminals in a context where shippers have choices which include whether or not to use the facilities is proposed to address research question II. The overall problem was decomposed using Lagrangian relaxation technique into a linked facility location sub-problem and a mode choice sub-problem. Key features of the model were illustrated through model properties and numerical examples. The mode choice problem was solved using modified Bregman's algorithm and cast as nested conditional probabilities of modal and IMT usage suitable for forecasting and policy testing. Two algorithms were proposed to solve the linked facility location and mode choice sub-problems; complete enumeration and a heuristic algorithm. The heuristic was shown to be bounded by $O(p\tau^2 T_B)$,

which means the running time grows linearly as the number of facilities to locate (p) increases for a fixed number of candidate IMT locations τ with T_B as the running time of the mode choice sub-problem, MCP. In terms of solution quality, the proposed heuristic algorithm returns the optimal solution on all the tested instances of the problem. Although, the heuristic has not been proved to return the optimal solution for every instance of the problem, it has been shown by numerical examples to work very well.

Chapter 6 Variable cargo demand models

6.1 Background

The primary objective for location an IMT is for it to serve as a transfer node between two modes (e.g., rail and trucks) where containers (or other loading units) arriving at the terminal on one mode (e.g., trains) are transferred to another mode (trucks) for onward journey to another cargo destination. These terminals over the years have evolved to perform value added or auxiliary activities such as warehousing, empty container storage and other cargo related activities (IPART 2011; Meyrick 2006). As noted in DoFD (2011) and Meyrick (2006) revenues from these auxiliary activities are vital for the viability and sustainability of many inland intermodal terminals. This is partly due to the high setup and operation costs associated with these terminals, especially those located in metropolitan areas, and also the need for it operate continuously throughout the year.

Chapters 1 and 2 show how the choice of mode depends on where the cargo is destined and also that intermodal transport may not feasible or cost competitive if the cargo destination is too close to the cargo origin. Conversely, the cargo destination depends on modal accessibility. That is, the cargo destination must be connected to the transport network and must be accessible by at least one available mode of transport. These reveal a 'natural' link between cargo distribution and mode choice, where the choice of mode is conditioned by the choice of cargo destination, whilst the mode choice influences the choice of cargo destination through accessibility measures. This makes the cargo distribution accessibility-sensitive and the easier a destination can be reached by available modes of transport the higher the quantity of cargo it attracts. It could be said in loose economic terms that travel cost to a zone is the price to pay to have access to participate in freight related activities (e.g., warehousing or storage) in that zone

and that the quantity of cargo attracted to that zone is non-increasing with travel cost or non-decreasing with accessibility.

Furthermore, the increase in cargo volumes and their distributions induce the need for freight facilities including intermodal terminals, which in turn promote intermodal transport. Intermodal transport increases modal options for shippers and hence influences cargo distributions. The goal in this chapter is to replace the fixed matrix with variable cargo demand models such that changes in cargo distribution patterns are allowed to influence terminal locations, whilst terminals locations influence cargo distributions through accessibility measures.

Two problems are added to the entropy framework; cargo production and cargo distribution problems. The generation and distribution problems together with the mode choice problem in Chapter 5 are developed within the entropy framework and are linked through accessibility measures such that changes in say the accessibility or cost of using IMTs influences the distribution of modes, which in turn influences the distribution and production of cargo (containers) whilst cargo production and/or distribution conditions modal and IMT demands. These problems together form the cargo flow problem (CFP) and is connected to the facility location problem (FLP) and together determines the most likely locations and usage of intermodal terminals.

The rest of the chapter is organised as follows; Section 6.2 presents the key assumptions underlying the cargo production and distribution models; the proposed methods for developing the cargo production and distribution models and how they can be expressed as constraints within the entropy framework is presented in Section 6.3. Section 6.4 presents the entropy-based mathematical formulation of intermodal terminals location with variable cargo demand problem (IMTL+VDP) and incorporates the cargo production, distribution, mode choice and facility location problems. Algorithm for solving the formulated problem is presented in Section 6.5. Finally, the implementation of the model in practice is discussed in Section 6.5, followed by the conclusions in Section 6.6.

6.2 Methodology

6.2.1 Assumptions

The assumptions used in the previous chapters also hold under this chapter with the exception of the fixed cargo matrix. In addition to these assumptions, the cargo generation and distribution models are assumed to be governed by a set of factors such that changes in these factors change the quantity of cargo generated and distributed in the study area. These factors are presented in the model by a set of variables with the modal accessibility or access to multiple modes variable (capturing changes in transport network conditions) being one of them. The other factors are assumed to be location (cargo origin or destination) specific factors such as land-use, industry-specific, demographic and socio-economic factors.

The set of location specific factors governing the production of cargo is represented by $\mathcal{G}$ and the variable g_{ik} representing the quantity of variable type $k \in \mathcal{G}$ associated with cargo production zone $i \in \mathcal{O}$ with $\bar{g}_k$ presenting weighted averages over all production zones and the weights been the observed cargo flows. Similarly, the set $\mathcal{H}$ represent the set of factors explaining the consumption of cargo at a given location with the variable a_{jl} representing the quantity of variable type $l \in \mathcal{H}$ associated with cargo destination or consumption zone $j \in \mathcal{D}$ and $\bar{a}_l$ presenting weighted averages over all cargo destination zones. Also, consistent with the previous chapters, the unit of analysis is the cargo measured in TEUs.

Decision Variables

The key outputs of the model are the flow variables W_{itj}, V_{itj}, U_{ij} for determining the demands for regional, metropolitan and road alone transport respectively and the location variables Y_t, which determines the locations to select for the developments of IMTs. Adding to the list are Q_{ij} for determining the distributions of cargo (in TEUs) from production zone $i \in \mathcal{O}$ to consumption zones $j \in \mathcal{D}$ and Q_i for determining the quantity of cargo produced by each production zone $i \in \mathcal{O}$.

6.2.2 The Cargo production and distribution models

Two main methods of incorporating cargo production and distribution models into the entropy framework are presented followed by discussions on their merits and limitations. The first

method is the Poisson cargo attraction method and the second is the weighted mean method. Each of these two methods is described below.

6.2.2.1 Poisson cargo attraction method

To incorporate cargo distribution model into the entropy framework, the study assumes that the quantity of cargo (in TEUs per day) attracted to each destination zone (TEUs per day) is independent Poisson random variables with density:

$$P(Q_j|d_j) = \frac{e^{-d_j}d_j^{Q_j}}{Q_j!}; Q_j = 0,1,2,\dots \tag{6.1}$$

where d_j are the mean quantity of cargo (in TEUs per day) arriving at destination $j \in \mathcal{D}$. Based on the fact that the possible values of $d_j \geq 0$, the 'natural' choice of function to link the mean cargo arrival d_j and the explanatory variables in set $\mathcal{H}$ is the log-linear function (Cameron and Trivedi 2013):

$$d_j = \exp\left(\sum_{l\in\mathcal{H}} \theta_l a_{jl}\right) \tag{6.2}$$

where θ_l is the weight or importance associated with variable a_{jl}. The term $a_j = \sum_{l\in\mathcal{H}} \theta_l a_{jl}$ can be considered as overall attractiveness of destination j as cargo destination zone. Given a set of independent observations, the goal is to find the set of parameters θ that makes the probability density function in (6.1) as likely as possible in re-producing the observed data. This can be done by constructing a log-likelihood function and maximising it with respect to θ.

$$\Lambda_{\text{MLE}} = \sum_{j\in\mathcal{D}} \ln P(Q_j|d_j) = \sum_{j\in\mathcal{D}} \{Q_j \ln d_j - d_j - \ln Q_j!\} \tag{6.3}$$

The first-order condition for optimal Λ_{MLE} with respect to the parameter vector θ yields:

$$\sum_{j\in\mathcal{D}} (Q_j - d_j)a_{jl} = 0; \ \forall l \in \mathcal{H}$$

alternatively,

$$\sum_{j \in \mathcal{D}} Q_j a_{jl} = \sum_{j \in \mathcal{D}} d_j a_{jl}; \quad \forall l \in \mathcal{H} \tag{6.4}$$

This can be expressed as weighted averages with the weights being the observed cargo flows as follows:

$$\sum_{j \in \mathcal{D}} Q_j a_{jl} = Z \bar{a}_l; \quad \forall l \in \mathcal{H} \tag{6.5}$$

where $\bar{a}_l = \frac{1}{Z} \sum_{j \in \mathcal{D}} d_j a_{jl}$ is the weighted average of variable $l \in \mathcal{H}$.

Corollary 6.1: If $Q_j = \sum_{i \in \mathcal{O}} Q_{ij}$, then following from (6.4) the following relation holds:

$$\sum_{j \in \mathcal{D}} \sum_{i \in \mathcal{O}} Q_{ij} a_{jl} = \sum_{j \in \mathcal{D}} d_j a_{jl}; \quad \forall l \in \mathcal{H} \tag{6.6}$$

Or from (6.5) we have:

$$\sum_{j \in \mathcal{D}} \sum_{i \in \mathcal{O}} Q_{ij} a_{jl} = Z \bar{a}_l; \quad \forall l \in \mathcal{H} \tag{6.7}$$

Equation (6.6) or (6.7) establishes the relationship between the demand for cargo at a given destination and the factors or destination specific characteristics driving the demand.

Corollary 6.2: If Q_i is the quantity of cargo (in TEUs per day) produced in zone $i \in \mathcal{O}$, and since Q_i and Q_{ij} are related through $Q_i = \sum_{j \in \mathcal{D}} Q_{ij}$, it implies Q_i are also Poisson distributed. Thus, based on (6.4), explaining the mean cargo production q_i by the set of production factors $\mathcal{G}$ the following relation also holds:

$$\sum_{i \in \mathcal{O}} Q_i g_{ik} = \sum_{i \in \mathcal{O}} q_i g_{ik}; \quad \forall k \in \mathcal{G} \tag{6.8}$$

Or in weighted averages we have:

$$\sum_{i \in \mathcal{O}} Q_i g_{ik} = Z \bar{g}_k; \quad \forall k \in \mathcal{G} \tag{6.9}$$

Equations (6.5), (6.7) and (6.9) are particularly useful when observed cargo flows q_{ij}, d_j or q_i respectively are not available for all zones. In such cases the weighted averages can be computed based on only zones where observed data are available. This leads to a more general formulation of the variable cargo demand models.

6.2.2.2 Weighted mean method

Let $f_l(a_{jl})$ be the data generation function for attraction variable $l \in \mathcal{H}$ for each cargo consumption zone $j \in \mathcal{D}$ with expected value $\bar{a}_l$; $l \in \mathcal{H}$. Then by definition, the expected value of the function $f_l(a_{jl})$; $l \in \mathcal{H}$ can be expressed as:

$$\mathbb{E}[f_l(a_{jl})] = \bar{a}_l = \sum_{j \in \mathcal{D}} p_j f_l(a_{jl}); \quad \forall l \in \mathcal{H} \tag{6.10}$$

where p_j is the probability of realising demand Q_j; $j \in \mathcal{D}$ based on the influence of attraction variable $l \in \mathcal{H}$. See Jaynes (1982) for the use of similar expression. Using the frequentist definition of probability $p_j = \frac{Q_j}{Z}$, and using the fact that $Q_j = \sum_{i \in \mathcal{O}} Q_{ij}$, Equation (6.10) can be re-written as follows:

$$\sum_{j \in \mathcal{D}} \sum_{i \in \mathcal{O}} Q_{ij} f_l(a_{jl}) = Z \bar{a}_l; \quad \forall l \in \mathcal{H} \tag{6.11}$$

The next step is the choice of the function $f_l(a_{jl})$, which can take several forms. The reader is referred to the papers by (Patil and Rao 1978; 1986) on various forms of specifying function $f_l(a_{jl})$. This study focusses on a simple specification of $f_l(a_{jl})$ called the sized based function (Patil and Rao 1978):

$$f_l(a_{jl}) = a_{jl}; \quad \forall l \in \mathcal{H}, j \in \mathcal{D} \tag{6.12}$$

By inserting Equation (6.12) into (6.11) produces the Poisson equation in (6.7):

$$\sum_{j \in \mathcal{D}} \sum_{i \in \mathcal{O}} Q_{ij}\, a_{jl} = Z\bar{a}_l; \quad \forall l \in \mathcal{H} \tag{6.13}$$

Thus, the Poisson method can be seen as a special case of the weighted mean method since the weighted mean method allows several choices of function $f_l(a_{jl})$ describing the available information on the variable of interest.

Corollary 6.3: If the function $f_k(g_{ik})$ generates the data for cargo production variable $k \in \mathcal{G}$ for each production zone $i \in \mathcal{O}$ with expected value $\bar{g}_k; k \in \mathcal{G}$ over all production zones, then based on (6.13) the following must be satisfied:

$$\sum_{i \in \mathcal{O}} Q_i g_{ik} = Z\bar{g}_k; \quad \forall k \in \mathcal{G} \tag{6.14}$$

6.2.2.3 Summary of available evidence

The relevant information and assumptions about the inland containerised system are summarised in the following constraints:

1. *Conservation of cargo flows.* These constraints guarantee that the total quantity of cargo originating from a given zone must equal the sum of cargo arriving at all destination zones from that origin:

$$\sum_{j \in \mathcal{D}} Q_{ij} = Q_i \quad \forall\, i \in \mathcal{O} \tag{6.15}$$

2. *Validation constraint I.* These constraints ensure that the observed quantity of cargo arriving at each destination from all origins equals the estimated equivalent by the model.

$$\sum_{i \in \mathcal{O}} Q_{ij} = d_j \quad \forall\, j \in \mathcal{D} \tag{6.16}$$

157

3. *Validation constraint II*. These constraints ensure that the observed quantity of cargo generated by each production zone equals the estimated equivalent by the model.

$$Q_i = q_i; \quad \forall \, i \in \mathcal{O} \tag{6.17}$$

4. *Validation constraint III*. This constraint ensures that the observed total quantity of cargo in the system equals the sum over all estimated cargo generated by the production zones. It is important to note that if q_i are observed and constraint (6.17) is satisfied, then (6.18) is automatically satisfied and the Lagrangian parameters associated (6.18) will be zero.

$$\sum_{i \in \mathcal{O}} Q_i = Z \tag{6.18}$$

5. *Exploratory constraints*: Constraints (6.9) and (6.7) explaining the factors governing the production and distributions of cargo respectively.

6.3 Incorporating variable cargo demand

Based on the existing information summarised in the constraints above, the entropy facility location problem in Chapter 5 is extended to include the following constraints:
1. Constraints (6.7) and (6.9)
2. Constraints (6.15) to (6.18) and the following non-negativity constraints:

$$Q_{ij} \geq 0; Q_i \geq 0; \quad \forall \, i \in \mathcal{O}; j \in \mathcal{D} \tag{6.19}$$

The extended problem is referred to as the entropy maximising facility location with variable demand problem (EMFL+VDP). The next section presents various way of handling the above formulated problem.

6.4 Solution to EMFL+VDP

The solution to EMFL+VDP is an extension of the solution to EMFLP in Chapter 5 to include the cargo production and distributions sub-problems. Following the decomposition procedure employed in Chapter 5, the EMFL+VDP is decomposed into FLP and CFP where the CFP

comprises the MCP in Chapter 5 and cargo production and distribution sub-problems. The solution to the MCP is the same as those developed in Chapter 5 except that the variable q_{ij} are now replaced with Q_{ij}; $\forall i \in \mathcal{O}; j \in \mathcal{D}$ to reflect the fact the distribution of cargo is no longer assumed fixed. The focus here is therefore on the solutions to the cargo production and distribution sub-problems. To solve these problems, a Lagrangian equation of CFP is constructed and applying the Karush-Kuhn-Tucker (KKT) optimality conditions with respect to the distribution variables Q_{ij} and the production variables Q_i; $\forall i \in \mathcal{O}; j \in \mathcal{D}$. The resulting Lagrangian equation is essentially the linear combination of the Lagrangian equation Λ_{LMCP} formed to solve the MCP in Chapter 5 and the new constraints developed above:

$$
\Lambda_{\text{LCFP}} = \Lambda_{\text{LMCP}} + \sum_{l \in \mathcal{H}} \theta_l \left(\sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} Q_{ij} \, a_{jl} - A_l \right) + \sum_{k \in \mathcal{G}} \phi_k \left(\sum_{i \in \mathcal{O}} g_{ik} Q_i - G_k \right) \tag{6.20}
$$

$$
+ \sum_{i \in \mathcal{O}} \gamma_{0i} \left(\sum_{j \in \mathcal{D}} Q_{ij} - Q_i \right) + \sum_{j \in \mathcal{D}} \gamma_{1j} \left(\sum_{i \in \mathcal{O}} Q_{ij} - d_j \right)
$$

$$
+ \sum_{i \in \mathcal{O}} \gamma_{2i}(Q_i - q_i) + \gamma_3 \left(\sum_{i \in \mathcal{O}} Q_i - Z \right)
$$

where $A_l = Z\bar{a}_l$ and $G_k = Z\bar{g}_k$. The parameters $\theta_l; \forall l \in \mathcal{H}$, $\phi_k; \forall k \in \mathcal{G}$ are the Lagrangian multipliers associated with cargo distribution constraint (6.7) and production constraint (6.9) and constants $\gamma_{0i}, \gamma_{1j}, \gamma_{2i}, \gamma_3; \forall i \in \mathcal{O}; j \in \mathcal{D}$ are the Lagrangian multipliers associated with constraints (6.15) to (6.18) respectively. Λ_{LCFP} is first optimised with respect to the distribution decision variables Q_{ij} followed by the cargo production decision variables Q_i leading to probability distributions for determining the production and distribution of cargo in the study area.

6.4.1 Cargo distribution model

The KKT conditions for a maximum Λ_{LCFP} with respect to the cargo distribution variable Q_{ij} are:

$$
\sum_{l \in \mathcal{H}} \theta_l a_{jl} + \gamma_{0i} + \gamma_{1j} + \eta_{ij} = 0 \; ; \; \forall \; i \in \mathcal{O}; j \in \mathcal{D} \tag{6.21}
$$

where η_{ij} are the Lagrangian multipliers associated with constraint (5.4) in Chapter 5. Making η_{ij} the subject from (6.21) we have:

$$e^{-\eta_{ij}} = e^{\gamma_{0i}}\exp\left(\gamma_{1j} + \sum_{l \in \mathcal{H}} \theta_l a_{jl}\right) \tag{6.22}$$

It was shown in Chapter 5 through Equation (5.35) that the parameters η_{ij} can be expressed in terms of modal accessibility:

$$e^{-\eta_{ij}} = Q_{ij}e^{-L_{ij}}$$

The above equation means that the parameters η_{ij} can be eliminated from (6.22) and making Q_{ij} the subject we have;

$$Q_{ij} = e^{\gamma_{0i}}\exp\left(\gamma_{1j} + L_{ij} + \sum_{l \in \mathcal{H}} \theta_l a_{jl}\right) \tag{6.23}$$

The Lagrangian parameters γ_i can be estimated by using Equation (6.23) to enforce constraint (6.15):

$$e^{\gamma_{0i}} = Q_i \frac{1}{\sum_{j \in \mathcal{D}} \exp\left(\gamma_{1j} + L_{ij} + \sum_{l \in \mathcal{H}} \theta_l a_{jl}\right)} \tag{6.24}$$

Equation (6.24) can be inserted into Equation (6.23) to produce the cargo distribution model:

$$Q_{ij} = Q_i \Pr(Q_{ij}); \ \forall \, i \in \mathcal{O}, j \in \mathcal{D} \tag{6.25}$$

where $\Pr(Q_{ij})$ is the probability of realising demand Q_{ij} between cargo production zone $i \in \mathcal{O}$ and consumption zone $j \in \mathcal{D}$:

$$\Pr(Q_{ij}) = \frac{\exp\left(\gamma_{1j} + L_{ij} + \sum_{l \in \mathcal{H}} \theta_l a_{jl}\right)}{\sum_{j \in \mathcal{D}} \exp\left(\gamma_{1j} + L_{ij} + \sum_{l \in \mathcal{H}} \theta_l a_{jl}\right)}; \ \forall \, i \in \mathcal{O}, j \in \mathcal{D} \tag{6.26}$$

Equation (6.26) can be generalised to allow differential degree of sensitivities of cargo distribution and modal distributions to changes in transport network variables such as travel time and costs by introducing a sensitivity parameter $0 < \lambda_D \leq 1$ into equation (6.26) as follows:

$$\Pr(Q_{ij}) = \frac{\exp\left(\gamma_{1j} + \lambda_D L_{ij} + \sum_{l \in \mathcal{H}} \theta_l a_{jl}\right)}{\sum_{j \in \mathcal{D}} \exp\left(\gamma_{1j} + \lambda_D L_{ij} + \sum_{l \in \mathcal{H}} \theta_l a_{jl}\right)} \tag{6.27}$$

The parameter λ_D is also called the structural parameter and together with the logsums L_{ij} links the mode choice model developed in Chapter 5 to the cargo distribution model in (6.27) such that changes in the choice of mode influence the distribution of cargo whilst cargo distribution conditions the choice of mode through the evaluation of Q_{ij}.

The term L_i in Equation (6.28) is derived from the denominator of (6.27) and forms the overall accessibility of cargo origin $i \in \mathcal{O}$ to all available cargo destinations and can also represents the cargo generation power of zone $i \in \mathcal{O}$.

$$L_i = \ln \sum_{j \in \mathcal{D}} \exp\left(\gamma_{1j} + \lambda_D L_{ij} + \sum_{l \in \mathcal{H}} \theta_l a_{jl}\right) \; ; \; \forall \, i \in \mathcal{O} \tag{6.28}$$

The Lagrangian parameters γ_{0i} in Equation (6.24) can then be expressed in terms of the cargo generation power of each production zone L_i; $i \in \mathcal{O}$ by using Equation (6.28):

$$e^{\gamma_{0i}} = Q_i e^{-L_i} \tag{6.29}$$

The parameters γ_{1j} acts like alternative specific constants and can be estimated iteratively using Corollary 6.4 below.

Corollary 6.4: The Lagrangian parameters γ_{1j} can be computed iteratively with the kth iterated values are evaluated using equation:

$$\gamma_{1j}{}^{k} = \gamma_{1j}{}^{k-1} + \ln\left(\frac{\tilde{Q}_j{}^{k-1}}{d_j}\right) \; ; \gamma_{1j}{}^{0} = 0; \; \forall \, j \in \mathcal{D}; k = 1,2,\dots \tag{6.30}$$

where $\tilde{Q}_j{}^{k-1}$ is the total cargo attracted to destination $j \in \mathcal{D}$

$$\tilde{Q}_j{}^{k-1} = \sum_{i \in \mathcal{O}} Q_{ij}{}^{k-1}$$

Proof 6.4: The proof follows directly from Proposition 4.6 in Chapter 4.

6.4.2 Cargo production model

The KKT conditions for a maximum Λ_{LCFP} with respect to the cargo production variable Q_i are:

$$\sum_{k \in \mathcal{G}} \phi_k g_{ik} - \gamma_{0i} + \gamma_{2i} + \gamma_3 = 0 \; ; \; \forall \; i \in \mathcal{O}; j \in \mathcal{D} \tag{6.31}$$

Substituting the parameters γ_{0i} in Equation (6.29) into (6.31) and making the decision variable $Q_i; i \in \mathcal{O}$ the subject we have:

$$Q_i = \exp\left(\gamma_3 + \gamma_{2i} + L_i + \sum_{k \in \mathcal{G}} \phi_k g_{ik}\right); \; \forall \, i \in \mathcal{O} \tag{6.32}$$

Similar to the distribution model in (6.27), Equations (6.32) can be generalised to also allow for differential degree of sensitivity of cargo production, distribution and mode choice with respect to changes in network conditions. This can be achieved by introducing the sensitivity parameter λ_G such that $0 < \lambda_G \le \lambda_D \le 1$ into equation (6.32) as follows:

$$Q_i = \exp\left(\gamma_3 + \gamma_{2i} + \lambda_G L_i + \sum_{k \in \mathcal{G}} \phi_k g_{ik}\right); \; \forall \, i \in \mathcal{O} \tag{6.33}$$

The parameters γ_{2i} are origin specific constants and can also be estimated iteratively using corollary 6.5 below:

Corollary 6.5: The Lagrangian parameters γ_{2i} can be computed iteratively with the kth iterated values are evaluated using equation:

$$\gamma_{2i}{}^{k} = \gamma_{2i}{}^{k-1} + \ln\left(\frac{\tilde{Q}_i{}^{k-1}}{q_i}\right) \quad ; \gamma_{2i}{}^{0} = 0; \ \forall \ i \in \mathcal{O}; k = 1,2,\dots \tag{6.34}$$

where $\tilde{Q}_i{}^{k-1}$ is the modelled or estimated cargo produced in zone $i \in \mathcal{O}$ through Equation (6.33) during iteration $k - 1$.

Proof 6.5: The proof follows directly from Proposition 4.6 in Chapter 4.

Similarly, the parameter γ_3 can be estimated using corollary 6.6 to ensure that the estimated total cargo in the system equals the observed total.

Corollary 6.6: The Lagrangian parameters γ_3 can be computed iteratively with the kth iterated values are evaluated using equation:

$$\gamma_3{}^{k} = \gamma_3{}^{k-1} + \ln\left(\frac{\tilde{Z}^{k-1}}{Z}\right) \quad ; \gamma_3{}^{0} = 0; \ \forall \ i \in \mathcal{O}; k = 1,2,\dots \tag{6.35}$$

where $\tilde{Z}^{k-1}$ is the modelled or estimated total cargo produced in the study area during iteration $k - 1$ and where:

$$\tilde{Z}^{k-1} = \sum_{i \in \mathcal{O}} Q_i{}^{k-1} ; k = 1,2,\dots \tag{6.36}$$

Proof 6.6: The proof follows directly from Proposition 4.6 in Chapter 4.

6.4.3 Parameter estimation in the CFP

Several simpler models can be derived from the models developed in (6.25) and (6.33) depending on the availability of data. The key ones are summarised as follows:

Model I (CFP₁): No available information on cargo distribution q_{ij}, production q_i, total cargo in the system Z or the quantity of cargo arriving at each destination d_j and no known factors governing the flows of cargo. In such situations, the structural parameters λ_D, λ_G can each be set to 1 and the Lagrangian parameters $\theta_l, \phi_k, \gamma_{0i}, \gamma_{1j}, \gamma_{2i}$ and γ_3 set to zero. The production model in (6.38) and the distribution model in (6.37) will be explained by only transport network conditions which is expressed in terms of modal (road alone and intermodal) access to each destination or cargo production zone:

$$\Pr(Q_{ij}) = \frac{\exp(L_{ij})}{\sum_{j\in\mathcal{D}} \exp(L_{ij})}; \ \forall\, i \in \mathcal{O}, j \in \mathcal{D} \tag{6.37}$$

$$Q_i = \exp(L_i); \ \forall\, i \in \mathcal{O} \tag{6.38}$$

Model II (CFP₂): Only information on total cargo Z in the system is available. Here, the resulting cargo distribution model is the same as that under CFP₁ or Equation (6.37). The production model accounts for the observed Z by incorporating the Lagrangian parameter γ_3 in Equation (6.38). Thus, the resulting distribution models in (6.37) is explained by only transport network conditions whilst the production model (6.39) is expressed in terms of both network conditions and a constant.

$$Q_i = \exp(\gamma_3 + L_i); \ \forall\, i \in \mathcal{O} \tag{6.39}$$

Model III (CFP₃): Only information on cargo production q_i is available. Observing q_i also means Z is observed since by definition $Z = \sum_{i\in\mathcal{O}} Q_i$. Again, the distribution model in Equation (6.37) remains unchanged but the production model is updated with the new information as Equation (6.40):

$$Q_i = \exp(\gamma_3 + \gamma_{2i} + \lambda_G L_i); \ \forall\, i \in \mathcal{O} \tag{6.40}$$

It can be observed that both γ_3 and γ_{2i} are constants so they can be combined into one constant: $\tilde{\gamma}_{2i} = \gamma_3 + \gamma_{2i}$. The parameter γ_3 can be normalised to zero ($\gamma_3 = 0$) resulting in $\tilde{\gamma}_{2i} = \gamma_{2i}$. Thus, estimating $\tilde{\gamma}_{2i}$ can be treated as estimating γ_{2i}. Equation (6.40) therefore simplifies to become:

$$Q_i = \exp(\tilde{\gamma}_{2i} + \lambda_G L_i); \; \forall \, i \in \mathcal{O} \tag{6.41}$$

Model IV (CFP₄): Available data on cargo production in each zone q_i with information on production factors $\mathcal{G}$. Again, observing q_i also means Z is observed since by definition $Z = \sum_{i \in \mathcal{O}} Q_i$. The distribution model is the same as that of CFP₄ or Equation (6.37). The resulting production model becomes:

$$Q_i = \exp\left(\tilde{\gamma}_{2i} + \lambda_G L_i + \sum_{k \in \mathcal{G}} \phi_k g_{ik}\right); \; \forall \, i \in \mathcal{O} \tag{6.42}$$

The parameters λ_G, ϕ_k in the model can be estimated using Poisson quasi-maximum likelihood estimator (QMLE) (Cameron and Trivedi 2013) and the $\tilde{\gamma}_{2i}$ estimate the same way as in CFP₃.

Model V (CFP₅): Only information regarding the quantity of cargo arriving at each destination d_j are available. Knowing d_j also implies that Z is known. The resulting production model is same as Equation (6.39) under CFP₂. The distribution model in (6.27) with $\lambda_D = 1$ reduces to:

$$\Pr(Q_{ij}) = \frac{\exp(\gamma_{1j} + L_{ij})}{\sum_{j \in \mathcal{D}} \exp(\gamma_{1j} + L_{ij})} \tag{6.43}$$

The estimation of the parameters γ_{1j} were described under Corollary 6.4.

Model VI (CFP₆): Available data on the quantity of cargo arriving at each destination d_j and information on distribution factors $\mathcal{H}$. The resulting production model is the same as Equation (6.39) under CFP₂. The distribution model in (6.27) with $\lambda_D = 1$ reduces to:

$$\Pr(Q_{ij}) = \frac{\exp(\gamma_{1j} + L_{ij} + \sum_{l\in\mathcal{H}} \theta_l a_{jl})}{\sum_{j\in\mathcal{D}} \exp(\gamma_{1j} + L_{ij} + \sum_{l\in\mathcal{H}} \theta_l a_{jl})} \qquad (6.44)$$

The parameters θ_l representing the weight or importance associated with each attraction parameter $l \in \mathcal{H}$ and can be estimated by enforcing constraint (6.13):

$$f(\theta_l) = \sum_{j\in\mathcal{D}} \sum_{i\in\mathcal{O}} Q_{ij} a_{jl} - A_l = 0; \forall\, l \in \mathcal{H}$$

or

$$f(\theta_l) = \sum_{j\in\mathcal{D}} \sum_{i\in\mathcal{O}} Q_i \left(\frac{\exp(\gamma_{1j} + L_{ij} + \sum_{l\in\mathcal{H}} \theta_l a_{jl})}{\sum_{j\in\mathcal{D}} \exp(\gamma_{1j} + L_{ij} + \sum_{l\in\mathcal{H}} \theta_l a_{jl})} \right) a_{jl} - A_l = 0; \forall\, l \in \mathcal{H} \qquad (6.45)$$

The functions $f(\theta_l)$ are continuous and differentiable with respect to θ_l and can be optimised using Newton Raphson's or Hyman (1969) methods to estimate θ_l. Several numerical examples show that Hyman method is more computational efficient and stable.

Model VII (CFP7): Available data on the quantity of cargo distributed between zones q_{ij} with available information on production $\mathcal{G}$ and distribution factors $\mathcal{H}$. Knowing q_{ij} means d_j, q_i and Z are known. The resulting models could be considered as full information models. The resulting production model is Equation (6.33) and the distribution model is Equation (6.27). The parameters in these models can be estimated using maximum likelihood estimator (MLE) or the Poisson quasi-MLE (QMLE) or other appropriate estimators such as Bayesian. The estimated parameters include the structural parameters λ_D, λ_G.

Table 6.1 provides a summary of the above seven production and distribution models. As shown in Table 6.1 other combinations of production and distribution models can also be achieved depending on data availability.

Table 6.1: Summary of cargo production and distribution models

	Data availability	Production model	Distribution model	Comments
1	No information and cargo production, distribution and total cargo in the system	$Q_i = \exp(L_i)$	$Q_{ij} = Q_i \dfrac{\exp(L_{ij})}{\sum_{j \in \mathcal{D}} \exp(L_{ij})}$	Car production and distribution models expressed in terms of only transport network variables
2	Only information on total cargo Z in the system is available	$Q_i = \exp(\gamma_3 + L_i)$	$Q_{ij} = Q_i \dfrac{\exp(L_{ij})}{\sum_{j \in \mathcal{D}} \exp(L_{ij})}$	Same as (1) but with a production constant to ensure that the total cargo estimated equals the observed.
3	Only information on cargo production in each zone q_i is available	$Q_i = \exp(\tilde{\gamma}_{2i} + \lambda_G L_i)$	$Q_{ij} = Q_i \dfrac{\exp(L_{ij})}{\sum_{j \in \mathcal{D}} \exp(L_{ij})}$	Same as (2) but with a production constant for each production zone
4	Available data on cargo production in each zone q_i with information on production factors $\mathcal{G}$.	$Q_i = \exp\left(\tilde{\gamma}_{2i} + \lambda_G L_i + \sum_{k \in \mathcal{G}} \phi_k g_{ik}\right)$	$Q_{ij} = Q_i \dfrac{\exp(L_{ij})}{\sum_{j \in \mathcal{D}} \exp(L_{ij})}$	Same as (3) in addition to variables explaining the production of cargo in each zone
5	Only information regarding the quantity of cargo arriving at each destination d_j are available	$Q_i = \exp(\gamma_3 + L_i)$	$Q_{ij} = Q_i \dfrac{\exp(\gamma_{1j} + L_{ij})}{\sum_{j \in \mathcal{D}} \exp(\gamma_{1j} + L_{ij})}$	Same as (2) in addition to attraction constant for each cargo destination zone
6	Available data on the quantity of cargo arriving at each destination d_j and information on distribution factors $\mathcal{H}$	$Q_i = \exp(\gamma_3 + L_i)$	$Q_{ij} = Q_i \dfrac{\exp(\gamma_{1j} + L_{ij} + \sum_{l \in \mathcal{H}} \theta_l a_{jl})}{\sum_{j \in \mathcal{D}} \exp(\gamma_{1j} + L_{ij} + \sum_{l \in \mathcal{H}} \theta_l a_{jl})}$	Same as (5) in addition to variables explaining the attraction of cargo to each destination zone
7	Full information available	$Q_i = \exp\left(\tilde{\gamma}_{2i} + \lambda_G L_i + \sum_{k \in \mathcal{G}} \phi_k g_{ik}\right)$	$Q_{ij} = Q_i \dfrac{\exp(\gamma_{1j} + L_{ij} + \sum_{l \in \mathcal{H}} \theta_l a_{jl})}{\sum_{j \in \mathcal{D}} \exp(\gamma_{1j} + L_{ij} + \sum_{l \in \mathcal{H}} \theta_l a_{jl})}$	Full behavioural production and consumption models

It has been demonstrated that the CFP can be reduced to several models depending on the availability of data. The CFP model comprises the mode choice problem (MCP), the variable cargo demand problem (VDP), which consists of the cargo production problem (CPP) and the cargo distribution problem (CDP). The MCP was discussed in Chapter 5 and was shown to be governed by the cost sensitivity parameter β and the IMT capacity constraint parameters ψ_t. The parameters governing the CDP and the CPP have been discussed above and the estimation of these parameters depends on data availability. As shown above the parameters in the CFP are inter-dependent, where the evaluated value of one is required to solve the other. The modified Bregman's algorithm (A_1) used in Chapter 4 for estimating the parameters in the MCP is expanded to include the estimation of parameters in the distribution and production problems. The expanded Bregman's algorithm for solving the CFP is presented as algorithm A_4. This algorithm will also prove useful in the model application stage.

Algorithm A₄: Modified Bregman's algorithm for solving the CFP

1. Initialisation:

   For a given set of located IMTs $\mathcal{K}$ with size p and starting cost sensitivity parameter $\beta = \frac{1}{\bar{c}}$, where $\bar{c}$ can be the average transport budget and $\psi_t = 0; \forall t \in \mathcal{K}, \lambda_D = \lambda_G = 1, \gamma_3 = 0, \gamma_{2i} = 0; \forall i \in \mathcal{O}$ and $\gamma_{1j} = 0; \forall j \in \mathcal{D}$

2. Logsums Update

   2.1 Update logsums over all located IMTs ℓ_{ij} using Equation (5.28) in Chapter 5

   2.2 Update the logsums L_{ij} over all transport modes using Equation (5.30) in Chapter 5

   2.3 Estimate the parameters in the CDP depending on the choice of distribution model (CFP₁ to CFP₇)

   2.4 Update the associated logsums L_i over all destination zones using Equation (6.28)

   2.5 Estimate the parameters in the CPP depending on the choice of production model (CFP₁ to CFP₇)

3. Flows Update

   3.1 Update the quantity of cargo produced in each origin zone depending on the choice of production model (CFP₁ to CFP₇)

   3.2 Update the distribution of cargo between zones depending on the choice of production model (CFP₁ to CFP₇) together with Equation (6.25).

   3.3 Update the demand for each mode; $X_{ij}, V_{itj}, W_{istj}$ using Equations (5.25), (5.24) and (5.23) respectively in Chapter 5.

   3.4 Update the demand for the located terminals using intermodal transport demands V_{itj}, W_{istj} or Equation (5.22b) in Chapter 5.

4. Update model parameters

   4.1. Update β from equation (5.36) using Newton Raphson or Hyman's method (Hyman 1969)

5. Update capacity constraints parameters

5.1. Update capacity constraint parameters associated with the CPP; γ_3 or γ_{2i}

5.2. Update capacity constraint parameters associated with the CDP; γ_{1j}

5.3. Update the Lagrangian multipliers $\psi_t; \forall\, t \in \mathcal{K}$ for IMT capacity constraints

6. Repeat steps (2)-(5) until convergence is achieved.

6.4.4 Connecting the solutions of FLP and CFP

It has been demonstrated that once the original problem is decomposed into FLP and CFP, each sub-problem can be solved to optimality given the required inputs from the other. Algorithm A_4 for solving the CFP can be embedded in the exact algorithm (A_2) (complete enumeration) and the heuristic algorithm (A_3) (entropic greedy) developed in Chapter 5 to solve the overall problem, where in each case algorithm A_1 for solving the MCP is replaced with algorithm A_4 for solving the CFP. The running time of both algorithms (A_2 and A_4) was shown to be dominated by the running time of the algorithm A_4 for the CFP. Algorithm (A_2) guarantees an optimal solution for all instances that are computational feasible. However, it is not suitable to solve large problem instances, especially problems involving regional intermodal transport use. Algorithm (A_3) on the other hand does not guarantee an optimal solution for all problem instances but was shown to be computational very efficient for solving large problem instances. The solution quality of this algorithm was shown in Chapter 5 to be very good and comparable to that of Algorithm A_2.

6.5 Model application

Once the best locations of the IMTs are determined, the coded transport network can be updated with the new located IMTs. The revised network then goes into the application version of the CFP called the cargo flow model (CFM) for forecasting future terminal demands and testing of various policies. The operation of CFM comprises only a few steps of algorithm A_4 since at this stage the location of the required number of terminals to develop are known and comprises step 2 (only 2.1, 2.2 and 2.4 are required), step 3, and step 5.3. Step 5.3 is only required if a located terminal reached its handling capacity when forecasting or testing policies.

The operation of CFM as shown in Figure 6.1 works as follows; It first extracts transport variables from the revised network to compute the cost of intermodal transports and road direct transport. These cost variables go into the terminals choice model (TCM) where accessibilities (logsum) to intermodal transport variables are computed. These accessibility

measures go into the mode choice model (MCM) where the access to multiple modes of transport variable are computed, which in turn go into the cargo distribution model (CDM) as variables. The CDM combines the accessibility variables with other destination specific variables to construct cargo origin logsums or cargo generation power for each cargo origin and go up the tree into the cargo production model (CPM) as additional variable.

Once the CPM received the logsums from the CDM, it computes the quantity of cargo produced in each origin zone. The cargo produced are distributed to their various consumption or destination zones by the CDM. The output from the CDM go into the MCM to determine the contribution of each mode (road alone and intermodal transport) in moving cargo from their production zones to their consumption zones. The TCM then takes the demand for intermodal transport to determine the demand or usage of each located terminal. At this stage, the quantity of cargo (in TEUs) produced by each origin zone, distribution of cargo in the study area, the contribution of each mode in the transport task and the usage of each located terminal are known.

Changes in transport (road) network conditions due to the terminals can be captured by linking the CFM with transport network models, which form the supply side of transport as shown in Figure 6.2. This supply-demand loop is important for general traffic impacts assessments and to ascertain if the road network around the located terminals have enough capacity to handle the extra traffic that the terminals will bring. It will also be helpful to ascertain if the located terminals help to solve congestion or general traffic problems in the study area, especially around the port. The operation of the supply-demand loop is shown in Figure 6.2, where the outputs from the CFM is first converted into trip matrices for assignment. These matrices can be combined with non-containerised vehicle (buses, cars, etc.) trip matrices and assigned to the road network. For example, the truck legs of intermodal transport chain can be combined with trip matrices from road alone transport to form the containerised trip matrices. The non-containerised trip matrices can be assumed to come from the existing transport model of the study area with suitable traffic assignment models (Bliemer et al. 2017; Bliemer et al. 2014; Bell 1995).

The combined matrices form the demand side of transport and are made to interact with the supply side of transport by assigning them to the transport network. The assignments may alter the conditions of the road network, for example in terms of changes in the time it takes by

each mode to move cargo from one location to another on the network. Changes in network conditions can be feedback into the CFM to produce a new set of matrices. The model is therefore, iterative and is iterated until the supply (assignment)-demand (matrices) equilibrium (convergence) is achieved. The network changes going into the CFM could induce changes in the usage of the located terminals, the share of each mode in the transport tasks and the production and distributions of cargo in the study area.
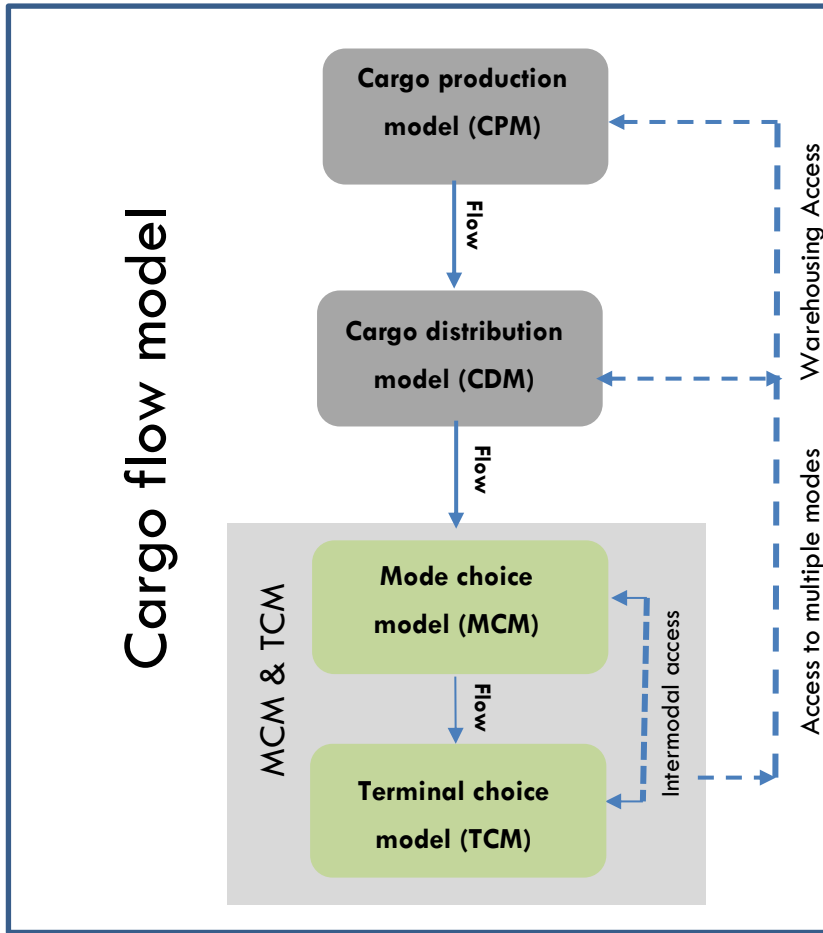
Figure 6.1: Cargo flow model (CFM) architecture
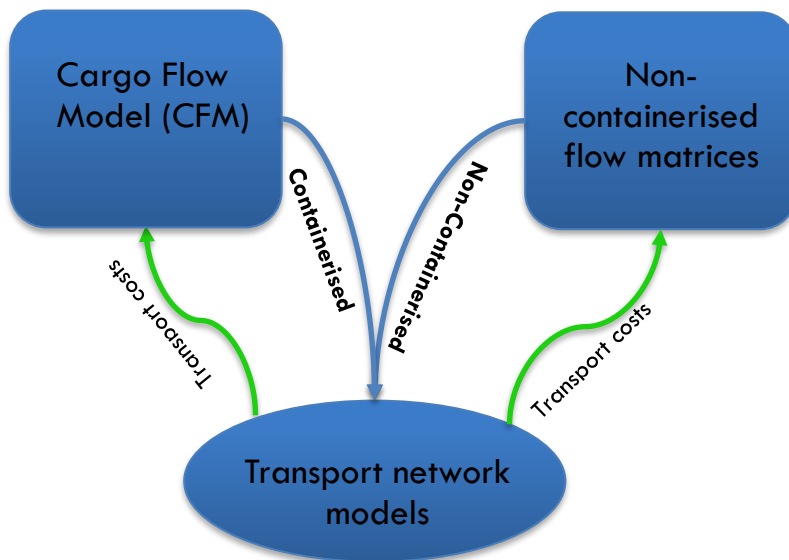
Figure 6.2: Demand-supply loop

6.6 Conclusions

This chapter deals with incorporating variable cargo demand in the entropy framework. Two methods were proposed for developing these models as constraints within the entropy framework; Poison method and weighted mean method. It was shown that the Poison method is a special case of the weighted mean method. For each method, two linked models were developed and incorporated; the cargo production model and cargo distribution model. Based on data availability, several special cases of the combined model (production and distribution models) were also investigated and how each can be solved to optimality were discussed. The overall problem was solved by adapting the complete enumeration algorithm for small instances or the heuristic algorithm for larger problem instances. Finally, the study provided practical ways of implementing the models in practice.

Chapter 7 Case study

7.1 Background

The primary objective of this chapter is to apply the models developed in the previous chapters to a real transport network and actual freight demand in the study area. The study area is the Sydney Greater Metropolitan Area (GMA), Australia. General geographical and demographic characteristics of the Sydney GMA and its transport network are provided together with data used for the modelling exercise. The data were used to build the entropy maximising facility location model suitable for appraising the viability of terminals in Sydney GMA and quantify shippers' responses to the located intermodal terminal(s). Once the best location(s) of the required number of terminals are determined, they are then coded in the transport network for forecasting and testing of various policies using the forecasting version of the model described in Chapter 6. The model can be used to examine the efficacy of alternative intermodal terminals in Sydney GMA and the sensitivity of the results to alternative assumptions about terminal user fee, rail and road transport costs and other drivers of cargo distribution and terminal usage.

The rest of the chapter is organised follows; Section 7.2 presents the detail description of the study area followed by analysis of relevant data. The methodological framework and construction of model variables are presented in Section 7.3. Analysis of results including locations analysis, sensitivity testing and policy testing are presented in Section 7.4. Finally, the conclusions are presented in Section 7.5.

7.2 Data

Data collected for this empirical exercise came from several sources and includes import containerised data and their distribution in the Sydney GMA, congested road network data, cost data, candidate intermodal container terminals and land use data. Section 7.2.1 provides a detail description of the study area. Analysis of import container and their distribution in study area together with candidate intermodal terminal locations and features are presented in Section

7.2.2. Section 7.2.3 discusses the transport network data required for the construction of cost variables and analysis of land use data required for the production and distributions models is presented in Section 7.2.4.

7.2.1 Sydney greater metropolitan area

The study area for the case study is defined to cover the whole of Sydney GMA in Australia (see Figure 7.1). The study area is divided into freight analysis zones where cargo can be seen as coming from one zone and destined to another zone as shown in Figure 7.2. The zones are connected together with a computer description of the existing road and rail networks at the level of detail appropriate to the zone system. The transport network constitutes the supply of transport and includes the geographical locations of the intermodal terminals while the movements of cargo between the zones and the type of modes used in the transport tasks constitute the demand for transport. The interactions between demand and supply forms the economic foundation of the proposed entropy maximising facility location model, described in detail in Section 7.3.2.

The study area was divided into 80 cargo destination zones in addition to the special zones designed for candidate IMT locations. These zones correspond to the Statistical local area (SLA), which is an Australian Standard Geographical Classification (ASGC) defined area and consists of one or more Collection Districts (CDs). SLAs cover, in aggregate, the whole of Australia without gaps or overlaps (ABS 2001). The CDs are the basic building block of ASGC and designed for use in the Census of Population and Housing as the smallest unit for collection, processing and output of data (ABS 2001). The zone system also corresponds to the zone system for the MetroScan-TI model (see Section 7.2.2), which provided transport network data in terms of congested travel times and distances by mode in the study area. Figure 7.3 explores the relationship between population size (Australian Bureau of Statistics) and quantity of cargo destined to each SLA. As would be expected areas with large population generally attract less cargo than those with small population. The added advantage of adopting this zone system is that released public data such as data on population, job, and other business and industry data relevant for modelling do not have to be altered, effectively, eliminating errors associated aggregation and disaggregation of these type of data. It also makes it easier to interpret model results as they can directly be linked to known or tangible features in the study area.
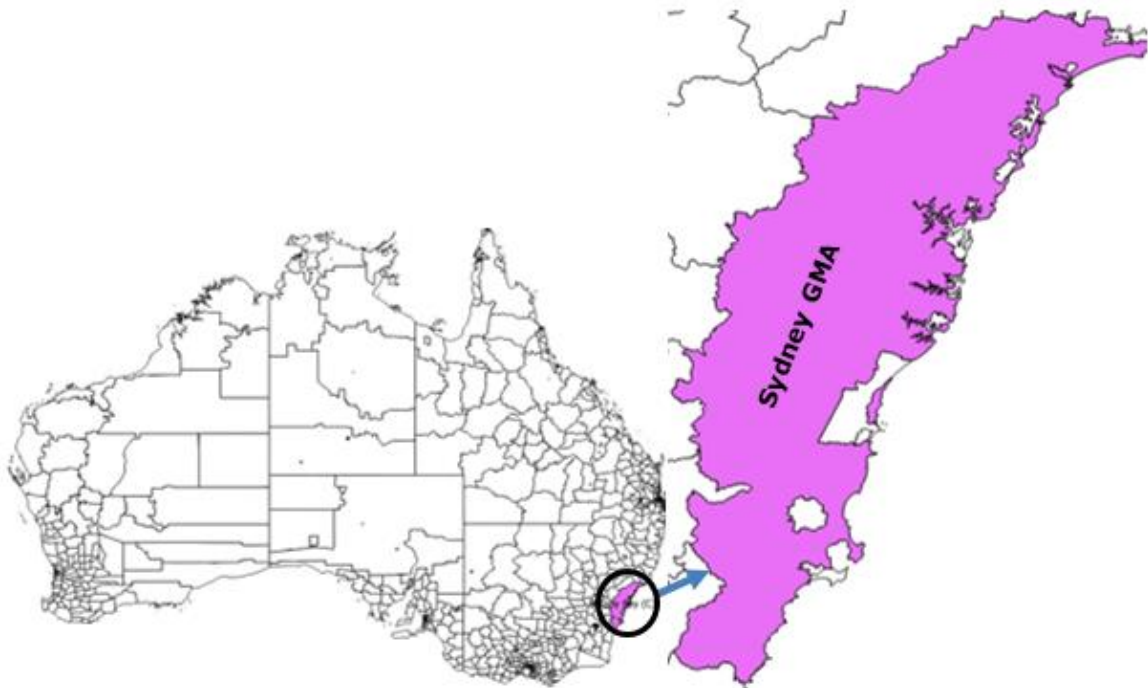
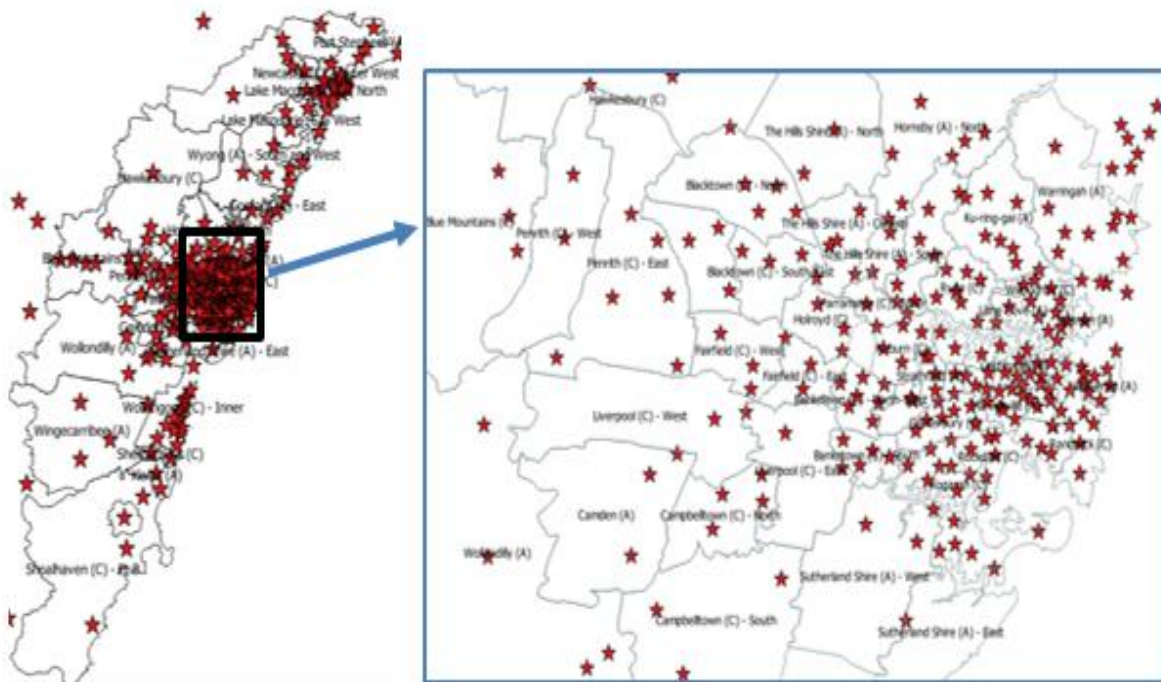**Figure 7.1: Geographical location of Sydney GMA within Australia**

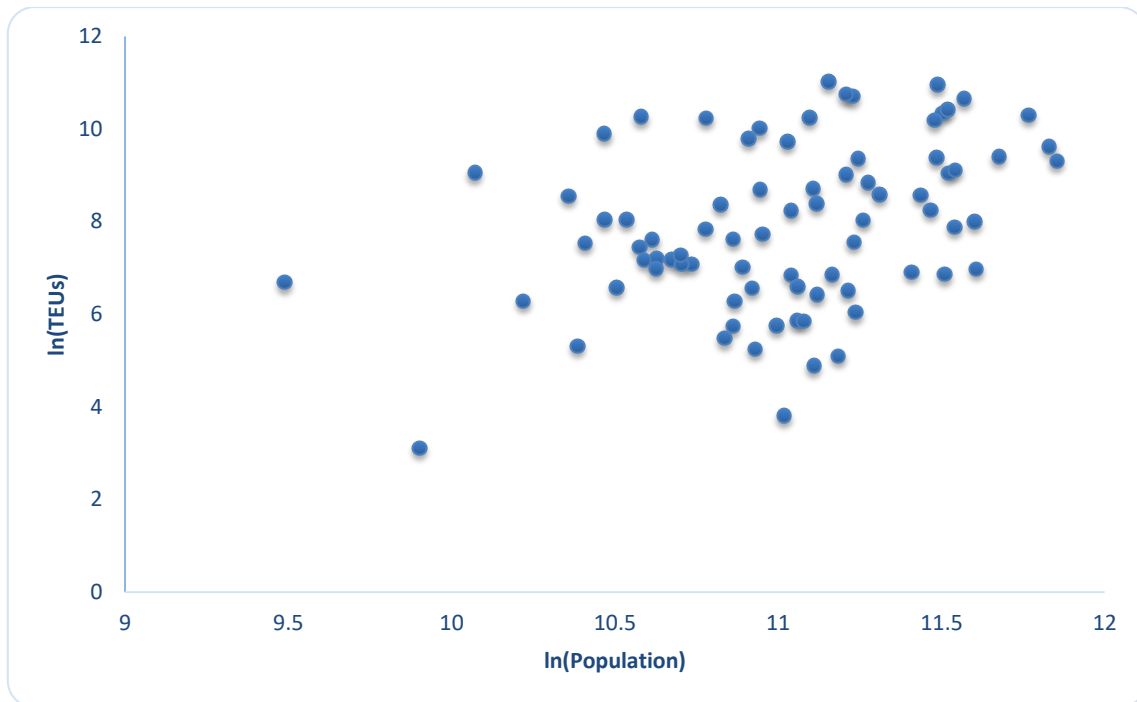**Figure 7.2 : Cargo destinations zones and delivery postcodes**

Figure 7.3 : Log-log relationship between cargo destination and population size

7.2.2 Import container flows and candidate IMTs

This section presents sources and descriptive statistics of the main data used for the modelling exercise; import containers and their distribution in the study area and candidate intermodal terminals and features. The data on import containers and their delivery postcodes within the study area were obtained from Australian Bureau of Statistics (ABS 2011). As shown in Figure 7.4 a total of about 0.83 million TEUs of cargo were imported in 2009-2010 out of which about 0.73 million TEUs representing about 88% were destined to the Sydney GMA (the study area). The distribution of the imported cargo to their reported postcodes are shown in Figure 7.2, where most of the cargo are concentrated within a 50kilometer radius of the port. The concentration of flows has contributed to the current congestion and associated problems especially around the seaport as most of the imported cargo (about 86%) are transported by road (Shipping Australia 2011). The brighter side of this concentration of cargo is that it makes intermodal transport option more promising as flows can be consolidated at a terminal away from congested areas and then distributed locally. The location of the terminal provides the opportunity to use more sustainable modes like rail for the movements of cargo between the port and the terminal.
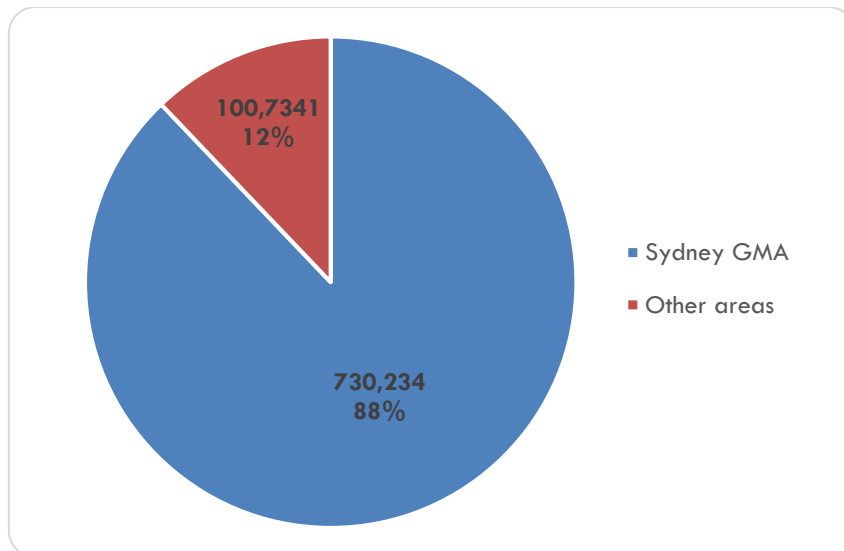
Figure 7.4 : Share of imported cargo to the study area

For the modelling exercise, the distribution of cargo are required at the study zone level rather than at the postcode level. Using the QGIS[5] tool each postcode was assigned to a zone, and the cargo flows aggregated to the zonal level. The geographical distribution of the aggregated flows at the zonal level and key cargo destinations are shown in Figures 7.2 and 7.5. The size of the ball in Figure 7.5 corresponds to the quantity of cargo assigned to that zone. For example, Blacktown comprises three study zones and together contributed about 13% of the total imported cargo destined to the study area, followed by Fairfield with a 10% share. The candidate IMTs with their fixed cost and handling capacities in Table 7.1 were derived from the national intermodal study, Australia (Meyrick 2006) and can also be found in (Piyatrapoom et al. 2006). The geographical locations of the 9 candidate IMTs and given names are shown in Figure 7.6 and were recommended by a national study (Meyrick 2006). The choice of the locations was influenced by their proximity to the rail and major road networks and in this case study, all candidate terminals are assumed to have access to the rail network. The national study reported a total annual cost (including capital and operations costs) of $5,053,688 for a terminal with a handling capacity of 150,000 TEU per annum. This equates to about $34 per TEU. For simplicity, all candidate terminals were assumed to have these features.

---

[5]QGIS is a free and open source geographical information system

Table 7.1: IMT features

ID	Names	Capital and operational cost ($ per year)	Handling capacity (TEUs per year)
1	Enfield	5,053,688	150,000
2	Yennora	5,053,688	150,000
3	Camellia	5,053,688	150,000
4	Eastern Creek	5,053,688	150,000
5	Moorebank	5,053,688	150,000
6	Ingleburn	5,053,688	150,000
7	Minto	5,053,688	150,000
8	Villawood	5,053,688	150,000
9	Chullora	5,053,688	150,000
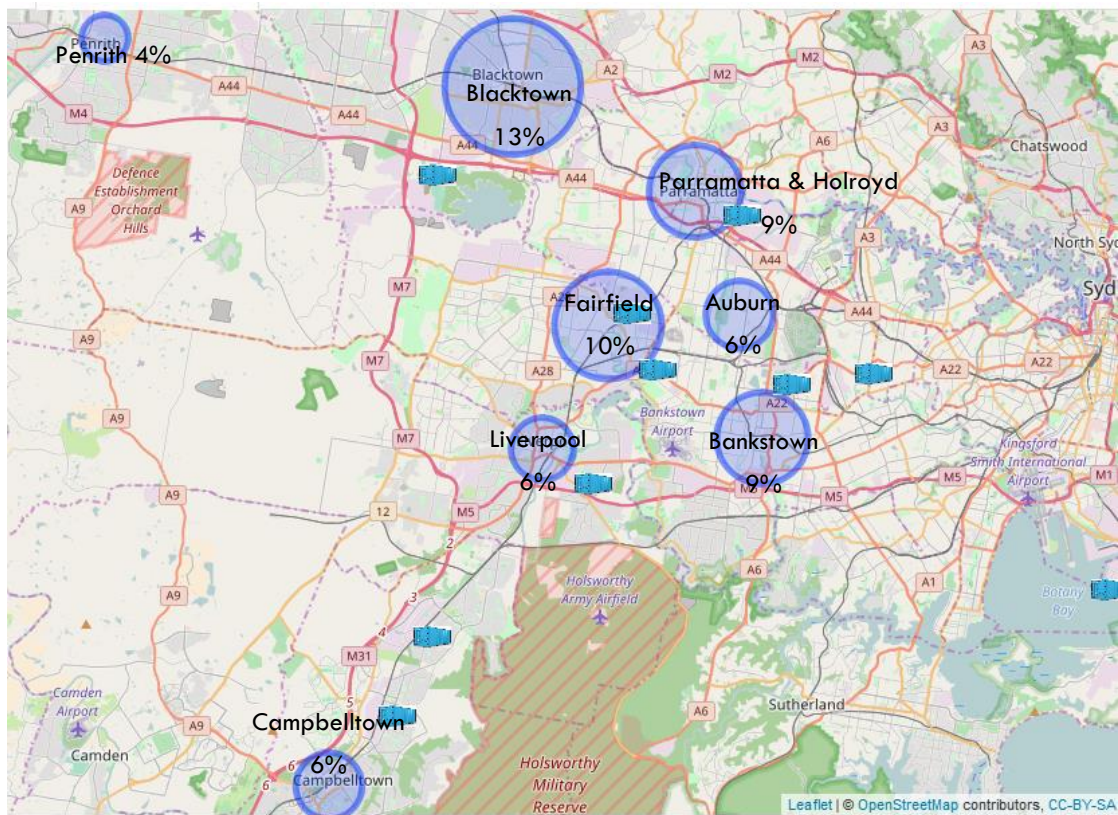
Figure 7.5 : Key cargo destinations
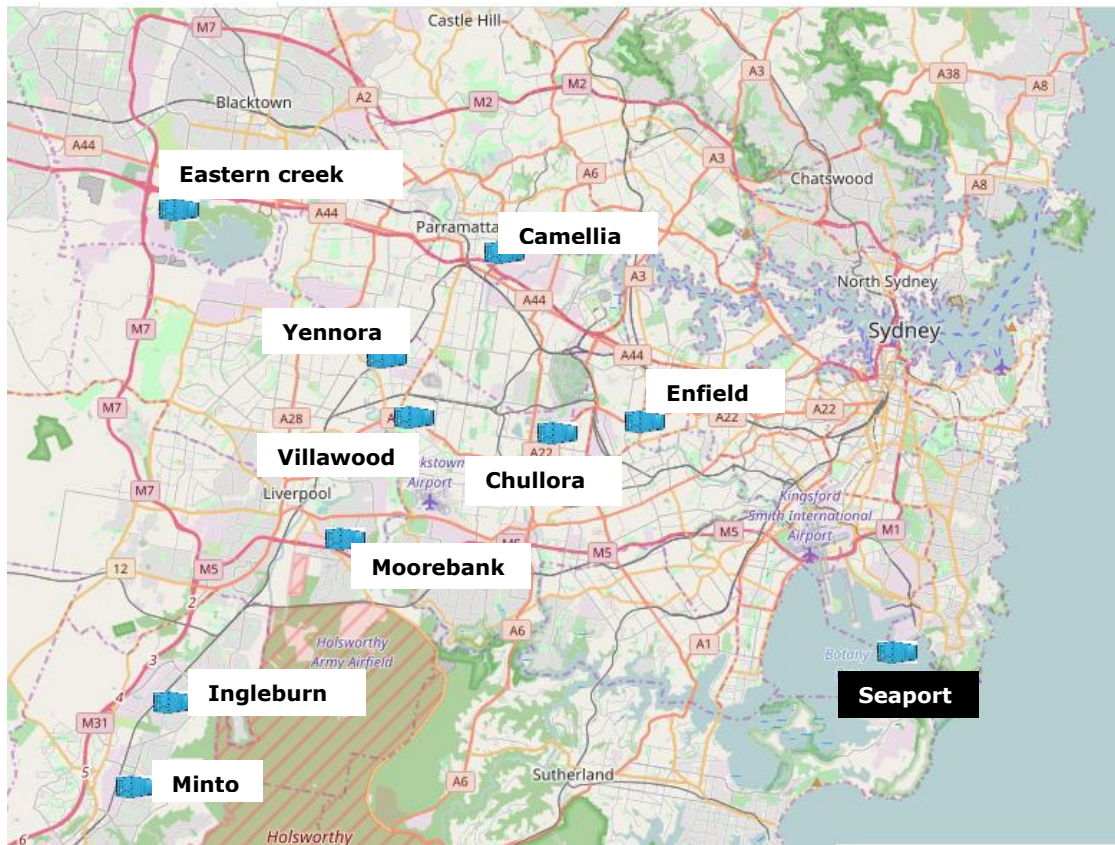
Figure 7.6 : Geographical locations of candidate IMTs and names

7.2.3 Transport cost variables

Transport variables such as travel times and distances between the study zones were obtained from the MetroScan-TI model. MetroScan-TI is a land use, transport and environment strategy and metropolitan infrastructure scanning tool that encompasses all modes of transport and is developed by ITLS, University of Sydney (Ellison et al. 2017). To the best knowledge of the author, it is the only model in the study area that allows the interactions between passenger modes and freights modes on the road network. This makes it suitable for providing reliable transport network variables such as congested travel times and distances, which are in turn used to construct the cost of container movements from the port to the various study zones and also between candidate intermodal terminals and final container destinations. Figure 7.7 reveals the inverse relationship between truck travel times extracted from the MetroScan-TI model and cargo flows (TEUs). As expected, on average the farther a freight analysis zone is from the port the less cargo it attracts and vice versa. The variable ln(TEUs) on the y-axis of Figure 7.7 represents the natural logarithm of the total quantity of cargo (TEUs) destined to each zone.
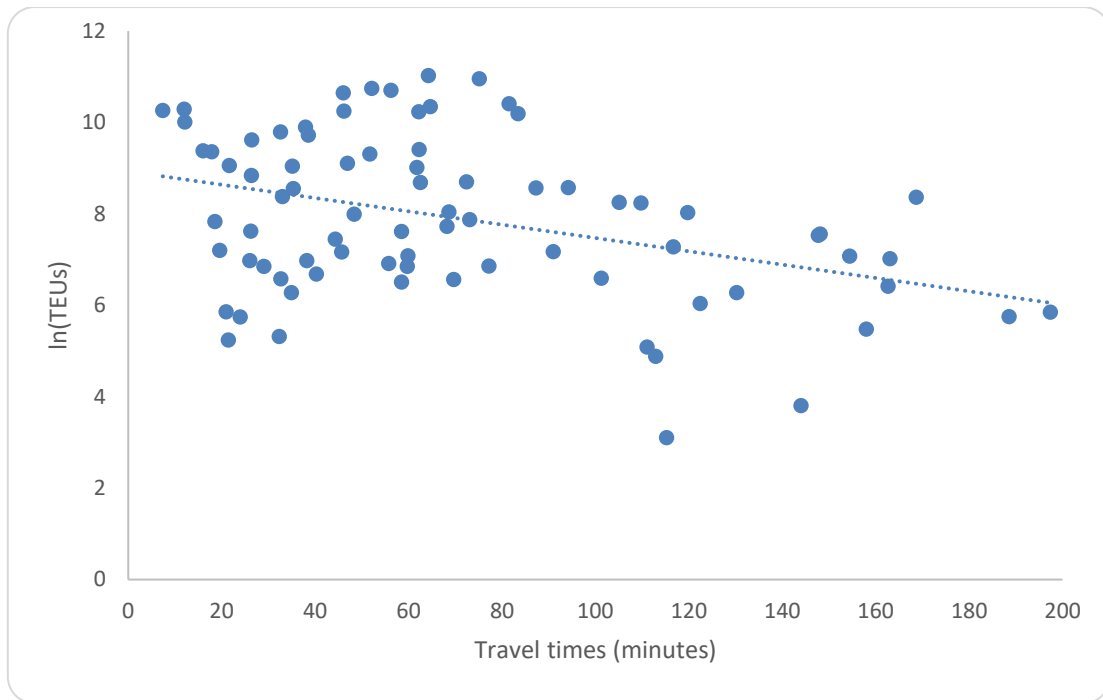
Figure 7.7 : Relationship between Cargo destinations and Travel time

7.2.4 Land use data

Land use data at the zonal level includes data on employment by industry and occupation from the Australian Bureau of Statistics (ABS) business counts by both employment and revenue, and land-use data derived from ABS meshblock data. All these data are available at the SLA-level required by the model and so need no further aggregation or disaggregation. Several land use data were analysed during the study but two emerged very strongly with respect to explaining the distribution of cargo in the study area; number of employees in manufacturing jobs and the number of employees in warehousing and storage activities. The strong relationships between these variables and cargo attraction (in TEUS) to each zone are shown in Figures 7.8. and 7.9. Figures 7.8 and 7.9 reveal log-log relationship between cargo destined to a zone and the number of employees in manufacturing, and warehousing and storage in that zone respectively. In general, the figures show that areas with high manufacturing and/or warehousing and storage jobs attracts more imported cargo and vice versa.
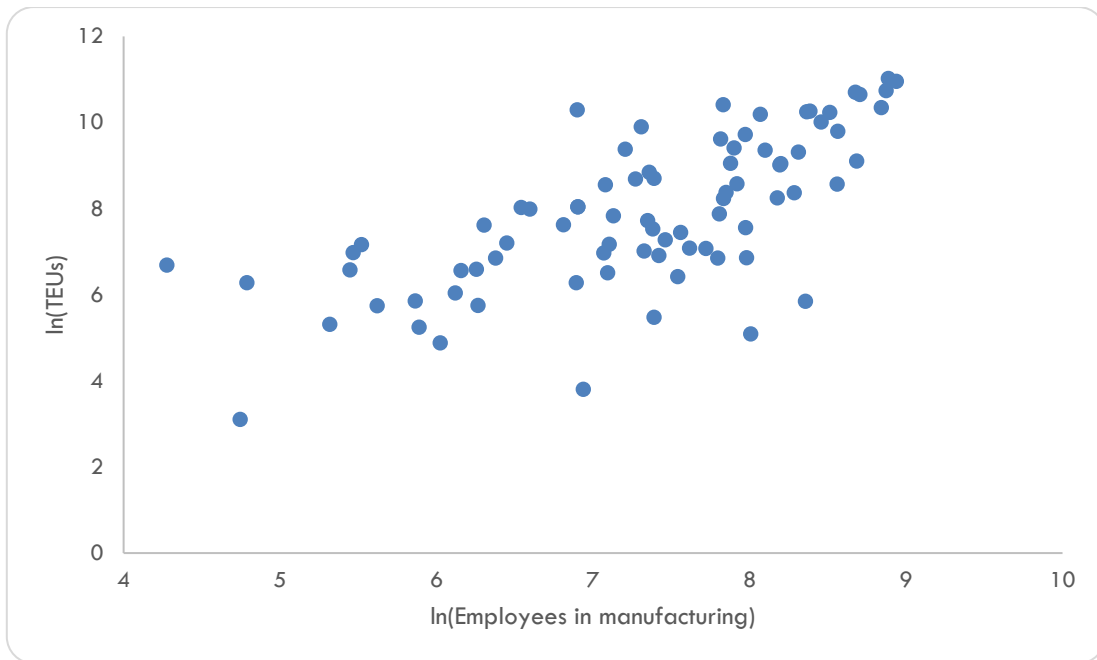
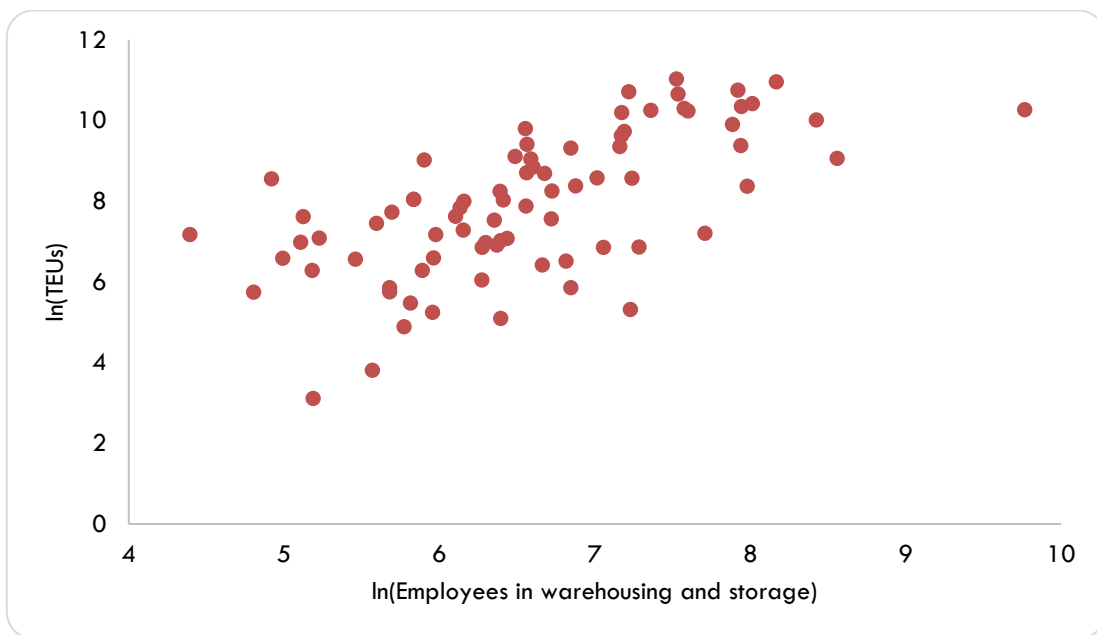**Figure 7.8 : Cargo attraction and access to manufacturing activities**

Figure 7.9 : Cargo attraction and access to warehousing and storage activities

7.2.5 Cost of container movements in Sydney GMA

The monetary costs of moving containers from the port to sample cargo destinations in Sydney GMA by truck were available in 2001 and shown in Table 7.2 below. This data provides a vital relationship between the cost of container movements by truck and network travel times and/or distances. These data were obtained from Access economics (2003) and available online.

Three separate regression models were investigated to gauge the best relationship between the cost of truck movements and transport variables. The transport variables investigated are travel time (min), distance (km) and generalised time (min). The generalised time (min) is a linear combination of time and distance expressed in time units, where the distance is converted into time unit by multiplying it by the ratio of vehicle operating cost ($ per km) and driver's value of time savings ($ per min). The three models are presented in Table 7.3, where model 2 (using time as transport variable) emerged as the statistical favourite. The results support the notion that urban congestion (expressed in travel times) is one of the main drivers of intermodal transport use. Thus, the decision on which mode of transport to use depend, in part, on how congested the transport system is and where the congested points are. The estimated parameters in model 2 with a fixed cost of $163.52 per TEU and a variable cost of $1.84 per TEU per minute were therefore carried forward in computing the monetary of truck movement between any points on the network. Sydney freight council (SFC 2007) reported the cost of rail to Sydney GMA to be between $80-$100. Similar range of values were reported in IPART (2007) and Shipping Australia (2011). The cost of $100 was carried forward and used in the modelling exercise and was assumed to be the cost of rail to each candidate IMT followed by some sensitivity analysis about this value during forecasting and policy testing.

Table 7.2: Cost of cargo movements by trucks (2001-2002)

Sample destinations	Distance (km, 1-way)	Travel time (min, 1-way)	Fare ($ per TEU)
Botany	1	2	160
City and Eastern Suburbs	11	20	200
South Sydney	15	22	200
Southern Suburbs	15	22	200
Inner West	21	32	220
Liverpool	32	40	230
South West	53	47	250
Central West	34	45	250
Industrial West	43	46	250
Blacktown	46	44	250
North Shore	30	42	250
Penrith	63	62	280
NW Sydney	37	59	280
Wollongong PO	84	84	320
Newcastle PO	172	154	440

Table 7.3: Estimation of truck cost parameters

Variables	Model 1 Coefficient	t-stats	Model 2 Coefficient	t stat	Model 3 Coefficient	t stat
Fixed cost ($)	184.97		163.52		169.60	
t-stats		31		67		49
Distance (Km)	1.53					
t-stats		15				
Travel time (min)			1.84			
t-stats				44		
Generalised time (min)					1.28	
t-stats						30
Parameters	2		2		2	
No of observations	13		13		13	
R Square	94.53%		99.34%		98.56%	
Adjusted R Square	94.11%		99.29%		98.45%	

7.3 Methodology

7.3.1 Background

The various data assembled for the modelling exercise came from various sources and collected or derived at different time periods. The import cargo distribution data was collected in 2009/10, land use data were available in 2011, the transport network data such as travel time and distances were obtained from the 2011 base year MetroScan-TI model. The parameters governing the cost of truck and rail modes were available in 2001 and 2006 respectively. The study adopted 2011 as the modelled year, the unit of analysis is TEU and the modelling time horizon is one year. The one year modelled period was adopted to fully capture seasonality and other variations in cargo flows over the year. This required all data to be converted to 2011. The following assumptions were used to do the conversions:

1. The import cargo flows in 2009/10 were converted to 2011 using total import cargo in 2011 and assuming that the proportion or distribution of cargo to various destinations in Sydney GMA in 2009/2010 is unchanged in 2011. Thus, if 13% of imported cargo are destined to say Blacktown in 2009/10, we expect Blacktown to capture the same 13% in 2011. The data for 2011 import cargo were obtained from ABS (2011).

2. The fixed and variable cost of using truck (see Table 7.3) were adjusted by a factor of 1.33 (price adjustment from 2001 to 2011). This factor was accessed from the website of the Reserve Bank of Australia (www.rba.gov.au/calculator).

3. The $100 cost of rail in 2006 was also converted to 2011 by a scale factor of 1.16 also derived from the Reserve Bank of Australia.

4. The other data sets such as land use and transport network variables were already in 2011 and needed no conversion.

7.3.2 Model structure and assumptions

The structure of the model adopted for the empirical exercise is a reduced version of the general structure discussed in Section 6.5. In particular, only the metropolitan containerised market is considered (see Chapter 4) and the use of import cargo data obviate the need for the cargo production model at least for the base year. The two transport modes available for the transport tasks in this market are the road alone and metropolitan intermodal transport (see Chapter 4). The overall problem comprises linked facility location problem for determining the best locations of terminals in the Sydney GMA, and the CFP which determines the usage of the located terminals. Here, the CFP comprises the variable cargo demand problem (VDP) and the mode choice problem (MCP). The solution to the CFP called the cargo flow model (CFM) is converted into a three-level nested logit model of cargo distribution model (CDM), the mode choice model (MCM) and the terminal choice model (TCM) as shown in Figure 7.10 and carried forward in forecasting and testing of various policies to promote the use of the located terminals. Additionally, the forecasting model (CFM) is not linked to the transport network so there is no looping between demand and supply. The implementation of the demand-supply was considered outside the scope of this study and for this study, only one terminal is required to be located in the Sydney GMA.
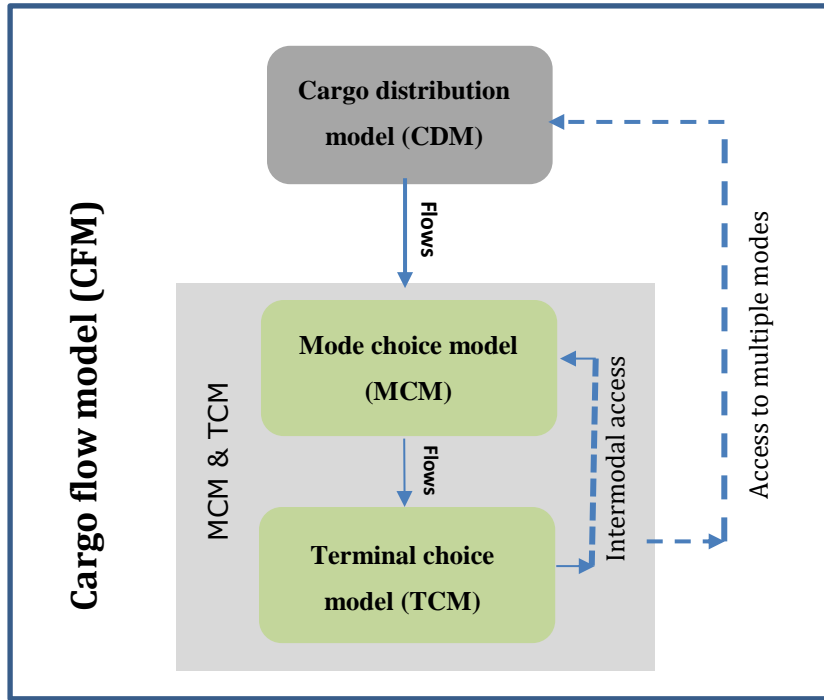
Figure 7.10 : Forecasting model architecture

7.3.3 Factors governing the mode choice model

The choice of mode is governed by the generalised cost incurred in the use of the mode and the transport budget. The cost of using truck between any two points on the network is expressed as:

$$c_{ij} = c_0 + \rho t_{ij}; \; \forall \; i \in \mathcal{O}; j \in \mathcal{D} \tag{7.1}$$

where c_0 is the estimated fixed cost, which in 2011 prices is \$217.54 per TEU, and ρ is the variables cost, estimated to be \$2.50 per TEU per minute and t_{ij} is the truck travel time between locations i and j on the network. Similarly, the total cost of intermodal transport is

$$c_{itj} = c_{it} + c_t + c_{tj}; \; \forall \; i \in \mathcal{O}; j \in \mathcal{D}, t \in \mathcal{T} \tag{7.2}$$

c_{itj} is the cost of rail and was fixed at \$116.00 at 2011 prices, c_t is the transfer or terminal user fee and since values of this variable are the same across candidate terminals it was set to zero during the location analysis. The optimal value to charge to attain the highest revenue was later

determined through optimal fee analysis. c_{tj} is the truck cost for final delivery and takes the form of Equation (7.1).

In computing the transport budget, the study assumes that shippers individually would not choose to increase their transport costs because a new IMT becomes available, so they would not do so collectively. It seems reasonable to therefore use the existing average transport cost of the study area (cost of road alone transport) to compute the budget. The total transport cost in the study area in absence of intermodal transport can be computed using:

$$c = \sum_{i \in O} \sum_{j \in D} c_{ij} q_{ij} \tag{7.3}$$

7.3.4 Factors governing the distribution of import cargo

The cargo distribution model is explained by four main variables: natural logarithm of the number of employees (labourers) in manufacturing as a proxy to access to manufacturing businesses and agglomeration; the natural logarithm of number of people employed in warehousing and storage industry, which is expected to quantify the benefits of performing warehousing activities at the located IMTs. The third variable (accessibility) captures access to key markets which in this study were identified as zones in the two main central business districts (Sydney and Parramatta) with set Ξ representing the collection of zones in these markets. The fourth variable is the logsum from the mode choice model as a measure of access to multiple modes of transport and composed of both road and intermodal accessibility measures. The access to key markets variables (third variable) were constructed as:

$$Access_j = \ln\left(\sum_{m \in \Xi} e^{-\beta c_{jm}}\right)$$

where c_{jm} is the cost (\$ per TEU) of using road alone transport between the cargo destination zone j and a key market zone $m \in \Xi$.

7.4 Analysis of Results

7.4.1 Estimated parameters

The analysis first considered the sensitivity of modal decisions to changes in the generalised costs of each mode. The estimated cost sensitivity parameters for the location of each candidate IMT in turn are shown Table 7.4 and Figure 7.11. The figure reveals different degrees of sensitivities depending on the location choice of each IMT with Eastern creek emerging as the location least sensitive to changes in generalised cost and Moorebank as the most sensitive location with cost sensitivity parameter of 0.2009. An important outcome for this result is that even within the same study area (Sydney GMA) the cost of freight transport could be valued differently depending on the locations of terminals. The corresponding estimated parameters for the distribution model are presented in Table 7.4. As expected the weight or importance of the two accessibility variables (access to key markets and modal access) increases with decreasing sensitivity to modal costs. This inverse relationship makes Eastern Creek the location with the greatest access to the key cargo markets (Sydney and Paramatta) and provides the greatest modal access to cargo destinations. Eastern Creek also has the greatest access to manufacturing activities. The location of IMT at Eastern Creek makes cargo distribution most sensitive to jobs in manufacturing and least sensitive to jobs in warehousing and storage compared with other candidate locations. For example, if Eastern Creek IMT is opened, a 1% increase in the number of manufacturing jobs (or warehousing and storage jobs) in a zone will increase the quantity of cargo attracted to that zone by about 0.91% (or 0.33%) compared with 0.89% (or 0.38%) increase if Moorebank was opened instead. Similar analysis can be made for the other candidate IMT locations.

Table 7.4: Estimated factors governing mode and cargo distribution (ordered by decreasing entropy)

| ID	IMTs	Entropy	Demand (TEUs)	Cost sensitivity (Beta)	Access to manufacturing	Access to warehousing and storage	Access to key markets	Modal access
4	Eastern Creek	3,983,884	170,857	0.0206	0.9079	0.3305	0.4116	0.1003
3	Camellia	3,006,532	159,981	0.0425	0.8902	0.3507	0.2521	0.0186
6	Ingleburn	1,538,224	51,862	0.0434	0.8905	0.3508	0.2484	0.0176
7	Minto	1,487,379	50,454	0.0439	0.8903	0.3511	0.2459	0.0170
1	Enfield	136,763	1,287	0.1125	0.8876	0.3687	0.1028	0.0021
9	Chullora	250,046	2,102	0.1223	0.8878	0.3699	0.0945	0.0018
8	Villawood	594,263	10,419	0.1824	0.8878	0.3747	0.0635	0.0011
2	Yennora	986,043	61,337	0.1872	0.8879	0.3750	0.0618	0.0011
5	Moorebank	228,098	7,660	0.2009	0.8877	0.3757	0.0577	0.0010
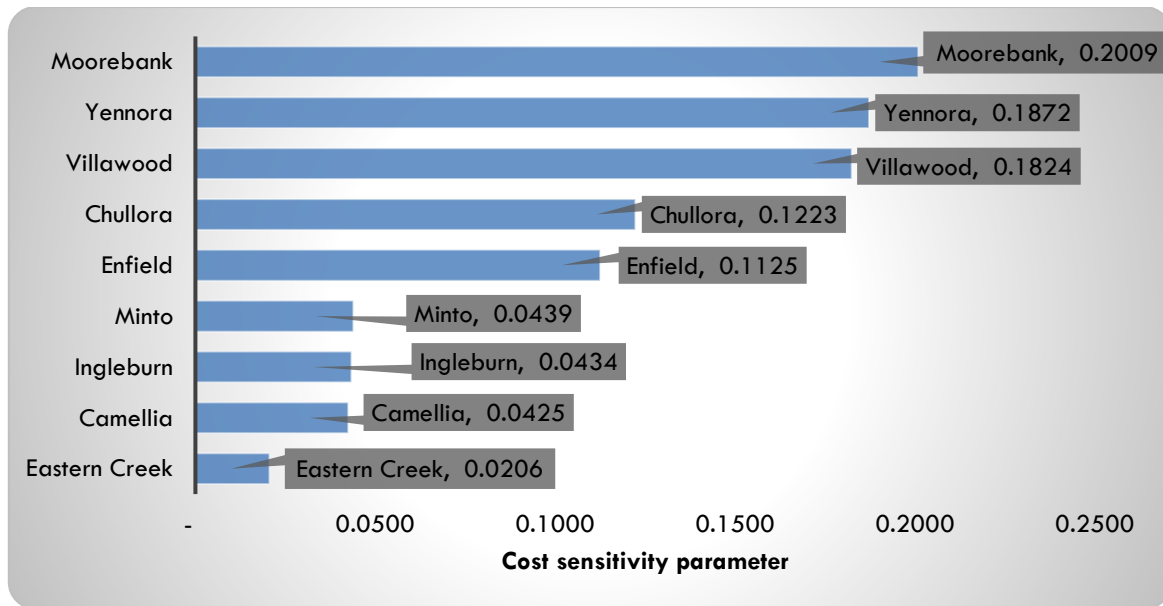
Figure 7.11 : Cost sensitivity parameter associated with each candidate IMT

7.4.2 Selecting the best IMT location

The previous section identified the key factors governing the mode choice and distribution model and how the location decisions of intermodal terminals influence the magnitude or impacts of these factors on modal decisions and cargo distributions. This section finds the most promising location among the candidate IMT locations. The selection criterion is maximum entropy which based on Proposition 4.7, the IMT with the maximum entropy is also the location which provides the largest consumer surplus or welfares for users. The results for the location of one IMT is presented in Table 7.4 and Figure 7.12. The table shows that the location that generates the largest consumer surplus also attracted the largest cargo demand and this location is Eastern Creek. Camellia is also very promising and closely follows Eastern Creek both in terms of consumer surplus and demand.

The distribution of demands of each candidate IMT location in shown in Figure 7.12 with Enfield being the least promising location. Figures 7.13 and 7.14 present the key areas in the study area where the demands for the two most promising IMTs (Eastern Creek and Camellia) are greatest. Figure 7.13 shows Eastern Creek obtaining almost half of its demand from cargo destinations in Blacktown (the largest cargo destination), Fairfield contributing about 14% and Penrith, and Paramatta and Holroyd contributing 12% each. The cargo destination zones within these four areas contribute about 87% of Eastern Creek's demand or market for intermodal transport. Camellia attracts demand from more areas than Eastern Creek,

although it also has Blacktown as its biggest market. Paramatta and Holroyd also contributes significantly (about 35%) to Camellia's demand. The other notable areas of demand are Fairfield (9%), Penrith (9%) and Auburn (5%). These five areas contribute about 95% of Camellia's demand. Thus, Blacktown, Fairfield, Penrith and Paramatta and Holroyd could be considered as the main catchment areas for Eastern Creek IMT whilst Blacktown, Fairfield, Penrith, Paramatta and Holroyd and Auburn are the catchment areas for Camelia IMT.
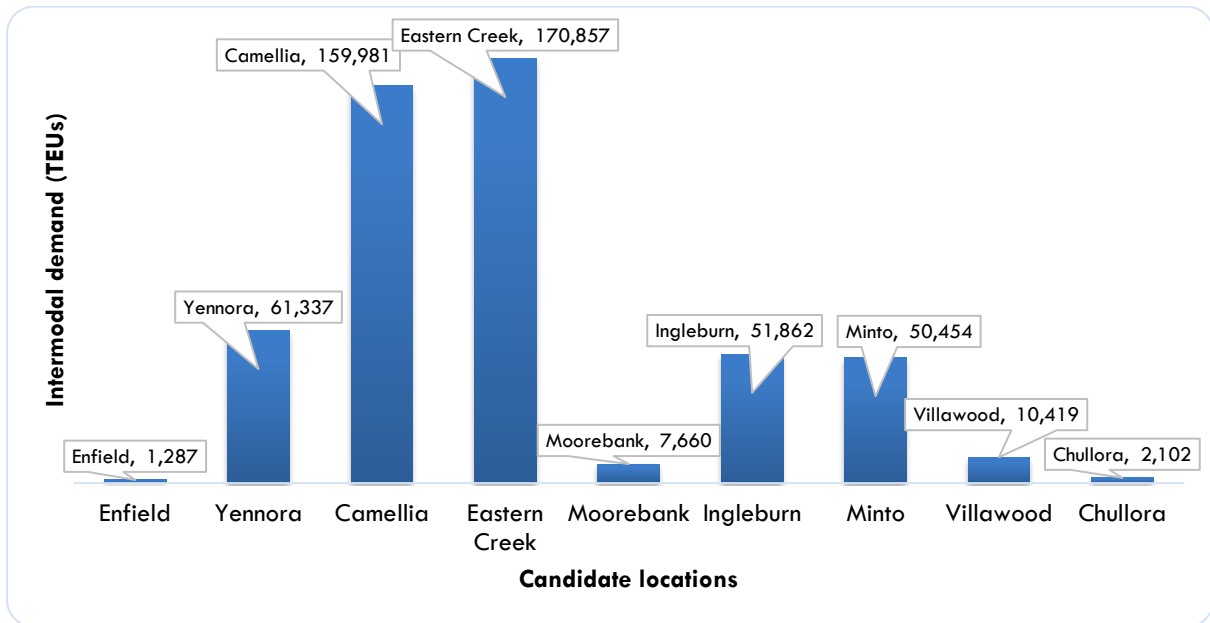
Figure 7.12 : Estimated demand for each candidate IMT location
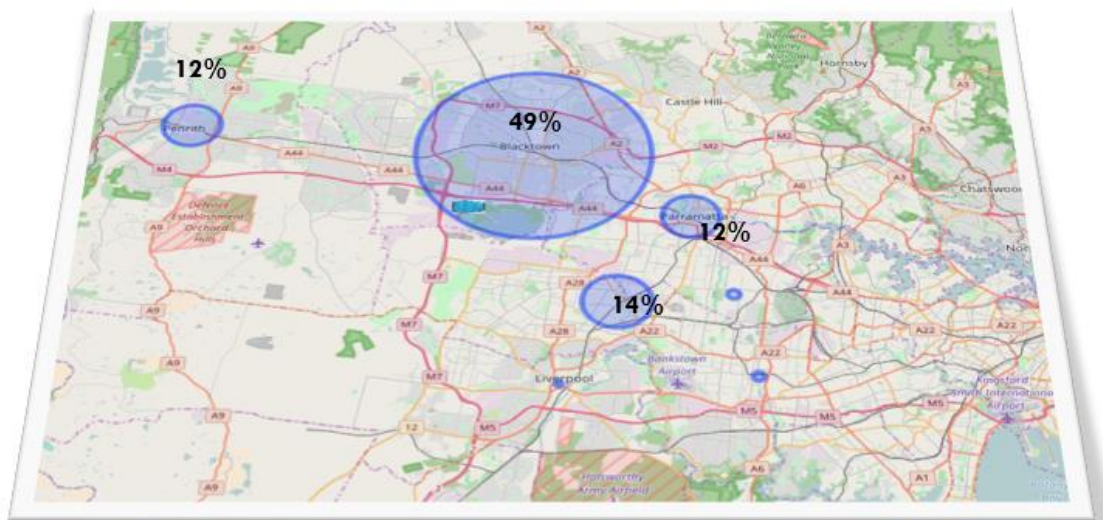
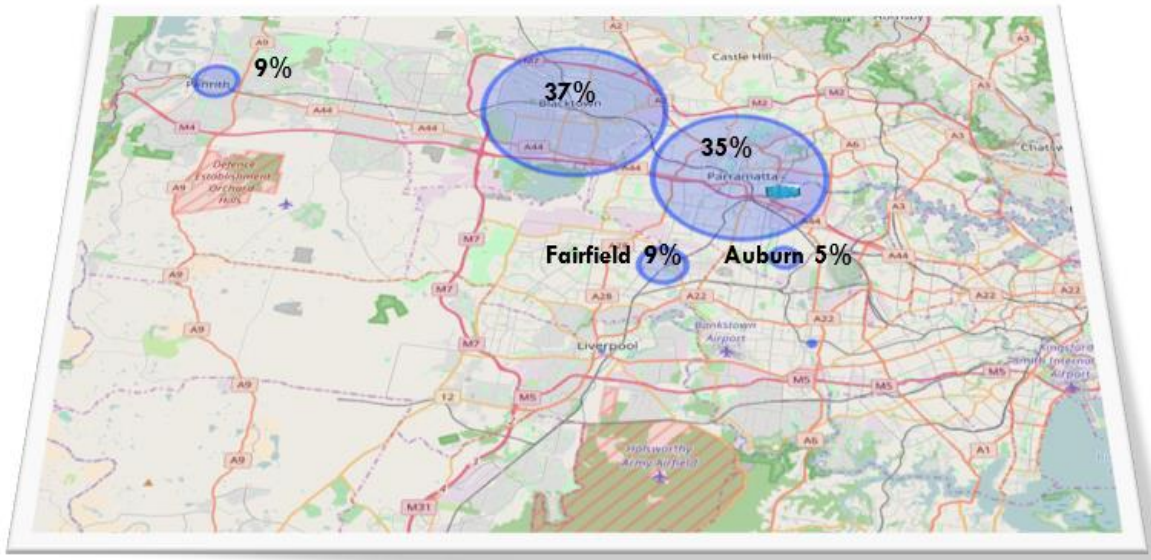**Figure 7.13 : Key Market for located (Eastern Creek) IMT**

Figure 7.14 : Key markets for Camellia IMT

The analysis above shows Eastern Creek and Camellia competing for almost the same markets. To test this, two IMTs were simultaneously located (*p*=2) with Eastern Creek and Camellia emerging as the best locations, having the largest entropy and hence the largest consumer surplus and demand. The contribution of each key market in the demand for the two IMTs are shown in Figure 7.15. The Figure shows Eastern Creek capturing about 79% of intermodal transport demand to Blacktown, 61% to Penrith, 56% to Campbelltown and 43% to Fairfield. Camellia on the other hand takes the bigger share of intermodal transport demand to Parramatta and Holroyd (86%), Auburn (91%), Bankstown (91%), Liverpool (71%) and Fairfield (57%). The location of the two IMTs increased the overall intermodal transport demand from 174,408 TEUs (Eastern Creek) to 255,756 TEUs, about 48% increase in intermodal transport demand and representing about 28% of the total cargo demand in the study area. Camellia contributes about 51% of the total intermodal transport demand and the rest (41%) by Eastern Creek. Although the results show some level of competition between these two IMTs, they can be developed simultaneously with minimum market saturation and also the fact that the overall demand for intermodal transport improved significantly from 19% to about 28%.

Similar analysis was carried out for the simultaneous location of 3 and 4 IMTs with the market shares of each optimal IMT shown in Figures 7.16 and 7.17 respectively. All the results show Eastern Creek as consistently captive to the Blacktown and Penrith markets, Camellia controls the intermodal market to Auburn, Parramatta and Holroyd. For the 3 located IMTs

(*p=3*), Ingleburn IMT controls the markets in Campbelltown and Liverpool, whilst for the four (*p=4*) located IMTs, Villawood has taken over the markets in Liverpool, Bankstown and Fairfield but Ingleburn still controls the intermodal market in Campbelltown.
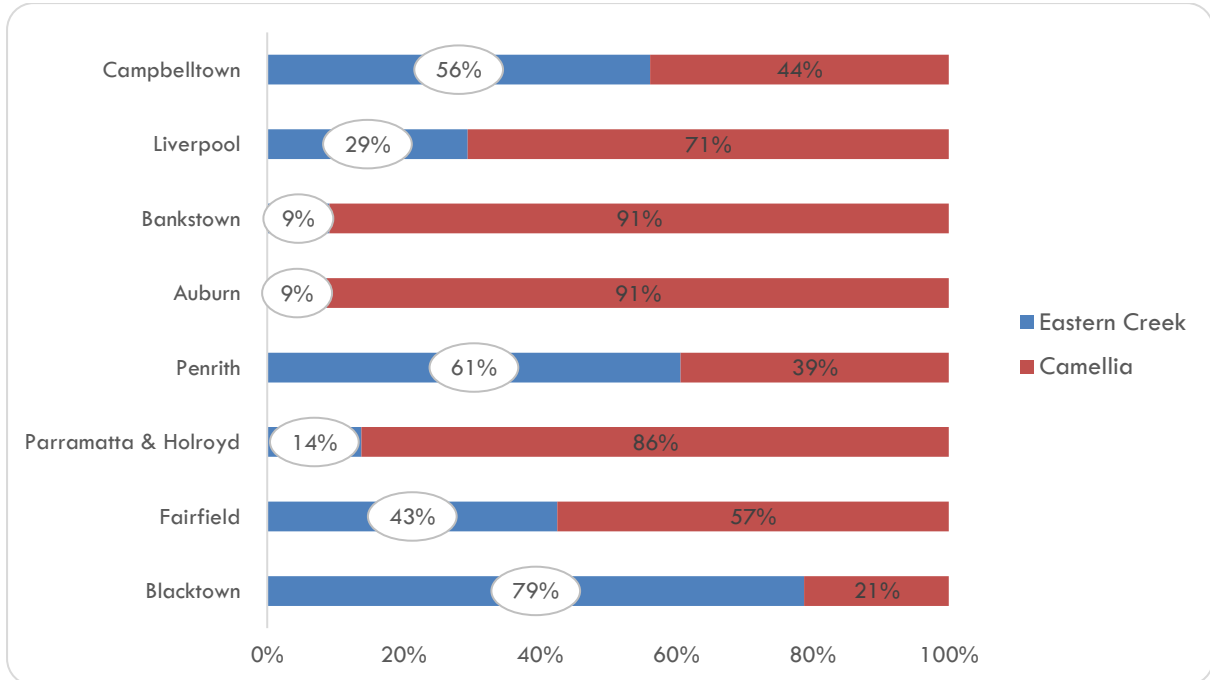

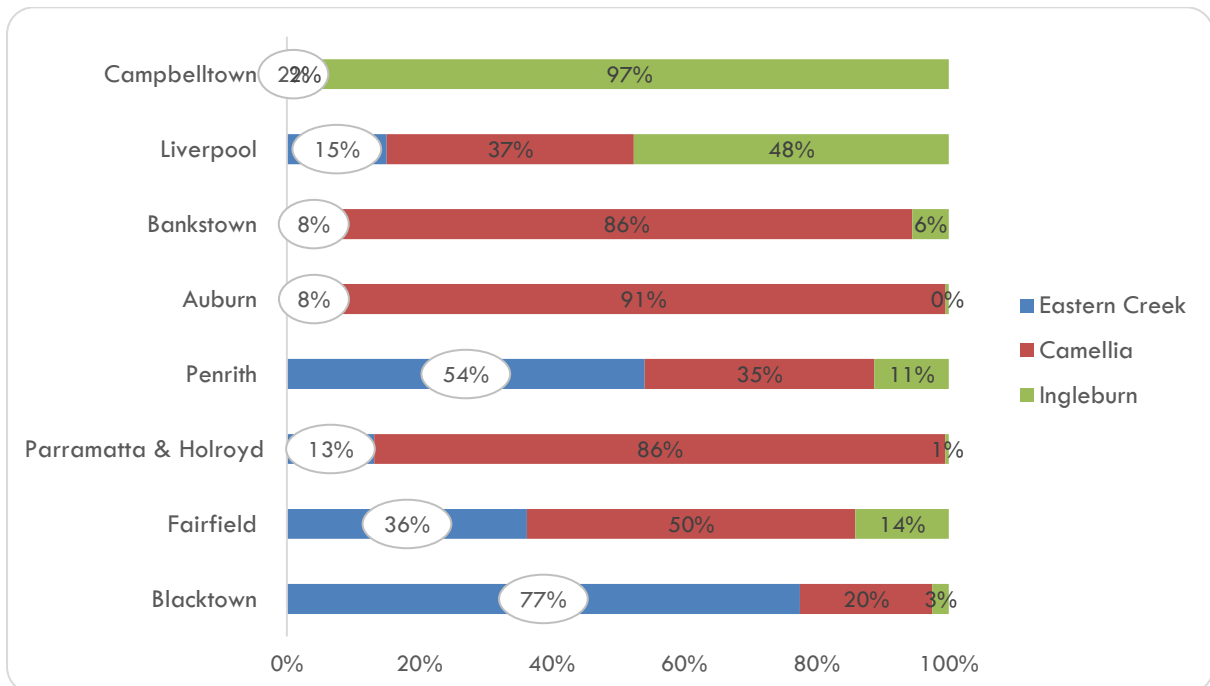**Figure 7.15: Market share for Eastern Creek and Camellia IMTs**

Figure 7.16: Market share for Eastern Creek, Camellia and Ingleburn IMTs
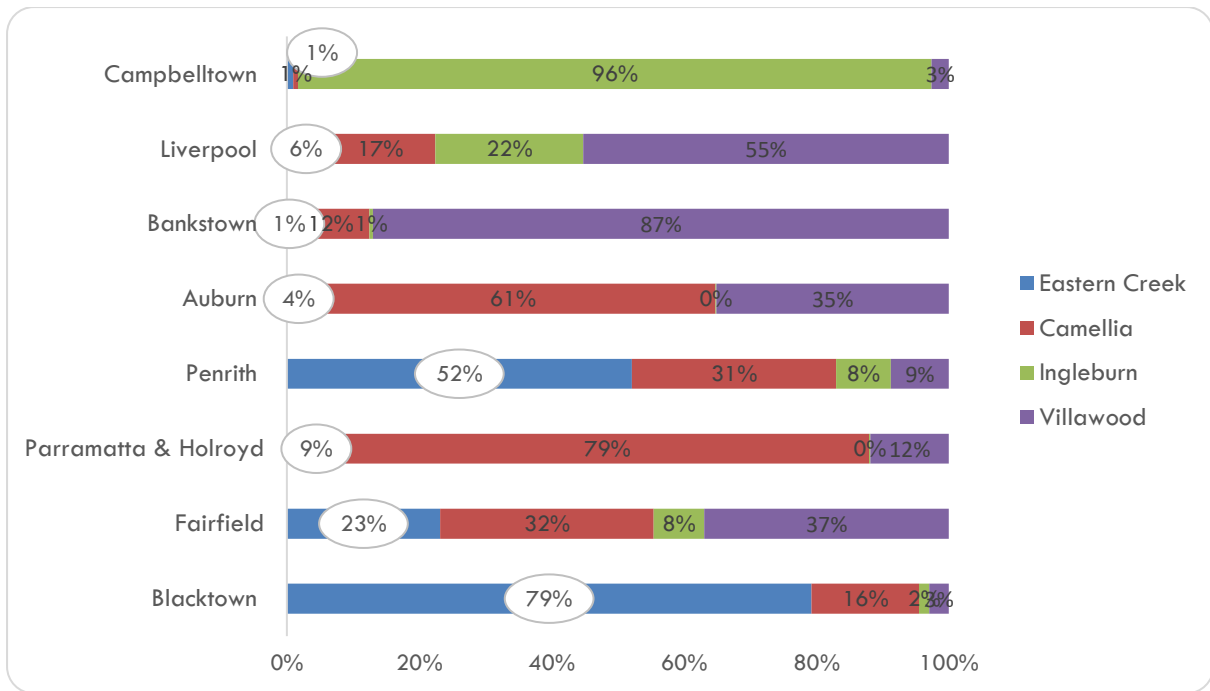
Figure 7.17: Market share for the four located IMTs

The overall contribution of each located IMT in total intermodal transport demand are shown in Figure 7.18 with the contribution of Eastern Creek decreasing from 100% under *p=1* (number of required IMTs to locate) to 32% under *p=4*. The figure also shows that contribution to intermodal transport usage by Camellia is on average higher than Eastern Creek for $p = 3$. These results are expected since as shown in Figures 7.13 (Eastern Creek) and 7.14 (Camellia) both IMTs have Blacktown as their biggest market and so they are expected to share the intermodal market in Blacktown if both are located. Additionally, Camellia has more catchments areas than Eastern Creek as shown in Figures 7.13 to 7.17. It is also worth noting that Eastern Creek is among the best places to locate IMTs for all $p \geq 1$. Similar observation is true for the Camellia IMT for $p \geq 2$. These imply that if the volume realized is currently not enough to justify the running and setup costs of two IMTs, then Eastern Creek could be developed first and add Camelia when cargo volumes grows.
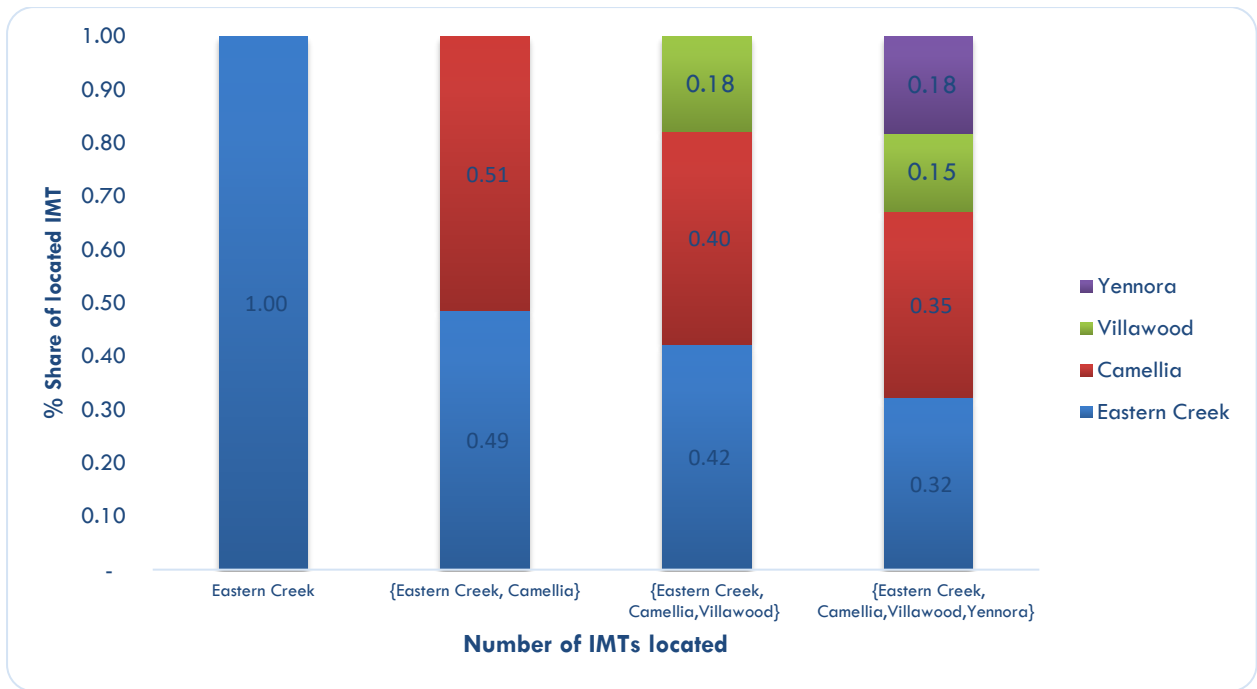
Figure 7.18: Robustness of optimal IMT selection

7.4.3 Impacts of cost variables on location decisions

The various components of the generalised cost of each mode were derived from existing data in 2001 and updated with factors to reflect the modelled year 2011. Thus, these cost variables may not accurately reflect market conditions and the construction of the generalised costs may contain some measurement errors. The cost of rail in particular, is lower than expected, whilst the fixed cost of final delivery by truck along the intermodal transport chain is higher than expected. This section investigates how variations in each cost component influence location decisions and demands. Figure 7.19 illustrates how the variations in rail cost influence each candidate intermodal demand. The rail cost was varied from a low value of $100 with a 10% increment in turn to a maximum of $200. The figure shows that low rail cost (less than $110) favours Camellia over Eastern Creek whilst higher cost favours Eastern Creek more. It can be seen from the figure that with the exception of Eastern Creek, the demand for the remaining candidate IMTs and hence intermodal demand falls sharply with increasing rail cost. This makes the viability and sustainability of these IMTs highly vulnerable to higher than expected rail cost. The figure also shows the reduction in intermodal mode (Eastern Creek) share with increasing rail cost-a decrease of 23% mode share under $100 rail cost to less than 0.5% under $200 rail cost. In summary, Figure 7.19 was used to demonstrate the significance of rail cost in intermodal location decisions and demands. A rail cost of less than $110 suggest Camellia

is more promising but the sensitivity test indicate it is more vulnerable to variations in rail costs. The intermodal demand for Eastern Creek is much more robust to variations in rail costs and hence a safer choice for the location of intermodal terminal among the candidate set.

The other important variable considered for sensitivity analysis is the fixed cost associated with the cost of final delivery by truck along the intermodal transport chain. It should be expected that this fixed cost can be significantly lower than the fixed cost of road alone transport. However, for the location exercise both costs were fixed at $218. This section investigates how lowering this fixed cost associated with intermodal transport affects the location decision and demand for the candidate IMTs. Starting with the same fixed cost as road alone transport, this fixed cost (intermodal fixed cost) was reduced in turn and impacts on IMT demands is shown in Figure 7.20. As expected all the candidate IMTs are promising promoters of intermodal transport with low fixed intermodal transport cost. The figure also shows the vulnerability or unsustainability of some IMT locations (especially, Moorebank, Enfield and Chullora) due to variations in fixed intermodal transport cost. Overall fixed transport cost of less than $210 makes Camellia more promising whilst higher values makes Eastern Creek location more desirable.
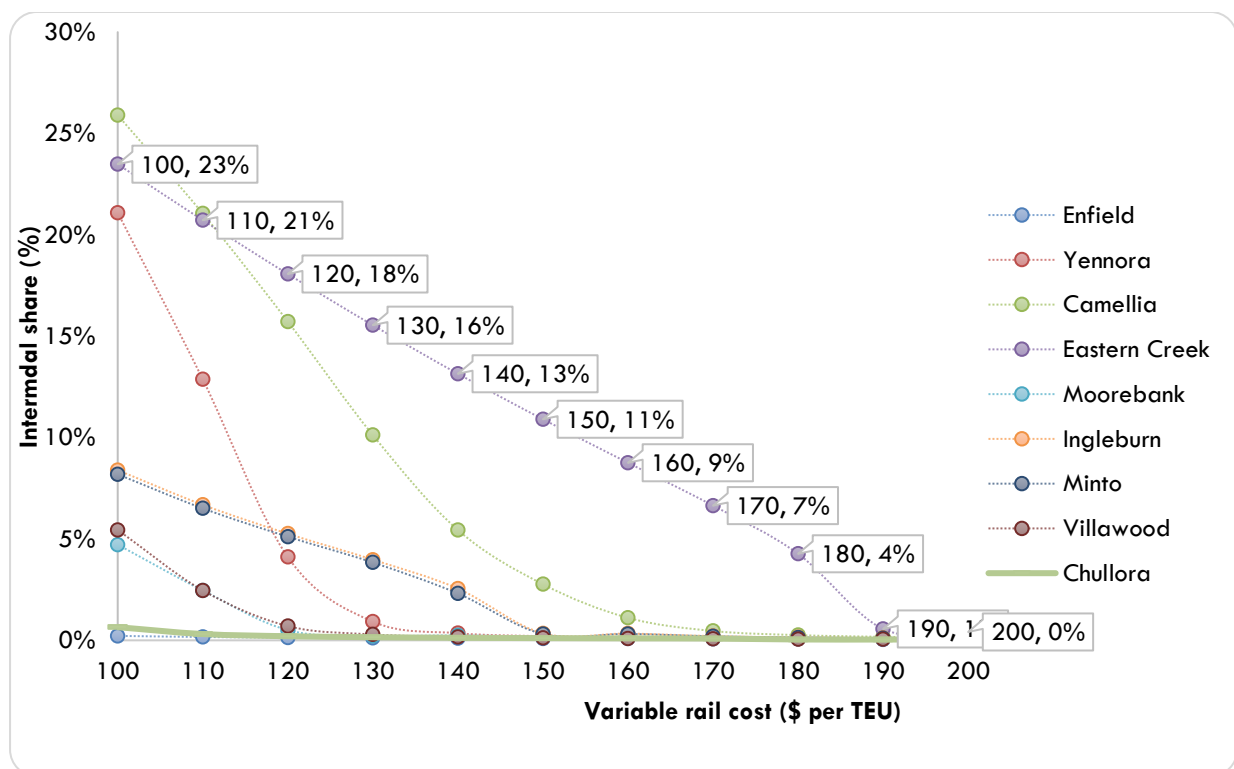

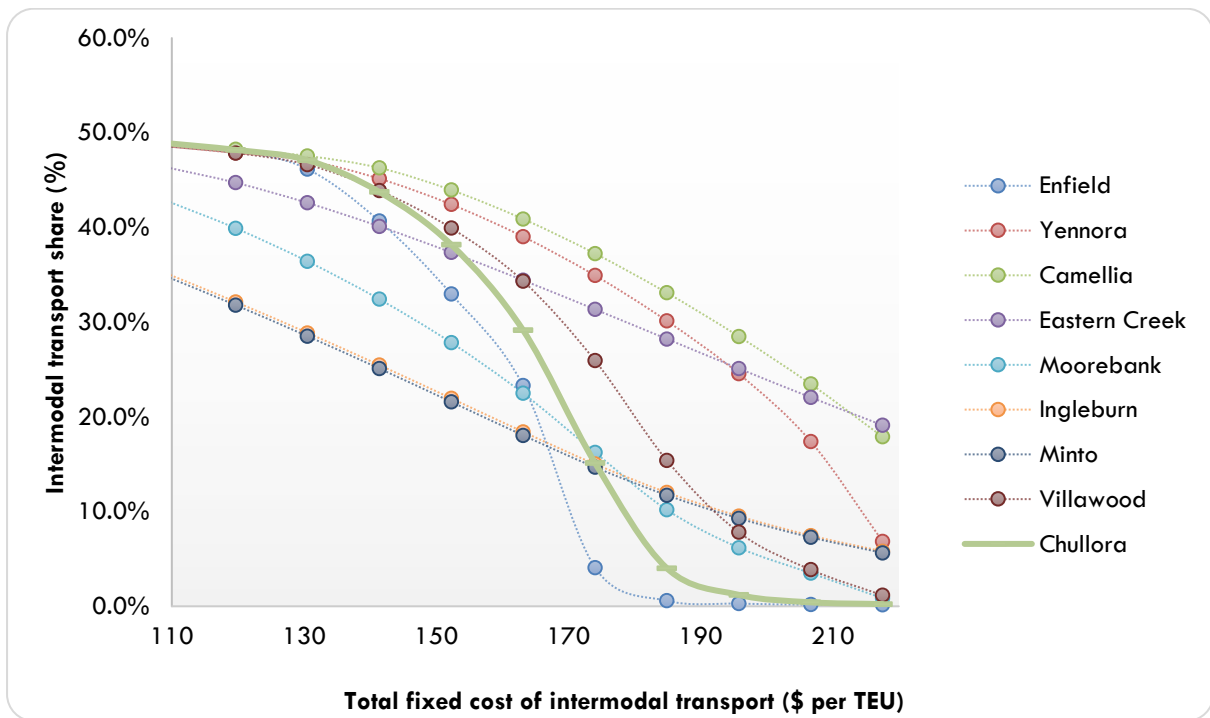**Figure 7.19: Impacts of rail cost on location decisions**

195

Figure 7.20: Impacts of rail plus fixed road costs on location decisions

This section looked at the sensitivity of the cost variables governing the choice of mode and hence the choice of locations for the development of intermodal terminals. The analysis shows that the demand for each candidate location is generally vulnerable to changes in cost variables. But the degree of vulnerability is less severe for Eastern Creek, making Eastern Creek, the most promising and economically safest place among the candidate IMT locations to develop an intermodal terminal. Eastern Creek is therefore carried forward for further analysis, forecasting and policy testing. The next section determines the optimal terminal user fee to charge at Eastern Creek.

7.4.4 Optimal IMT Charge Analysis

Based on the analysis in the section above, Eastern Creek IMT was carried forward to determine the best fee to charge for using the terminal as a transfer node. To do this, different level of charges ($ per TEU) were tested and observed the impact on revenue and demand levels. It is expected that as the terminal user fee is increased from zero, those that used the terminal (or intermodal transport) would pay more, which (initially at least) would lead to more revenue despite the revenue lost from the few users who chose to avoid the use of intermodal transport. However, as the user fee level increase further, there would be a user fee level above

which increases in fee would not lead to more revenue because those diverting away would lead to loss of more revenue than that gained from the extra money paid by the remaining users. It was expected that there would therefore be an optimum user fee level which would give the maximum revenue.

As shown in Figure 7.21, the optimal fee to charge at the terminal is $70.00 ($ 35 for each lift assuming two lifts are required). This figure compares favourably with average costs of $60 (Shipping Australia 2011), $60-$80 reported by the Sea Freight Council (SFC 2007) and $50 reported in IPART (2007). The $70 charge is expected to reduce traffic by about 57% compared to the zero charge scenario (i.e., from 170,857 TEUs to 74,053 TEUs) and generate revenue of about $5.2 million per year. Clearly, IMT charges have significant impacts on its use and could be considered as an important area for government intermodal-oriented policy interventions. A fee of say $40 (see Figure 7.21) could reduce demand by only 36% relative to the zero charge. Thus, any level of government subsidy is expected to have significant positive impact on IMT usage as shown in Figure 7.21. It may be that the terminal user fee adopted may be below or above the optimal value but it is important to establish the revenue consequences of this decision. The optimal user terminal fee of $70 will be coded into the model and used in forecasting and policy testing in the next section.
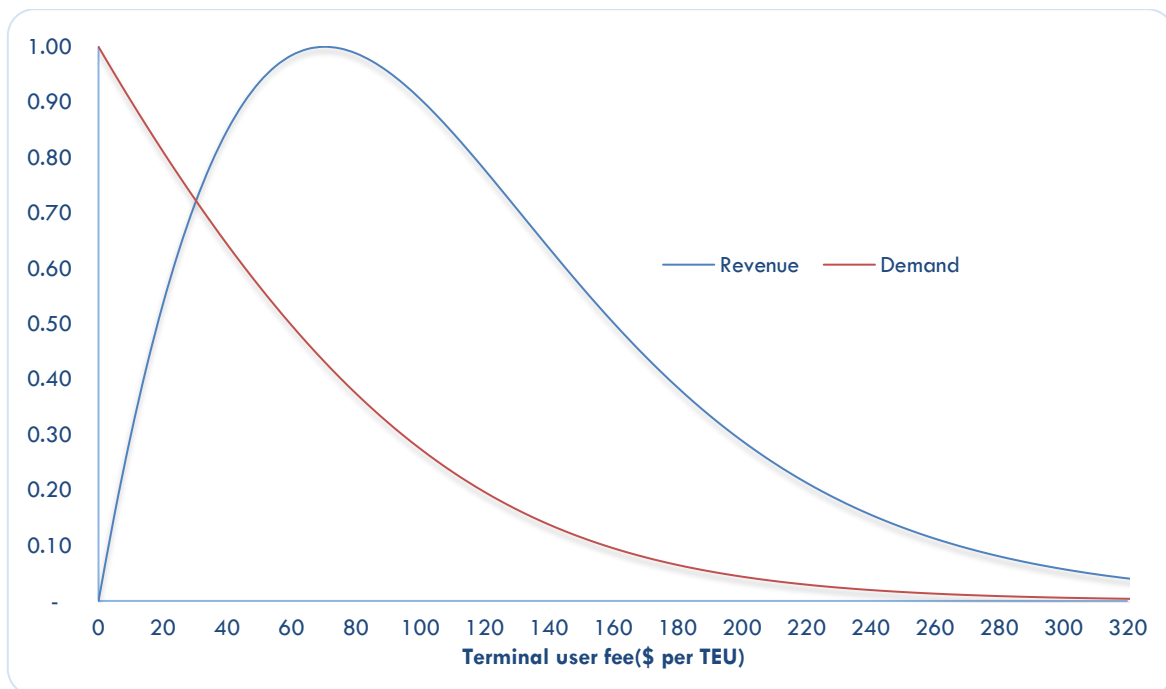

Figure 7.21: Optimal charge of using Eastern Creek ($ per TEU)

197

7.4.5 Eastern Creek as promising IMT location

Here, the Eastern Creek IMT is assumed located and coded as part of the transport network of the study area, where containers can be transferred between modes and where warehousing and storage activities can be performed. This section provides detail discussion on the factors governing both the distribution and mode choice models with the Eastern Creek being the transfer node along the intermodal transport chain. The cost sensitivity parameter governing the mode choice model was estimated to be 0.0206 and the estimated parameters governing the cargo distribution model are presented in Table 7.5. The results in the table shows expected positive marginal utilities for increasing access to manufacturing and warehousing businesses, accessibility to the key markets and multiple modes of transport (mode choice logsums). For example, the positive value (0.9079) associated with manufacturing indicates that zones with high manufacturing jobs are more likely to be cargo (container) destinations and may be indicating the existence of agglomeration of freight related businesses in the area. This is also true for zones with high warehousing and storage jobs. The access to the key markets variable is also positive (0.4116) indicating that all things being equal zones with easy access to the key markets (Sydney and Paramatta) are more attractive container destinations. This analysis is also true for zones with access to multiple modes of transport derived from the mode choice model. All the estimated parameters except access to multiple modes parameter are significant at 95% confidence interval.

Looking at the magnitude of the estimated variables, access to manufacturing businesses has the biggest influence on the distribution of containers in the metropolitan region and hence the location and use of IMTs. The second most important factor is the access to key markets. Together, access to manufacturing, warehousing and storage may be revealing the agglomeration effect associated with the containerised trade. This makes agglomeration (access to freight-related business) and access to key markets the key drivers in the choice and usage of intermodal terminals. Access to multiple modes, though not significant at 95% confidence interval (though significant at 85%) has the expected sign and provides the important link between the mode choice model and the cargo distribution model. The low variability in this variable may be linked to the two fixed costs components (fixed rail cost and the fixed truck cost) of the intermodal transport cost. The study therefore concluded it is important to carry this variable forward in further analysis.

Table 7.5: Estimated factors governing the cargo distribution model

Variables		Distribution model	
		Coefficient	t stats
Access to Manufacturing		0.9079	
	t-stats		6.1
Access to Warehousing and Storage		0.3305	
	t-stats		2.3
Access to key Markets		0.4116	
	t-stats		2.5
Access to Multiple Modes		0.1003	
	t-stats		1.3
No of Estimated Parameters		80	
No of observations		4	
R Square		98.7%	
Adjusted R Square		98.6%	

Figure 7.22 reveals the catchment areas of Eastern Creek in the study area. Bigger markets are represented by darker blue, implying Blacktown South and South-West are the biggest markets, followed by Penrith-East and Fairfield-West. Essentially, zones in Blacktown, Penrith, Fairfield, Paramatta, Holroyd and The Hills Shire form the market for the Eastern Creek terminal. Figure 7.23 shows the savings in truck-km travelled due to the development of the Eastern Creek terminal. The figure shows that the total-km travelled by trucks in the Sydney GMA will be reduced by about 8.4% if the Eastern terminal is open. The darker green areas in the figure represent bigger savings in truck-km travelled. Figure 7.22 corresponds very well with Figure 7.23 indicating that the market areas for the Eastern Creek terminal contributes more to the savings in total truck-km travelled. Thus, the location of the Eastern Creek terminal is expected to have significant positive impacts on the environment and also leads to less damage to roads as many container trucks will be taken off the road network.
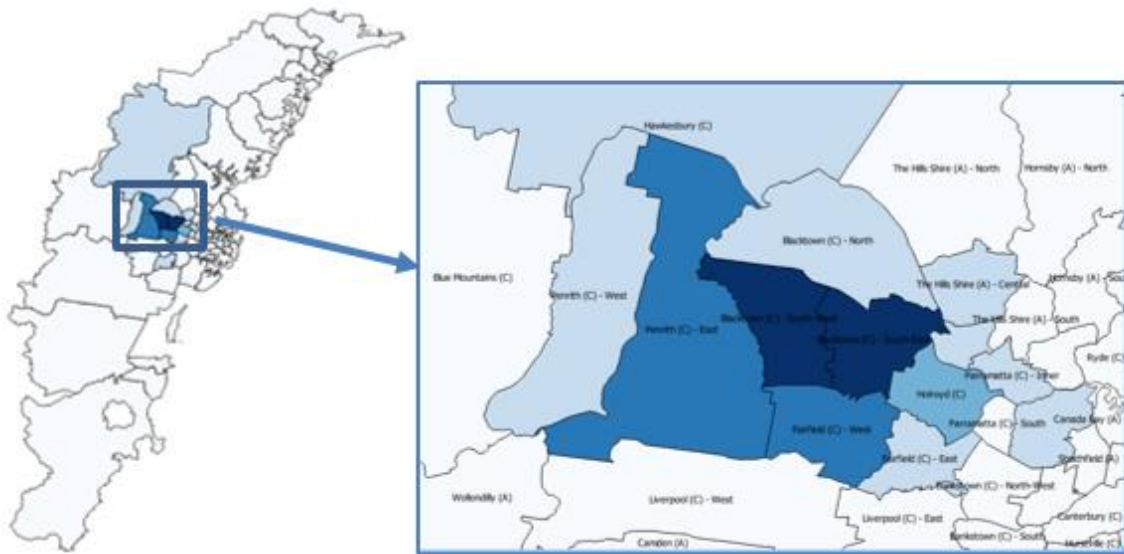
Figure 7.22: Main markets for Eastern Creek terminal

Total truck kilometres travelled	
No IMT (TEU-KM)	16,576,456
IMT (TEU-KM)	15,181,841
% Savings	8.41%
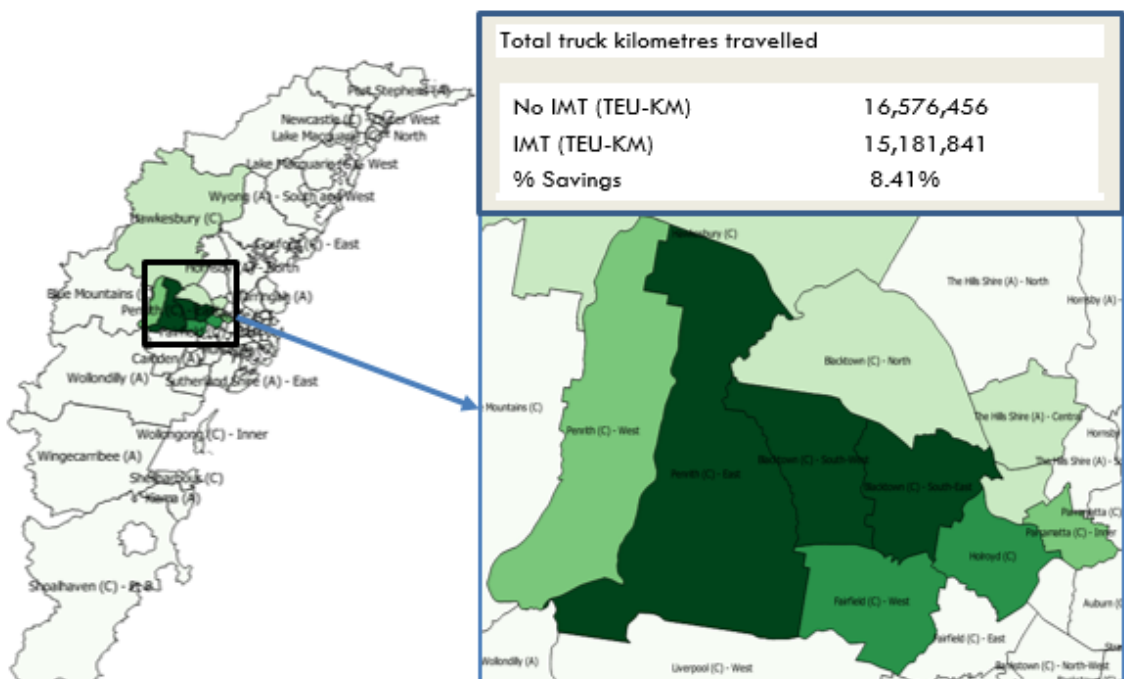
Figure 7.23: Reduction in truck kilometres travelled due to Eastern Creek

7.4.6 Forecasting and Policy Testing

The forecasting process starts by first extracting network variables (e.g., travel time and distances between the freight analysis zones) from the network using existing model of the study area. The network variables go into the mode choice model, where they are used to construct the generalised costs of each mode of transport. The parameters (fixed and variable

road cost, rail cost and terminal user fees) may be updated to reflect the forecast or policy year. The mode choice model uses the generalised cost values to compute the logsums; access to key markets and modal access to each cargo destination zone. The accessibility measures then go into the distribution model where they are combined with land use variables (also updated to reflect the forecast or policy year) and forecast port throughput to determine the distribution of cargo in the study area and hence the quantity of cargo reaching each destination. These outputs then go into the model choice model to determine the demand for each mode of transport (road vs intermodal) and the demand for intermodal transport is used to determine the demand for each located intermodal terminal. Thus, once, the port throughput of the forecast year is known and the other inputs variables are known, the demand and revenue for the located IMTs can be readily predicted and can be used in cost and benefit analysis.

This section, however, focusses on investigating plausible policies that the government can put in place to support intermodal transport. These policies can be linked to the variables used in the mode and distribution models such that impacts of these on intermodal transport usage could be measured through changes in these variables. The policies to be tested are grouped into land use and transport policies. Detailed discussions on the policies under each group are presented in the sections below.

7.4.6.1 Land use policies

Land use policies generally comprises policies whose impacts on intermodal transport usage can be translated into the number of people employed in manufacturing, warehousing and storage jobs. Some of these policies include encouraging shippers to co-locate their warehousing and storage activities near or to the located terminal either through re-zoning or some form of tax rebate or tax exemptions for participating shippers. Additionally, due to the relatively easy access to key markets from the terminal, large manufacturing and retail companies can site their distribution and warehousing centres close to the terminal to at least benefit from access to multiple modes of transport such that imported goods can be transported to the local markets more cheaply or used as inputs for local manufacturing (IPART 2007; Solomon 2014).

The Eastern Creek terminal itself can supply a range of value-added activities such as warehousing and storage of empty containers. All these freight activities are expected to

generate jobs, and the number of manufacturing and warehousing jobs can be used to quantify the amount of intermodal transport usage. In modelling terms, the zone encompassing Eastern Creek (Blacktown (C) - South-East) was used for this testing. The number of employees in manufacturing in the zone was increased by 1% in turn and the results are presented in Figure 7.24. The increase in jobs produced two responses; the re-distribution of cargo to the reference zone and more use of Eastern Creek as transfer node. If the jobs are created within the Eastern Creek IMT, then intermodal demand reduces to rail demand as there will be no need for final delivery by trucks. The figure shows that increasing manufacturing jobs in Blacktown (C) - South-East by 1% has the potential of attracting an extra 587 TEUs (about 0.8% increase) of cargo to the zone, out of which about 248 TEUs (about 42%) would be transported through the terminal or make use of intermodal transport (about 0.3% increase in intermodal transport demand). This makes the impacts of manufacturing jobs on cargo distribution more severe than mode switch.

Similar analysis was conducted on the impact of jobs creation in warehousing and storage on mode switch and cargo re-distribution and the results are presented in Figure 7.25. The results as expected in Figure 7.25 are not as strong as that of manufacturing. A 1% increase in warehousing jobs in Blacktown (C) - South-East can attract about 215 TEUs cargo to the zone, with less than half (92 TEUs) been transported intermodally. Thus, jobs created in manufacturing are expected on have more impacts on cargo distribution and modal switch than creating equivalent number of jobs in warehousing and storage. In general, a 1% change in the number of manufacturing jobs to a zone will result in about 0.8% change in the quantity of cargo expected to be attracted to that zone. For warehousing, the change is lower at about 0.3% for a 1% change in the number of warehousing and storage jobs as shown in Figure 7.25. In both scenarios about 42% of the increase in demand will be transported intermodally.
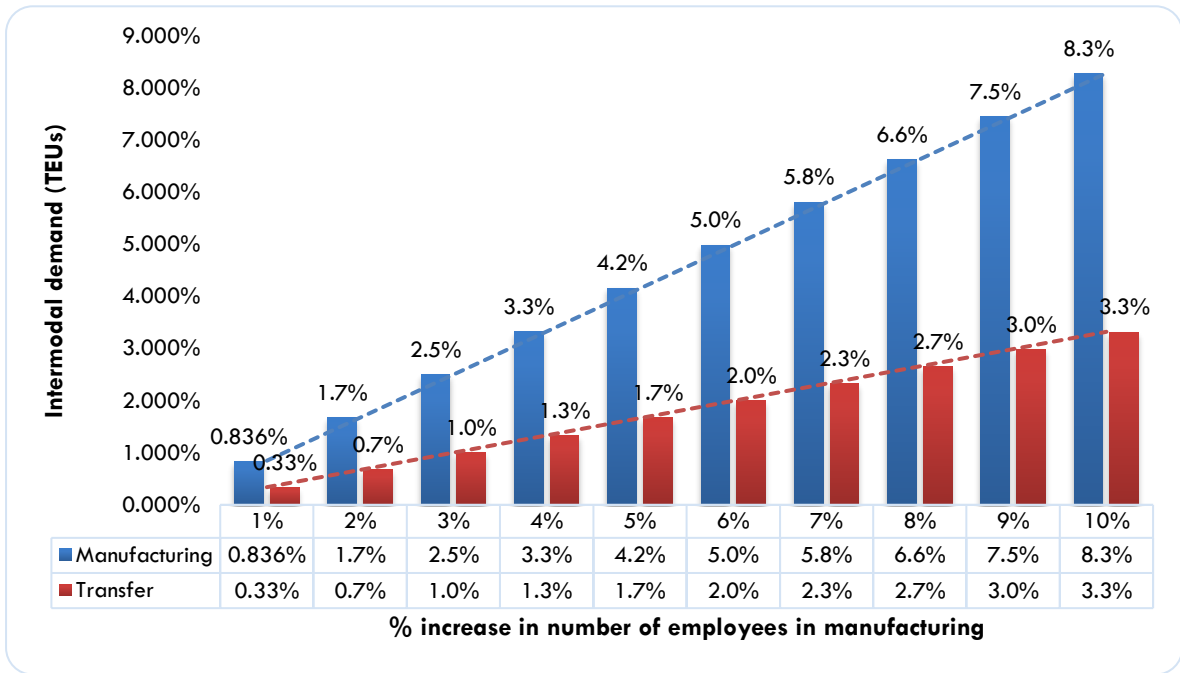
Figure 7.24: Impacts of manufacturing activities on the IMT usage

	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%
Manufacturing	0.836%	1.7%	2.5%	3.3%	4.2%	5.0%	5.8%	6.6%	7.5%	8.3%
Transfer	0.33%	0.7%	1.0%	1.3%	1.7%	2.0%	2.3%	2.7%	3.0%	3.3%

% increase in number of employees in manufacturing
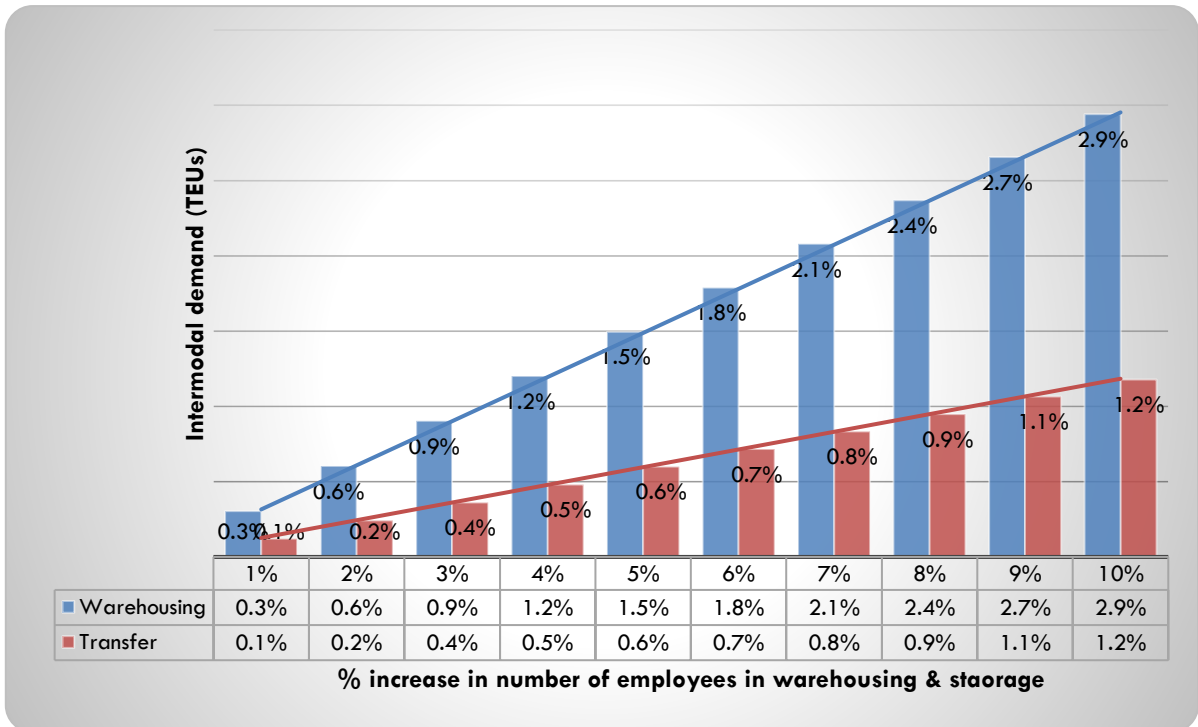
Figure 7.25: Impacts of warehousing and storage activities on the IMT usage

	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%
Warehousing	0.3%	0.6%	0.9%	1.2%	1.5%	1.8%	2.1%	2.4%	2.7%	2.9%
Transfer	0.1%	0.2%	0.4%	0.5%	0.6%	0.7%	0.8%	0.9%	1.1%	1.2%

% increase in number of employees in warehousing & staorage

7.4.6.2 Transport policies

The next set of policies relate to changes in the transport network conditions. In particular, the study investigated polices that reduce the cost of rail, cost of final delivery along the intermodal transport chain or penalise the use of road alone transport through some form of road pricing or restricting truck access to certain parts of the road network. Policies with the potential of reducing the cost of rail, may include investing in dedicated freight rail lines between the port and the terminal for the provision of uninterrupted services between the port and the located terminal; better alignments of stevedores' operation time windows and rail paths to improve the flexibility of train operations (Shipping Australia 2011) and standardizing the length of rail sidings at port terminals to reduce or avoid the need to split trains accessing port terminals, which causes unnecessary delays at ports. The government can also subsidise rail service providers to encourage them to take loads that may not be commercial viable to take but can help achieve a broader policy objective of increasing rail model share.

Policies with the potential of reducing the cost of final delivery along an intermodal transport chain may include encouraging as many shippers as possible to co-locate their warehousing and storage activities near to the terminal. The success of this policy depends critically on the geographical locations of the terminals with respect to cargo origins/destinations, which is why a model is useful to determine the best places to locate these facilities. Another policy option is for the government to relax restrictions on the use of high productivity vehicles on certain part of the road network for intermodal transport users in order to reduce the cost of pickups and final deliveries through economies of scale. The restrictions can be relaxed for certain time of day or during the weekends (IPART 2007). The government can also invest in new roads or upgrade existing roads, especially intermodal connector roads to improve the connectivity between the terminals and the highway network.

Finally, the government can implement policies that can specifically target road alone transport users in order to make intermodal transport more competitive. Some of these policies include imposing a direct 'congestion' charge on road alone transport users accessing the port at certain of time of the day to dissuade some to switch to intermodal transport. Another policy may be to restrict truck access to the parts of the road network that are vital for truck container movements to and from the port. The restriction could be put in place permanently or could be operational during certain times of the day or week. The impacts of these or equivalent policies

on intermodal demand can be quantified through their contribution in reducing the cost of intermodal transport.

Policies on rail were captured in the model by varying the cost of rail and observing the impacts on intermodal usage as shown Figure 7.26. The figure shows that a 10% increase in the cost of rail (from $100 to $110) will decrease intermodal demand by about 13% (from 80,770 TEUs to 92,866 TEUs), indicating the strong impact rail cost has on intermodal demand. A rail cost at $250 has the potential of rendering the terminal almost unsustainable or unviable as demand is expected to fall below 15,000 TEUs, which with the $70 terminal user fee translate to an annual revenue of less than $ 0.5 million. In Table 7.1 a viable terminal is expected to generate an annual revenue of more than $5million (Meyrick 2006).
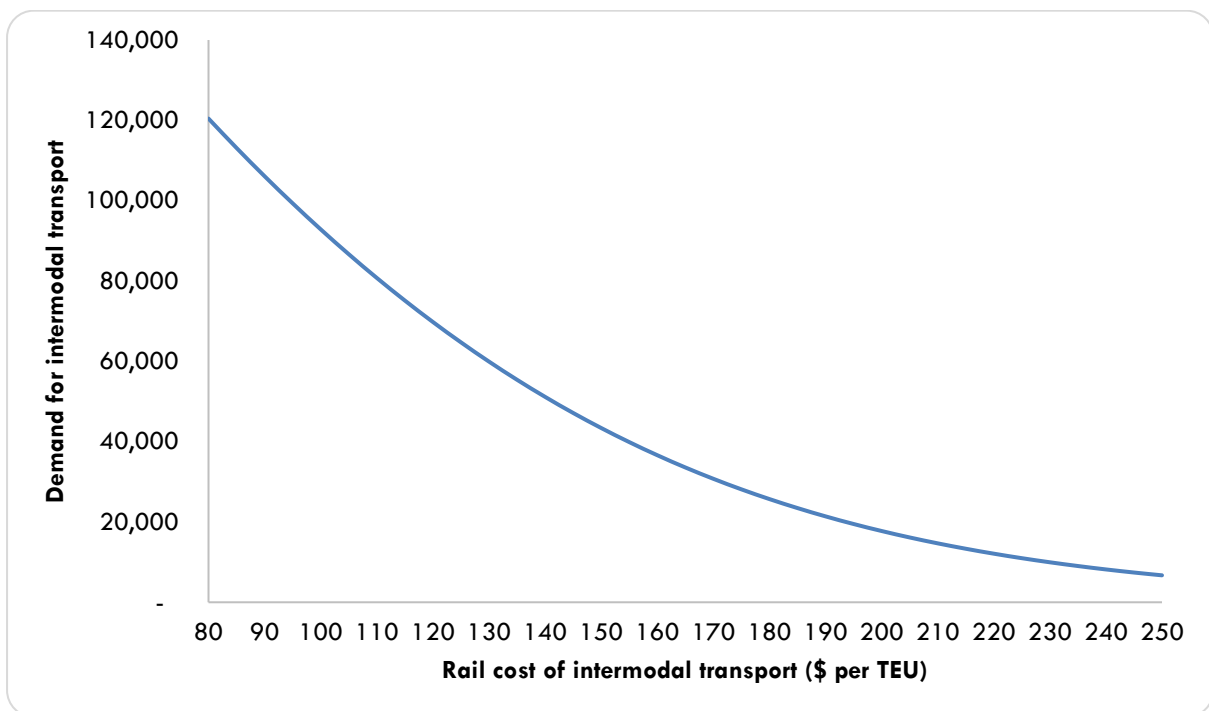

Figure 7.26: Impacts of rail cost on the IMT usage

The second variable of interest is the cost of final delivery along the intermodal chain. The total truck cost component (both fixed and variable for each origin-destination) was reduced gradually by 1% in turn to as high as 20% and the impact on intermodal demand from each reduction is shown in Figure 7.27. It can be seen from the figure that for example, a 10% reduction in the cost of final delivery has the potential of increasing intermodal demand by

almost 50% (from 74,053 TEUs to 110,419 TEUs). The results make it a very attractive policy to pursue with the potential of significantly reducing the cost of final delivery.
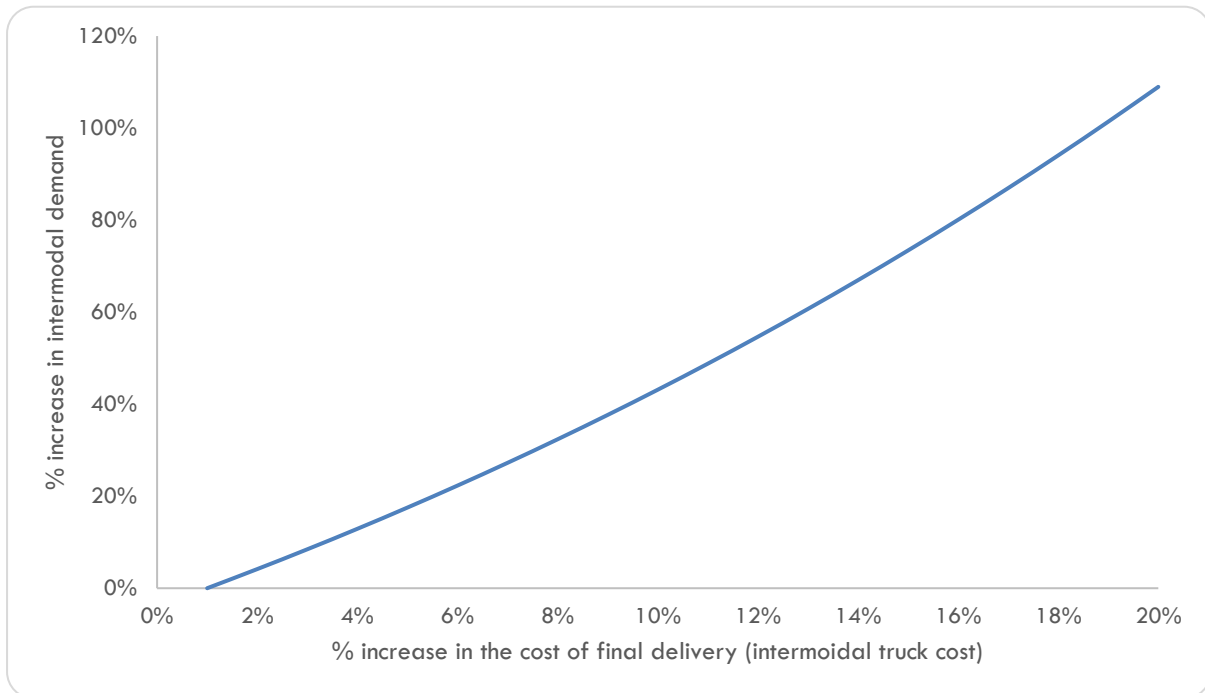

Figure 7.27: Impacts of final intermodal delivery costs on the IMT usage

Finally, the study turns to the cost of road alone transport to ascertain how changes in each cost component affects intermodal demand. Implementing the policy on road pricing or port congestion charge is expected to increase the fixed cost component of road alone cost, whilst general congestion due to increasing traffic demands can be reflected in the variable cost (function of travel time). The outcomes on varying these two cost components are shown in Figure 7.28. The figure shows that the impact of some form of road pricing on intermodal demand is stronger than that of the cost of congestion. For example, a 5% increase in the fixed cost (e.g., through congestion charge) of road can result in about 17% (from 74,053TEUs to 86,528 TEUs) increase in intermodal demand whilst a similar increase in the variable cost of road (e.g., delays in traffic) has the potential of increasing intermodal demand by 14% (from 74,053 TEUs to 84,094 TEUs). These results in general, suggest a positive future outlook for the located IMT as congestion in the study area is set to get worse in the future. In other words, the cost of transporting containers by trucks to and from the port is set to go up significantly due to congestion and/or through some form of road pricing.
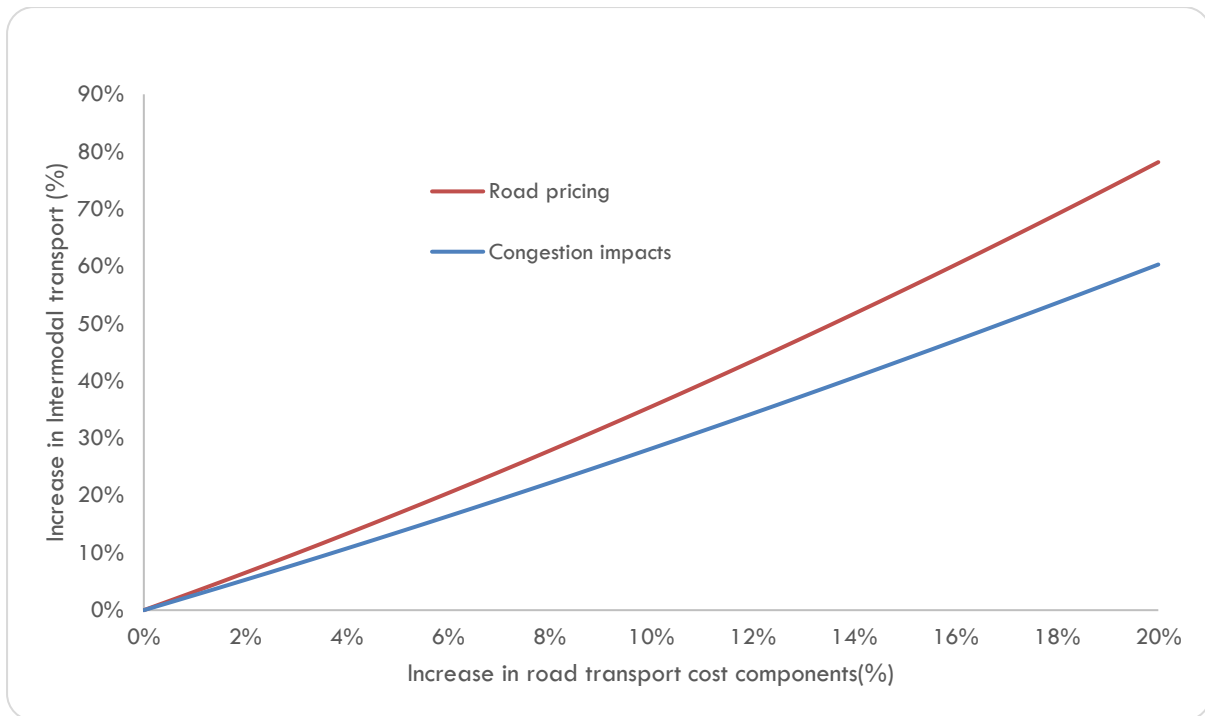
Figure 7.28: Impacts of road pricing and congestion on IMT usage

7.5 Conclusion

This case study demonstrates some of the key features of the proposed entropy framework for locating multi-user intermodal terminals. The entropy framework allows the facility location problem to be linked to a behavioural mode and distribution problems such that the facility location problem can be seen as conditioning the distribution of cargo and mode choice whilst mode and cargo distribution influence intermodal location decisions. Data used for this exercise were synthesised from several sources including Sydney freight council (SFC), existing integrated transport model of the study area (MetroScan-TI), transport for NSW (TfNSW), and the Australian Bureau of Statistics (ABS). The model has been shown to produce intuitive and realistic results both in terms of locating the facilities and testing of various policy options.

The results of the models have several implications for policy makers. One important outcome is the effect of container distribution on intermodal location and usage, accounted for through the cargo distribution model. The distribution model has been shown to be governed by four important policy variables, two (access to manufacturing and warehousing variables) of which reveal the existence of agglomeration (i.e., clusters of industries in a single location) associated with containerized trade. The results suggest that the access to manufacturing has the biggest impact on the distribution of containers in the metropolitan region of Sydney. This

result is consistent with recent study by Chandler et al. (2015) who provided empirical evidence of the strong relationship between the number of manufacturing businesses and the location of intermodal freight terminals in the US. Similar conclusion was also reached in the recent study by Tsekeris (2016). Thus, the catchment area (market) of a located intermodal terminal is more likely to include areas with high manufacturing businesses. Also, of importance to policy makers are the results that container destinations must have convenient access to key markets. The access to key market variable, measured as function of truck travel cost from container destinations to the key markets emerged with positive value and statistically significant, suggesting that easy access from container destinations to key markets is essential. Finally, proximity and convenient access to multiple modes of freight transport also have significant effect on container distribution, although its impact is relatively small compared with access to manufacturing or warehousing businesses.

There is strong evident to suggest that the best place to locate an intermodal terminal in Sydney GMA is in Eastern Creek partly, due to its proximity to major container destinations (e.g., Blacktown) and hence its access to manufacturing and warehousing businesses. The results show that performing auxiliary activities such as warehousing at the terminal will provide a significant source of revenue due to potential container re-distribution and increase in the use of intermodal transport. Policy testing shows a positive future outlook for the located terminal at Eastern Creek as intermodal terminal use is set to go up significantly partly due to worsening traffic congestion in the study area especially around the seaport and also through government interventions to achieve certain policy objectives such as promoting sustainable modes of transport like rail or solving the problem of truck driver shortages or promoting local enterprises. In summary, the results of this empirical work have shown that the proposed entropy framework of locating intermodal terminals has the potential of working very well in practice. The empirical work has also addressed the research question IV that motivated the work, by identifying the factors governing the distribution of cargo and the choice of mode in the metropolitan containerised market and how changes in these factors alter the distribution pattern and the location and use of intermodal container terminals.

Chapter 8 Conclusions

8.1 Background

A new method for finding the best locations for multi-user facilities is proposed. The method is underpinned by the principle of entropy maximisation. The proposed entropy framework allows the linking of a location model with behaviour models to simultaneously determine the location and usage of the facilities. An equivalence has been shown between entropy and shipper welfare, following from the random utility interpretation of the discrete choice model, which in turn is a consequence of entropy maximisation. The method has been successfully applied to locate inland intermodal terminals to maximise the benefit to shippers. This chapter provides a summary of the models proposed, the algorithms developed and illustrations of how the models were implemented using numerical examples and a case study. The chapter is organised in the following way. The fulfillment of the research objectives and questions are concluded in Section 8.2. Section 8.3 details the key contributions made in this study to the literature and in practice. Finally, the limitations of this study and the directions for further research are outlined in Section 8.4.

8.2 Fulfilment of research objective and questions

The objective of this research was to develop mathematical models suitable for planning the development of inland intermodal terminals to maximise shipper utility or welfare. The hinterland is the area behind the seaport where cargo to the port (export cargo) originate or cargo from the port (import cargo) are destined or where cargo productions and/or consumptions take place. Two hinterland containerised transport markets were identified; metropolitan transport market and the regional transport market. The metropolitan market comprises the import and export (IMEX) markets where imported cargo are transported to their various destination areas in the hinterland and cargo from production areas in the hinterland are transported to the port for export. Two modes of transport are available for the transport tasks in this market; road alone transport and metropolitan intermodal transport. The

metropolitan intermodal transport comprises a combination of rail (or high carrying capacity mode) for the movements of cargo between the seaport and inland terminals and trucks for local pickups and deliveries. The use of the port requires only one intermodal terminal to be used along the intermodal chain. The economies of this mode as discussed in Chapters 1 and 4 are based on the economies of scale of using rail and potential savings in transport costs as a result of avoiding congestion around ports and between ports and delivery/pickup centres in the hinterland. This mode of freight transport is particularly attractive for city ports like Sydney with limited space for physical expansion but experiencing continuous growth and poor or inadequate transport infrastructure connecting the port to the hinterland.

The regional market on the other hand comprises the movements of cargo between their production and consumption areas in the hinterland without making use of the seaports. Two modes of transport are also available for the transport tasks in this market; the road alone transport and regional intermodal transport. Regional intermodal transport comprises two terminals as transfer nodes along the intermodal transport chain. Here, the cargo are first consolidated at a terminal close to the cargo origin using trucks and then transported by rail (or a high capacity mode) to another terminal close to cargo destination areas where the cargo are finally distributed by trucks to their various destinations. The economics of this mode is based on both the economies of scale of using rail and the economies of distance due to the long distance separating the areas of cargo production and cargo consumption as discussed in Chapters 1 and 5.

In addition to the potential cost savings by individual shippers, governments or local authorities can use the promotion of intermodal transport as a tool for achieving certain policy objectives such as achieving a pre-defined target of rail mode share in the transport tasks, some percentage reduction of truck related greenhouse gas emissions, promotion of local enterprise and economic development or as a more sustainable alternative to road network expansions. As discussed in Chapter 7, government promotion of this mode can take several forms including some form of road pricing, subsidies and general network improvements, especially connector links to the terminals.

The problem of locating intermodal terminals is analogous to the classical facility location problem and comprises two linked sub-problems; the location problem and the allocation problem. As discussed in Chapter 1 and 2, the location problem determines the

locations of the required number of terminals on the network. The allocation problem on the other hand determines the usage of the located terminals. These two sub-problems share a common objective, namely the optimisation of shipper welfare or benefit, which is measured by entropy. It was demonstrated that the allocation problem is a mode choice problem with the choice alternatives being road alone and intermodal transport, whereas the choice of intermodal transport leads to a choice of one or more of the located terminals. It also turns out that the choice of mode depends on cargo destination. Conversely, cargo destination must be connected to the transport network or, more generally, must be accessible by at least one mode of transport. It is expected and also demonstrated in Chapter 7 that destinations that are more accessible by multiple modes of transport are more likely to attract more cargo. Thus, the problem of determining the production and distribution of cargo conditions the choice of mode (mode choice problem) whilst the production and distribution of cargo is influenced by modal accessibility.

The overall intermodal location problem therefore comprises three linked sub-problems; the facility location problem, the mode choice problem and the cargo production and/or distribution problems. Analogous to the classical facility location problem, the allocation problem now comprises linked mode choice and cargo production and/or distribution problems. The combined problem was referred to as the cargo flow problem. This reduced the intermodal location problem into linked facility location problem which determines the location of the terminals and the cargo flow problem, which determines the demand or usage of the terminals. The mathematical formulation of the problem was based on the principle of entropy maximisation. The entropy principle and its suitability for solving the problem were discussed in Chapter 3. The method was shown to provide a universal means of constructing probability distribution about a system based on all known information about it (system). Additionally, it allows probabilities to be updated when new information about the system become available, making the entropy probability distribution describing the system at any given time reflects all known properties about the system. The system in this instance is the containerised transport system. The thesis investigated four broad research questions and the fulfilment of each question is assessed in the following subsections.

8.2.1 Locating intermodal terminal in metropolitan markets

Research question 1: Given the distribution of containerised cargo and candidate terminal locations on the transport network, what are the best places to locate p intermodal terminals to best serve the metropolitan containerised market?

The answer to the above question was model based. Under the assumption of fixed cargo flow matrix (demand) the research question was cast as a mathematical problem – the metropolitan intermodal terminal location problem (MIMTLP) – with an entropy objective function to be optimised and a set of constraints imposed by the metropolitan market in Chapter 4. It was also shown in Chapter 4 that maximising entropy is equivalent to maximising the shippers' expected utility or welfares. Thus, the best p terminals are selected on the basis of providing shippers with the largest consumer surplus**.**

The fixed demand assumption means that the cargo flow problem of MIMTLP reduces to a mode choice problem. The formulated problem was solved by decomposing it into a facility location sub-problem and a mode choice sub-problem using Lagrangian relaxation technique. For any given set of located terminals, the solution to the mode choice problem was proved to exist and is unique. The mode choice sub-problem was then converted into a two-level nested logit model with the upper model determining the demand for each mode (intermodal versus road alone) and the lower model determining the demand of the located terminals. The parameters in the resulting nested logit model were estimated using a modified version of the Bregman's algorithm in Chapter 4. Once it was shown that the mode choice sub-problem can be solved, an enumeration algorithm that embeds the solution of the mode choice problem was proposed for selecting the required places to locate the p intermodal terminals. The algorithm is a complete enumeration algorithm and guarantees an optimal solution and was considered to be computationally feasible since there are limited number of places in the metropolitan region where terminals can be located.

Once the best locations of the required number of terminals are determined, the nested logit model (probability distribution) derived from the mode choice sub-problem contains the properties of the metropolitan market. It was shown that the resulting nested logit model can then be carried forward to forecast future demand for the located terminals and use to test for various policies to promote their usage. The model developed can be used to locate a given

number of terminals in any metropolitan region and hence fulfilled research question I. An application of this model in practice was illustrated in Chapter 7.

8.2.2 Locating intermodal terminals in regional markets

Research question II: Given the distribution of containerised cargo and candidate terminal locations on the transport network, what are the best places to locate p intermodal terminals to best serve the regional containerised market?

Research question III: Given limited demand for intermodal transport in either or both markets, what are the best places to locate p intermodal terminals to serve both the metropolitan and the regional containerised markets?

These research questions were answered together in Chapter 5, where the model developed in Chapter 4 was extended and made suitable for also locating terminals to serve the regional containerised market. The extended problem was made flexible to locate terminals to serve the metropolitan containerised market or the regional containerised market or both markets through the incorporation of user-defined switch variables. Important properties of this generalised problem were also presented in a form of propositions. Solution techniques employed in solving the extended problem were similar to those employed in solving the metropolitan intermodal terminal location problem and involves first using Lagrangian relaxation to decompose the problem into facility location and mode choice sub-problems. Here, the mode choice sub-problem was converted into a three-level nested logit model of main mode choice (road alone versus intermodal transport), intermodal transport choice (metropolitan intermodal transport vs regional intermodal transport) and intermodal terminal choice (determines the demand for located terminals). Bregman's algorithm was then extended to estimate the parameters governing the behaviour of the three-level nested logit model.

Due to the complexity associated with the generalised problem, the complete enumeration algorithm developed to answer research question I was found not to be computationally efficient enough for answering research questions II and III. By exploiting some properties of the entropy function, a fast heuristic algorithm with embedded Bregman's algorithm was proposed and its solution quality was demonstrated through extensive numerical

examples. The proposed heuristic algorithm was shown to be computationally efficient and its computational complexity grows only linearly with respect to the number of terminals to locate. In terms of solution quality of the heuristic, it returned the optimal solution on all test instances. Although the heuristic has not been proven to return optimal solution for every instance of the problem, it has been shown by extensive numerical examples to work satisfactorily. In summary, the complete enumeration or the heuristic algorithm with embedded Bregman's algorithm for solving the mode choice algorithm provides adequate answers to research questions II and III. A specific application of the model was illustrated using test data where 2,3,4, and 5 terminals were located in turn to serve either or both markets in Chapter 5.

8.2.3 Variable cargo demands

Research question IV: What are the key factors governing the distribution of cargo in the containerised markets (metropolitan or regional) and to what extent do changes in these factors alter the distribution pattern and the location of intermodal container terminals?

This question was answered through model enhancement and a case study implementation of the models developed. The fixed cargo demand matrix assumption was relaxed in Chapter 6 to allow cargo production and/or distribution to be influenced by modal accessibility and land use factors, whilst the production and/or distribution of cargo continues to condition the choice of mode. In addition to the inherent link between the cargo distribution problem and mode choice problem, the cargo distribution problem is required to determine the potential use of the terminal for warehousing and storage activities.

The model used in answering research question III was extended to also include a variable cargo demand model. Two methods were proposed to allow the incorporation of cargo demand models as constraints within the entropy framework; Poisson method and weighted mean method. It turns out that the Poisson method is a special case of the weighted mean method as shown in Chapter 6. The overall problem was called intermodal terminal location with variable cargo demand problem. The overall problem was again decomposed into facility location and cargo flow sub-problems. Here, the cargo flow sub-problem comprised the mode choice problem and the cargo production and/or distribution problem, and are connected through accessibility measures. Several special cases of the cargo flow problem depending on data availability were presented and the ways each can be solved to optimality were also

discussed. The overall problem was solved by adapting the complete enumeration algorithm for small instances and the heuristic algorithm for larger problem instances.

The model was implemented using the Sydney greater metropolitan area (GMA) as a case study in Chapter 7. The study found Eastern Creek as the most promising location for the development of intermodal terminal in the study area, with container destinations in Blacktown as its key markets. Other significant catchment areas were found to be Penrith, Paramatta and Fairfield. These four areas together represent over 86% of the market for Eastern Creek. The location of two terminals yielded Eastern Creek and Camellia as the best solutions and together increase intermodal transport share from 19.5% to about 29%. The catchment areas for these two terminals are Blacktown, Fairfield, Parramatta and Holroyd, Penrith, Auburn, and Bankstown. Together, these areas captured over 90% of the intermodal markets, with Eastern Creek controlling the market in Blacktown and Penrith whilst the other markets are largely controlled by Camellia. The test for the location of three and four terminals simultaneously were also conducted. However, for forecasting and policy testing, the location of only one terminal was assumed to be of interest, which from the location analysis yielded Eastern Creek as the best location.

The choice of mode was governed by the generalised cost and together with the shadow price and sensitivity parameter form the utilities of the available modes where for each origin-destination pair a mode with higher utility has higher probability of been used for the transport task. The distribution of cargo in the study area was found to be governed by four main factors; access to manufacturing activities, access to warehousing and storage activities, access to key markets, which in the study represent the two central business districts – Sydney and Paramatta – and finally access to multiple modes of transport, which is a combined measure of access by road alone and access by intermodal transport modes. Once Eastern Creek was determined as the best terminal location and coded into the transport network as a transfer facility, the mode choice and distribution model were carried forward for policy testing. The study found that charging terminal user fee of $70 per TEU yielded the maximum revenue though at the expense of losing about 44% usage compared with a zero fee. This is one of the several areas that the government or local authority can intervene to boost intermodal demand. For example, it was shown that a reduced terminal user fee of $50 (subsidy by government) will reduce intermodal demand by just 20% rather than the 44% decrease under the $70 fee.

Policies relating to improving the cost of intermodal transport or making intermodal transport more competitive were also investigated. These include implementing policies with the potential of reducing the cost of rail in the overall cost of intermodal transport. Such policies may include investing in a dedicated freight rail lines between the port and the terminal and/or subsidising the cost of rail use long the intermodal chain. The study shows that a 10% increase in the cost of rail has the potential of decreasing intermodal demand by about 13%, assuming all other factors remains unchanged. Several values of rail costs were tested and the impacts of intermodal demand were presented in Chapter 7. The other important area that the government can support is the cost of final delivery or local pickups, which are essentially the truck cost component of intermodal transport cost. Government can improve the connector roads to the located terminals and improve the connectivity between the terminals and the major highways through road expansion and/or investments in new roads. The government can also give tax incentives to truck companies who make their vehicle available for use along the intermodal transport chain. It was shown that reducing the cost of final delivery along the intermodal transport chain by just 10% can trigger about 43% increase in intermodal demand.

Finally, the study investigated the impacts of policies such as road pricing and natural growth in traffic congestion on intermodal demand. The road pricing policy and congestion effect were implemented by varying the fixed cost component and the variable cost component of the overall road alone transport cost respectively. It was shown that a 5% increase in the fixed cost of road alone transport can boost intermodal demand by about 17% whilst a 5% increase in the variable cost can increase intermodal demand by about 14%. The later outcome suggests a positive outlook for the located terminal as congestion in the study area is set to worsen in the future.

The study also investigated the impact of the distribution model on intermodal demand. As expected, the study found the impacts of changes in the factors governing the distribution model on intermodal terminal demand in two ways; through cargo re-distribution and through modal switch. For example, a 1% increase in the number of manufacturing jobs in Blacktown (C)-South-East (the zone encompassing Eastern Creek) has the potential of attracting an extra 587 TEUs (about 0.8% increase) of cargo to the zone, out of which about 248 TEUs (about 42%) would be transported through the terminal or make use of intermodal transport. This makes the impacts of manufacturing jobs on cargo distribution more severe than mode switch. For warehousing, the impact is expected to be lower at about 0.3% increase in cargo to the zone

for a 1% increase in the number of warehousing and storage jobs. In conclusion, the study found that among the plausible policies tested, those directed at reducing the cost of final delivery has the greatest impact on intermodal demand, albeit based on this case study. This cost component of intermodal transport can undo any cost advantage due to economies of scale of rail if not checked and kept to a minimum.

8.3 Thesis contributions

This thesis made several methodological contributions to the literature. Perhaps the most important contribution is the recognition that behavioural models can be embedded in a facility location problem to determine the best locations of multi-user facilities in general and intermodal terminals in particular. This was achieved by an entropy framework, which comprises an objective function capturing all possible states of the decision variables subject to a set of constraints representing all known information about the system of interest. In the intermodal terminal location application, the objective function captures all possible states of modal flows and the constraints comprised a linked facility location problem, a mode choice problem and a cargo production/distribution problem. The study has shown the equivalence between entropy and shipper welfare and thus, by maximising entropy, we maximise shippers' welfares. The framework can in general be expanded to include other relevant problems, which for the purpose of this study can include the problem of empty container distribution.

As a way of generalisation, the framework is applicable to locating a facility whose usage is determined by the choice outcomes of many individual decision makers. The proposed framework is also suitable for applications with the following features; first, applications where it is almost impossible to track every potential user of the facility and extract the factors governing his/her decision making process; second, available data for modelling are aggregate in nature or are both aggregate and disaggregate with each explaining some aspect of the system under consideration. Lack of data in sufficient detail and quality for modelling is common in the freight sector as potential facility users, shippers, carriers or other logistics providers are generally unwilling to provide data at the level of detail required to fully understand their choice processes, due to fear of losing competitive advantage of their businesses; third, acknowledging the fact that even if we have all the information about each decision maker, many of them cannot be quantified and included in the modelling process. These and other

related factors make the proposed entropy framework for determining facility location and usage more desirable.

The second important contribution within the framework is the general construction of the mode (Chapter 4 and 5) and cargo production and distribution (Chapter 6) problems and how they can be embedded within the entropy framework as constraints. The solution to the entropy model are probability distributions describing aspects of decision makers' behaviour. For example, the mode choice component describes modal decisions of the decision makers in an aggregate fashion, whilst the distribution model describes the choice of cargo destinations. Several properties of the entropy model were presented in a form of propositions including a general method of dealing with capacity constraint when constructing the probability distributions.

The third contribution relates to the development of algorithms for solving the formulated problems. Although several solution approaches may exist, the study exploit the structure of the framework and found that the formulated problem can be solved efficiently by separating the location aspect of the problem from the demand/usage part (also called the cargo flow problem) of the problem using Lagrangian relaxation technique. The study shows that the cargo flow part of the problem can then be converted into a nested logit choice model to explain the choice behaviour of facility users and the parameters governing this model can be estimated using the proposed modified Bregman's algorithm. To solve the overall model, two main general solution algorithms were proposed; complete enumeration and a more efficient entropic greedy heuristic algorithm with Bregman's algorithm embedded in both for solving the cargo flow problem.

The complete enumeration algorithm was proposed to deal with small to medium sized problems and proved to be very useful and feasible for locating terminals to serve the import and export container market. It also provided a benchmark for gauging the quality of the proposed heuristic. The geographical region making up the regional intermodal market are usually large and encompass a whole country or several countries or economic regions. The entropic greedy heuristic was developed to solve large problems of this kind. The study shows that the computational time of this algorithm grows only linearly with the number of terminals to locate making it very efficient for solving large problem instances. As a heuristic, there was no formal proof of it returning an optimal solution for every instance, but the extensive

numerical examples show that the quality of solutions are comparable to those found using complete enumeration.

The fourth contribution deals with the implementation of a case study based on the proposed model. Specifically, the model was used to determine the best places in Sydney GMA to locate terminals to serve the import containerised market. The full model comprised linked location model, mode choice model and the cargo distribution model. The study found Eastern Creek as the most promising location for the development of intermodal terminal in the study area, with container destinations in Blacktown as its key markets. Other significant catchment areas were found to be Penrith, Paramatta and Fairfield. The choice of mode was governed by the generalised cost and together with the shadow price and sensitivity parameter form the utilities of the available modes where for each origin-destination a mode with higher utility has higher probability of been used for the transport task. The distribution of cargo in the study area was found to be governed by four main factors; access to manufacturing activities, access to warehousing and storage activities, access to key markets, which in the study represent the two central business districts – Sydney and Paramatta – and finally access to multiple modes of transport, which is a combined measure of access by road and access by intermodal transport.

It is important to note that the data for the case study were derived from secondary sources and the source of each data set is provided in the thesis. This study is not responsible for the quality or accuracy of these data as several attempts made to verify the accuracy of these data proved futile. The verification process includes comparing data from several sources and contacting representatives of relevant bodies but none was satisfactory. Thus, the results and conclusions reached by this empirical analysis are conditioned on the assumption that the data are accurate. Readers must therefore be cautious about wider applications of the results.

8.4   Limitations and future direction

This thesis has developed a policy-oriented and flexible model to help develop terminals to promote shipper welfare. Important features of the model and how it works in practice have been illustrated through a case study implementation. The study so far has not considered the movements and management of empty containers in the hinterlands. Proper management of empty containers is generally considered as one of the important factors driving the

development and usage of intermodal terminals (AHRCR 2007; Meyrick 2006). In addition to the empties providing extra revenue for the terminal, the terminals can serve as a market for empty containers where importers can return (off-hire) empties and exporters can hire empties for the export of their products. The model can be readily extended with appropriate constraints to also capture the empty container movements in the study area. The extension can be done by either extending existing constraints to also capture the flows of empties and/or add new set of constraints to both reflect the flow of empties and their relationships with non-empty container flows.

Another important extension that can benefit the literature is to extend the model or develop another model to determine the optimal capacity to create at the located terminals in the opening year and subsequent years as demand grows. This problem is generally known in the literature as capacity expansion (or staging) problem. The objective of this problem is to determine how much capacity to add and when. One significant reason for considering such a problem is that the cost of adding capacity typically comprises a fixed cost and a variable cost, which is a function of the amount of capacity to create. Creating large capacities at a given time benefits from economies of scale, and hence a reduction in overall cost and although the capacity may not be needed immediately it is expected to be needed eventually to meet an increased future demand. However, creating excess unused capacity may tie up much needed capital that could yield more returns when invested elsewhere. The outcome of this model will also help to plan the capacity expansion more efficiently as capital and other administrative requirements can be made in good time before the terminal reaches it capacity. The forecast demand for the located terminal(s) at any given year can be estimated using the cargo flow part of the problem (converted into a nested logit of mode, production and distribution models) given that the inputs to the model can be forecast/known for the given year. The forecast demand are in turn used to determine the required capacity for the given year.

In an application context, the case study did not include assigning the demands by modes from the entropy model to the transport network to determine potential changes in network conditions due to the located terminals and how these changes can alter the demand and possibly the optimal locations of the terminals. This process will also help to identify the roads, especially connector roads, that need expanding to accommodate the extra growth in traffic and where such expansions are not physically possible due to local opposition or lack of available affordable lands, the location decisions can be altered. As discussed in Section 6.5,

the demand for each mode can be converted into matrices for assignment. These matrices can then be added to matrices from other non-containerised modes such as passenger cars and busses and then assigned to the road network. It is important to note that the model developed is not restricted or limited by any particular transport model or assignment method, although an opportunity exists to explore the effects of several assignment methods (for a classification, see Bliemer et al. 2017) on location decisions and demands. In practice, any existing transport model of the study area can be used for this exercise. The transport model provides as output the travel times and distances, which reflect the prevailing network conditions, which then go into the proposed entropy model as inputs and outputs matrices for assignment to the transport network.

Furthermore, implementing the model to locate terminals to serve the regional market and/or both the regional and metropolitan markets will be desirable. A case study that allows these three scenarios (metropolitan, regional or both) to be investigated can benefit both the literature and practice. The outcome of this investigation as illustrated in Chapter 6 can lead to the selection of locations that are neither the optimal locations for the metropolitan nor the regional market but best for serving both markets.

Chapter 9 References

Abdinnour-Helm, S., 1998. A hybrid heuristic for the uncapacitated hub location problem. European Journal of Operational Research 106 (2–3), 489–499.

Access Economics and Maunsell, 2003. Port Botany expansion, trade and capacity study, prepared for Sydney Port Corporation. Available from http://www.sydneyports.com.au/__data/assets/pdf_file/0015/2751/App_D.pdf [26 July, 2014]

Adler N. 2005. The effect of competition on the choice of an optimal network in a liberalized aviation market with an application to western Europe, *TranspSci*, vol. 39, pp 58–72

Adler N., Smilowitz K. 2007. Hub-and-spoke network alliances and mergers: price-location competition in the airline industry, Transportation Research Part B, vol 41, pp 394–409.

Abdinnour-Helm, S., 2001. Using simulated annealing to solve the p-hub median problem. International Journal of Physical Distribution & Logistics Management 31 (3), 203–220.

Alumur, S. A., Kara, B. Y., & Karasan, O. E. (2009). The design of single allocation incomplete hub networks. Transportation Research Part B, 43, 936–951.

Alumer, S., Kara, B. 2008. Network hub location problems: The state of the art, *European Journal of Operational Research*, vol. 190, pp 1-21.

Anas, A., 1981. Discrete choice theory, information theory and the multinomial logit and Gravity models, Transp Res B, Vo. 17B, No I. pp 13-23.

Arnold, P., Peeters D., Thomas, I., 2004. Modelling a rail/road intermodal transportation system, *Transportation Research E*, vol. 40, pp. 255-270.

Arnold, P., Marchand, H., Peeters, D., Thomas, I. 2001. Pour une localisation optimale des centres de transbordement intermodaux entre r_eseaux de transport: formulation et extensions, *The Canadian Geographer*, vol. 45, pp. 427-436.

Aros-Vera F., Marianov, V., Mitchell, J.E., 2013, 'p-Hub approach for the optimal park-and-ride facility location problem', European Journal of Operational Research, 226, pp 277–285.

ATRI (America transport research institute), 2014. Critical issues in the trucking industry 950 North Glebe Road Arlington, VA 22203. Retrieved: 21 Oct, 2016, http://www.intrucking.org/uploads/3/7/6/7/37679031/atri_2014_top_industry_issues_report_final.pdf

Australian Bureau of Statistics 2001. 2001 Census of Population and Housing - Geographic Areas. Available from http://www.abs.gov.au [March, 2016]

Australian Bureau of Statistics 20011. 2001 Information Paper: Experimental Statistics on International Shipping Container Movements. Available from http://www.abs.gov.au [March, 2015]

Australia's House of Representative Commettee report (AHRCR), 2007. *Intermodal Facilitie* . Available from <www.aph.gov.au>. [20 March, 2014]

Aykin, T., 1995. The hub location and routing problem. European Journal of Operational Research 83, 200–219

Aykin, T., 1990. On a quadratic integer program for the location of interacting hub facilities. European Journal of Operational Research 46 (3), 409–411.

Aykin, T. 1988. On the location of hub facilities, *Transportation Science,* vol. 22, No. 2, pp. 155–157.

Ayres, R. and Kneese, A. 1969. Production, consumption, and externalities, The American Economic Review 59(3): 282–297.

Batty M., 2010. Space, Scale and Scaling in Entropy Maximising, Geographical Analysis 42 (2010) 395–421.

Bayes T., 1763. An Essay towards solving a Problem in the Doctrine of Chances, Phil. Trans. Roy. Soc, 330, (1763).

Beasley JE., 1990. OR-Library: distributed test problems by electronic mail, Journal of the operational research society 41 (11), 1069-1072.

Bell, MGH., 1995. Alternative to Dial's logit assignment algorithm, *Transportation Research B* 29B(4) 287–295.

Bell, M.G.H. 1983, 'The estimation of an origin–destination matrix from traffic counts', *Transportation Science* Vol. 17, No. 2, pp. 198–217.

Benth, F.E., Groth, M., 2009. The minimal entropy martingale measure and numerical option pricing for the Barndorff-Nielsen-Shephard stochastic volatility model. Stoch. Anal. Appl. 2009, 27, 875–896.

Bera, A.K.; Park, S.Y., 2009. Optimal Portfolio Diversification Using the Maximum Entropy Principle. In Proceedings of the Second Conference on Recent Developments in the Theory, Method, and Applications of Information and Entropy Econometrics, Washington DC, WA, USA, 23 September 2009.

Bernhofen DM, El-Sahli Z and Kneller R, 2014. Estimating the effects of the container revolution on the world trade. CESifo Working Paper Series 4136. CESifo, Centre for Economic Studies and Ifo Institute. Munich.

Bernoulli, D. (1777), Mem. St Petersburg Acad., Acta Acad. Petrop., pp. 3–33; English translation, 'The most probable choice between several discrepant observations', Biometrika, 45, 1–18 (1961). Reprinted in Pearson & Kendall (1970), pp. 157–167. Questions taking the mean of several observations; discussed in Chapter 7.

Bureau of Infrastructure, Transport and Regional Economics, 2016, Why short-haul intermodal rail services succeed, Report 139. Retrieved: 21 November, 2016, https://bitre.gov.au/publications/2016/rr_139.aspx.

Bureau of Infrastructure, Transport and Regional Economics, 2014, Why short-haul intermodal rail services succeed, Report 139. Retrieved: 21 November, 2016, https://bitre.gov.au/publications/2016/rr_139.aspx

Bliemer, M.C.J., Raadsen, M.P.H., Brederodeb L., Bell, M.G.H.,Wismans L., Smith M., 2017.Genetics of traffic assignment models for strategic transport planning, Transport Reviews, vol.37:1, pp. 56-78.

Bliemer, M.C.J., Raadsen, M.P.H., Smits, E-S. Zhou, B. Bell, M.G.H., 2014. Quasi-dynamic traffic assignment with residual point queues incorporating a first order node model. Transportation Research Part B: Methodological, 68, pp.363–384.

Boland, N., Krishnamoorthy, M., Ernst, A.T., Ebery, J., 2004. Preprocessing and cutting for multiple allocation hub location problems. European Journal of Operational Research 155 (3), 638–653

Boltzmann, L., 1872. *Lectures on Gas Theory*, translated by S.G. Brush, University of California Press, Berkeley.

Bontekoning, Y.M., Macharis, C, and Trip, J.J. 2004. Is a new applied transportation research field emerging? A review of intermodal rail-truck freight transport literature, *Transport Research Part A*, vol. 38, No. 1, pp. 1-34.

Boyd, S., Vandenberghe, L., 2009. Convex Optimisation, Cambridge University Press, Cambridge, United Kingdom.

Broeze, F., 2002. The globalisation of the oceans: Containerisation from the 1950s to the present. St. Johns, NF, Canada: International Maritime Economic History Association.

Bretthorst, L., 1988. Bayesian Spectrum Analysis and Parameter Estimation, III Lecture Notes in Statistics 48, Springer-Verlag, New York, New York, 1988.

Bretthorst L., 1990. An introduction to parameter estimation using Bayesian probability theory. In: P. F. Fougere (ed.), Maximum Entropy and Bayesian Methods, 53-79.

Brouwer, L. E. J., 1910. Uber eineindeutige stige Transformationen von Flachen in Sich, Mathematische Annalen 67, 176–80.

Buchen, P.W. Kelly, M., 1996. The maximum entropy distribution of an asset inferred from option prices. J. Financ. Quant. Anal. 1996, 31, 143–159.

Burden R.L., Faires J.D., 2010. *Numerical Analysis: Ninth edition, Brooks/Cole Cengage Learning.*

Burg, J.P.,1978. The Relationship Between Spectra and Maximum Likelihood Spectra, Analysis, Edited by: Childers, D.G., IEEE pp. 132-133, 1978.

Cameron A.C., Trivedi P.K., 2013. *Regression Analysis of Count Data: Second edition,* Cambridge University Press.

Campbell, J., and O'Kelly, M.E., 2012. Twenty-Five years of hub location research, *Transportation Science*, vol. 46, pp. 153-169.

Campbell, J.F., Ernst, A.T., Krishnamoorthy, M., 2005a. Hub arc location problems: Part I – Introduction and results. Management Science 51 (10), 1540–1555.

Campbell, J.F., Ernst, A.T., Krishnamoorthy, M., 2005b. Hub arc location problems: Part II – Formulations and optimal algorithms. Management Science 51 (10), 1556–1571.

Campbell, J.F., 1996. Hub location and the p-hub median problem. Operations Research 44 (6), 1–13.

Campbell, J. 1994. Integer programming formulations of discrete hub location problems, *European Journal of Operational Research,* vol. 72, pp. 387–405.

Campbell, J.F., 1992. Location and allocation for distribution systems with transshipments and transportation economies of scale. Annals of Operations Research 40, 77–99.

Caticha A., 2012. *Entropic inference and the foundations of Physics*, Dept. of Physics, University at Albany-SUNY.

Caticha A., Giffin A., 2006. Updating Probabilities, Bayesian Inference and Maximum Entropy Methods in Science and Engineering, ed. by A. Mohammad-Djafari, AIP Conf. Proc. 872, 31 (2006).

225

Chandler, D., Zou, B., Bernstein, S., Benedict, A., Shankar, P., 2016. Metrics for cargo-oriented development: Measuring the dynamic interaction of community economic development & freight system efficiency, TRB 95th Annual Meeting, Washington DC, January, 2016.

Chen, J.F., 2007. A hybrid heuristic for the uncapacitated single allocation hub location problem. Omega 35, 211–220.

Clarke M., Wilson A. G., 1985. The Dynamics of Urban Spatial Structure: The Progress of a Research Programme, Transactions of the Institute of British Geographers, Vol. 10, No. 4 (1985), pp. 427-451.

Clausius, R., 1865, "Ueber verschiedene für die Anwendung bequeme Formen der Hauptgleichungen der mechanischen Wärmetheorie", *Annalen der Physik*, 125: 353–400.

Contreras I., Diaz J. A., Fernandez E., 2011. Branch-and-price for large-scale capacitated hub location problems with single assignment. INFORMS Journal on Computing, 23:41–55.

Cox R.T., 1946. Probability, Frequency and Reasonable Expectation, Am. J. Phys. 14, 1.

Crainic T.G., Kim K.H.,2006. Intermodal transportation. In: C. Barnhart and G. Laporte (Eds.), Handbook in OR & MS, Vol. 14.

Cudahy, B.J., 2006. The container revolution: Malcolm McLean's 1956 innovation goes global. TR News 246, 5-9.

Donoso, P., Grange L., 2010. A Microeconomic Interpretation of the Maximum Entropy Estimator of Multinomial Logit Models and Its Equivalence to the Maximum Likelihood Estimator, Entropy 2010, 12, 2077-2084.

Department of Finance and Deregulation, 2011. *Information paper on Moorebank Intermodal Terminal Project*, Moorebank project office, The Australian Government.

Diaconis P., Zabell S. L., 1982. Updating Subjective Probabilities, J. Am. Stat. Assoc. 77, 822 (1982).

Drezner, Z., 1984. The p-Centre Problem-Heuristic and Optimal Algorithms, The Journal of the Operational Research Society, Vol. 35, No. 8, pp. 741-748.

Ebery J., 2001. Solving large single allocation p-hub problems with two or three hubs, European Journal of Operational Research 128 (2), pp 447–458.

Eiselt H.A., Marianov V.A. 2009. Conditional p-hub location problem with attraction functions, Computer and Operations Research vol. 36, pp 3128–35.

Elhedhli, S., Hu, F.X., 2005. Hub-and-spoke network design with congestion. Computers and Operations Research 32, 1615–1632.

Ellison, R.B., 2014. Understanding dynamic responses to mitigation policies for intra-urban road freight emissions, PhD. Thesis, The Institute of Transport and Logistics Studies (ITLS), University of Sydney.

Ellison R, Teye C and Hensher DA 2017 'Commodity-based heavy vehicle model for Greater Sydney', *5th International Choice Modelling Conference*, Cape Town, South Africa, 5th April 2017

Ernst, A.T., Jiang, H., Krishnamoorthy, M., 2005. Reformulations and computational results for uncapacitated single and multiple allocation hub covering problems. Unpublished Report, CSIRO Mathematical and Information Sciences, Australia.

Ernst, A., Hamacher, H., Jiang H., Krishnamoorthy, M., Woeginger, G., 2002. Uncapacitated single and multiple allocation p-hub center problems. Unpublished Report, CSIRO Mathematical and Information Sciences, Australia.

Ernst, A.T., Krishnamoorthy, M., 1999. Solution algorithms for the capacitated single allocation hub location problem. Annals of Operations Research 86, 141–159.

Ernst, A.T., Krishnamoorthy, M., 1998a. Exact and heuristic algorithms for the uncapacitated multiple allocation p-hub median problem. European Journal of Operational Research 104, 100–112.

Ernst, A.T., Krishnamoorthy, M., 1998b. An exact solution approach based on shortest-paths for p-hub median problems. Informs Journal on Computing 10 (2), 149–162.

Ernst AT., Krishnamoorthy M., 1996. Efficient algorithms for the uncapacitated single allocation p-hub median problem, Location Science Vol 4, No 3, pp 139-154, 1996.

Evans, S. P., 1976. Derivation and Analysis of Some Models for Combining Trip Distribution and Assignment, Transportation Research B 10, 105–22.

Evans, S. P., 1973. A Relationship Between the Gravity Model for Trip Distribution and the Transportation Model of Linear Programming, Transportation Research 7, 39–61.

European Commission, 2011. *WHITE PAPER - Roadmap to a Single European Transport Area – Towards a competitive and resource efficient transport system*, pp. 30, Brussels.

European Commission, 2006. Intermodal freight terminals- In search of efficiency to support intermodality growth, Brussels.

European Conference of Ministers of Transport, United Nations Economic Commission for Europe & Statistical Division and European Union Eurostat. (1997), Glossary for transport statistics. http://www1.oecd.org/cem/online/glossaries/.

Farahani RZ., Hekmatfar M., Fahimnia B., Kazemzadeh N., 2014. Hierarchical facility location problem: Models, classifications, techniques, and applications, *Computers & Industrial Engineering*, vol.68, pp. 104-17.

Fisk, C., 1985. Entropy and information theory: are we missing something? Environment and Planning A, 1985, volume 17, pages 679-710.

Feldman, Richard, and Marilyn Gross, Transportation Expressions, Washington, DC: Bureau of Transportation Statistics, U.S. Department of Transportation, 1996.

Frieden, B.R., 1972. Restoring with Maximum Likelihood and Maximum Entropy, J. Opt. Soc. Am. 62, 511.

Garey, M. R. and D. S. Johnson, 1979. Computers *and Intractability: A Guide to the Theory of NP-Completeness*, W.H. Freeman and Co., New York.

Gauss, C.F., 1809. Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium. Hamburg: Perthes et Besser. English translation by C.H. Davis, reprinted by Dover, New York (1963).

Ghane-Ezabadi, M., Vergara, H., A., 2016. Decomposition approach for integrated intermodal logistics network design, Transport Research Part E, 89, 53-69.

Gibbs J. W., 1902. Elementary Principles in Statistical Mechanics, (1902); Reprinted in Collected works and commentary, (Yale University Press 1936), and (Dover Publications, Inc. 1960).

Gilboa, I., Lieberman O., Schmeidler D., 2008. On the definition of objective probabilities by empirical similarity, Synthese, Springerlink.com, DOI 10.1007/s11229-009-9473-4.

Giffin A., 2008. Maximum entropy: The universal method for inference, PhD. Thesis, College of Arts & Sciences, Department of Physics, University at Albany, State University of New York.

Golob, T.F., Regan, A.C., 2000. Impacts of highway congestion on freight operations: Perceptions of trucking industry managers, Transportation Research Board 79th Annual Meeting, Washington DC, January, 2000.

Hakimi SL., 1965. Optimum distribution of switching centers in a communication network and some related graph theoretic problems. Oper Res 13:462–475.

Hakimi, S. L. 1964. Optimum Location of Switching Centers and the Absolute Centers and Medians of a Graph, *Operations Research*, vol. 12, pp. 450-459.

Harney, H.L.,2003. Bayesian Inference: Parameter Estimation and Decisions Parameter Estimation and Decisions, *Springer-Verlag Berlin Heidelberg,* New York, pp. 10.

Hensher, D. A., Rose, J. M. and Greene, W. H. 2015. *Applied Choice Analysis*. 2nd edition, Cambridge: Cambridge University Press.

Hensher, D., Figliozzi, M.A.: Behavioural insights into the modeling of freight transportation and distribution systems. Transp. Res. B 41(9), 921–923 (2007).

Hensher, D.A. and T.F. Golob (1998). Searching for policy priories in the formulation of a freight transport strategy: An analysis of freight industry attitudes toward policy initiatives. Transportation Research, E - Logistics and Transportation, in press.

Hensher, D.A., G. Chow and J. King (1996). Assessment of freight-related industry needs, perceptions and expectations in NSW, Parts I and II, Report prepared for the Roads and Traffic Authority of NSW, Institute of Transport Studies, University of Sydney.

Hills, P.J., 1996. What is induced traffic? Transportation 23: 5-16, 1996.

Hutt, E.E., 2011, 'Stochastic Hub and Spoke Networks', PhD. Thesis, Homerton College, Judge Business School, University of Cambridge.

Hyman G.M., 1969. The calibration of trip distribution models. Environment and Planning 1, 105-112.

IPART (Independent Pricing and Regulatory Tribunal of New South Wales 2008, Reforming Port Botany's links with inland transport), 2008. Review of the interface between the land transport industries and the stevedores at Port Botany. Final Report of inquiry. Retrieved: 21 Feb, 2015,.https://www.ipart.nsw.gov.au/Home/Industries/Special-Reviews/Reviews/Port-Botany

Ishfaq R, Sox C., 2012. Design of intermodal logistics networks with hub delays, European Journal of Operational Research, 220, 629-641.

Ishfaq R, Sox C., 2011. Hub location–allocation in intermodal logistic networks, European Journal of Operational Research, 210, 213–230.

Jaynes E. T. 1988. The relationship of Bayesian and maximum entropy methods. In: Erickson G.J., and Smith C.R. (eds), Maximum Entropy and Bayesian Methods in Science and Engineering (Vol 1), 25-29.

Jaynes E. T. 1985. Bayesian Methods: General Background, in Maximum Entropy and Bayesian Methods in Applied Statistics, J. H. Justice (ed.) (Cambridge UP, 1985) and at http://bayes.wustl.edu.

Jaynes, E. T.,1980. Marginalization and Prior Probabilities. In: Bayesian Analysis in Econometrics and Statistics, A. Zellner, ed., North-Holland Publishing Company, Amsterdam, 1980.

Jaynes, E.T., 1957. Information Theory and Statistical Mechanics, *Physical Review*, vol. 106, no. 4, pp. 620-630.

Jeffreys H., 1939. Theory of Probability, Oxford U. Press, Oxford.

Kara, B.Y., Tansel, B.C., 2003. The single-assignment hub covering problem: Models and linearizations. Journal of the Operational Research Society 54, 59–64.

Kara, B.Y., Tansel, B.C., 2000. On the single-assignment p-hub center problem. European Journal of Operational Research 125, 648–655.

Kaufman L., Broeckx F., An algorithm for the Quadratic Assignment Problem using Benders' decomposition, European J. Oper. Res. 2 (1978) 204-211.

Khamsi, M. A., and W. A. Kirk. 2001. An Introduction to Metric Spaces and Fixed Point Theory. New York: John Willy and Sons, INC. Press

Kim H, O'Kelly ME., 2009. Reliable p-hub location problems in telecommunication networks. Geographical Anal. 41(3):283–306.

Klincewicz, J.G., 1996. A dual algorithm for the uncapacitated hub location problem. Location Science 4 (3), 173–184

Klincewicz, J.G., 1992. Avoiding local optima in the p-hub location problem using tabu search and GRASP. Annals of Operations Research 40, 283–302.

Klincewicz, J.G., 1991. Heuristics for the p-hub location problem. European Journal of Operational Research 53, 25–37.

Kolmogorov, A. N. (1933), Grundbegriffe der Wahrscheinlichkeitsrechnung, Ergebnisse der Math. (2), Berlin; English translation, Foundations of the Theory of Probability, Chelsea Publishing Co., New York (1950). Described in Appendix A.

Kullback, S., and Leibler, R. A., 1951. On information and sufficiency. *Annals of Mathematical Statistics* 22, 79-86.

Lamond, B., and Stewart, N. F., 1981. Bregman's Balancing Method, Transportation Research Part B. vol. 15, pp. 239-248.

Laplace, P. S. (1812), Th´eorie Analytique des Probabilit´es, 2 vols., Courcier Imprimeur, Paris; 3rd edition with supplements, 1820. Reprinted in Laplace (1878–1912), Vol. 7.

Laplace, P. S., 1772, M´emoire sur les solutions particuli`eres des ´equations diff´erentielles et sur les in´egalites s´eculaies des plan`ates Oeuvres completes, 9, 325

Levinson, M., 2006. The Container Shipping and the Economy: Stimulating Trade and Transformations Worldwide. TR News 246, 11-14.

Lin C. C, Lee S. C., 2010. The competition game on hub network design, Transportation Research Part B, vol. 44, pp 618–29.

Lin, C-C., Chiang, Y-I., Lin S-W, 2014. Efficient model and heuristic for the intermodal terminal location problem, *Computers & Operations Research*, vol. 51, pp. 41-51.

Lin, C. C., Lin, J. Y., & Chen, Y. C., 2012. The capacitated p-hub median problem with integral constraints: An application to a Chinese air cargo network. Applied Mathematical Modeling, 36(6), 2777–2787.

Limbourg, S., Jourquin, B., 2009. Optimal rail-road container terminal locations on the European network. Transportation Research Part E Logist. Transp. Rev. 45, 551-563.

Liu, Y., Liu, C.J., Wang, D.H., 2011. Understanding atmospheric behaviour in terms of entropy: a review of applications of the second law of thermodynamics to meteorology. Entropy 2011, 13, 211–240.

Lium A-G, Crainic TG, Wallace SW., 2009. A study of demand stochasticity in service network design. Transportation Sci. 43(2):144–157. Location Sci. 4(3).

Lüer-Villagra A., Marianov V.A. 2013. Competitive hub location and pricing problem, European Journal of Operations Research, vol. 231, pp 734–44.

Marianov V., Serra D., ReVelle C., 1999. Location of hubs in a competitive environment. European Journal of Operations Research, vol. 114, pp 363–71.

Meyrick, S., 2006. *National Intermodal Study, Prepared for the Department of Transport and Regional Services.* Available from <http://trove.nla.gov.au/work/26851390?q&versionId=32351714>. [18 March, 2014]

McFadden, D. 1974. Conditional logit analysis of qualitative choice behaviour, in P. Zarembka, ed., *Frontiers in Econometrics*, Academic Press, New York, pp. 105–142.

Moseley, G. W., 1973. Lorries and the Environment, Technical report, Department of the Environment, London.

Murchland, J.D. 1977, The multi-proportional problem, Manuscript JDM-263, draft 1, University College London Transport Studies Group

National Cooperative Highway Research Program Report 586. *Rail Freight Solutions to Roadway Congestion-Final Report and Guidebook*, Transportation Research Board, Washington D.D, 2007.

Nierat, P., 1997. Market area of rail-truck terminals: pertinence of the spatial theory', *Transportation Research A*, vol. 31, pp. 109–127.

Norris, B.,1994. Intermodal Freight: an Industry Overview. FHWA, Washington, D.C.

O'Kelly ME, Bryan D., 1998. Hub location with flow economies of scale. Transportation Research Part B 32(8):605–616.

O'Kelly, M.E., Bryan D., Skorin-Kapov D., Skorin-Kapov J., 1996. Hub network design with single and multiple allocation: a computational study, Location Science, Vol. 4, No. 3, pp. 125-138.

O'Kelly, M.E., Skorin-Kapov, D., Skorin-Kapov, J., 1995. Lower bounds for the hub location problem. Management Science 41 (4), 713–721.

O'Kelly, M.E. 1987. A quadratic integer program for the location of interacting hub facilities, *European Journal of Operational Research*, vol. 32, pp. 393–404.

Park, J., Ben-Akiva, M. Boldoc, D., 1995. Railroad Marketing Support System Based On the Freight Choice Model, PhD. Thesis, The Center for Transport Study, MIT.

Patil G P., Rao, C. R..,1986. On discrete weighted distributions and their use in model choice for observed data, communication statistics theory methods., 15(3), 907-918.

Patil G P., Rao, C. R.., 1978. Weighted distributions and size biased sampling with applications to wildlife populations and human families Biometrics, 34, 179 189.

Piyapatroomi, Noppadol and Bunker, Jonathan M. and Ferreira, Luis (2006) Freight Intermodal Terminal Systems for Port of Brisbane, Melbourne and Sydney. Technical Report 2006-1, Transport Research, School of Urban Development, Queensland University of Technology.

Pirkul, H., Schilling, D.A., 1998. An efficient procedure for designing single allocation hub and spoke systems. Management Science 44 (12), 235–242.

Planck M., 1901. On the Law of Distribution of Energy in the Normal Spectrum, Annalen der Physik, 4,553.

Qu, Y., Bektas, T., Bennell, J., 2016. Sustainability SI: Multimode Multicommodity Network Design Model for Intermodal Freight Transportation with Transfer and Emission Costs, Networks and Spatial Economics, 16, 303- 329.

Racunica, I. and Wynter, L., 2005. Optimal Location of Intermodal Freight Hubs Transport Research Part B, 39, 453- 477.

Rahimi, M., Asef-Vaziri, A., Harrison, R., 2008. An inland port location-allocation model for a regional intermodal goods movement system. Maritime Economics and Logistics, 10, 362-379.

RAND Europe (G. de Jong, M. Pieters, A., Daly, I., Graafland, E., Kroes, C., Koopmans) (2005) Using the logsum as an evaluation measure, Literature and Case study, Prepared for AVV Transport Research Centre, Report WR-275-AVV, RAND Europe, Leiden.

Rodrigue J-P., and Slack, B. 2013. Intermodal Transportation and Containerization. In: Rodrigue, J-P, (2013) (ed), *The Geography of Transport Systems*, Third Edition, London.

Jana, P., Roy, T.K., Mazumder, S.K., 2007. Multi-objective mean-variance-skewness model for portfolio optimisation. Appl. Math. Optim. 2007, 9, 181–193.

Sasaki M., Campbell J., Krishnamoorthy M., Ernst A.T., 2014, 'A Stackelberg hub arc location model for a competitive environment', Computers & Operations Research vol. 47, pp 27–41.

Sasaki M., 2005, 'Hub network design model in a competitive environment with flow threshold', J Oper Res Soc Jpn, vol. 48, pp 158–71.

Sasaki M., Fukushima M., 2001, 'Stackelberg hub location problem', J Oper Res Soc Jpn vol. 44, pp 390–402.

Sasaki, M., Suzuki, A., Drezner, Z., 1999. On the selection of hub airports for the airline hub-and-spoke system. Computers & OR 26, 1411–1422.

Savage S.,2008. Flaw of averages. San Jose Mercury News.

Savage, Leonard J. (1950). "The Role of Personal Probability in Statistics." Econometrica, 18, 183–184. (Abstract from a paper presented at the 1949 Econometric Society meetings.)

Shannon, C. E. and Weaver, W., 1949. The Mathematical Theory of Communication, Univ. of llinois Press, Urbana, IL, 1949.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal* 27, 379-423 and 623-656

Schilling D.A., Jayaraman V., Barkhi R., 1993. A review of covering problems in facility location problem, Location Science, 1(1), pp. 25-55.

Sea Freight Council of NSW 2005. Sydney's Intermodal Solution, Report prepared by Sd+D for the Sea Freight Council.

Seth A., Kapur J., 1990. A comparative assessment of entropic and non-entropic methods of estimation. In: P. F. Fougere (ed.), Maximum Entropy and Bayesian Methods, 451-462.

Shipping Australia Limited, 2011. *Metropolitan Intermodal Terminals Study*. Available from <http://shippingaustralia.com.au/publications/Intermodal_Terminals_WEB.pdf > [21 Feb, 2014].

Shore J. E., Johnson R. W. 1980. Axiomatic derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross Entropy", IEEE Trans. Inf. Theory IT-26, 26 (1980); "Properties of Cross-Entropy Minimisation, IEEE Trans. Inf. Theory IT-27, 26(1981).

Skorin-Kapov, D., Skorin-Kapov, J., 1994. On tabu search for the location of interacting hub facilities. European Journal of Operational Research 73, 502–509.

Sim T., Lowe T., Thomas B., 2009. The stochastic p-hub center problem with servicelevel constraints. Computers & Operations Research, 36:3166–3177.

Skorin-Kapov D., Skorin-Kapov J., O'Kelly, M.E.,1996. Tight linear programming relaxations of uncapacitated p-hub median problems, European Journal of Operational Research 94, 582-593.

Slack, B. 1998. Intermodal transportation. In: Hoyle, B., Knowles, R. (Eds.), *Modern Transport Geography*, Wiley, C., Chichester.

Smith, K., Krishnamoorthy, M., Palaniswami, M., 1996. Neural versus traditional approaches to the location of interacting hub facilities. Location Science 4 (3), 155–171.

Sohn, J., Park, S., 2000. The single allocation problem in the interacting three-hub network. Networks 35 (1), 17–25.

Sohn J., Park S., 1998. Efficient solution procedure and reduced size formulations for p-hub location problems, *European Journal of Operational Research 108, pp. 118-126.*

Solomon, M B 2014, "'Siding' up to intermodal. As intermodal traffic grows, more companies look to plant their DC site flags alongside rail facilities", DC Velocity, 6 January

Sorensen, K., Vanovermeire, C., and Busschaert, S. 2012. Efficient metaheuristics to solve the intermodal terminal location problem, *Computers & Operations Research*, vol. 39, pp. 2079-2090.

Stanimirovic J., 2010. A genetic algorithm approach for the capacitated single allocation p-hub median problem, Computing and Informatics, Vol. 29, 2010, 117–132.

Teye, C., Bell, M.G.H., Bliemer, M.C.J., 2016. Urban intermodal terminals: The entropy maximising facility location problem, Transportation Research Board 95th Annual Meeting, Washington DC, January, 2016.

Teye, C., Bell, M.G.H., Bliemer, M.C.J., 2015. Lagrangian Algorithm for Locating Intermodal Container Terminals in Metropolitan Areas. *Conference of Australian Institutes of Transport Research*, Melbourne, Australia, February, 2015.

Topcuoglu, H., Corut, F., Ermis, M., Yilmaz, G., 2005. Solving the uncapacitated hub location problem using genetic algorithms. Computers & OR 32 (4), 967–984.

Train, K. ,2009. *Discrete Choice Methods and Simulation: Second edition,* Cambridge University Press.

Tribus M., 1978. Thirty Years of Information Theory", The Maximum Entropy Formalism, R.D. Levine and M. Tribus (eds.) (MIT Press, Cambridge 1978).

Truesdell, C., 1966. Six Lectures on Modern Natural Philosophy, SpringerVerlag, Berlin.

Tsekeris, T., 2016. Interregional trade network analysis for road freight transport in Greece, Transportation Research Part E 85 (2016) 132-148.

UNCTAD (United Nations Conference on Trade and Development), 2013. Review of Maritime Transport. UNCTAD/RMT/2013: New York and Geneva.

United Nations (1992) Transportation Financial/Economic Planning Model, Volume 3, Inland Container Deports with Model Split Module, ST/ESAP/1240.

Usta, I.; Kantar, Y.M., 2011. Mean-variance-skewness-entropy measures: a multi-objective approach for portfolio selection. Entropy 2011, 13, 117–133.

Wagner B., 2008. A note on location of hubs in a competitive environment', European Journal of Operations Research, vol. 184, pp 57–62.

Wagner, B., 2004. Model formulations for hub covering problems. Working paper, Institute of Operations Research, Darmstadt University of Technology, Hochschulstrasse 1, 64289 Darmstadt, Germany.

Wang, J. J., & Cheng, M. C. (2010). From a hub port city to a global supply chain management center: A case study of Hong Kong. Journal of Transport Geography, 18, 104–115.

Wang J., 2007. The β-reliable median on a network with discrete probabilistic demand weights. Operations Research, 55:966–975.

Williams H.C.W.L, 1975. A comparison of some calibration techniques for double constrained models with an exponential cost function, Transportation Research, vol. 10, pp 91-104.

Williams H.C.W.L, 1977. On the formation of travel demand models and economic evaluation measures of user benefit, Environment and Planning A, vol. 9, pp 285-344.

Williams P. M., 1980. Bayesian Conditionalization and the Principle of Minimum Relative Information, Brit. J. Phil. Sci. 31, 131 (1980).

Wilson, A. G., 2010. Entropy in Urban and Regional Modelling: Retrospect and Prospect, Geographical Analysis 42 (2010) 364–394 r 2010.

Wilson, A. G., 1972. Some Recent Developments in Micro-Economic Approaches to Modelling Household Behaviour, with Special Reference to Spatio-Temporal Organisation, In Papers in Urban and Regional Analysis, 216–36, edited by A. G. Wilson. London: Pion.

Wilson, A. G., 1971. A Family of Spatial Interaction Models and Associated Developments, Environment and Planning 3, 1–32.

Wilson, A. G., 1970. Entropy in Urban and Regional Modelling, London: Pion.

Wilson, A. G., 1967. A Statistical theory of spatial distribution models, Transportation Research. Vol. 1, pp. 253-269.

Yang, T. H. (2009). Stochastic air freight hub location and flight routes planning. Applied Mathematical Modeling, 33, 4424–4430.

Zellner, A., 1987. An Introduction to Bayesian Inference in Econometrics, John Wiley and Sons, New York, 1971; Second edition 1987.

Zhang, M., Wiegmans, B., Tavasszy, L., 2013. Optimisation of multimodal networks including environmental costs: a model and findings for transport policy, Computers in industry, 64, 136-145.