Transferability Approach for Freight Data Measurement and Model Development in Resource-Constrained Regions

THESIS

Submitted in partial fulfilment

of the requirements for the degree of

DOCTOR of PHILOSOPHY

by

BALLA BHAVANI SHANKAR

ID No. 2019PHXF0418H

Under the Supervision of

Prof. Prasanta Kumar Sahu

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI

2024

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

CERTIFICATE

This is to certify that the thesis titled **Transferability Approach for Freight Data Measurement and Model Development in Resource-Constrained Regions** submitted by **Mr. Balla Bhavani Shankar** ID No **2019PHXF0418H** for the award of Ph.D. of the institute embodies original work done by him under my supervision.

Prasant Sahu

Signature of the Supervisor
Name in capital letters: PROF. PRASANTA KUMAR SAHU
Designation: Associate Professor
Date: 16-02-2024

DECLARATION

This is to certify that the thesis titled **Transferability Approach for Freight Data Measurement and Model Development in Resource-Constrained Regions** is based on my own research work and has been carried out under the guidance and supervision of **Prof. Prasanta Kumar Sahu**, Associate Professor, Department of Civil Engineering, BITS Pilani, Hyderabad Campus.

The data and information which I have used from various sources have been duly acknowledged. I declare that this work has not been previously submitted by me to any other university/institute for the award of any other degree or diploma.

Date:16-02-2024 **Signature of the Candidate**

Place: Hyderabad **ID No. 2019PHXF0418H**

Dedicated

to

My Mentor

Prof. Prasanta Kumar Sahu

&

All My

Teachers

ACKNOWLEDGEMENTS

I am profoundly thankful and hold great respect for **Prof. Prasant Kumar Sahu**, who was my Ph.D. supervisor. Being his second Ph.D. student has been a true blessing. He has been an exceptional teacher, imparting to me a wealth of knowledge and expertise in academic research. I am grateful for the time he devoted, the ideas he shared, and the invaluable advice he provided, greatly enriching my Ph.D. journey making it both productive and intellectually stimulating. His dedication and commitment to hard work have been a remarkable example of what can be achieved through patience and persistent efforts. Throughout my doctoral studies, he has been a source of constant encouragement, offering a balanced perspective, delightful companionship, and a profusion of outstanding guidance and instruction. I am sincerely grateful for his profound impact on my academic and personal growth. His mentorship has been instrumental in shaping my research abilities and fostering my development as a researcher.

I extend my heartfelt gratitude to the esteemed members of the departmental advisory committee (DAC): **Prof. Sridhar Raju** and **Dr. Durgesh Vikram**. Their generous commitment of time and active participation in my progress seminars provided me with invaluable feedback, enabling me to realise my research potential. I would also like to thank **Dr. Bandhan Bandhu Majumdar**, NIT Durgapur, for his valuable input on my research. Additionally, I would like to express my thanks to **Dr. Agnivesh Pani**, IIT (BHU) Varanasi, for his invaluable contributions and insights, which greatly enhanced the quality and depth of my research. His unwavering support and constructive input have been instrumental in shaping the success of my academic journey, and I am genuinely grateful for his guidance.

I am grateful to my dear friend, **Mr. Aitichya Chandra**, for his support throughout my research journey. Additionally, I would like to convey my heartfelt thanks to my brother, **Mr. Sai Naveen**, for his continuous assistance throughout my doctoral studies. I also thank my colleagues and dear friends – **Dr. Mallikarjun Patil**, **Mr. Siddardha Koramati,** and **Mr. Naveed Farooz -** for their unconditional support. I also extend my gratitude to all my friends who have supported me during my Ph.D. journey. Additionally, I would like to express my gratitude to the establishment owners, whose active participation and cooperation during the survey provided the foundational data for this thesis. I am also profoundly thankful to all the survey enumerators, including my friends and students, whose dedicated efforts contributed significantly to the data collection process. Their commitment and assistance have been invaluable to the success of this research. Furthermore, I acknowledge the anonymous reviewers who diligently reviewed my papers and provided thought-provoking and comprehensive comments. Their feedback played a crucial role in reshaping my perspectives and enhancing the quality of my work.

In conclusion, I am deeply grateful to my parents, namely, my mother, **Ms. Vara Lakshmi**, and my father, **Mr. Rambabu**. Their limitless and unwavering support has been the bedrock of my academic voyage. I'd also like to express my heartfelt gratitude to my late grandmother, **Ms. Bhadramma**, and my late grandfather, **Mr. Nooka Raju**, for inspiring me to pursue doctoral studies.

ABSTRACT

In the contemporary era, the scarcity of freight demand models is evident compared to the prevalence of activity-based passenger demand models. These models' pivotal components, namely freight generation (FG) and freight trip generation (FTG) models, play a crucial role in predicting truck patterns and the impacts of truck movements. Accurate predictions of FG and FTG through these models are of utmost significance to transportation authorities and planning agencies. In addition, these models are essential tools for forecasting local, regional, state, and national freight movements, facilitating facility planning, designing policy interventions, and implementing investment schemes. Despite their significance, the shortage of freight models compared to passenger demand models arises due to the inadequate availability of freight data at aggregated and disaggregated levels. The scarcity of data is attributed to various factors, such as the data's proprietary nature, limited resources, and the absence of commodity flow surveys due to the lack of public freight data sources. Moreover, political priorities also contribute to the limitations in gathering comprehensive freight data. The scarcity of modelling efforts is particularly noticeable at the disaggregate level, which is vital for understanding the quantitative aspects of freight movements. Nonetheless, one helpful approach known as spatial transferability is practised to overcome the data scarcity constraints. This approach involves applying a model developed for one region to another, which can help address some data limitations and extend the utility of existing freight models. In order to overcome these real-time problems, this thesis contributes to freight literature on three fronts: (i) model estimation (Which modelling methodology accurately estimates FG at industrial, regional and state levels? Which establishment typology increases the accuracy of FG models?), (ii) model application (How effective are urban models using various transfer methods when utilised in suburban regions? How do various modelling methodologies perform in the context of spatial transferability? What is the influence of sample size on spatial transferability of the FG models?), and (iii) measurement (How do we minimise the efforts of freight data collection for the regions that are constrained in terms of resources?). This thesis attempts to augment the literature base in freight transportation planning by addressing these research questions, as explained below.

First, different modelling approaches were employed to develop freight generation (FG) models. The limitations encountered in traditional parametric modelling methods, such as ordinary least squares (OLS), necessitated diverse approaches. These limitations arise from the realistic possibility of violating fundamental assumptions like linearity or data distribution. The complexity of the problem is further compounded by the need to model multiple industry sectors within the freight system and the potential use of different explanatory variables. Currently, there is a lack of clear guidance on selecting the appropriate modelling methodology for a given case. Although non-parametric models address the limitations of traditional FG models, they have rarely been investigated for their predictive capabilities. This study aims to investigate this frequently researched question by comparing the performance of various modelling methodologies for predicting FG. The methods evaluated include OLS, weighted least squares (WLS) regression, robust regression (RR), seemingly unrelated regression (SUR), multiple classification analysis (MCA), and Support Vector Regression (SVR). Models using these approaches were developed at the state, regional (encompassing urban and suburban areas), and industrial levels. Subsequently, the models were validated to determine the most suitable modelling approach for each level. The validation results indicate that non-parametric SVR models are preferable for developing FG models. On the other hand, MCA models show higher precision in predicting FG for suburban models. Overall, the comparison and interpretation of the results suggest that non-parametric models outperform parametric models in predicting FG.

Second, a novel establishment typology was introduced in the freight demand modelling framework. The typology was done using latent class cluster analysis, and the factors considered were ‘a priori’ industrial classification, employment, fleet ownership, period of formation, and commodity value density. The models of each cluster were developed using RR and MCA. These models were compared with ‘a posteriori’ (a novel) industry class models in terms of prediction ability. It is noticed that the introduction of establishment typology in the framework has improved the prediction ability of the models.

Third, the focus shifts to the spatial transferability approach for FG models. Examining the spatial transferability of FG models is a substantial research requirement, with an objective to facilitate the use of previously formulated models in new application contexts, regardless of the availability of local data. This research endeavour is particularly significant for planning agencies in large and rapidly developing economies like India to minimise the enormous freight survey costs in regions with limited institutional capacity and resources for freight data collection. Given the vast geographic expanse of most Indian cities, an essential research question concerning transferability pertains to the accuracy of the suburban freight activity estimates given by the models developed for urban areas and vice versa. This study answers this research question by (i) comparing the effectiveness of transferability with respect to the direction of transfer and (ii) determining the suitability of various models for spatial transfer. Primary data gathered from the establishments in various Indian cities form the basis for this study. Using this data, a series of FP models were developed for urban and suburban establishments to understand the differential influence of geography and industry segments on the model coefficients. The findings of the transferability assessment provide evidence of asymmetry in the direction of transferability. Furthermore, the findings reveal that urban FP models exhibit a greater degree of transferability to suburban regions. The investigation highlights that the models are more effectively transferable upon updating the parameters, as opposed to the direct transfer of the model without any parameter updates.

Fourth, an investigation was conducted to ascertain the factors influencing transferability. Given that demand models require substantial time, cost, and data resources, it is essential to analyse the effects of sample size on the transferred model in a region. The insights gained from such analysis can lead to resource savings in freight data collection programs. While conventional models like OLS regression have been previously assessed for transferability, non-conventional models' predictive ability and transferability remain understudied. Therefore, understanding whether non-conventional models exhibit greater transferability than conventional ones is crucial for planning agencies to adopt more reliable modelling approaches. This study investigated the spatial transferability of freight production models using OLS, RR, and MCA. The results from the transferability assessment reveal that MCA models exhibit higher transferability.

Fifth, the study evaluated transferability by varying the sample sizes to examine the extent of transferability. MCA models demonstrated a minor deviation among the models, indicating their preference for transferability when dealing with small sample sizes. In addition, there was a negligible effect on the extent of transferability when the sample size varied.

Finally, this study aims to develop a transferability-based framework to streamline freight data collection efforts at a disaggregate level. The emphasis is on examining the transferability of freight generation (FG) models to offer recommendations on the degree of

transfer, transfer direction, and sample size determination for new regions. FG models were developed for different regions of the nation. After model development, the models were assessed for spatial transferability using different transfer methods. Multidimensionality scaling was utilised to identify new regions which are geographically similar to the study regions in terms of population density, road density, seaport proximity, number of establishments, and land value. Subsequently, K-means clustering was performed on the identified geographically similar regions to group them into four clusters. The sample size was determined based on interpolating the transferability results across the regions with similar demographics. The importance of this research lies in its potential to minimise the expenditure and time of freight data collection drastically.

Keywords: Freight logistics; Freight generation; Spatial transferability; Sample size; Parametric models; Non-parametric models

TABLE OF CONTENTS

Acknowledgements....................v

Abstract.................... vii

Table of Contents.................... xi

List of Tables.................... xvi

List of Figures.................... xviii

List of Abbreviations.................... xix

Part I: Background....................1

Chapter 1: Introduction....................2

1.1 Background and Motivation....................2

1.1.1 Freight Generation....................4

1.1.2 Freight Trip Generation....................5

1.1.3 Spatial Transferability....................5

1.2 Research Objectives....................6

1.3 Research Questions and Contributions of this Thesis....................7

1.4 Thesis Outline....................8

Chapter 2: Literature Review....................10

2.1 Model Estimation....................10

2.1.1 Explanatory Variables....................10

2.1.2 Modelling Methods....................12

2.2 Model Application....................16

2.2.1 Metrics for Spatial Transferability Assessment....................16

2.2.2 Passenger Demand Models....................18

2.2.3 Freight Literature....................20

2.3 Measurement 21

2.4 Summary...22

Chapter 3: Research Design...23

3.1 Data Collection Program 23

3.1.1 Sampling Frame Development...23

3.1.2 Pilot Survey...24

3.1.3 Sample Selection, Recruitment and Documentation...........................25

3.1.4 Data Collection 25

3.2 Methodological Framework 26

Part II: Model Estimation 27

Chapter 4: Estimation of Freight Generation Models 28

4.1 General 28

4.2 Methodology...28

4.2.1 Parametric Modelling Approaches...31

4.2.2 Non-parametric Modelling Approaches......................................32

4.2.3 Model Validation 33

4.3 Results and Discussion 34

4.3.1 Estimation of Freight Generation Parametric Models 34

4.3.2 Estimation of Freight Generation Non-Parametric Models 37

4.3.3 Internal Validation of Freight Generation Models..............................40

4.3.4 External Validation of Freight Generation Models.............................40

4.4 Research Implications 42

4.5 Summary...47

Chapter 5: Establishment Typology in Freight Modelling Framework........49

5.1 General 49

5.2 Methodology...50

5.2.1 Period of Formation Effect and Business Age..54

5.2.2 Value Density of Commodities..54

5.2.3 Freight Generation Models 54

5.3 Results and Discussion..54

5.3.1 Establishing Typology 54

5.3.2 Profiling Latent Segments of Establishments 56

5.3.3 Comparing Forecasting Accuracy...58

5.4 Summary...60

Part III: Model Application 62

Chapter 6: Spatial Transferability and Transfer Methods 63

6.1 General 63

6.2 Methodology..64

6.3 Results and Discussion..66

6.3.1 Preliminary Analysis...66

6.3.2 Estimation of Freight Production Models...68

6.3.3 Transferability Assessment...70

6.4 Summary...75

Chapter 7: Effect of Sample Size on Spatial Transferability 79

7.1 General 79

7.2 Methodology..80

7.2.1 Model Estimation..82

7.2.2 Spatial Transferability Assessment...84

7.2.3 Effect of Sample Size..84

7.3 Results and Discussion..85

7.3.1 Descriptive Statistics...85

7.3.2 Freight Production Models..85

7.3.3 Transferability Assessment of Freight Production Models......................90

7.3.4 Examination of Sample Size Effect 90

7.4 Summary..95

Part IV: Measurement...98

Chapter 8: Framework for Reducing Freight Survey Resources Using Spatial Transferability..99

8.1 General 99

8.2 Methodology...100

8.2.1 Model Estimation...101

8.2.2 Model Application 102

8.2.3 Sample Size Determination..104

8.3 Results and Discussion 105

8.3.1 FG Modelling...105

8.3.2 Application-based Approach...105

8.3.3 Estimation-based Approach 110

8.3.4 Summary of Spatial Transferability of Freight Generation Models 110

8.3.5 Exploring Geographically Similar Regions 113

8.3.6 Sample Size Determination..113

8.3.7 Examination of Transferability Strategy..115

8.4 Research Implications 119

8.5 Summary..120

Part V: Conclusions...123

Chapter 9: Conclusions and Recommendations 124

9.1 Specific Conclusions 124

9.1.1 Model Estimation...124

9.1.2 Model Application 126

9.1.3 Measurement 128

Chapter 10: Specific Contributions 129

Chapter 11: Future Scope of Work 131

References 133

List of Publications 144

Brief Biography of the Candidate 146

Brief Biography of the Supervisor 147

LIST OF TABLES

Table 1.1 Summary of past FG/FTG studies 3

Table 2.1 Summary of past FG/FTG studies 17

Table 2.2 Some key studies from literature 18

Table 4.1 Descriptive statistics of data 30

Table 4.2 Summary of parametric freight production models....35

Table 4.3 Summary of parametric freight attraction models 36

Table 4.4 MCA table of FG rates....38

Table 4.5 Summary of support vector regression models....39

Table 4.6 Validation of FP models 43

Table 4.7 Validation of FA models....44

Table 5.1 Measurement LCCA model for identifying the establishment typology....56

Table 5.2 Parameters and Z values of the estimated LCCA model with covariates...57

Table 5.3 Employment-based freight production models....58

Table 5.4 Employment-based freight attraction models....59

Table 5.5 Freight production rates....59

Table 5.6 Freight attraction rates 60

Table 6.1 Summary statistics....67

Table 6.2 FP models based on geographical location....69

Table 6.3 FP models for '*a posteriori*' segmentation of establishments....69

Table 6.4 Transferability assessment of area-based FP models 71

Table 6.5 Transferability assessment of employment-based FP models....71

Table 6.6 TI for combined transfer and joint context estimations....76

Table 7.1 Descriptive statistics of variables 86

Table 7.2 Summary of OLS and RR models 87

Table 7.3 Summary of Multiple Classification Analysis (MCA)...............................89

Table 7.4 Summary of best transferable modelling approaches................................93

Table 7.5 Summary of best transfer cases 94

Table 7.6 Transferability assessment of recommended cases 95

Table 8.1 Summary of primary data used in modelling...100

Table 8.2 Summary of FG models...106

Table 8.3 Summary of naïve transfer of FG models..107

Table 8.4 Summary of updated model transfer of FG models.................................109

Table 8.5 Summary of transferable models through estimation-based approach.....110

Table 8.6 Summary of transferability assessment results..112

Table 8.7 Comparison between conventional and transferability approaches..........117

LIST OF FIGURES

Figure 1.1 Thesis outline....9

Figure 3.1 Study cities 24

Figure 3.2 Outline of methodological framework 26

Figure 4.1 Details of a ‘posteriori’ industrial segments....29

Figure 4.2 RMSE values for various FG models on external validation....45

Figure 4.3 MAE values for various FG models on external validation....46

Figure 5.1 Outline of the proposed approach for freight demand estimation....52

Figure 5.2 Graphical representation of LCCA model....53

Figure 6.1 Scatter plots for FP models of urban and suburban establishments....68

Figure 6.2 Transferability assessment of urban and suburban FP models....73

Figure 6.3 Transferability assessment of a posteriori segmentation FP models....74

Figure 6.4 Summary of transferability assessment....74

Figure 7.1 Comparison of FP rates across various industries....88

Figure 8.1 Framework to determine required sample size for new regions....101

Figure 8.2 Map of MDS and KMCA....114

Figure 8.3 Flowchart to determine sample size for new regions 116

Figure 8.4 Comparison of minimum sample size requirements 118

Figure 8.5 Comparison of survey durations....118

Figure 8.6 Comparison of survey costs....119

LIST OF ABBREVIATIONS

1	ANN	Artificial Neural Networks
2	CK	Central Kerala
3	CV	Coefficient of Variation
4	EBFS	Establishment-based Freight Survey
5	FA	Freight Attraction
6	FG	Freight Generation
7	FP	Freight Production
8	FTA	Freight Trip Attraction
9	FTG	Freight Trip Generation
10	FTP	Freight Trip Production
11	GFA	Gross Floor Area
12	GIS	Geographical Information System
13	GLZM	Generalised Linear Model
14	HYD	Hyderabad
15	INR	Indian Rupees
16	ISIC	International Standard Industrial Classification
17	JAI	Jaipur
18	KMCA	K-Means Clustering Algorithm
19	LCA	Latent Class Analysis
20	LCCA	Latent Class Cluster Analysis
21	MAE	Mean Absolute Error
22	MAPE	Mean Absolute Percentage Error
23	MDS	Multidimensionality Scaling
24	NAICS	North American Industry Classification System
25	NE	Number of Employees
26	NK	North Kerala
27	OLS	Ordinary Least Squares
28	R^2	Coefficient of Determination
29	RATE	Relative Aggregate Transfer Error
30	RMSE	Root Mean Square Error

31	RR	Robust Regression
32	SD	Standard Deviation
33	SE	Standard Error
34	SVR	Support Vector Regression
35	TI	Transfer Index
36	TR^2	Transfer R^2
37	USD	US Dollars
38	WLS	Weighted Least Squares
39	WRMSE	Weighted Root Mean Square Error
40	ρ	Coefficient of correlation

Part I: Background

Chapter 1: Introduction

1.1 Background and Motivation

Spatial transferability is the practice of applying or transferring a model from one geographic region to another, with or without the use of data specific to the second region (Balla & Sahu, 2023; Holguín-Veras et al., 2013; Pani, Sahu, & Bhat, 2021). Transferring models to various spatial contexts has been extensively studied and well-documented (Sikder et al., 2013). This practice has the potential to reduce the expenses and duration of freight surveys in regions with limited or negligible investments in data collection. Establishment-based freight survey (EBFS), commodity flow survey (CFS), vehicle trip diary, roadside interview survey, driver survey, parking survey, vehicle observation survey, vehicle traffic count survey, supplier survey, service provider survey, freight operator survey, and global positioning system survey are among the widely deployed freight data collection techniques (Allen et al., 2012). Out of all these survey techniques, the establishment-based freight survey has a relatively higher cost and time per completed response. Lawson et al. (2012) reported the typical cost of collecting a response in the United States of America (USA) as USD 120 (USD means US Dollar). The cost of a complete response in Canada was approximated to be USD 190, while the cost of a complete response in Paris, France was a whopping USD 500 (Toilier et al., 2016). The United States of America, Canada, and France have gross domestic product (GDP) per capita, which is exceedingly higher than the global average (USD 12,234.8), as per the records of the World Bank for the year 2021 (The World Bank Group, 2022). GDP per capita USA was USD 70,248.0; in Canada, it was USD 51,987.9; in France, it was USD 43,659.0 (The World Bank Group, 2022). These GDP per capita figures presented in Table 1.1 indicate that the developed nations can allot resources to various activities, such as freight surveys, for the provision of critical information for transportation planning, logistics, and commerce and trade.

In contrast to developed nations, developing nations such as India have limited institutional capacity to conduct freight surveys despite being one of the fastest-growing economies in the world. The country's relatively lower GDP per capita of USD 2,256.6 in 2021 (The World Bank Group, 2022) poses a significant ordeal for the freight transportation and logistics sector. However, India's vigorous industrialization, rapid urbanization, and rising consumer spending power have contributed to the expansion of national freight movements that are estimated to increase at a rate of 9.7% through 2031-2032 and reach beyond 13,000 billion tonne-kilometres (National Transport Development Policy Committee, 2014). Nonetheless, due to the sparsity of resources, the nation falls short in terms of the freight data required to fit accurate freight demand models (Kamboj et al., 2022). The dearth of accurate freight demand models constrains the development of efficient transportation planning and logistics systems, making it difficult for enterprises to optimize their supply chain operations.

Table 1.1 Summary of past FG/FTG studies

Country Name	Gross domestic product per capita in 2021 (The World Bank Group, 2022)
USA	USD 70,248.0
Canada	USD 51,987.9
France	USD 43,659.0
India	USD 2,256.6
Global average	USD 12,234.8

In past studies, several investigations were done to find the direction and extent of transferability among the study regions. However, the applicability of these transferability conclusions to newer regions or regions lacking freight data remains unexplored. This knowledge gap complicates the reliable prediction of the freight demand in emerging nations. It is necessary to have solid estimates of freight demand because the freight establishments are such huge freight generators that the stakeholders (i.e. the residents and governing bodies of the surrounding area) are concerned about the negative externalities that the freight activities may cause (Mohapatra et al., 2021; Pani et al., 2018; Sahu et al., 2020, 2022). Consequently, demystifying the concept of transferring a model to an inexperienced region will assist policymakers and planners in mitigating the associated externalities, planning environmentally friendly infrastructure for sustainable growth, and managing freight movements. Moreover, the past studies did not explore the transferability results in the context of sample and modelling methodology. Investigating this research shortcoming can benefit planners and researchers in minimising the budget, time and personnel required for freight surveys.

In summary, this thesis serves as a valuable resource that bridges the gap between academic research and practical applications in the realm of freight activity. By contributing insights into freight generation, it aims to empower decision-makers across both the public and private sectors to make well-informed choices that can lead to a more efficient and sustainable transportation system, ultimately benefiting the economy and society as a whole.

1.1.1 Freight Generation

The freight generation, or FG, is bi-conceptual (Balla et al., 2023; Sahu & Pani, 2020). The quantity produced and the quantity attracted by an establishment are considered. Freight production (FP) is the process of producing tonnes of freight by the establishment, and freight attraction is the process of attracting tonnes of freight (FA) by the establishment. FG is a crucial part of the transportation and logistics sector because it helps to determine the demand for freight transportation services. FG can result from various factors, including the level of economic activity in a particular area, the production and demand for goods within a particular industry, consumption patterns, seasonal factors, and the presence of infrastructure. FG models can be used to estimate demand for freight, and population exploitation is increasing the need for precise models. Despite this, there have been more truck-related activities in the manufacturing sectors to meet the public's demand for a particular good despite the absence of FG models. In short, these models can help design an efficient freight transportation system.

The efficient movement and timely availability of goods, including raw materials and finished products, is made possible with the help of efficient freight transportation, which plays an essential role in cities and metropolitan areas worldwide (Chandra et al., 2023; Oka et al., 2019; Ortúzar & Willumsen, 2011). Because of the high concentration of economic activity in these regions, consumers have convenient access to a diverse selection of goods and services that can be obtained relatively quickly (Chandra et al., 2022; Holguín-Veras et al., 2016). The quality of freight transport services available in an area significantly impacts the costs incurred by businesses and the dependability of the delivery of goods, both of which are essential for commercial activity (Kuzmyak, 2008; Lindsey et al., 2014; Pani et al., 2019). The significance of freight transportation is expected to continue to increase due to the expansion of online commerce and the implementation of just-in-time inventory practices (Taniguchi et al., 2016). However, freight activity is also associated with negative externalities, such as increased levels of air pollution, noise, and congestion, which affect society and the environment. In urban areas, delivery trucks significantly contribute to traffic congestion, and freight transportation significantly contributes to climate change due to the energy it consumes and the greenhouse gas emissions it produces (Pani, Sahu, & Holguín-Veras, 2021).

Accurate forecasts of past and future freight activity are required to solve logistics and transportation issues successfully. A complete understanding of freight generation (FG), which includes the total tonnage and number of trips, is essential for infrastructure development and policy planning. Because understanding makes it easier to allocate resources efficiently and allows for quick, well-informed decision-making, this knowledge is crucial for meeting the logistical demands of households and businesses. By providing insightful information about the origins of freight activity within the transportation system, this thesis aims to advance the field of study in this context significantly. By conducting thorough analyses and utilising relevant data, this study aims to shed light on the patterns, trends, and factors that affect freight movement. This work aims to provide actionable information to decision-makers in the public and private sectors by gaining a deeper understanding of the underlying dynamics. These insights can be used by public sector stakeholders, such as government agencies and urban planners, to design and implement transportation infrastructure projects that can effectively handle the anticipated freight volume. They can decide whether to invest in constructing new highways, railroads, ports, and other vital transportation infrastructure.

Furthermore, a thorough understanding of freight activity can help create laws and policies that support environmentally friendly and sustainable freight practices. The results of this thesis can be precious for private sector organisations, including companies engaged in supply chain management and logistics. Companies can optimise their operations and streamline their supply chains by understanding the patterns of freight movement and locating potential bottlenecks. This understanding of freight patterns and bottlenecks can result in reduced costs, quicker deliveries, and higher customer satisfaction.

1.1.2 Freight Trip Generation

Freight trip generation (FTG) pertains to the number of trips conducted by freight or commercial vehicles (such as trucks and vans) to or from a particular establishment. FTG encompasses two distinct but interrelated concepts: freight trip production (FTP) and freight trip attraction (FTA). FTP denotes the count of outbound trips originating from an establishment. At the same time, FTA signifies the tally of inbound trips directed toward that establishment. Estimating FTG is as crucial as estimating FG in transportation infrastructure planning, environmental impact assessment, economic analysis, and policy formulation.

1.1.3 Spatial Transferability

FG models are one of the numerous freight demand models available. The phrase "spatial transferability of FG models" refers to the suitability of using models developed with data and information from one geographical region to forecast tonnage in another region (Holguín-Veras et al., 2013; Klodzinski & Al-Deek, 2003; Pani, Sahu, & Bhat, 2021). These models were developed using regional data and information. This topic is of considerable interest from both a theoretical and a practical standpoint. Theoretically, an examination of a model's performance in various contexts should shed light on its ability to generate credible forecasts under various conditions. For regions that cannot afford to invest in extensive data collection procedures, the ability to transfer models from one region to another can provide significant cost and time savings. From a practical standpoint, this is something that can be said. Developing FG models typically necessitates expensive data collection via various surveys, which frequently exceeds the budget of planning organisations tasked with this task (Allen et al., 2012; Pani & Sahu, 2019b, 2022; Samimi et al., 2013).

Consequently, only large metropolitan areas with substantial budgets can manage such extensive data collection programmes. Due to a lack of resources, smaller and medium-sized regions must import models from larger regions. Transferring models reduces the costs associated with data collection and the time and effort required for model development and estimation procedures. This issue is critical in the context of FG models, which rely heavily on abundant data inputs, skilled personnel, and substantial time investments.

1.2 Research Objectives

The main objectives of this research are as follows:

1. Developing freight generation (FG) models at the industrial, regional, and state levels using various methodologies.
2. Evaluating the transferability of FG models across multiple regions using various transfer methods.
3. Examining the factors affecting the extent of transferability of FG models
4. Developing a framework to minimise the freight survey resources using a transferability approach.

The objectives can be broadly categorised into (i) model estimation, (ii) model application, and (iii) measurement. In short, objectives are categorised into 3Ms (MMM). The initial "M" denotes model estimation, referring to the development of FG models. The second "M" involves model application, wherein the developed models are applied to other regions; this application is also referred to as spatial transferability. The final "M" pertains to measurement, involving planning strategies for data collection. This thesis aims to develop resource effficient solutions to estimate the freight demand through spatial transferability approach. This shortage will be accomplished by addressing these research objectives as well as the associated research needs. Freight transport planners, operators, trucking companies, business establishments, and policy analysts who advise governments on urban and regional transportation planning for various planning horizons are the people who are expected to use the findings of this research.

1.3 Research Questions and Contributions of this Thesis

The research questions in the thesis have been carefully addressed and presented below. These questions aim to contribute significantly to the existing body of literature in the field by offering detailed explanations within specific contexts.

Which modelling methodology accurately estimates FG at industrial, regional and state levels?

The freight data from Kerala State were used as the basis for the development of FG models. These models were developed for (i) a variety of industries that were classified utilising different industrial segments, (ii) urban and suburban regions, and (iii) the state of Kerala. These models were developed with parametric and non-parametric modelling approaches. Then, they were validated to determine which approach was superior.

Which establishment typology increases the accuracy of FG models?

Based on fleet ownership, value density of commodity, year of formation, industry sector, and employment, the establishments were aggregated into different classes. Models were developed for these classes, and these models were compared with those developed with a priori industry classification (already existing classification system) to understand whether the rational method of aggregation helps improve the models' prediction ability.

How effective are urban models using various transfer methods when utilised in suburban regions?

Based on their proximity to urban or suburban areas, the businesses in Kerala were divided into urban and suburban categories. The urban and suburban FG models were developed, and their ability to be transferred between locations was evaluated. Different transfer methods were utilised in conducting this evaluation. These methods included naive transfer, combined transfer estimation method, and joint context estimation.

How do various modelling methodologies perform in the context of spatial transferability?

Both parametric and non-parametric modelling approaches were utilised to develop the FG models for use in different regions of Kerala, Jaipur, and Hyderabad. The spatial transferability of these models was evaluated using the naive transfer method to compare the modelling approaches.

What is the influence of sample size on spatial transferability of the FG models?

The FG models were developed for North Kerala, Central Kerala, Jaipur and Hyderabad. These models were developed, and their ability to be transferred was analysed. In this evaluation, the effect of sample size on the extent of transferability was evaluated by varying the sample size.

How do we minimise the efforts of freight data collection for the regions that are constrained in terms of resources?

In order to provide an answer to this question, we proposed a novel method for determining the size of the sample. This size was reduced as much as possible by using different transfer methods. In addition, we used various statistical techniques to locate regions that are geographically analogous to one another. We saved time, money, and human resources by cutting down the sample size for the freight survey.

1.4 Thesis Outline

The thesis is organised into five parts, and the outline is illustrated in Figure 1.1. This chapter belongs to Part I, which describes the research motivation, objectives, and research questions addressed in our research. The subsequent chapter in Part I presents an extensive literature review of the different research objectives addressed in this study. After the literature review, the research design is discussed. Part II consists of two chapters on the estimation of FG models. In chapter 4, the FG models were developed at various industrial, regional and state levels. This chapter discusses and compares the parametric and non-parametric modelling methods. In chapter 5, the novel establishment typology in freight demand modelling framework was used, and models were developed for these typologies. Part III consists of two chapters that briefly discuss the spatial transferability assessment. Chapter 6 discusses FG model transferability (model application) and compares different transfer methods. Chapter 7 discusses the different factors that influence the extent of transferability of FG models. In part IV, we presented a chapter that mainly focuses on reducing efforts for freight surveys. This chapter presents a novel methodology based on the transferability assessment results to determine the sample size. In addition, we compared our approach with the conventional approach of sample size determination. Part V consists of key conclusions, contributions, limitations and recommendations for further study.

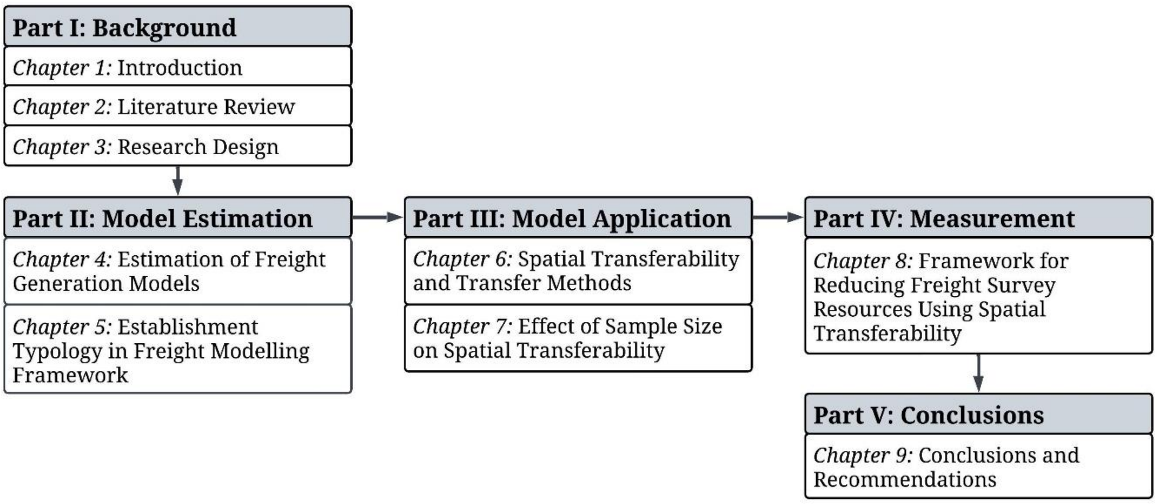


Figure 1.1 Thesis outline

Chapter 2: Literature Review

2.1 Model Estimation

In order to develop accurate freight demand models, it is essential to have a comprehensive understanding of the various contributing factors that influence freight generation (FG) and freight trip generation (FTG) patterns. This understanding needs to be analysed at aggregated and disaggregated levels to capture the intricacies of the freight movement process. Below is a review of the findings related to the impact of explanatory variables, which will further enrich the ongoing discussion:

2.1.1 Explanatory Variables

Numerous studies have examined freight generation (FG) and freight trip generation (FTG) using a diverse range of explanatory variables. Noteworthy findings from some of these studies indicate a positive correlation between FG/FTG and microscopic factors related to the establish which include business size variables such as employment and area (Balla et al., 2023; Holguín-Veras et al., 2014; Pani et al., 2018; Sahu & Pani, 2020). There are a few more studies which showed positive correlation of less extent to factors such as years in business (Pani et al., 2018, 2020), number of items (Alho & Silva, 2014), number of vendors (Günay et al., 2016), share of transport cost (Ha & Combes, 2015), existence of a supply chain (Sánchez-Díaz, 2018) and warehouse availability (Alho & Silva, 2014).

It is to be noted that floor area is seldom used as the sole explanatory variable in FG/FTG; instead, it is used in combination with employment. A comparative analysis between FG models based on area and those based on employment has uncovered an important insight: in cities characterized by dense commercial activities where acquiring land is a challenge, the use of area as an indicator for representing business size tends to yield skewed results (Pani et al., 2018). This finding highlights the significance of the macro-level factors, such as the city type, on the performance of micro-level parameters. Notably, a statistically significant relationship exists between FG/FTG by an establishment and various macroscopic factors in its vicinity. These factors include land use patterns (Holguín-Veras et al., 2012b), socio-demographic characteristics (Ortúzar & Willumsen, 2011), industrial attributes (Cantillo et al., 2014; Novak et al., 2011), and network characteristics (Sánchez-Díaz et al., 2016). Understanding and incorporating these macro-level variables is essential for developing more robust and accurate models.

Holguin-Veras et al. (2002) found that FG and freight trip generation FTG models are better at explaining things when they have more variables that are similar to the location of establishments. Freight activity is affected by how close a business is to major traffic generators or major arterial roads (Sánchez-Díaz et al., 2016) and ports (Patil & Sahu, 2017), but this has not yet been the subject of a detailed statistical study. One multizone attribute termed "port influence" shows how close an establishment is to a port and how that affects its location. This attribute showed a positive correlation with FG (Novak et al., 2011). Freight activity is also thought to be affected by a place's commercial appeal, which is often measured by its land value. This assumption fits with the idea that premium spaces would have more FG and FTG than isolated areas in suburban settings (Pani et al., 2018; Sánchez-Díaz et al., 2016). Attributes like the number of stores, salespeople, and other land use variables are also used to figure out how appealing a certain area is (Comi et al., 2013). Conversely, the unemployment rate within a zone has been found to exhibit a negative association with FTG (Garrido & Mahmassani, 2000).

FTG has been positively associated with the per capita income and the block size of a zone, which is measured by the road length divided by the number of intersections (Miodonski & Kawamura, 2012). These results show that these macro-level characteristics help to explain why FTG variations happen. Additionally, various modelling approaches, including Ordinary Least Squares (OLS) and spatial econometric models for FTG (Sánchez-Díaz et al., 2016), pooled regression models for FG (Sahu & Pani, 2020), and spatial regression models for FG (Novak et al., 2011), have provided further support for the idea that macroscopic characteristics, even when used as proxy variables, enhance the predictive capability of FG and FTG models. These large-scale characteristics can help us figure out how each establishment uses FG and FTG. Also, (Sánchez-Díaz et al., 2016) found that there is spatial autocorrelation among retail establishments. This means that retail businesses in areas with a lot of retail employment tend to have more inbound (FTA) trips than those in areas with less retail employment. This spatial correlation shows how bigger factors in the environment have an effect on freight activity. These large-scale characteristics can help us figure out how each establishment uses FG and FTG. The advancement of Geographical Information Systems (GIS) and the availability of open-source computer programs with satellite imagery, such as Google Earth, offer opportunities to incorporate locational spatial effects into FG and FTG models.

2.1.2 Modelling Methods

Most freight studies have used linear regression models to estimate FG and FTG at industry-, city- and state levels. In the United States (Jaller et al., 2015), employment-based FTG models for various industries (based on the North American Industry Classification System (NAICS)) with 243 establishments' data were built using the OLS approach. The proportion of variation explained by the employment variable in predicting the trips was between 1.1% to 65.4%. Also, it was noticed that the goodness of fit of the attraction models was better than the production models. Using OLS regression, Sahu and Pani (Sahu & Pani, 2020) estimated FG rates (tons per employee and tons per m^2) for 13 International Standard Industrial Classification (ISIC) industrial classes (sample size varied from 10 to 96). The proportion of variation explained by employment in estimating FP ranged from 59% to 91%, and FA ranged from 49% to 93%. In the case of area-based FG models, it was noticed that the proportion of variation explained by the explanatory variable in predicting FP ranged from 29% to 93%, and FA ranged from 61% to 95%. Sánchez-Díaz (2017) developed linear regression models to estimate FTG with observations ranging from 16 to 54 for various commercial sectors with employment and area as explanatory variables. In estimating the freight trips, the proportion of variation explained by employment was between 19% and 53%. For area-based FTG models, the proportion of variation explained by the explanatory variable was between 15% and 45%. Oliveira et al. (2021) developed a linear regression model with a sample size of 434 to estimate the weekly deliveries in the seven historic cities of Brazil. The proportion of variation explained by the explanatory variables in predicting the deliveries was 16%. It was observed that the influence of area is negative on the number of trips, and the logical reason for this was not explained. In the city of Quito (Puente-Mejia et al., 2020), FTG was estimated for different categories of retailers using linear regression models with a sample size between 6 and 227 by employing retail area and number of employees as explanatory variables. It was found that employment was the better variable in predicting trips. The proportion of variation explained by the employment variable in predicting freight trips ranged from 41% to 83%. In Colombia, Gonzalez-Calderon et al. (2021) modelled the empty trips made by different types of trucks using a sample of 135,564 trucks. The commodity groups were considered explanatory variables in these models, and the modelling methodology used was OLS. From the models, it was clear that the trucks transporting certain types of commodities, which needed unique technology to be carried, were returning empty. The proportion of variance explained by the kind of commodity in estimating the empty truck trips ranged from 20.5% to 83.2%. The models that estimated the empty trips made by 2-axle single-unit trucks had a better fit. In New York City (Campbell et al., 2018), to assess the parking demand, the freight and service trips were estimated for various commercial sectors in the city using the OLS modelling approaches (linear and exponential models) as per the suggestions given in NCFRP 37 (Holguín-Veras et al., 2016). The model results showed that implementing specific programmes like Off-Hour Deliveries and staggered deliveries could reduce parking needs during peak hours.

Gonzalez-Feliu and Peris-Pla (2017) developed aggregate and disaggregate employment-based FTA models for various retailing activities in urban zones of Lyon City using OLS. In these models, the proportion of variation explained by the explanatory variables in predicting the trips was between 52% and 82%. Bastida and Holguín-Veras (2009) developed OLS models for the boroughs of New York City - Manhattan and Brooklyn for estimating the FTG of receivers and carriers. The Manhattan sample had 180 receivers and 192 carriers; the Brooklyn sample had 200 receivers and 139 carriers. The goodness of fit of the carrier models in both boroughs was better than that of the receiver models. The proportion of variation explained in the receiver models by explanatory variables in estimating the freight trips was 61.1% and 80.9% in Manhattan and Brooklyn, respectively. Sánchez-Díaz (2017) quantified the freight needs in Gothenburg City by developing FTA models with a sample size of 195 using robust linear regression. The proportion variation explained by employment in estimating weekly truck trips was 47%; the area in estimating weekly truck trips was 40%. In India (Pani et al., 2018), city-specific linear regression models were developed to estimate FP and FA. The number of observations used in developing these models ranged from 30 to 85. Also, in the same study (Pani et al., 2018), state-level models with a sample size of 338 were built to estimate FG. In both city-specific and state-level models, employment, gross floor area, and a new variable termed ‘years in business’ were used as the explanatory variables. In these models, the proportion of variation in FG explained by the explanatory variables ranged from 58% to 72%.

Some studies adopted modelling approaches like MCA, generalised linear model (GLZM), partition method, ordered logit, ordered probit, negative binomial regression, and artificial neural networks (ANN). For Lisbon City (Alho & Silva, 2017), average weekly deliveries were estimated using MCA with the data from 604 retail establishments. The employment and the industry categories were considered explanatory variables. The proportion of variation explained by these variables in predicting the deliveries was 44.5%. Sahu and Pani (Sahu & Pani, 2020) also used MCA to estimate the FG rates per establishment at different employment- and area levels across 13 ISIC industries. The employment explained the variations in FG rates from 53.1% to 56.1%; the area explained the variations in FG rates from 43.5% to 57.5%. Alho and Silva (2014, 2017) predicted weekly deliveries considering employment, area and the industry category using GLZM to counter the effect of non-normality and heteroskedasticity presence in the data. They also modelled FTG using the partition method, which allows the explanatory variables to be a mixture of numeric and categorical variables. In addition, the ordered logit model was used, considering FTG as an ordered variable. It was concluded that the contribution of the explanatory variable varied based on the methodology chosen. In New York City (Bastida & Holguín-Veras, 2009), the number of deliveries received and shipped per establishment was estimated using MCA. MCA analysis identified employment, commodity type, and industry segment as better explanatory variables in estimating freight trips. However, MCA models were not validated internally or externally to check the percentage of error in the results. Pani et al. (2022) proposed ordered probit, a quantitative approach, for understanding how the logistic decisions in various industries were influenced by the variables - sourcing approach, fleet ownership, area, employment, and business age. It was shown that the usage of own trucks increased with an increase in employment. Also, the study revealed a strong positive correlation between business age and the frequency of truck trips. The machinery and chemical product industries were positively correlated with the outsourced truck trip frequencies.

Sánchez-Díaz (2017) considered freight weights to be ordered levels and developed area-based and employment-based FG models using the ordered logit model for Gothenburg City, Sweden. The area model showed that the retailing establishments in perishable goods attract more freight. In comparison, the employment models indicated that retail establishments in non-perishable goods, health care and the public sector attract less freight tonnage. Using negative binomial regression, Pani et al. (2021) estimated area-based and employment-based FTP models for different industries. In this technique, the predicted variable shows the characteristics of discrete distribution, which is more appropriate for modelling FTG, a count variable (not a continuous variable). It is not appropriate to use OLS because, in this technique, the predicted variable is assumed to be continuously distributed between $-\infty$ to $+\infty$. It was concluded that the larger establishments produce more trips than smaller establishments, and the number of trips produced increased at a marginal rate. Kulpa (2014) estimated FTG for light and heavy trucks in the Kraków metropolitan area, Poland, with a relatively smaller sample size of 50 observations using multiple regression and ANN. From the ANN models, it was observed that the employment variable had more influence on the freight trips. In most of the cases, ANN models performed better.

FG and FTG were modelled using several approaches in the previous research efforts; however, a few studies examined the prediction ability of the models. In Latin America (Puente-Mejia et al., 2020), for different retailers in various zones, FTG models were validated internally using the mean absolute percentage error (MAPE); the MAPE values were not greater than 156%, and the most negligible value was 32%. It was noticed that the model developed for the retailer type of service/religious/municipal was superior compared to other retailer-type models. Pani et al. (2018) analysed the performance of FG (FP and FA) models at two levels – city and state. The city-specific FG models were internally validated using root mean square error (RMSE), and the state-level FG models were externally validated using MAPE. In city-specific single variable models, it was observed that most of the models with employment as an explanatory variable had lower RMSE than that of models with the area as an explanatory variable in a particular city. These lower values of RMSE showed that employment-based models were superior in performance in most cities. At the state level, it was observed that the MAPE values of FP models are lower than that of FA models. Jaller et al. (2015) validated the employment-based FTG (FTP and FTA) models for various industry sectors using RMSE. The model validation results showed that among the FTG models, the FTA models had better prediction ability than the FA models.

Some studies tried to find the best modelling approach for estimating FG/FTG. Alho and Silva (2014) considered the correlation coefficient between the observed and predicted values to assess the predictive ability of the FTG models. FTG models developed with OLS had a correlation coefficient greater than 0.72, whereas, for FTG models developed using GLZM, the correlation coefficients were more significant than 0.64. It was clear that the OLS models outperformed the GLZM models. In another study, Alho and Silva (2017) validated the FTG models using RMSE, mean absolute error (MAE), percentage of correct predictions, and Spearman correlation coefficient. Evidently, the prediction abilities of the non-parametric models developed using MCA and partition method were of the same magnitude. The performances of the parametric models, which included GLZM and ordered logit, were inferior to the non-parametric models. In a study by Kraków by Kulpa (2014), the FTG models developed using OLS and ANN were cross-validated using the MAE. It was observed that the ANN models gave lower MAE values compared to the OLS models.

The most relevant recent studies related to FG/FTG are summarised in Table 2.1. It is abundantly clear from these studies that the predictive power of FG models is frequently disregarded. In addition, some research has focused only on non-parametric models as a means of estimating FG/FTG. This study aims to compare the various modelling methodologies that can be used to predict FG and to contribute to improving the chances of a more accurate representation of the freight demand estimation at the state, regional, and industrial segment levels.

2.2 Model Application

2.2.1 Metrics for Spatial Transferability Assessment

In the age of constrained survey assets and ever-developing demand for disaggregated travel information, examining the spatial transferability of models has become a pillar of travel demand analysis. An overview of previous studies shows that various measures could be used to assess the transferability of the estimated models. For instance, Atherton & Ben-Akiva (1976) used transferability test statistics to assess the transferability of the work-trip modal-split model. They also used the Bayesian update process to enhance transferability. Taking forward their study, updating techniques like the transfer scale approach and combined transfer estimator were developed to improve the prediction accuracy of the transferred models (Ben-Akiva & Bolduc, 1987). Koppelman & Wilmot (1982) used several measures such as transfer index (TI), transfer ρ^2, root mean square error (RMSE), aggregate prediction statistic and relative aggregate transfer error (RATE) to assess the transferability of disaggregate choice models.

Table 2.1 Summary of past FG/FTG studies

Paper	Study Area	FG/FTG Modelling Approach	Validation
Gonzalez-Calderon et al. (2021)	Colombia	FTG – OLS	-
Bastida and Holguín-Veras (2009)	Manhattan, Brooklyn (New York, USA)	FTG – OLS, MCA	-
Campbell et al. (2018)	Troy, New York City (USA)	FTG – OLS	-
Oliveira et al. (2021)	Brazil	FTG – OLS	-
Gonzalez-Feliu and Peris-Pla (2017)	Lyon (France)	FTG – OLS	-
Pani et al. (2022)	Kerala State (India)	FTG – Ordered Probit	-
Pani et al. (2021)	Jaipur and Kerala State (India)	FG – OLS; FTG – Negative binomial regression	-
Sahu and Pani (Sahu & Pani, 2020)	Kerala State (India)	FG – OLS, MCA	-
Puente-Mejia et al. (2020)	Latin American cities	FTG – OLS	Internal validation using MAPE
Alho and Silva (2017)	Lisbon (Portugal)	FTG – MCA, Partition method, GLZM, Ordered logit	Internal validation using RMSE, MAE, percentage of correct predictions and Spearman correlation coefficient
Sánchez-Díaz (2017)	Gothenburg (Sweden)	FTG – Robust regression; FG - Ordered logit	Internal validation using RMSE
Jaller et al. (2015)	Manhattan, New York, New Jersey (USA)	FTG – OLS	External validation using RMSE
Kulpa (2014)	Kraków and Poznań (Poland)	FTG – Constant Trip Rates, OLS, ANN	External validation using MAE and MAPE
Alho and Silva (2014)	Lisbon (Portugal)	FTG – OLS, GLZM	Internal validation using correlation coefficient
Pani et al. (2018)	Kerala State (India)	FG – OLS	Internal validation using RMSE; External validation using MAPE

FG: Freight generation, FTG: Freight trip generation, OLS: Ordinary Least Squares, WLS: Weighted Least Squares, GLZM: Generalised linear model, ANN: Artificial neural networks, MAPE: Mean Absolute Percentage Error, MAE: Mean Absolute Error, SSE: Sum of Squares Error, RMSE: Root Mean Square Error, MAE: Mean Absolute Error

The transfer ρ^2 is used to test the goodness of fit of the estimation context model when transferred to the application context (Galbraith & Hensher, 1982). To assess the transferability of the ordered response model, Agyemang-Duah & Hall (1997) employed weighted root mean square error (WRMSE), transfer pseudo R^2, aggregate prediction statistic and root of sum of residual error. They further updated the model parameters of the estimated model by using the scaling updating technique. McArthur et al. (2011) used the relative number of wrong predictions and standardized root mean square error to analyse the transferability of parameters in gravity models. Transfer R^2 is used to assess the extent of transferability of linear regression models by Wilmot (1995). The key results obtained from some of the past studies are summarized in Table 2.2. From these studies, it can be interpreted that the sample size significantly impacts model transferability. Models with small sample sizes have low TI values.

Table 2.2 Some key studies from literature

Reference	Model	Study Area	Results
Wilmot (1995)	Trip Generation Models	Different cities in South Africa	TR^2 = -0.44 to 0.80
Santoso and Tsunokawa (2005)	Mode Choice Models	Urban and suburban areas of Ho Chi Minh City, Vietnam	TI = 0.45 to 0.99
Karasmaa (2007)	Mode Choice Models	Helsinki Metropolitan Area and Turku region in Finland	TI = 0.80 to 0.95
Kuo and Tang (2011)	Mode Choice Models	Ho Chi Minh City, Vietnam and Phnom Penh City, Cambodia	TI = 0.01 to 0.99
Sikder et al. (2014)	Time-of-day Choice Models	Different counties in San Francisco Bay Area of California	TI = -0.21 to 0.94

2.2.2 Passenger Demand Models

A few studies in transportation planning focused on the transferability of the travel demand models built for developed economies. Some critical past studies (Fu et al., 2006; Johnston, 2007; Sikder et al., 2013) suggested that models can be transferable between regions with contextual similarities in terms of socio-demographic factors such as household income, auto-ownership, household structure, key employment type/industries, employment pattern, commuting pattern and degree of urbanization (Schiffer, 2012; Wafa et al., 2015). Institute of Transportation Engineers manual (Tian et al., 2019) uses the suburban passenger trip data for estimation context. It transfers the trip generation model as an application context for predicting trips in urban areas. In the US (Bowman et al., 2017), using logistic regression, Activity-Based

Travel Forecasting Models were modelled for 13 metropolitan regions, and the transferability of these models among the regions was assessed. Another study in the US emphasised the transferability of tour-based time-of-day choice models developed for various counties in the San Francisco Bay Area of California using logistic regression (Sikder et al., 2014). Most of these studies assessed the transferability of the logistic regression models. Later, a few studies in the US started focussing on more robust modelling approaches and transferability. Tang et al. (2018) used neural networks to build mode choice models for several regions in the US. They examined the extent of transferability of these models. Sikder and Pinjari (2013) assessed the transferability of the travel demand models developed for the US using a multiple discrete-continuous extreme value approach. Apart from the US, some studies in developed countries explored the transferability of passenger demand models. In Finland, Karasmaa (2007) evaluated the transferability of mode and destination choice models developed for the Helsinki Metropolitan Area and Turku region, and the modelling approach used was logistic regression. Assessment of the spatial transferability of activity-based models that were generated using the Albatross was carried out in the Netherlands (Arentze et al., 2002). Hensher and Ton (2000) explored the spatial transferability of commuters' mode choice models between the cities of Sydney and Melbourne, and the modelling approaches used were neural networks and nested logit. The number of times people went shopping was calculated using ordered response models in the Toronto Metropolitan Area, located in Canada (Agyemang-Duah & Hall, 1997). These models were tested to see if they could be transferred to other regions. Wilmot (1995) updated trip generation models developed using linear regression. Full transfer and partial transfer of the models were done. The full transfer includes the transfer of the entire model (also known as naïve transfer); partial transfer includes updating some parameters based on local data and transferring it (Rose & Koppelman, 1984). Kawamoto (2003) developed person-based trip generation models using OLS for two Brazilian cities, São Paulo and Bauru and assessed the model transferability. In each of the research mentioned above, it was discovered that the applicability of the findings was restricted to the study regions.

Among the transferability studies in transportation planning, a limited number of studies tried to explore the effect of sample size on the extent of transferability. Sikder et al. (2013) proved that the extent of transferability is a function of the sample size. They assessed for transferability of choice models, which were developed by pooling the data from various regions; data pooling increased the sample size. These pooled data models showed better transferability than those developed with regional or local data. Karasmaa (2007) focused on investigating the impact of sample size on different transfer methods, including Bayesian updating, combined transfer estimation, transfer scaling, and joined context estimation. The model transferability was tested using six different sizes of sample sets. It was noticed that the joint context estimation method proved to be a better transfer method if both estimation and application context data were available. Santoso and Tsunokawa (2005) tried to improve the transferability of mode choice models by updating the model parameters, and the updating was done using different sets of sample sizes. It was noticed that the transferability of the model improved when a larger sample size was used for updating the models. These are the studies in which the effect of sample size on passenger demand models' transferability was analysed. It is to be noted from the literature that analysing the extent of transferability of freight demand models as a function of sample size is still unexplored in freight studies.

2.2.3 Freight Literature

Only a limited number of prior studies have been dedicated to investigating the transferability of freight demand models. An investigation of the transferability of the linear regression models that were used to forecast freight trips was carried out by Holguin-Veras et al. (2013). The primary objective was to assess whether these linear regression models could be effectively applied in different scenarios. The models were updated using various transfer methods to enhance the transferability. However, it is crucial to acknowledge that the scope of this investigation was confined to specific study areas or regions. Consequently, it remains uncertain whether these models can be extrapolated to novel regions beyond the ones studied. A similar limitation was noticed in a study by Pani et al. (2021). They developed models to estimate the freight tonnage and trips generated by India's various industries. The quantity was modelled using linear regression, while the number of trips was modelled using negative binomial regression. All these models' transferability to various study regions across the nation was evaluated. The transferability between regions within a state, i.e., intrastate transferability, was observed to be superior to interstate transferability. The geographical location of these regions could be the cause. The assessment of transferability across coastal regions was performed, and it was determined that the extent of transferability was enhanced. Comparable evaluations were conducted between these coastal and landlocked regions, and it was determined that the extent of transferability was somewhat diminished. However, the applicability of these transferability findings to regions that were not included in the study has not yet been investigated. In addition, the results of these studies were limited to the estimation of FG/FTG in the regions, and the sample size requirement in geographically similar regions was undetermined.

2.3 Measurement

Freight surveys primarily aim to fulfil the data requirements associated with modelling urban freight movement at three levels: a) city level, b) establishment level, and c) supply chain level (Kriegel et al., 2011). The freight survey techniques currently in practice include Establishment-based freight survey (EBFS), commodity flow survey (CFS), freight operator survey, stakeholder survey, administrative survey, roadside intercept survey, truck traffic count survey, vehicle trip diary survey, vehicle tracking (GPS) based survey, license plate survey, and parking survey. Among these techniques, EBFS has gained significant popularity for developing freight demand model systems due to its capacity to capture decision-making behaviour within the freight system (Allen et al., 2012; Toilier et al., 2018). Two primary reasons account for its prevalence. The first reason is that freight movement decisions are taken by establishments in the freight system (Pani & Sahu, 2019a). Therefore, surveying the establishments that produce and consume freight is necessary to reflect the economic behavioural theory and how the decisions are affected by policy interventions, land use changes, and logistical provisions. This approach parallels the household survey in passenger transportation (Hunt et al., 2006). The other reason is the feasibility of surveying the freight origin nodes (production units or shippers), which are fewer than the end consumers or retailing firms (Mccabe et al., 2013). In numerous instances, the practice of EBFS is carried out concurrently with roadside intercept surveys or GPS surveys. This simultaneous implementation is undertaken to enhance the precision of data concerning the shipment size and truck routes (Allen et al., 2012). In regions where conducting EBFS is not feasible, models like FRETURB are used to characterize urban freight operations based on two inputs: establishment file and zoning file (Toilier et al., 2018).

Indeed, the studies in freight transportation planning have primarily focused on documenting various survey methodologies to streamline and improve the efficiency of freight data collection at the national level. While these efforts have been valuable in enhancing data collection practices, the studies have not extensively explored or investigated statistical approaches to reduce the resources required for freight surveys.

2.4 Summary

Previous freight studies have typically estimated freight generation (FG) or freight trip generation (FTG) using conventional approaches like OLS regression. However, some studies have explored non-conventional modelling techniques as well. Nevertheless, there remains a need for further investigation to compare and evaluate these different modelling methodologies to identify the most suitable approach for accurate FG estimation. Additionally, the existing studies have often focused on individually developing city-specific, regional, or industrial models. However, it is essential to conduct comparative analyses at various levels to determine the best modelling choice for each specific context. This comprehensive evaluation can provide valuable insights into selecting the most appropriate modelling approach for FG estimation at different geographical and industrial scales.

The literature suggests that relatively few transportation planning studies specifically addressed passenger demand model transferability. Similarly, there has been a smaller number of studies focusing on the transferability assessment of freight demand models. However, the existing freight studies have not thoroughly explored the factors that influence transferability, such as the choice of modelling methodology and the sample size used in the models. Additionally, these studies have tended to provide general transferability results without considering the extrapolation of these results to new regions that lack sufficient freight data.

Furthermore, the current research on freight transferability has been primarily focused on determining the extent of transferability. However, it has not extensively investigated how survey resources can be conserved through the use of transferable models. There is a need for more comprehensive research to understand the potential savings in terms of time and cost that can be achieved by utilising transferable freight demand models in regions with limited data availability.

Chapter 3: Research Design

3.1 Data Collection Program

The nationwide freight data collection program was conducted through the Establishment-based Freight Survey (EBFS) method. Skilled enumerators performed face-to-face interviews with employees who possessed knowledge of the establishment's freight logistics. The gathered information encompassed various key aspects, namely (i) essential establishment details, such as gross floor area (GFA), number of employees (NE), and the products manufactured; and (ii) freight characteristics, including freight generation (FG), freight trip generation (FTG), and trip frequency. Additionally, comprehensive data was meticulously recorded concerning each freight trip, including its origin, destination, shipment size, and truck type. The data collection effort spanned across several cities in Kerala, including Kannur, Kochi (also known as Cochin), Kottayam, Kozhikode (also known as Calicut), Malappuram, Palakkad, and Thrissur. Moreover, data was also collected from landlocked cities such as Jaipur and Hyderabad. The study cities are geo-coordinated in Figure 3.1.

3.1.1 Sampling Frame Development

The database on the number of establishments registered in each city can be assessed through the Udyam list (Government of India, 2022). During sampling frame development, we faced the following issues – (i) some establishments in the list were not operative, and (ii) the unavailability of addresses of establishments. The information about the establishments was accessed through Commercial Tax Departments and local search websites to deal with these issues. The non-functional establishments were removed from the list. The final sampling frame consisted of following number in the study cities – Kannur (4,761), Kochi (13,045), Kottayam (4,983), Kozhikode (6,900), Malappuram (11,984), Palakkad (5,330), and Thrissur (7,167), Jaipur (31,725), and Hyderabad (97,733).

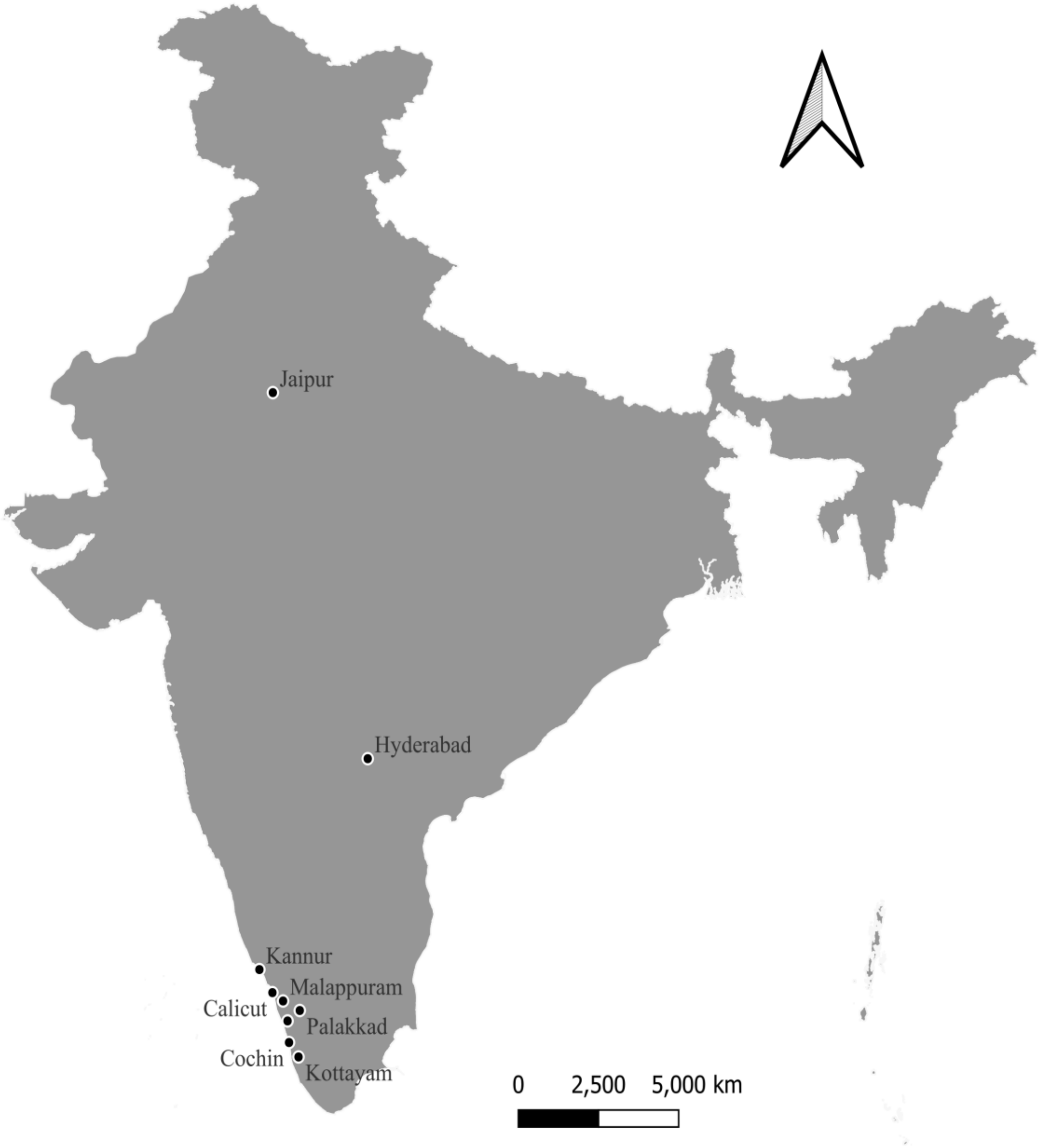


Figure 3.1 Study cities

3.1.2 Pilot Survey

The effectiveness of the designed questionnaires was assessed by conducting pilot surveys in the study cities. The study cities underwent pilot surveys to assess the questionnaire's efficacy and suitability. It was observed that face-to-face interviews yielded a higher response rate making it a better data collection method in this context. Based on the insights from the pilot survey, the questionnaire was modified to be shorter and more precise to improve the response rate. The interviewers were trained regarding the various aspects of the EBFS before the data collection stage based on the experiences from the pilot surveys. The interviewer training covered the following aspects: (a) content briefing, (b) survey administration, (c) interviewing procedures and (d) probing techniques.

3.1.3 Sample Selection, Recruitment and Documentation

Simple random sampling was adopted in this study since auxiliary information was missing for many records in the sample frame. The required minimum sample size (Raosoft Inc, 2004) for the study area was calculated using the following formula.

$$n = \left[\frac{\dfrac{z^2 \times p(1-p)}{e^2}}{1 + \left(\dfrac{z^2 \times p(1-p)}{e^2 N} \right)} \right] \quad 3.1$$

Where *n* is the minimum sample size required; N is the population; *z* is the z score (*z* = 1.96 if we consider a confidence level of 95%); *p* is the proportion of the population, which is taken as 0.5 by default; *e* is the margin of error which is by default considered as 0.05 or 5%.

Samples were drawn from the database until the required number of completed surveys was obtained. Initially, we used a web-based (email) strategy to contact the establishments in the overall sample. Unfortunately, this strategy yielded no responses. Consequently, we chose a telephone-based strategy. We called each establishment in the sample and asked to speak with the logistics manager or someone in a managerial position who could serve as the survey respondent. After receiving a positive response, a follow-up call was placed to the designated contact person to inquire about their willingness and availability to participate in the survey. When reluctance was initially expressed during the initial contact, we inquired about a convenient time for subsequent callbacks.

3.1.4 Data Collection

Following the completion of interviewer training, a data collection schedule was created based on the pre-scheduled appointment timeslots of businesses within the study area. Interviewers, organised into teams, were strategically assigned to various regions within the study area in order to maximise the number of interviews conducted each day. During these interviews, participants were encouraged to provide thorough responses by being informed of the significance and intent behind each question. However, we encountered a significant obstacle, when enquired for information regarding the tonnage requirements of establishments. The majority of establishments measured or weighed their shipments at varying intervals, such as weekly, monthly, or annually, relying primarily on third-party logistics providers. To address this issue, we gathered tonnage data according to the records maintained by the establishments and converted it into equivalent weekly or daily tonnage. Regular checks were performed throughout the data collection process to ensure that responses were complete. In cases where data was missing or entries were ambiguous, we promptly identified and clarified these issues by contacting the respective establishments via telephone. The data collection process consisted of multiple phases, resulting in a total of 943 completed interviews across the study cities.

3.2 Methodological Framework

To achieve the primary objectives of the thesis, we broadly divided the objectives as MMM, as depicted in Figure 3.2. The 3Ms method comprises three key steps: "Model Estimation" entails the development of freight data models, "Model Application" focuses on transferring these developed models spatially, and "Measurement" involves the collection of freight data at the establishment level. We gathered data from the designated study areas and employed modelling techniques to facilitate the transfer of the models. The findings from this transferability assessment will aid in determining the minimum sample requirement (which is once again “measurement”) for regions that lack sufficient freight data. Therefore, in the end, we get to the methodology illustrated in Figure 3.2. For the outline of Parts II, III, and IV, refer to section 1.4: Thesis Outline

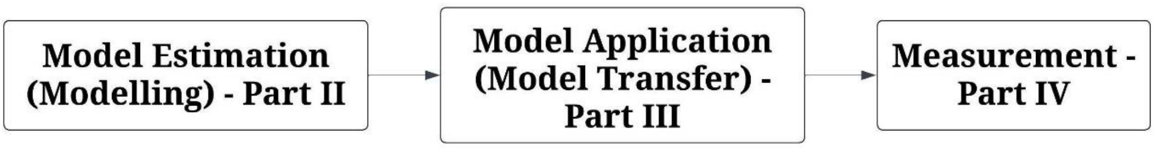


Figure 3.2 Outline of methodological framework

Part II: Model Estimation

Chapter 4: Estimation of Freight Generation Models

4.1 General

There are relatively a few freight demand models available. However, many studies concentrate on the passenger travel demand model in developing countries. The following are some of the factors that are contributing to the difficulties experienced by developing countries in the freight demand models' estimation: (i) there are no freight survey practises in place as a result of limited resources, and (ii) passenger traffic is given higher priority than freight traffic. Freight generation (FG) models are bi-conceptual and are included in the freight demand models. FG encompasses freight production (FP) and freight attraction (FA). The majority of the studies in the past did not validate the models. However, some of these studies attempted to estimate the FG models. In addition, few studies investigated the accuracy of different modelling methodologies. Also, a limited number of studies have developed FG models at various levels. This study makes an effort to address these research gaps by (i) estimating FG models using parametric and non-parametric modelling methodologies, (ii) developing FG models at state, regional, and industrial levels, and (iii) determining which modelling methodology is the most effective at each level.

4.2 Methodology

The establishment-based freight survey (EBFS) was used to collect data on freight transportation from 432 businesses in the districts of Kannur, Kochi, Kottayam, Kozhikode, Malappuram, Palakkad, and Thrissur. When developing these models, we considered FP, FA, the total gross floor area (GFA) of the establishment and the total number of employees (NE) who work there. GFA and NE were considered to be explanatory variables. The establishments were classified according to (i) their geographical location and (ii) a posteriori segmentation of the industries, and this method was proposed by Pani and Sahu (2019a). The posteriori segmentation involves categorising industries according to a novel approach derived from statistical evidence. The establishments were divided into urban and suburban categories, determined by their respective geographical locations. The establishments were categorised into three industrial segments using a posteriori segmentation scheme (Pani & Sahu, 2019a). Further details of the industrial segmentation are presented in Figure 4.1.

Industrial Segment 1	**Industrial Segment 2**	**Industrial Segment 3**
ISIC 16: Wood, wood products, furniture, and fixtures ISIC 24-25: Basic metal, alloy, metal products	ISIC 11: Beverages ISIC 13:Textile mills ISIC 17-18: Paper, paper products, and printing ISIC 22: Plastic and rubber products ISIC 29-30: Transportation equipment	ISIC 10: Food products ISIC 14: Wearing apparel ISIC 20-21: Basic chemicals, chemical products, and pharmaceuticals ISIC 23: Non-metallic mineral products ISIC 26-28: Machinery and equipment ISIC 32: Other manufacturing industries

ISIC : International Standard Industrial Classification

Figure 4.1 Details of a 'posteriori' industrial segments

The key statistics for the sampled establishments' FP, FA, GFA, and NE at state-, urban-, suburban- and industrial segment levels are presented in Table 4.1. In Kerala State, the average weekly FP is 21.3 tons; the average weekly FA is 15.2 tons; the average GFA is 773.7 m^2; the average NE is 28. In the data set, 327 establishments are located in urban regions, and the remaining are located in suburban regions. The average weekly FP and FA of the urban establishments are 19.6 tons and 14.4 tons, respectively; the average weekly FP and FA of the suburban establishments are 26.8 tons and 17.8 tons, respectively. At the regional level (urban and suburban), it is observed that the FP and FA of suburban establishments are more compared to urban establishments. Also, the averages of business size variables (GFA and NE) of the suburban establishments are more. It is also observed that the average FP and FA of suburban establishments are more than the average FP and FA of an establishment in Kerala. The data set includes 101 industrial segment 1 establishments, 94 industrial segment 2 establishments, and 237 industrial segment 3 establishments. At the industrial segment level, the average weekly FP range from 11.6 tons to 40.3 tons; the average weekly FA range from 11.5 tons to 23.8 tons. The average weekly FP and FA of industrial segment 1 is more than the other industrial segments. Also, these average values are more than the average values of establishments in Kerala State.

Table 4.1 Descriptive statistics of data

Variables	**FP**	**FA**	**GFA**	**NE**
Units	tons/week	tons/week	m^2	No.
Kerala State (n=432)				
Minimum	0.6	0.7	53.0	1
Maximum	173.1	104.2	4599.4	200
Average	21.3	15.2	773.7	28
Coefficient of Variation	1.1	0.8	0.9	1.1
Urban Establishments (n=327)				
Minimum	0.6	0.7	53.0	1
Maximum	173.1	67.3	4333.5	200
Average	19.6	14.4	749.8	26
Coefficient of Variation	1.1	0.8	0.9	1.1
Suburban Establishments (n=105)				
Minimum	1.9	1.2	65.3	3
Maximum	136.5	104.2	4599.4	185
Average	26.8	17.8	848.0	31
Coefficient of Variation	1.0	0.9	0.9	1.0
Industrial Segment 1 (n=101)				
Minimum	1.4	0.7	122.4	4
Maximum	136.5	104.2	4559.4	83
Average	40.3	23.8	800.0	22
Coefficient of Variation	0.7	0.7	0.8	0.7
Industrial Segment 2 (n=94)				
Minimum	1.4	1.2	65.3	3
Maximum	173.1	55.8	4333.5	180
Average	25.2	16.0	890.8	33
Coefficient of Variation	1.0	0.8	0.9	1.0
Industrial Segment 3 (n=237)				
Minimum	0.6	1.2	53.0	1
Maximum	94.2	50.0	4333.2	200
Average	11.6	11.5	712.9	28
Coefficient of Variation	1.0	0.8	0.9	1.1

The FG models were developed using parametric and non-parametric modelling methodologies. The parametric methods included ordinary least squares (OLS) regression, weighted least squares (WLS) regression, robust regression (RR), and seemingly unrelated regression (SUR). The non-parametric methods included multiple classification analysis (MCA) and support vector regression (SVR).

4.2.1 Parametric Modelling Approaches

The most commonly used methodology for FG estimation is OLS. We used OLS in developing FG models. The following is the model structure for OLS.

$$FG_i = \beta_0 + \beta_1 (X_j)_i + \varepsilon_i \tag{4.1}$$

Where, FG_i is freight generation (FP or FA) in tons by i^{th} establishment; X_j is the independent variable, j =1 & 2 (GFA & NE); β_0 is the intercept; β_1 is OLS estimator for the slope of the regression line; ε_i is a stochastic error term that is assumed to be normally distributed with a mean of zero and a variance of one.

One of the assumptions of the OLS is that the error term has a uniform variance (i.e., homoscedasticity). If this assumption is violated, we adopt WLS regression. In this regression method, the weights are incorporated into every observation to reduce the effect of non-uniform variance. The weights are incorporated as follows.

$$\frac{FG_i}{\sigma_i} = \beta_0^* \left(\frac{1}{\sigma_i}\right) + \beta_1^* \left(\frac{(X_j)_i}{\sigma_i}\right) + \frac{\varepsilon_i}{\sigma_i} \tag{4.2}$$

Where, β_0^* and β_1^* are transformed estimates by dividing equation 4.1 by the standard deviation (σ_i).

The OLS estimations lead to inaccurate estimates in the presence of outliers and influential observations. The effect of these observations can be minimised using RR. This regression technique includes M estimation, R estimation, and L estimation. We used the M estimation procedure given by Mendenhall and Sincich (2012).

SUR is a system of linear equations where the error terms are correlated. Every equation in consideration possesses its response variable and a distinct set of explanatory variables. Within these equations, each represents a legitimate regression equation suitable for estimation. The structure of the system of equations is as follows.

$$(FG_j)_i = \beta_0 + \beta_1 (GFA_i) + \varepsilon_{1i} \tag{4.3}$$

$$(FG_j)_i = \beta_2 + \beta_3 (NE_i) + \varepsilon_{2i} \tag{4.4}$$

4.2.2 Non-parametric Modelling Approaches

MCA is a widely used technique that does not impose any specific functional form. MCA yields outcomes similar to those obtained from multiple linear regression with dummy variables. In this approach, the average FG (dependent variable) values for each combination of cross-classified categories can be determined, irrespective of the goodness of fit. The explanatory variables, GFA and NE, are divided into four distinct bands (GFA: 0-400 m^2, 400-800 m^2, 800-1200 m^2, >1200 m^2; NE: 0-15, 15-30, 30-45, >45). The model structure of MCA at various levels is as follows.

For state level, $$FG_i = \beta_0 + \sum_{k=1}^{3} \delta_k \left(x_{j_k}\right)_i + \varepsilon_i \quad 4.5$$

For regional level, $$FG_i = \beta_0 + \sum_{k=1}^{3} \delta_k \left(x_{j_k}\right)_i + \theta(R)_i + \varepsilon_i \quad 4.6$$

For industrial level, $$FG_i = \beta_0 + \sum_{k=1}^{3} \delta_k \left(x_{j_k}\right)_i + \sum_{s=1}^{2} \theta_s (I_s)_i + \varepsilon_i \quad 4.7$$

Where FG_i is the freight generation (FP or FA) in tons by the i^{th} establishment in a geographical region R_j (*R = 1* for urban and *R = 0* for suburban) and in industrial segments I_s predicted using k^{th} level of the business size variable x_j ∀ j, *j=1 and 2* for NE and GFA, respectively, x_{j_k} is a dummy variable defined for GFA and NE level ∀ *k*, *k = 1 to 3* (the area level >1200 m^2, and the NE level >45 are the references); I_s is the dummy variable defined for industrial segments I_s $\forall\, s, s = 1\ to\ 2$ (industrial segment 3 is taken as reference); β_0 is the intercept; $\delta\ and\ \theta$ are the regression coefficients; ε_i is the stochastic error term such that $E(\varepsilon_i) = 0$.

SVR is a supervised machine learning algorithm that is dependent on kernel functions. The commonly employed kernel functions include linear, polynomial, sigmoidal, and radial basis. The SVR establishes a hyperplane (line/curve) that accommodates as many points as possible within the decision boundaries. These boundaries are defined by a distance of ε (allowable error) from the hyperplane. We used radial basis kernel functions for estimating FG considering ε = 0.1. The objective of the kernel SVR is to minimise the following objective function.

$$\frac{1}{2}w^T w + C\sum_{i=1}^{N}\varepsilon_i \qquad 4.8$$

subject to

$y_i(w^T x_i + b) \geq 1 - \varepsilon_i$

$\varepsilon_i \geq 0, i = 1, 2, \dots, N$

Where w is the weight vector of the separating hyperplane, the regularisation term or box constraint, C, is a hyperparameter that varies depending on the optimisation goal.

The radial basis kernel function is as follows.

$$K(x, u) = exp\left(-\frac{||x - u||^2}{\sigma^2}\right) \qquad 4.9$$

FG is estimated using the radial basis kernel function for ε and C. The grid search method combined with internal validation is used to determine the values of ε and C. The ε values range from 0 to 0.1; C values range from 1 to 100.

4.2.3 Model Validation

Once we developed parametric and non-parametric models for different levels (state, regional, and industrial), we performed model validation to determine the most effective modelling approach at each level. The validation process involved utilising the root mean square error (RMSE) and mean absolute error (MAE) metrics. These metrics, as defined in the literature, are always positive. Lower values of RMSE and MAE indicate greater prediction ability and superior model performance. The formulae for these metrics are as follows.

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_n - \hat{y}_{\boldsymbol{n}})^2} \qquad 4.10$$

$$MAE = \frac{1}{N}\sum_{n=1}^{N}|y_n - \hat{y}_{\boldsymbol{n}}| \qquad 4.11$$

Where, N is the number of observations, y is the observed FG, and ŷ is the predicted FG.

The dataset at all levels was split into training and validation datasets. The proportions of these datasets were 90% and 10% of the total dataset. We validated the models using internal and external validation techniques.

4.3 Results and Discussion

First, the models were estimated using parametric and non-parametric modelling approaches using 90% of the total dataset. Second, the estimated models were validated. The results and discussion of the estimated models and their validations are as follows.

4.3.1 Estimation of Freight Generation Parametric Models

Table 4.2 provides an overview of parametric FP models. It is clear from this table that each parametric model has a statistically significant result of at least 90%. As a result of the fact that all of the coefficients are positive, the model summary makes it abundantly clear that there is a positive correlation between FP and the business size variables (GFA and NE). The neoclassical economics theory of production agrees with this positive correlation, which is consistent with its predictions. According to this theory, the amount of output (i.e., quantity produced) is a function of the parameters that are considered inputs. These parameters include employment, land, and capital. Some of the freight studies provided an explanation that was very similar to this (Balla et al., 2021; Pani et al., 2018).

Table 4.2 Summary of parametric freight production models

Freight Production (FP) Models									
Area-based FP Models					Employment-based FP Models				
Term	OLS	WLS	RR	SUR	Term	OLS	WLS	RR	SUR
Kerala State									
Obs.	388	388	343	388	Obs.	388	388	337	388
Intercept	8.14***	8.46***	6.12***	17.51***	Intercept	12.68***	9.95***	6.64***	18.81***
GFA	1.65***	1.60***	1.16***	0.45***	NE	0.30***	0.41***	0.26***	0.08***
R^2	0.28	0.15	0.29	0.13	R^2	0.17	0.17	0.34	0.10
RMSE	*18.868*	18.870	20.047	20.681	RMSE	*20.192*	20.471	21.476	21.284
Kerala Urban Region									
Obs.	292	292	258	292	Obs.	292	292	255	292
Intercept	7.69***	8.12***	5.07**	16.81***	Intercept	12.21***	9.60***	6.24***	17.98***
GFA	1.59***	1.53***	1.11***	0.39***	NE	0.28***	0.39***	0.22***	0.07***
R^2	0.25	0.13	0.31	0.11	R^2	0.15	0.15	0.35	0.10
RMSE	*18.777*	18.782	20.055	20.471	RMSE	*19.946*	20.210	21.459	20.961
Kerala Suburban Region									
Obs.	94	94	87	94	Obs.	94	94	89	94
Intercept	10.14***	10.21***	8.96$^{\#}$	19.88***	Intercept	14.54***	11.53***	11.90$^{\#}$	21.43***
GFA	1.73***	1.73***	1.31**	0.59***	NE	0.34***	0.45***	0.26$^{\#}$	0.11**
R^2	0.34	0.20	0.27	0.19	R^2	0.22	0.20	0.18	0.12
RMSE	*18.865*	*18.865*	19.730	20.892	RMSE	*20.619*	20.919	21.376	21.828
Kerala Industrial Segment 1									
Obs.	91	91	82	91	Obs.	91	91	82	91
Intercept	24.06***	25.33***	20.75***	33.07***	Intercept	14.64***	17.16***	13.38***	28.28**
GFA	2.04***	1.88***	2.04***	0.93***	NE	1.19***	1.07***	1.35***	0.57***
R^2	0.25	0.17	0.34	0.18	R^2	0.47	0.32	0.57	0.34
RMSE	*23.510*	23.534	23.741	24.636	RMSE	*19.689*	19.784	19.947	21.990
Kerala Industrial Segment 2									
Obs.	85	85	75	85	Obs.	85	85	78	85
Intercept	7.06*	9.26***	7.82*	16.29***	Intercept	8.93**	5.68*	6.20**	17.26***
GFA	2.07***	1.79***	1.32***	1.06***	NE	0.51***	0.62***	0.52***	0.26***
R^2	0.44	0.25	0.48	0.34	R^2	0.44	0.39	0.62	0.33
RMSE	*20.029*	20.168	21.904	21.799	RMSE	*20.101*	20.462	20.233	21.855
Kerala Industrial Segment 3									
Obs.	214	214	186	214	Obs.	214	214	183	214
Intercept	5.22***	3.85***	2.52***	9.72***	Intercept	6.07***	4.31***	4.13***	9.98***
GFA	0.95***	1.15***	1.06***	0.34***	NE	0.21***	0.29***	0.21**	0.08***
R^2	0.28	0.25	0.54	0.17	R^2	0.36	0.31	0.58	0.22
RMSE	*9.856*	9.944	10.059	10.628	RMSE	*9.285*	9.600	9.529	10.303

Note: (1) Obs. denotes number of observations, (2) Parameter for GFA is in 100 m^2, where GFA is gross floor area of the establishment (3) FP (freight production) in tons per week, (4) *** p<0.001, ** p < 0.01, * p < 0.05, $^{\#}$ p < 0.1, (5) OLS, WLS, RR and SUR stand for ordinary least squares regression, weighted least squares regression, robust regression, and seemingly unrelated regression, (6) RMSE stands for root mean square error.

Table 4.3 Summary of parametric freight attraction models

Freight Attraction (FA) Models									
Area-based FA Models					Employment-based FA Models				
Term	OLS	WLS	RR	SUR	Term	OLS	WLS	RR	SUR
Kerala State									
Obs.	264	264	241	264	Obs.	264	264	247	264
Intercept	8.06***	7.41***	4.60***	12.86***	Intercept	10.17***	9.81***	7.76***	13.51***
GFA	0.88***	0.97***	1.13***	0.26***	NE	0.17***	0.18***	0.17***	0.05***
R^2	0.30	0.22	0.53	0.15	R^2	0.21	0.17	0.33	0.10
RMSE	*9.577*	9.597	9.858	10.533	RMSE	*10.154*	10.163	10.421	10.814
Kerala Urban Region									
Obs.	206	206	187	206	Obs.	206	206	187	206
Intercept	7.47***	7.38***	4.35***	12.27***	Intercept	9.51***	9.24***	6.84***	12.91***
GFA	0.92***	0.93***	1.11***	0.28***	NE	0.18***	0.19***	0.19***	0.05***
R^2	0.30	0.23	0.55	0.16	R^2	0.23	0.19	0.41	0.12
RMSE	*9.426*	9.427	9.662	10.375	RMSE	*9.882*	9.887	10.155	10.604
Kerala Suburban Region									
Obs.	58	58	52	58	Obs.	58	58	52	58
Intercept	10.15***	7.13***	3.74^	14.88***	Intercept	12.66***	12.30***	10.83***	15.55***
GFA	0.77***	1.16***	1.47^	0.21*	NE	0.14**	0.15**	0.13***	0.04^
R^2	0.27	0.28	0.52	0.13	R^2	0.13	0.12	0.17	0.06
RMSE	*9.960*	10.421	11.38	10.892	RMSE	*10.869*	10.875	11.081	11.288
Kerala Industrial Segment 1									
Obs.	56	56	46	56	Obs.	56	56	46	56
Intercept	20.93***	18.48***	15.88***	23.40***	Intercept	14.56***	13.77***	15.35***	22.45***
GFA	0.43^	0.72^	0.27**	0.15^	NE	0.45**	0.49**	0.16*	0.10^
R^2	0.35	0.34	0.53	0.10	R^2	0.18	0.14	0.12	0.07
RMSE	*16.977*	17.123	18.211	17.112	RMSE	*15.695*	15.707	17.376	16.681
Kerala Industrial Segment 2									
Obs.	58	58	55	58	Obs.	58	58	55	58
Intercept	6.32***	6.48***	4.25*	10.63***	Intercept	10.53***	8.92***	8.71***	12.63***
GFA	1.18***	1.16***	1.36***	0.67***	NE	0.16***	0.21***	0.13*	0.10***
R^2	0.56	0.45	0.70	0.46	R^2	0.22	0.21	0.25	0.19
RMSE	*8.050*	8.052	8.189	8.962	RMSE	*10.775*	10.913	11.119	10.971
Kerala Industrial Segment 3									
Obs.	152	152	139	152	Obs.	152	152	139	152
Intercept	4.03***	3.53***	2.74***	7.39***	Intercept	6.29***	6.05***	4.48***	8.71***
GFA	1.06***	1.13***	1.00***	0.60***	NE	0.20***	0.21***	0.19***	0.11***
R^2	0.55	0.42	0.72	0.45	R^2	0.48	0.34	0.66	0.39
RMSE	*6.311*	6.329	6.542	6.979	RMSE	*6.776*	6.783	7.067	7.339

Note: (1) Obs. denotes number of observations, (2) Parameter for GFA is in 100 m^2, where GFA is gross floor area of the establishment (3) FP (freight production) in tons per week, (4) *** p<0.001, ** p < 0.01, * p < 0.05, $^{\#}$ p < 0.1, (5) OLS, WLS, RR and SUR stand for ordinary least squares regression, weighted least squares regression, robust regression, and seemingly unrelated regression, (6) RMSE stands for root mean square error.

On a closer look at the summary of these parametric models, it is noticed that the performance of OLS models is prominently higher than that of all other parametric models. The FP rates of suburban establishments (tonnage produced per 100 m^2 and tonnage produced per 100 m^2) are significantly higher than those of urban establishments. There are a few hypotheses as to why suburban areas have higher FP rates, the most prominent of which are as follows: (i) the employees can be hired for lower wages because the cost of living is lower in suburban regions; and (ii) the land in suburban areas is more affordable, which enables more land to be acquired. Table 4.3 provides a concise summary of the FA models. The summary shows that all the models have a statistical significance level of at least 90%. It can also be seen that the variables of FA and business size positively correlate. Compared to all of the other parametric modelling methodologies, it is abundantly clear that OLS is superior in estimating FA.

4.3.2 Estimation of Freight Generation Non-Parametric Models

The MCA averages are presented in Table 4.4. The table shows that the FP and FA increase with an increase in GFA and NE. It is also evident that urban establishments produce and attract more freight compared to suburban establishments. This statement is in line with the findings of parametric models. Among the various industrial segments, the highest FP and FA are in segment 1 and the lowest in segment 3. The proportion of variance explained by area in predicting FG ranges from 0.21 to 0.42; the proportion of variance explained by employment in predicting FG ranges from 0.20 to 0.53.

The radial basis kernel function was used during the SVR model training process. We only used the radial basis kernel function because all the models that were developed using the other kernel functions (linear, polynomial and sigmoidal) demonstrated poor performance in terms of their ability to predict. In Table 4.5, the RMSE values for each model that was developed using SVR at various levels are presented.

Table 4.4 MCA table of FG rates

Freight Productions (FP) Rates						
Area Levels (m^2)	Kerala State (tons)	Geographical Location		Industrial Segments		
		Urban (tons)	Suburban (tons)	Segment 1 (tons)	Segment 2 (tons)	Segment 3 (tons)
0 – 400	11.13	9.97	14.98	32.54	17.28	5.82
400 - 800	19.65	18.46	23.47	36.46	21.19	9.73
800 - 1200	21.11	19.70	24.71	41.23	25.97	14.51
> 1200	39.23	38.02	43.03	55.92	40.66	29.20
R^2	0.21	0.21	-	0.42	-	-
RMSE	19.764	19.648	-	16.665	-	-
Area Levels (m^2)	Kerala State (tons)	Geographical Location		Industrial Segments		
		Urban (tons)	Suburban (tons)	Segment 1 (tons)	Segment 2 (tons)	Segment 3 (tons)
Employment Levels	Kerala State	Geographical Location		Industrial Segments		
		Urban	Suburban	Segment 1	Segment 2	Segment 3
0 – 15	12.19	11.58	14.76	33.66	13.52	3.92
15 – 30	23.76	22.86	26.05	40.77	20.63	11.03
30 – 45	26.80	25.76	28.95	49.79	29.65	20.05
> 45	40.79	39.94	43.12	66.93	46.79	37.19
R^2	0.20	0.21	-	0.53	-	-
RMSE	19.784	19.737	-	15.575	-	-
Freight Attraction (FA) Rates						
Area Levels (m^2)	Kerala State (tons)	Geographical Location		Industrial Segments		
		Urban (tons)	Suburban (tons)	Segment 1 (tons)	Segment 2 (tons)	Segment 3 (tons)
Area Levels (m^2)	Kerala State	Geographical Location		Industrial Segments		
		Urban	Suburban	Segment 1	Segment 2	Segment 3
0 – 400	9.08	8.60	10.83	17.13	10.93	7.20
400 - 800	13.42	12.94	15.17	20.08	13.88	10.15
800 - 1200	17.10	16.53	18.76	23.56	17.36	13.63
> 1200	26.13	25.66	27.88	33.13	26.92	23.20
R^2	0.27	0.27	-	0.38	-	-
RMSE	9.767	9.724	-	8.979	-	-
Area Levels (m^2)	Kerala State (tons)	Geographical Location		Industrial Segments		
		Urban (tons)	Suburban (tons)	Segment 1 (tons)	Segment 2 (tons)	Segment 3 (tons)
Employment Levels	Kerala State	Geographical Location		Industrial Segments		
		Urban	Suburban	Segment 1	Segment 2	Segment 3
0 – 15	10.22	9.98	11.45	19.25	10.58	7.43
15 – 30	9.50	16.06	17.53	23.96	15.29	12.14
30 – 45	17.31	16.88	18.35	26.24	17.56	14.41
> 45	26.00	25.71	27.17	36.16	27.48	24.33
R^2	0.22	0.22	-	0.38	-	-
RMSE	10.094	10.076	-	9.021	-	-

Note: (1) R^2 represents coefficient of determination, (2) RMSE stands for root mean square error, (3) FP and FA rates in tons

Table 4.5 Summary of support vector regression models

Dependent Variable	Independent Variable	RMSE
Kerala State		
FP	GFA	18.467
FP	NE	19.764
FA	GFA	9.480
FA	NE	9.954
Kerala Urban Region		
FP	GFA	18.765
FP	NE	19.709
FA	GFA	9.425
FA	NE	9.633
Kerala Suburban Region		
FP	GFA	18.798
FP	NE	19.488
FA	GFA	8.704
FA	NE	10.045
Kerala Industrial Segment 1		
FP	GFA	22.880
FP	NE	17.829
FA	GFA	16.057
FA	NE	16.383
Kerala Industrial Segment 2		
FP	GFA	20.049
FP	NE	19.362
FA	GFA	7.855
FA	NE	10.210
Kerala Industrial Segment 3		
FP	GFA	9.633
FP	NE	8.340
FA	GFA	6.183
FA	NE	6.625

Note: (1) FP and FA stand for freight production and freight attraction, (2) GFA and NE stand for Gross floor area and number of employees, and (3) RMSE stands for root mean square error

At the state level, the performance of SVR models is significantly superior to that of MCA models. The FG estimates provided by SVR are superior, despite the geographical location in which they are located (whether urban or suburban). The SVR models developed at the industrial level have a superior ability to make accurate predictions compared to the MCA models, especially in the case of industrial segment 3. Therefore, in contrast to the other non-parametric modelling approach (MCA), SVR has shown superior performance at various levels (state, regional and industrial levels).

4.3.3 Internal Validation of Freight Generation Models

According to the summary of the parametric models, OLS provides the most accurate estimates of FG. It was determined through the internal validation of the non-parametric models by using RMSE that the SVR method is the non-parametric method that should be used at the state level. At the regional level, SVR also performs consistently better than other non-parametric approaches (MCA), regardless of whether urban or suburban. On the other hand, MCA demonstrates superior performance in modelling FG for industrial segments 1 and 2. The SVR method is recommended for industrial segment 3.

4.3.4 External Validation of Freight Generation Models

Figure 4.2 represents the graphical representation of RMSE values, and Figure 4.3 represents MAE values. These figures display the RMSE and MAE values for each model. They are based on the data from Table 4.6, which summarises the validation of FP models and Table 4.7, which summarises the validation of FA models. The external validation results show that the RR method is the best among the parametric modelling methods, while the SVR method is the best among the non-parametric modelling approaches. When the values of RMSE and MAE are compared, it is evident that SVR is the superior method for modelling at the state level; it outperforms both parametric and non-parametric modelling approaches.

Regarding regional FG models, the non-parametric modelling approaches demonstrate improved predictability. To be more specific, SVR is considered to be the superior method for urban FG models. In contrast, MCA performs better for suburban FG models. SVR models produce better results, despite RR models having better prediction ability at the industrial level. When the models developed at various levels are considered, it is abundantly clear that a non-parametric approach such as SVR will be favoured when developing FG models. In addition, non-parametric models are consistently superior in making accurate predictions compared to parametric models.

The findings of the external validation demonstrate that the non-parametric modelling approaches perform better than the parametric modelling approaches when estimating FG. This finding is consistent with the studies that were conducted in the past (Chang, 2005; Mostafa, 2004; Xie et al., 2003; Zhang & Xie, 2008) in various subfields related to transportation engineering. Mostafa (2004), for example, utilised artificial neural networks (ANN) to forecast the marine traffic in the Suez Canal and compared its performance with the parametric approach like Autoregression Integrated Moving Average (ARIMA). According to the findings of this study, the ANN model, which is non-parametric, performed significantly better than the ARIMA model. Studies conducted in the other transportation domains came to the same conclusion: non-parametric models are superior to parametric models. When analysing the factors that influence accident frequency, Chang (2005) used both negative binomial regression and ANN. The results showed that the non-parametric ANN model performed significantly better. For mode choice modelling in the Bay Area of San Francisco, California, Xie et al. (2003) evaluated a parametric modelling approach - a multinomial logit model with a non-parametric modelling approach – decision tree and ANN. They discovered that the ANN model performed better in most scenarios. Zhang and Xie (2008) developed travel mode choice models in the San Francisco Bay Area using multinomial logit, ANN, and SVR. The SVR exhibited a superior ability to predict the outcomes of these three modelling methods.

Although these studies almost universally agree that the non-parametric models are superior, it is essential to remember that the modelling strategy used is contingent on the dataset. In other words, the modelling strategy used depends on the dataset. Kulpa (2014) developed models to estimate the number of truck trips taken in Poland using the parametric method - multiple linear regression and the non-parametric method – ANN. It was discovered through the findings of external validation that the modelling approach had varying degrees of superiority, and this was because two different datasets were used. From these findings, we concluded that parametric models sometimes produce superior outcomes. These findings highlight that the model performance highly depends on the data used in various modelling approaches. Consequently, arriving at conclusive generalisations regarding which modelling approach is appropriate for any dataset is challenging.

However, non-parametric models may be a preferable alternative for estimating FG in developing countries where freight data does not follow any particular trend or distribution. This lack of trend in freight data in developing economies can be attributed to the prioritisation of passenger traffic over freight traffic in transportation planning, which results in smoother movement for passengers but constrained movement for freight. In other words, passenger traffic is prioritised over freight traffic. In addition, the lack of resources in developing nations required to set up an efficient freight transportation system can lead to uncertainty in freight logistic decisions at the establishment level. This uncertainty can be detrimental to the establishment. As a consequence of this, employing non-parametric models could end up being the choice that is best suited to circumstances such as these.

4.4 Research Implications

The modelling methodologies presented in this study are essential for transportation planners and practitioners looking to adopt suitable approaches for forecasting FG. The increased demand for commodities, rapid urbanisation, the growth of e-commerce, globalisation, and flexible government policies in establishing industries have all increased the need for such forecasting. Predictions can be made using either a parametric or non-parametric approach. However, caution is required while selecting one over the other. It is essential to remember that the inferences drawn from this study regarding the performance of various modelling approaches are limited to the study dataset. It is suggested to use parametric modelling in situations in which the distribution of the data is known. However, non-parametric techniques should be used when the distribution is unknown.

The FG model is a beneficial instrument for transportation planners, as it can estimate the amount of freight tonnage in various regions. Accurate freight demand forecasting through non-parametric modelling approaches such as multiple classification analysis (MCA) can facilitate the smooth movement of commercial vehicles in sparsely populated regions with high industrial growth potential due to lower land value. Estimates from the model can be used to assist in locating significant freight corridors in the state of Kerala and other geographical regions in India that are comparable to it. This research contributes to the quantification of FG and the comprehension of its variation across a wide range of geographical locations and industrial types by employing an appropriate modelling approach.

Table 4.6 Validation of FP models

Validation method	Metric	Area-based FP Models						Employment-based FP Models					
		Parametric				Non-parametric		Parametric				Non-parametric	
		OLS	WLS	RR	SUR	MCA	SVR	OLS	WLS	RR	SUR	MCA	SVR
State-level Pooled Establishment Model													
Internal Validation	RMSE	*18.868*	18.870	20.047	20.681	19.764	*18.467*	*20.192*	20.471	21.476	21.284	19.784	*19.764*
	MAE	13.739	13.747	*12.631*	15.168	13.918	*12.605*	14.513	14.588	*12.933*	15.485	14.206	*13.026*
External Validation	RMSE	*26.880*	26.888	28.557	27.670	27.221	*26.082*	24.658	*23.616*	27.072	27.211	24.120	*22.840*
	MAE	16.813	16.827	*16.043*	18.138	17.637	*16.003*	16.384	15.838	*15.584*	18.219	17.020	*14.824*
Urban Establishment Model													
Internal Validation	RMSE	*18.777*	18.782	20.055	20.471	19.648	*18.765*	*19.946*	20.210	21.459	20.961	19.737	*19.709*
	MAE	13.223	13.227	*12.028*	14.600	13.831	*12.495*	13.897	13.901	*12.187*	14.806	14.113	*12.560*
External Validation	RMSE	13.557	*13.541*	14.693	14.405	26.521	*13.391*	12.742	*12.389*	14.605	14.426	23.625	*12.126*
	MAE	10.384	10.406	*9.888*	11.936	17.559	*9.276*	11.070	10.893	*10.199*	12.302	16.973	*10.093*
Suburban Establishment Model													
Internal Validation	RMSE	*18.865*	*18.865*	19.730	20.892	19.648	*18.798*	*20.619*	20.919	21.376	21.828	19.737	*19.488*
	MAE	14.995	15.005	*14.144*	16.062	13.831	*13.386*	16.120	16.294	*15.254*	16.783	*14.113*	14.259
External Validation	RMSE	47.120	*47.111*	49.353	47.729	*26.521*	47.705	42.240	*40.368*	45.805	46.163	*23.625*	44.847
	MAE	35.939	35.951	*35.051*	37.045	*17.559*	35.751	31.727	*29.908*	31.719	35.641	*16.973*	33.013
Industrial Segment 1 Model													
Internal Validation	RMSE	*23.510*	23.534	23.741	24.636	*16.665*	22.880	*19.689*	19.784	19.947	21.990	*15.575*	17.829
	MAE	16.818	*16.765*	16.992	17.527	*10.215*	15.496	15.558	15.589	*15.363*	16.293	*9.850*	13.749
External Validation	RMSE	*15.853*	16.320	16.065	19.538	24.072	*15.105*	17.252	17.614	*17.041*	19.793	21.041	*17.002*
	MAE	*11.788*	11.846	12.448	13.523	15.067	*10.663*	14.396	14.259	*14.205*	14.364	14.334	*14.206*
Industrial Segment 2 Model													
Internal Validation	RMSE	*20.029*	20.168	21.904	21.800	*16.665*	20.049	*20.101*	20.462	20.233	21.855	*15.575*	19.362
	MAE	14.393	14.292	*13.167*	14.660	*10.215*	13.629	12.839	12.759	*12.279*	14.332	*9.850*	10.457
External Validation	RMSE	*14.135*	14.349	14.370	15.666	24.072	*14.130*	10.008	9.556	*9.340*	13.354	21.041	*12.269*
	MAE	10.461	11.024	*8.226*	13.424	15.067	*8.097*	8.324	*7.400*	7.436	11.861	14.334	*8.570*
Industrial Segment 3 Model													
Internal Validation	RMSE	*9.856*	9.944	10.059	10.628	16.665	*9.633*	*9.285*	9.600	9.529	10.303	15.575	*8.340*
	MAE	6.291	6.228	*5.920*	7.168	10.215	*5.889*	5.947	5.884	*5.538*	6.957	9.850	*5.518*
External Validation	RMSE	5.666	5.878	*5.116*	5.935	24.072	*5.091*	4.828	5.058	*3.879*	5.656	21.041	*5.211*
	MAE	4.553	4.591	*3.984*	5.180	15.067	*3.104*	4.017	4.166	*3.411*	5.023	14.334	*4.193*

Table 4.7 Validation of FA models

		Area-based FA Models						**Employment-based FA Models**					
Validation method	**Metric**	**Parametric**				**Non-parametric**		**Parametric**				**Non-parametric**	
		OLS	**WLS**	**RR**	**SUR**	**MCA**	**SVR**	**OLS**	**WLS**	**RR**	**SUR**	**MCA**	**SVR**
State-level Pooled Establishment Model													
Internal Validation	RMSE	*9.577*	9.597	9.858	10.533	9.767	*9.480*	*10.154*	10.163	10.421	10.814	10.094	*9.954*
	MAE	6.978	6.914	*6.574*	7.887	7.106	*6.710*	*7.600*	7.603	7.236	8.149	7.685	*7.126*
External Validation	RMSE	16.199	*16.044*	16.167	17.595	15.397	*15.077*	16.904	*16.830*	17.367	17.748	15.762	*15.012*
	MAE	8.771	8.633	*8.274*	9.892	8.718	*8.590*	9.571	9.568	*9.125*	10.158	9.300	*8.989*
Urban Establishment Model													
Internal Validation	RMSE	*9.426*	9.427	9.662	10.375	9.724	*9.425*	*9.882*	9.887	10.155	10.604	10.076	*9.633*
	MAE	6.727	6.715	*6.311*	7.715	7.026	*6.295*	7.294	7.288	*6.874*	7.911	7.656	*6.838*
External Validation	RMSE	10.157	10.130	*10.114*	11.812	15.190	*10.110*	11.553	*11.512*	11.906	12.234	15.623	*11.512*
	MAE	6.482	6.464	*6.069*	7.896	8.628	*6.006*	7.743	7.745	*7.423*	8.248	9.341	*7.053*
Suburban Establishment Model													
Internal Validation	RMSE	*9.960*	10.421	11.380	10.892	9.724	*8.704*	*10.869*	10.875	11.081	11.288	10.076	*10.045*
	MAE	7.659	7.529	*7.409*	8.361	7.026	*6.500*	8.478	8.515	*8.144*	8.772	7.656	*7.024*
External Validation	RMSE	26.959	26.340	*26.247*	27.977	*15.190*	26.345	27.060	*26.957*	27.751	28.003	*15.623*	27.347
	MAE	15.261	15.148	*14.969*	15.654	*8.628*	15.525	15.147	15.186	*14.544*	15.693	*9.341*	14.796
Industrial Segment 1Model													
Internal Validation	RMSE	*16.977*	17.123	18.211	17.112	*8.979*	16.057	*15.695*	15.707	17.376	16.681	*9.021*	16.383
	MAE	11.863	11.872	*10.159*	11.917	*6.400*	8.843	10.531	10.530	*9.807*	11.581	*6.447*	9.694
External Validation	RMSE	10.139	10.529	*8.863*	10.022	14.121	*8.408*	10.353	10.484	*8.623*	9.949	14.213	*9.389*
	MAE	8.801	9.078	*7.609*	8.558	8.466	*7.655*	8.119	8.105	*7.346*	8.365	9.121	*7.834*
Industrial Segment 2 Model													
Internal Validation	RMSE	*8.050*	8.052	8.189	8.962	8.979	*7.855*	*10.775*	10.913	11.199	10.971	*9.021*	10.210
	MAE	6.336	6.372	*6.052*	7.483	*6.400*	6.679	8.181	8.250	*7.966*	8.455	*6.447*	6.675
External Validation	RMSE	7.622	7.637	*7.579*	8.257	14.121	*7.794*	*8.695*	8.741	8.849	8.877	14.213	*8.170*
	MAE	6.537	6.564	*6.074*	7.328	8.466	*6.757*	6.946	6.828	*5.877*	7.184	9.121	*5.592*
Industrial Segment 3 Model													
Internal Validation	RMSE	*6.311*	6.329	6.542	6.979	8.979	*6.183*	*6.776*	6.783	7.067	7.339	9.021	*6.625*
	MAE	4.495	4.486	*4.293*	5.212	*6.400*	4.510	4.936	4.907	*4.674*	5.498	6.447	*4.765*
External Validation	RMSE	6.504	6.578	6.706	*6.340*	14.121	*6.385*	6.912	6.957	7.080	*6.709*	14.213	*6.912*
	MAE	5.050	5.039	*4.617*	5.241	8.466	*4.087*	5.182	*5.168*	5.514	5.413	9.121	*5.096*

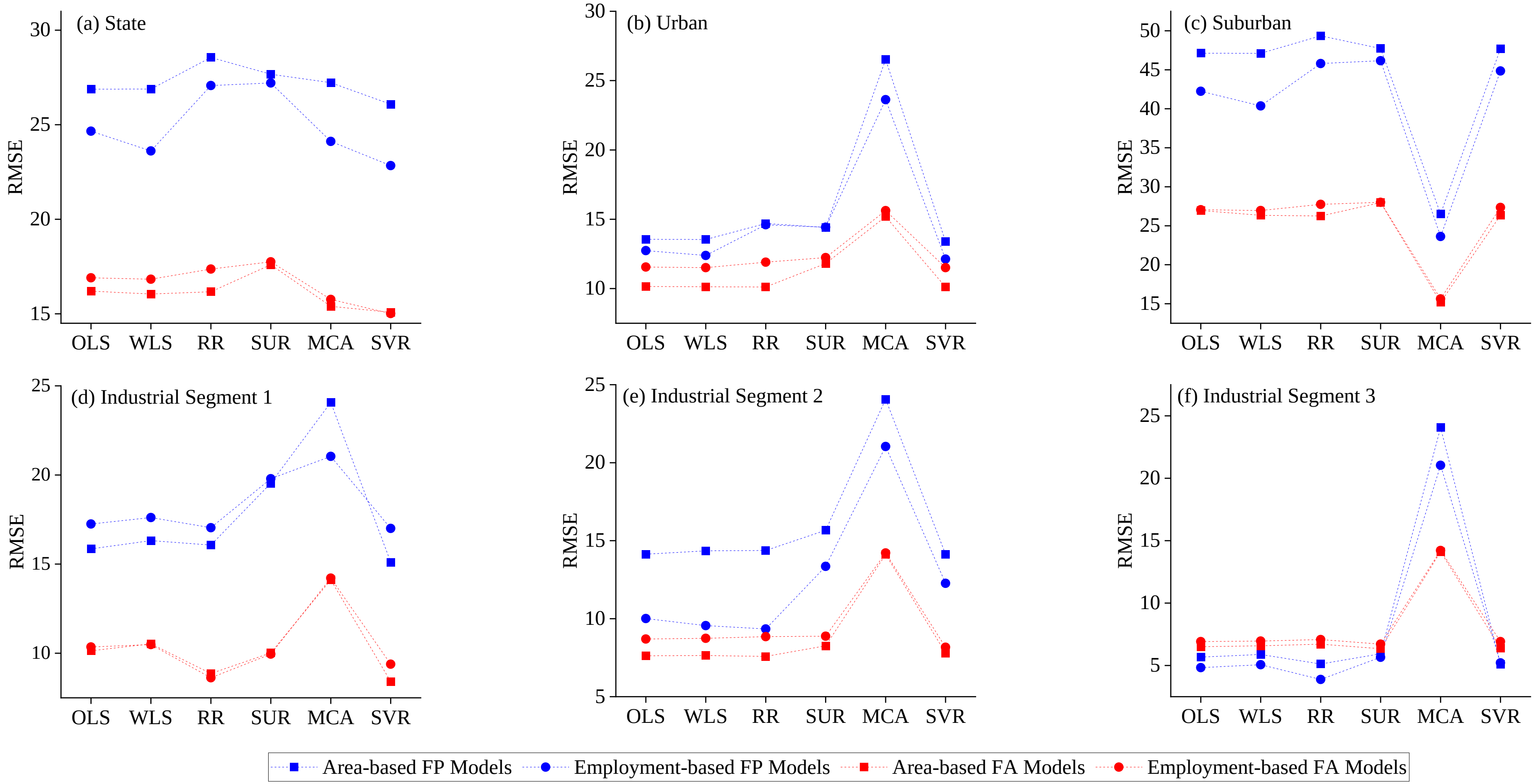


Figure 4.2 RMSE values for various FG models on external validation

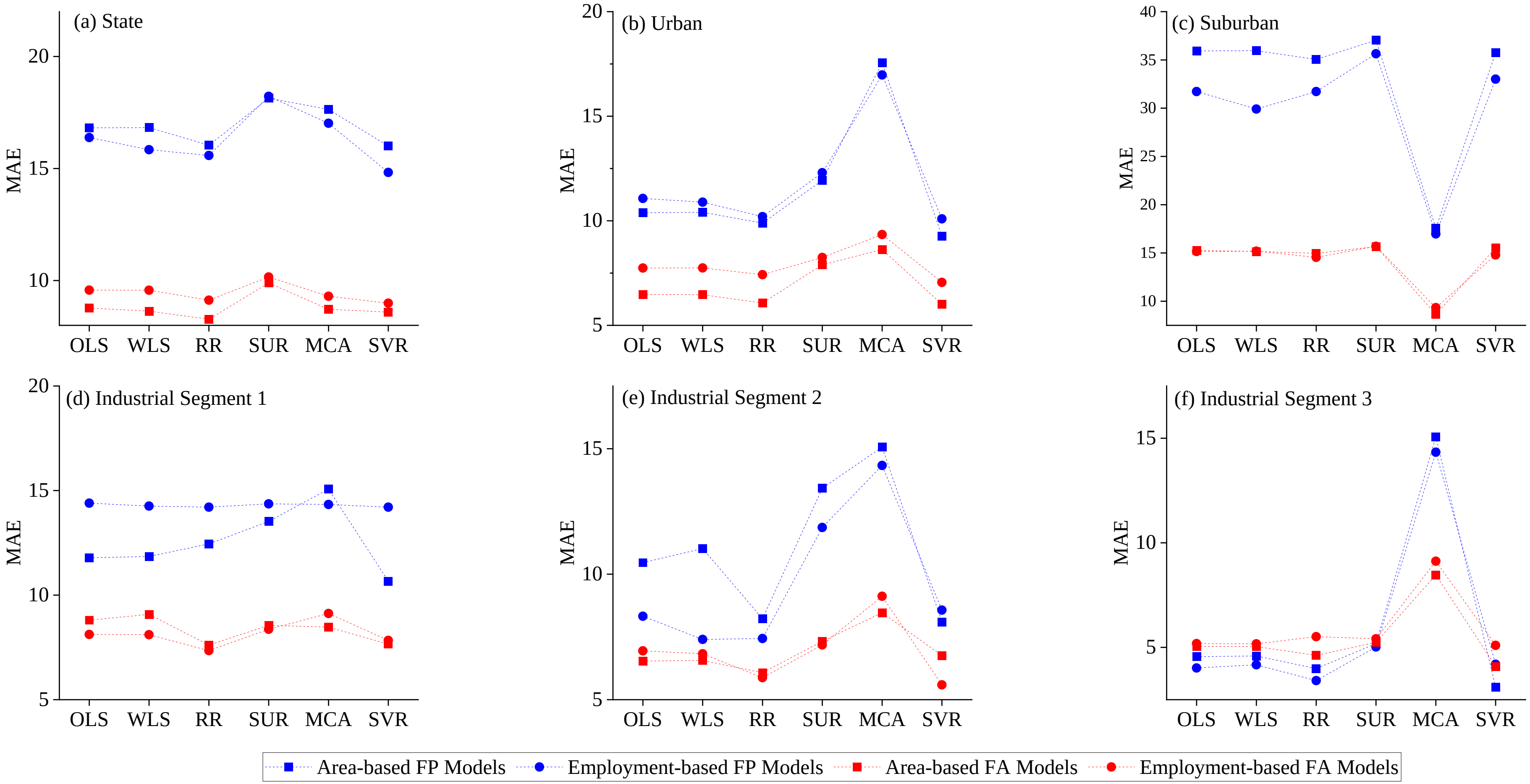


Figure 4.3 MAE values for various FG models on external validation

In transportation planning, selecting the appropriate modelling approach has several advantages, including the ability to pinpoint the most advantageous locations for truck terminals, freight consolidation centres, delivery hubs, and warehouses. In addition, accurate estimates are helpful when clustering industries and strategically selecting areas to locate businesses based on the availability of raw materials and the ability to acquire them. In addition to the benefits that planning provides, appropriate freight demand models can assist owners, managers, and employees in arriving at the best decisions regarding logistics. When applied at the establishment level, these models help minimise the overall cost of commercial vehicle movements, contribute to the facilitation of workforce planning, and provide insights into the mode preference for freight transportation. The freight demand models also support expansion plans by estimating the amount of freight generation based on information regarding employees and gross floor area. In addition, these models also help facilitate the growth and development of businesses.

4.5 Summary

The investigation discussed in this study aims to evaluate the accuracy of forecasts produced by various modelling approaches for calculating FG, specifically FP and FA, at various levels, including state, regional, and industrial segments. The data used for modelling were collected employing an EBFS carried out in seven cities in the Indian state of Kerala. Both parametric and non-parametric modelling techniques were utilised when developing a set of FP and FA models. These models take into account GFA and NE. The OLS, WLS, RR, and SUR are all examples of parametric modelling approaches. MCA and SVR are examples of non-parametric modelling approaches.

According to the findings of the model estimation, the FP and FA rates in establishments in suburban areas are higher. These higher rates can be attributed to two possible reasons: (i) the availability of cheap land in suburban areas, which enables larger establishment areas, and (ii) the recruitment of more workers who are willing to work for lower wages due to the lower cost of living in these areas. The higher FG rates associated with increased GFA and NE are consistent with the theory of production. This theory postulates that output quantity depends on land, employment, and capital. Establishments located in suburban regions have higher average FP and FA, which may result from heavily congested urban roads with insufficient width.

The models developed at various levels were validated using internal and external validation techniques. The validation results show that SVR models, a non-parametric modelling approach, perform better than other approaches when modelling at the state, regional, and level of specific industrial segments. In the case of suburban models, MCA, a non-parametric approach, displays a higher degree of accuracy in FG prediction. RR (parametric models) exhibit superior prediction ability in forecasting FG in specific industrial segments. Overall, comparing results and their interpretation point to the superiority of non-parametric models in FG prediction, with RR being the only parametric approach offering performance comparable to that of the non-parametric models. Nevertheless, SVR is the method for modelling FG when the establishments' freight data distribution is unknown. This choice of methodology is because SVR allows for more accurate predictions in the presence of data with unknown distribution. In addition, non-parametric techniques such as MCA offer a tabular representation of FG rates that the reader can easily understand.

The following are some of the most critical findings and contributions of this study: (i) the suitability of SVR as a superior non-parametric approach for modelling FG at all levels; (ii) the effectiveness of MCA for FG estimation in less populated regions; and (iii) the significant improvement in FG prediction ability achieved by employing non-parametric approaches. However, it is essential to keep in mind that the findings of this research contrasting parametric and non-parametric modelling approaches, as well as the conclusions drawn from those findings, are unique to the dataset that was utilised. Additional research with a wider variety of datasets is required to determine the general performance of particular modelling approaches. For instance, additional research is required to understand why non-parametric models perform poorly for specific industrial models. This question cannot be answered without further investigation. In some instances, the validation results show that the predictive capabilities of parametric and non-parametric models are very similar, indicating the requirement for further testing with additional datasets. In other instances, the validation results show no significant difference between the two types of models. When modelling FG with different approaches, it is recommended that future research should investigate the transferability of models developed in a variety of geographical contexts.

Chapter 5: Establishment Typology in Freight Modelling Framework

5.1 General

The ability to forecast and to plan proactively for evolving freight travel patterns is a critical research area for countries around the globe (Holguín-Veras et al., 2016; Pani et al., 2019). Even in an emerging economy like India, freight ton kilometres are growing at an average growth rate of 9.7% and is are expected to grow further (Pani et al., 2018). Transportation planners, terminal operators, policymakers, and transport authorities must be able to anticipate these changes in freight travel pattern so as to manage or design facilities that foster economic growth and minimize minimise their negative externalities. The importance of this subject has led to multiple studies covering different steps of freight demand models, such as demand generation, flow distribution, vehicle class selection, shipment size choice and route assignment (Gonzalez-Feliu & Sánchez-Díaz, 2019). The first step in modelling freight demand is to identify and estimate demand generation, typically in two alternate quantifications: freight generation (FG) which that quantifies the tonnage of shipments and freight trip generation (FTG) that which quantifies the number of truck trips (Holguín-Veras et al., 2016). At this stage, the choice of granularity for aggregation of freight data is an important and contentious issue faced by modellers. Depending upon the aim, scope and requirements of the model, the needs of data aggregation and in turn, the model accuracy can vary from one model or region to another.

Furthermore, other challenges in aggregation are the availability, quality and confidentiality issues of freight data which often relies on establishment surveys that tend to have proprietary information. For all these reasons, it is crucial to study the relationships of between alternate data aggregation schemes and the accuracy of resulting models to investigate the impact of data aggregation choices. Therefore, this study explores usage of a new establishment typology that captures the variation in fleet ownership, commodity value and establishments’ formation period. The potential of this typology to act as aggregation schemes in the freight demand modelling framework is assessed with respect to traditional classification systems in terms of forecasting accuracy. This study is an important contribution to the literature because the existing works largely focus on producing the model itself but not on the way freight data is aggregated (by means of classification systems or typologies) and its accuracy within a framework of comparative analysis.

Freight data aggregation is a crucial step of the freight demand modelling process for improving the prediction accuracy of model estimates and organising the models so that they complement the forecast requirements of land-use ordinances and policy interventions (Holguín-Veras et al., 2012a). Lately, there is a growing interest in identifying the most suitable classification system for aggregation, which provides homogenous classes that are internally consistent in terms of their demand and trip characteristics (Gonzalez-Feliu & Sánchez-Díaz, 2019; Holguín-Veras et al., 2012a; Pani & Sahu, 2019b). An examination of the literature reveals that the previous studies have predominantly used a priori segmentation approach rather than the data-driven a posteriori segmentation approach. Where segmentations are carried out, they are invariably based on land-use or industrial classifications, not on more complex, statistically derived latent clusters of relevant characteristics (Pani & Sahu, 2019a). The limitation of this approach is that it does not enable us to know whether the existing classification systems are indeed the most important determinants of travel patterns. This contrasts with market segmentation literature that starts from the premise that there is little point in addressing the average consumer or in this context, the average establishment in a particular industry sector or land-use class. Instead, different establishments must be treated in different ways because they are characterised by an inherent business life cycle stage and vehicle ownership evolution process that has an inherent linkage to their freight travel behaviour. This paper addresses this critical research gap by examining how establishments can be meaningfully grouped based on their fleet ownership patterns (size and composition), business age and stage (period of formation), and commodity type (value density). In addition, how these groups can be linked to the classification systems used in traditional approaches (industry sectors).

5.2 Methodology

The methodological approach adopted for developing the establishment typology is outlined in Figure 5.1. As shown, the traditional aggregation approach involves the usage of industrial classes (Holguín-Veras et al., 2016), land use classes (Holguín-Veras et al., 2012a) or ensembles of ‘a posteriori’ segments based on ‘a priori’ classes (Pani & Sahu, 2019a). The proposed approach involves the usage of novel variables that capture the period of formation effect (how business age and freight demand are correlated?), commodity type (how the capital and opportunity cost tied up with commodities determine freight demand?), and fleet ownership (how does the ability to have own-account fleet influence the freight demand?). This study uses latent class cluster analysis (LCCA) which captures the underlying patterns in establishment characteristics, to develop this typology. The main idea of LCCA is that a discrete latent variable can account for the observed associations between a set of indicators, such that, conditional on the latent class variable, these associations become insignificant (Kemperman & Timmermans, 2009). The resulting latent classes, in this context, are groups of establishments that exhibit more homogeneity as a cluster than the total sample from which they are drawn. A graphical representation of LCCA model used in the present study is given in Figure 5.2. As shown, there are two components to the latent class cluster model: (i) measurement model which estimates mixture densities $P(y_{it} = m_t|X)$ and (ii) structural model which estimates mixing weights $P(X|Z_i)$. These model components are explained as follows. While the measurement model is used for developing the establishment typology (step 2 in the proposed approach), structural model is used for allocating the establishments into the typology (step 1 in the proposed approach) based on traditional variables such as industry sector or business size indicators (employment or area). The impacts of proposed typology for freight data aggregation on forecasting accuracy using several statistical metrics.

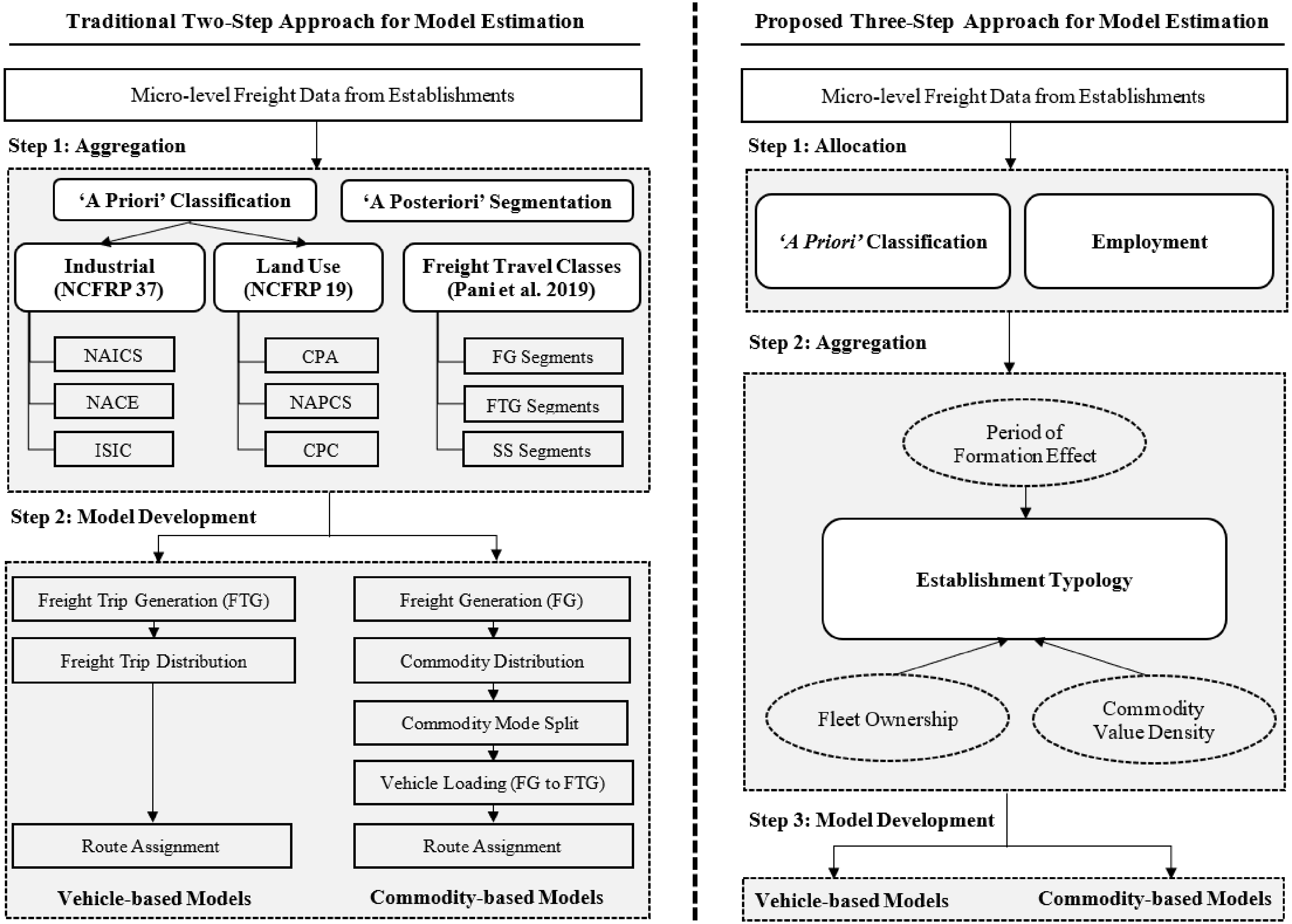


Figure 5.1 Outline of the proposed approach for freight demand estimation

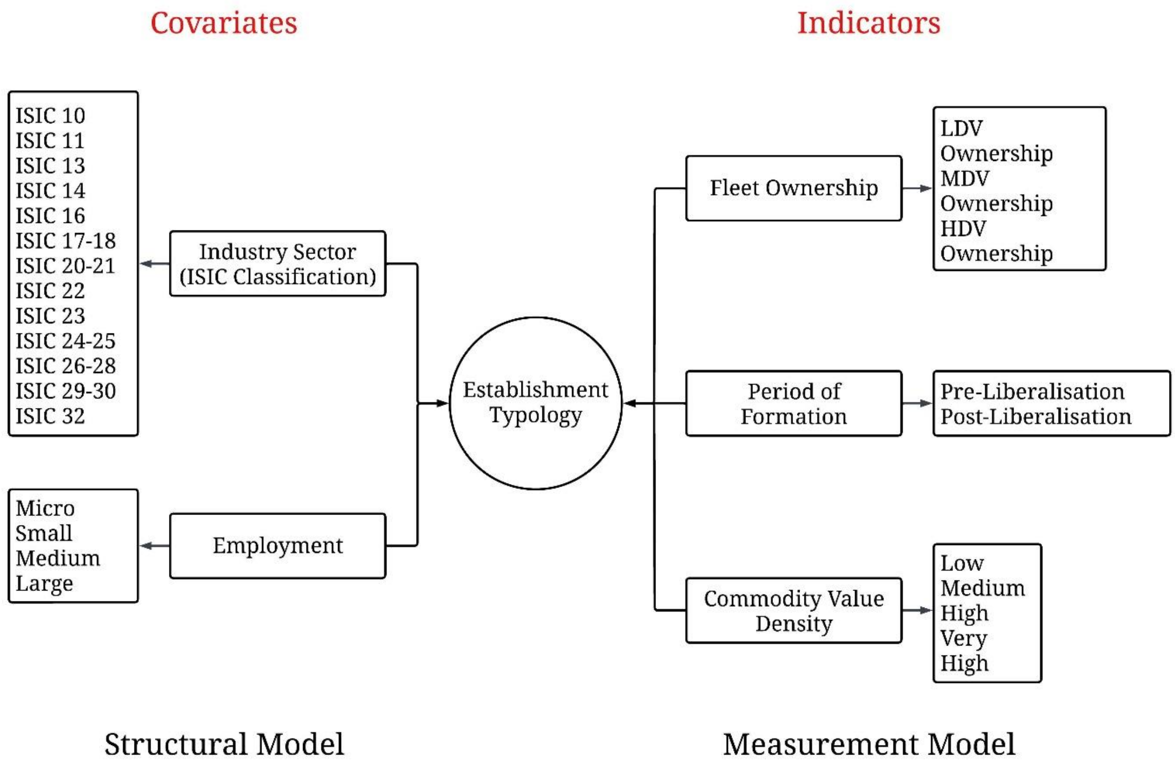


Figure 5.2 Graphical representation of LCCA model

This study uses data from an establishment-based freight survey (EBFS) conducted in Kerala, a Southern State of India. The survey coverage included shippers, such as manufacturing units, assembling companies and raw material production sites. A total of 432 establishments in Kerala provided usable information on their 1613 shipment records. More information on the data collection framework, sampling strategies, response rates, and sample representativeness can be found in Pani and Sahu (Pani & Sahu, 2019b). The survey dataset includes standard freight activity measures such as tons shipped, frequency of shipments, destination of shipment, etc., and establishment characteristics such as gross floor area, number of employees, business age, industry sector, and fleet ownership. The establishments were classified into thirteen sectors using the International Standard Industrial Classification (ISIC) system, ranging from ISIC 10 (food products) to ISIC 32 (other manufacturing products). More information on the industrial classes and their expansions can be found in Pani and Sahu (2019a). The truck fleet ownership data were classified according to their gross vehicle weight (GVW) into three types: (i) light-duty vehicles (LDVs) – GVW < 3.5 tons; (ii) medium-duty vehicles (MDVs) – 3.5 < GVW < 12 tons; (iii) heavy-duty vehicles (HDVs) – GVW > 12 tons. The vehicles included in LDVs are small trucks, pickups and utility vans. On the other hand, MDVs include delivery vans, and HDVs include tractor-trailers and rigid trucks. Fleet ownership values were recorded separately for each truck type as an ordinal variable, which can take up four values: 0, 1, 2 and 3+.

5.2.1 Period of Formation Effect and Business Age

Many recent freight studies have suggested an influence of business age on freight demand (Pani et al., 2018). This study captures this effect by dividing establishments into three based on their period of formation: (i) Post-independence period with low GDP growth (1947 – 1974); (ii) Post-Industrialisation period with medium GDP growth (1975 – 1990); and (iii) Post-liberalisation period with high GDP growth (1991 – present).

5.2.2 Value Density of Commodities

The value density of commodities reflects the amount of capital and opportunity costs tied up to them while in transit. Based on the decision tree analysis (CHAID) conducted using EBFS data, value density was divided into four categories: (i) Low (≤ 47.5), (ii) Medium (47.5 – 132.3), (iii) High (132.3 – 588.4), and (iv) Very High (≥588.4).

5.2.3 Freight Generation Models

The FG models were developed using robust regression (RR) and multiple classification analysis (MCA). The models were categorised into four categories – (i) regional model, (ii) establishment typology models, (iii) a posteriori classification models, and (iv) a priori classification models. The accuracy of four categories of models was compared.

5.3 Results and Discussion

5.3.1 Establishing Typology

The item-response probabilities presented in Table 5.1 are used to label the latent classes as the following: (i) Truck-Free Firms Post-Liberalisation with High Value Products, (ii) LDV Owners Post-Liberalisation with Very High Value Products, (iii) MDV-HDV Owners Post-Liberalisation with Low Value Products, (iv) HDV Owners Post-liberalisation with Medium Value products, and (v) LDV-MDV Owners Pre-Liberalisation with Very High Value Products. A brief discussion of the latent class profiles is given below.

Latent Class 1: Truck-Free Firms Post-Liberalisation with High Value Products

Establishments in this class do not possess truck ownership, as indicated by the high percentages of establishments lacking LDV (75%), MDV (79%), and HDV (99%). A significant portion of these establishments came into existence during the post-liberalisation period. In addition, 75% of the establishments manufacture high value density products.

Latent Class 2: LDV Owners Post-Liberalisation with Very High Value Products

54% of establishments in this class own LDVs and were started in the post-liberalisation period. The value density of the products manufactured by most establishments is exceptionally high.

Latent Class 3: MDV-HDV Owners Post-Liberalisation with Low Value Products

Within this class, establishments posses a minimum of one MDV (60%) or one HDV (63%). These establishments manufacture products with exceptionally high value density. These establishments were initiated post-liberalisation.

Latent Class 4: HDV Owners Post-liberalisation with Medium Value Products

This class primarily consists of establishments, nearly 87% established before the era of liberalisation. These establishments possess at least one HDV (79%) and produce medium-value-density products.

Latent Class 5: LDV-MDV Owners Pre-Liberalisation with Very High Value Products

This class encompasses establishments that own at least one LDV (93%) or one MDV (89%) and are involved in producing high-value-density products. These establishments were established during the post-liberalisation period.

Table 5.1 Measurement LCCA model for identifying the establishment typology

Indicators		Latent Class 1	Latent Class 2	Latent Class 3	Latent Class 4	Latent Class 5	Overall Sample
Cluster Size		**32.41%**	**28.94%**	**18.75%**	**13.66%**	**6.25%**	**100%**
Indicator Variables							
LDV	0	**0.75**	0.46	**0.94**	**0.95**	0.07	0.66
	1	0.18	**0.35**	0.06	0.05	0.14	0.18
	2	0.04	**0.14**	0.00	0.00	0.33	0.11
	3+	0.03	0.05	0.00	0.00	**0.46**	0.06
MDV	0	**0.79**	**0.88**	0.40	**0.81**	0.11	0.73
	1	0.15	0.09	**0.41**	0.14	0.16	0.15
	2	0.06	0.02	0.17	0.05	**0.45**	0.09
	3+	0.00	0.00	0.02	0.00	0.28	0.03
HDV	0	**0.99**	**0.87**	0.37	0.21	**0.68**	0.72
	1	0.01	0.11	**0.27**	0.23	0.22	0.13
	2	0.00	0.02	**0.25**	**0.32**	0.09	0.10
	3+	0.00	0.00	0.12	0.23	0.02	0.06
Value Density (INR/kg)	Low (≤ 47.5)	0.02	0.00	**1.00**	0.00	0.00	0.20
	Medium (47.5 – 132.3)	0.23	0.00	0.00	**0.81**	0.17	0.20
	High (132.3 – 588.4)	**0.75**	0.00	0.00	0.11	0.24	0.31
	Very High (≥588.4)	0.00	**1.00**	0.00	0.08	**0.59**	0.29
Cohort	Post-Liberalisation	**0.82**	**0.78**	**0.72**	**0.60**	0.13	0.72
	Pre-Liberalisation	0.18	0.22	0.28	0.40	**0.87**	0.28
Covariates							
Industry Sector	ISIC 10	**0.59**	0.00	0.00	0.00	**0.20**	0.23
	ISIC 11	0.10	0.00	0.00	0.00	**0.17**	0.05
	ISIC 13	0.00	0.05	0.00	0.00	**0.19**	0.03
	ISIC 14	0.00	**0.14**	0.00	0.00	0.09	0.04
	ISIC 16	0.00	0.00	**0.60**	0.00	0.00	0.12
	ISIC 17-18	0.06	0.00	0.04	0.00	0.03	0.03
	ISIC 20-21	0.00	**0.40**	0.00	0.00	**0.10**	0.10
	ISIC 22	**0.22**	0.00	0.00	0.08	0.00	0.09
	ISIC 23	0.02	0.00	**0.28**	0.02	0.00	0.06
	ISIC 24-25	0.00	0.00	0.08	**0.79**	0.00	0.12
	ISIC 26-28	0.00	**0.27**	0.00	0.00	**0.18**	0.08
	ISIC 29 - 30	0.01	0.02	0.00	**0.11**	0.00	0.02
	ISIC 32	0.00	0.12	0.00	0.00	0.03	0.03
Employment	Micro (1 to 10 Emp.)	**0.43**	**0.34**	**0.34**	0.03	0.00	0.32
	Small (11 to 50 Emp.)	**0.57**	**0.58**	**0.61**	**0.64**	0.00	0.55
	Medium (51 to 100 Emp.)	0.01	0.08	0.01	**0.31**	**0.67**	0.10
	Large (101+ Emp.)	0.00	0.00	0.04	0.02	**0.33**	0.03

5.3.2 Profiling Latent Segments of Establishments

The structural LCCA model estimated in this study assesses how well one can allocate establishments into the proposed typology based on the covariates and the model parameters. The LCCA parameters are effect-coded, and more helpful interpretations of models are presented in Table 5.2 using within-cluster distributions of the covariates. The advantage of this model is that it requires the input of variables used in the current modelling approach, such as industry sector and employment. For instance, the propensity to belong to latent class-1 “Truck-Free Firms Post-Liberalisation with High Value Products” decreases with employment greater than . Many similar interpretations can be made from these results and the LCCA model parameters.

Table 5.2 Parameters and Z values of the estimated LCCA model with covariates

Prediction of Indicators (Measurement Model)													
	Latent Class 1		**Latent Class 2**		**Latent Class 3**		**Latent Class 4**		**Latent Class 5**		***Wald***	***p*-value**	***R^2***
	β_X^t	*z*-value	β_X^t	*z*-value	β_X^t	*z*-value	β_X^t	*z*-value	β_X^t	*z*-value			
LDV Ownership	0.46	2.71	0.56	3.16	-1.42	-3.86	-1.57	-3.44	1.97	7.52	60.05	0.00	0.33
MDV Ownership	-0.36	-2.44	-0.96	-3.87	0.17	1.16	-0.48	-2.15	1.63	6.91	53.58	0.00	0.32
HDV Ownership	-2.77	-4.76	-0.49	-1.86	1.22	6.37	1.63	7.88	0.40	1.53	74.97	0.00	0.43
VD: Low	0.39	0.23	-1.05	-0.30	6.40	3.13	-2.73	-0.85	-3.01	-0.94	97.11	0.00	0.70
VD: Medium	1.62	1.10	-2.40	-0.71	-2.04	-0.61	2.36	1.51	0.47	0.30			
VD: High	2.81	2.09	-1.93	-0.66	-2.13	-0.74	0.41	0.28	0.84	0.58			
VD: Very High	-4.82	-1.83	5.39	2.82	-2.23	-0.76	-0.04	-0.03	1.70	1.16			
Post-Liberalisation	0.54	4.70	0.42	3.32	0.25	2.01	-0.02	-0.12	-1.19	-4.71	33.83	0.00	0.16
Pre-Liberalisation	-0.54	-4.70	-0.42	-3.32	-0.25	-2.01	0.02	0.12	1.19	4.71			
Intercepts of the Measurement Model													
	LDV Ownership		MDV Ownership		HDV Ownership								
	β_m^t	*z*-value	β_m^t	*z*-value	β_m^t	*z*-value							
0	1.98	8.81	1.79	12.11	2.30	9.37							
1	0.63	4.12	0.51	3.06	0.75	4.19							
2	-0.49	-3.10	-0.09	-0.57	-0.55	-3.49							
3+	-2.11	-7.32	-2.21	-7.03	-2.50	-7.90							
Wald	82.05		158.86		92.72								
p-value	0.00		0.00		0.00								
	Value Density												
	β_m^t	*z*-value											
Low	-1.01	-0.72											
Medium	0.32	0.27											
High	0.32	0.30											
Very High	0.37	0.34											
Wald	3.52												
p-value	0.00												
	Cohort												
	β_m^t	*z*-value											
Post-Liberalisation	0.23	2.89											
Pre-Liberalisation	-0.23	-2.89											
Wald	8.37												
p-value	0.00												
Prediction of Latent Class Membership (Structural Model)													
	Latent Class 1		**Latent Class 2**		**Latent Class 3**		**Latent Class 4**		**Latent Class 5**		***Wald***	***p*-value**	
	γ_{xr}	*z*-value	γ_{xr}	*z*-value	γ_{xr}	*z*-value	γ_{xr}	*z*-value	γ_{xr}	*z*-value			
Industry Sector													
ISIC 10	6.58	1.75	-2.21	-0.36	-1.59	-0.26	-3.50	-0.56	0.71	0.20	38.75	0.83	
ISIC 11	3.81	0.84	-2.77	-0.40	-1.27	-0.20	-1.92	-0.27	2.15	0.38			
ISIC 13	-1.74	-0.24	4.41	1.23	-1.50	-0.22	-2.76	-0.38	1.59	0.45			
ISIC 14	-2.44	-0.35	4.71	1.37	-1.65	-0.23	-2.25	-0.31	1.62	0.43			
ISIC 16	-3.67	-0.43	-3.81	-0.44	6.47	1.48	0.41	0.05	0.61	0.07			
ISIC 17-18	2.40	0.72	-4.04	-0.58	2.70	0.80	-1.99	-0.28	0.95	0.23			
ISIC 20-21	-2.64	-0.43	5.32	1.68	-2.04	-0.33	-2.03	-0.32	1.39	0.41			
ISIC 22	5.40	1.53	-1.99	-0.32	-2.04	-0.33	3.05	0.89	-4.41	-0.68			
ISIC 23	1.91	0.61	-3.41	-0.56	5.20	1.76	0.94	0.27	-4.64	-0.73			
ISIC 24-25	-4.62	-0.66	-5.28	-0.75	3.14	0.91	7.73	1.93	-0.97	-0.14			
ISIC 26-28	-2.01	-0.32	5.28	1.53	-2.13	-0.33	-2.11	-0.33	0.96	0.30			
ISIC 29-30	0.08	0.02	-0.10	-0.03	-3.21	-0.47	5.36	1.52	-2.12	-0.29			
ISIC 32	-3.04	-0.43	3.90	1.13	-2.08	-0.29	-0.94	-0.13	2.17	0.51			
Employment													
Micro	3.73	1.76	2.83	1.00	-0.13	-0.07	-3.53	-1.81	-2.90	-0.88	19.16	0.08	
Small	2.78	1.16	1.11	0.48	0.56	0.28	-0.99	-0.47	-3.47	-0.96			
Medium	-2.75	-1.16	-0.24	-0.10	-2.29	-0.87	2.55	1.18	2.73	1.30			
Large	-3.77	-1.01	-3.70	-0.91	1.86	0.58	1.97	0.64	3.63	1.35			
Intercepts	-1.18	-0.46	0.01	0.01	0.38	0.19	-0.26	-0.12	1.05	0.58	0.54	0.97	

5.3.3 Comparing Forecasting Accuracy

The suitability of the proposed typology in the modelling frameworks for FG was assessed by comparing its forecasting accuracy with traditional ‘a priori’ industrial classes. The study findings in this regard are expected to improve the understanding of the impacts of aggregation and reduce the cost of the freight data collection procedure.

Robust Regression

In Table 5.3 and Table 5.4, the summary of different categories of models is presented. Only some ISIC classes with larger sample sizes are presented in these tables. When we compare regional models to establishment typology models and assess their performance using correlation coefficients (ρ), we find that the latter category of models outperforms the former. The prediction accuracy of these establishment typology models is on par with that of a posteriori models. Therefore, it's important to highlight that the prediction accuracy of aggregation models, including establishment typology and a posteriori classification, is comparable. When we contrast these aggregation models with a priori classification models, we observe that some of these models excel in terms of predictive capability.

Table 5.3 Employment-based freight production models

	n	α	β	R^2	SE	SD	ρ
Regional Model							
Kerala State	432	6.60***	0.26***	0.34	7.9	22.9	0.43
Establishment Typology							
Latent Class 01	140	2.80**	0.36**	0.30	4.1	9.6	0.64
Latent Class 02	125	5.49***	0.14*	0.45	4.4	16.5	0.60
Latent Class 03	81	19.43***	1.29***	0.64	13.9	25.8	0.56
Latent Class 04	59	10.60**	0.15***	0.19	7.3	15.5	0.34
Latent Class 05	27	-	0.45***	0.80	19.2	30.7	0.05
A Posteriori Classification							
FP Segment-1	101	12.96***	1.34***	0.54	17.7	26.9	0.69
FP Segment-2	94	5.85**	0.52***	0.62	10.5	26.2	0.67
FP Segment-3	237	4.22***	0.19***	0.55	4.2	11.3	0.61
FTP Segment-1	327	5.18***	0.30***	0.43	7.3	23.4	0.50
FTP Segment-2	105	14.85*	0.12*	0.06	20.4	21.0	0.15
A Priori Classification							
ISIC 10	98	2.73***	0.348***	0.70	4.0	13.2	0.73
ISIC 16	50	-	2.46***	0.87	16.1	22.7	0.40
ISIC 20-21	44	3.96***	0.20***	0.61	4.0	11.3	0.74
ISIC 24-25	51	7.77*	1.46***	0.76	14.8	30.4	0.81

Note: ρ refers to the coefficient of correlation calculated between the actual and predicted values of FP

Table 5.4 Employment-based freight attraction models

	n	α	β	R^2	SE	SD	ρ
Regional Model							
Kerala State	308	7.69***	0.17***	0.324	6.3	12.6	0.43
Establishment Typology							
Latent Class 1	103	2.42***	0.32***	0.41	3.3	8.3	0.43
Latent Class 2	93	5.84***	0.17***	0.56	4.4	9.5	0.70
Latent Class 3	53	16.65***	0.19*	0.12	9.4	17.8	0.38
Latent Class 4	40	-	0.56***	0.62	6.5	7.6	0.25
Latent Class 5	19	-	0.27***	0.75	11.7	12.2	0.04
A Posteriori Classification							
FP Segment-1	101	-	0.85***	0.67	13.7	16.5	0.40
FP Segment-2	67	8.09**	0.15*	0.29	6.4	12.0	0.47
FP Segment-3	175	4.53***	0.18***	0.62	3.7	9.1	0.66
FTP Segment-1	224	5.70***	0.20***	0.52	5.1	12.1	0.51
FTP Segment-2	105	-	0.43***	0.52	14.1	13.4	0.29
A Priori Classification							
ISIC 10	75	3.66***	0.17***	0.53	3.0	8.5	0.49

Note: ρ refers to the coefficient of correlation calculated between the actual and predicted values of FA

Multiple Classification Analysis

Table 5.5 Freight production rates

Employment Level	Establishment Typology				
	Latent Class 1	Latent Class 2	Latent Class 3	Latent Class 4	Latent Class 5
	Freight production rates in tons				
0-15	7.022	3.822	37.185	12.571	27.741
15-30	14.394	11.194	44.557	19.943	35.113
30-45	21.954	18.754	52.117	27.503	42.673
45+	25.837	22.637	56.000	31.386	46.556
	Adj. R^2 = 0.47, Std. Error = 16.8, p-value <0.001, ρ = 0.69				
Employment Level	A posteriori classification based on FP			A posteriori classification based on FTP	
	Segment 1	Segment 2	Segment 3	Segment 1	Segment 2
	Freight production rates in tons			Freight production rates in tons	
0-15	15.858	32.832	3.590	10.305	14.746
15-30	22.073	39.047	9.805	22.267	26.708
30-45	32.107	49.081	19.839	28.477	32.918
45+	39.125	56.099	26.857	33.601	38.042
	Adj. R^2 = 0.44, Std. Error = 17.3, p-value <0.001, ρ = 0.66			Adj. R^2 = 0.18, Std. Error = 20.8, p-value <0.001, ρ = 0.43	

In Table 5.5 and Table 5.6, the FP rates and FA rates are presented respectively. From Table 5.5, it is seen that establishments belonging to latent class 3 have the highest FP rates. However, from Table 5.6, it is noticed that the FA rates of establishments from latent class 2 are higher. On comparision of MCA models of establishment typology with the MCA models of posteriori classification models, the prediction ability of former models is better.

Table 5.6 Freight attraction rates

Employment Level	Establishment Typology				
	Latent Class 1	Latent Class 2	Latent Class 3	Latent Class 4	Latent Class 5
	Freight attraction rates in tons				
0-15	11.193	21.396	7.542	7.412	13.314
15-30	15.471	25.674	11.820	11.690	17.592
30-45	17.388	27.591	13.737	13.607	19.509
45+	23.899	34.102	20.248	20.118	26.020
	Adj. R^2 = 0.34, Std. Error = 10.3, p-value <0.001, ρ = 0.60				
Employment Level	A posteriori classification based on FP			A posteriori classification based on FTP	
	Segment 1	Segment 2	Segment 3	Segment 1	Segment 2
	Freight attraction rates in tons			Freight attraction rates in tons	
0-15	9.957	19.356	7.001	8.700	12.791
15-30	14.295	23.694	11.339	14.675	18.766
30-45	16.527	25.926	13.571	16.677	20.768
45+	24.368	33.767	21.412	22.770	26.861
	Adj. R^2 = 0.34, Std. Error = 10.4, p-value <0.001, ρ = 0.58			Adj. R^2 = 0.21, Std. Error = 11.3, p-value <0.001, ρ = 0.46	

5.4 Summary

Forecasting and planning ahead for changes in how freight moves is an important area of global research. Even in developing economies like India, freight tonne kilometres are going up, and it is important for economic growth and minimising negative effects to be able to predict these changes. A lot of research has gone into making freight demand models, which include things like how demand is made, how flows are distributed, how vehicles are chosen, how the size of shipments is determined, and how routes are assigned. The first step in modelling freight demand is to figure out how demand is made, which is measured as freight generation (FG) and freight trip generation (FTG). These numbers are very important for figuring out the model's level of detail for aggregating data. This study comes up with a new way to cluster establishments based on fleet ownership, the value of the goods they sell, and how long they've been around. Previous studies have mostly used a segmentation method called "a priori segmentation," which is based on existing categories. This method may not be able to capture the complexity of travel patterns.

The freight data collected from Kerala is clustered into five classes based on latent class cluster analysis. These latent classes are as follows: (i) Truck-Free Firms Post-

Liberalisation with High Value Products, (ii) LDV Owners Post-Liberalisation with Very High Value Products, (iii) MDV-HDV Owners Post-Liberalisation with Low Value Products, (iv) HDV Owners Post-liberalisation with Medium Value Products, and (v) LDV-MDV Owners Pre-Liberalisation with Very High Value Products. FG models were developed for each latent class using robust regression (RR) and multiple classification analysis (MCA). Some RR models have shown that the aggregation has improved the prediction. However, MCA models have clearly stated that the models for latent classes are more accurate compared to the MCA models developed for a posteriori classification. Increase in the accuracy of the models due to rational aggregation of data can reduce the requirement of the data. With the essence for some data, the resources for data collection are minimised.

Part III: Model Application

Chapter 6: Spatial Transferability and Transfer Methods

6.1 General

Accurate estimation of freight demand is essential for efficient transportation systems in developing nations like India. However, the nation lacks established practices for conducting commodity flow surveys. The surveys are not conducted due to a lack of resources for conducting freight surveys. The practice of spatial transferability can overcome the resource limitation. In this practice, the already developed model in a region is transferred or applied to another region. By investigating the spatial transferability of the models, metropolitan planning organisations (MPOs) can potentially lessen the financial commitment they make to multiple data collection programmes that are required to develop freight modelling systems (Wafa et al., 2015). Although research on model transferability is limited in the field of freight demand, there are notable studies addressing transferability in the context of passenger trip generation and mode choice models (Agyemang-Duah & Hall, 1997; Sikder et al., 2013; Sikder & Pinjari, 2013; Wafa et al., 2015). By utilising estimated model parameters, MPOs can apply the models to local datasets with relatively smaller sample sizes. This practice allows them to plan facilities while adhering to time and financial constraints. This reduction in cost and time is beneficial for freight demand modelling systems, which heavily rely on data from establishment-based freight surveys (EBFS) or commodity flow surveys. The difficulties associated with EBFS, such as the necessity of extensive training for surveyors and lower rates of complete responses, can result in time delays and higher unit costs per data point when attempting to achieve the required sample size for accurately representing the population (Pani & Sahu, 2019a; Samimi et al., 2013). These problems highlight how important it is to investigate the transferability of freight demand models within the context of a landscape that is constantly evolving in terms of model estimation and diagnostic studies (Giuliano et al., 2010; Holguín-Veras et al., 2016; Pani et al., 2018; Pani & Sahu, 2019b).

It is necessary to conduct a comprehensive study investigating the spatial transferability of freight generation models across different regions to understand the level of transferability between similar or different states. Such research is crucial for planning agencies in developing countries like India as it can help reduce costs associated with conducting freight surveys in regions lacking institutional capacity and resources. Given the sprawling nature of Indian cities, an important research question is determining the extent of transferability when urban models are applied to suburban regions, or vice versa, for estimating freight activity.

This study aims to address this issue by offering two solutions: (i) comparing the relative effectiveness of transferability based on the direction of transfer and (ii) assessing which models can be successfully transferred and which cannot. The models were developed for urban and suburban regions using data from 432 establishments collected through the establishment-based freight survey (EBFS) conducted in seven cities in Kerala. In addition, the models were estimated using 'a posteriori' industrial segmentation scheme (Pani & Sahu, 2019a).

6.2 Methodology

The preliminary analysis was done, and it was found that the relation between the dependent and independent variables was linear. The preliminary analysis included a scatter plot and the coefficient of correlation calculation. The ordinary least squares (OLS) regression was adopted for estimating freight production (FP) for urban and suburban establishments. The explanatory variables considered for modelling were the gross floor area (GFA) and employment (NE). The models were developed without intercept. The logical explanation is that an establishment's economic activity is zero when the employment and area are zero. A similar logical explanation was given in past freight studies (Sahu & Pani, 2020).

After the models were developed, we assessed the models for transferability across different regions and industrial classes. The study uses four statistical metrics to assess transferability: Transfer R^2 (TR^2), Transfer Index (TI), weighted Root Mean Square Error (WRMSE), and relative aggregate transfer error (RATE).

1. Transfer R^2 (TR^2) measures the proportion of variation of the dependent variable in the application context data captured by the transferred model. The maximum value of TR^2 is 1, indicating complete transferability, while 0 suggests no explanatory power in the application context. Negative TR^2 values indicate that the transferred model's results are inferior to the mean of the application context data.

$$Transfer\ R^2\ (TR^2) = 1 - \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y - \bar{Y})^2} \quad 6.1$$

2. Transfer Index (TI) is defined as TR^2 divided by R^2 and quantifies how well the transferred model performs relative to the application context model. TI ranges between 0 and 1, where 1 signifies perfect transferability, and 0 indicates that the transferred model does not describe anything in the application context. Negative TI values are not recommended as they can lead to misleading outcomes. TI values greater than 1 indicate that the transferred model is superior to the local model.

$$Transfer\ Index\ (TI) = \frac{TR^2}{R^2} \quad 6.2$$

3. Weighted Root Mean Square Error (WRMSE) serves as an index to measure the relative error of the model transferred from the estimation context to the application context using local data.

$$Weighted\ Root\ Mean\ Square\ Error\ (WRMSE) = \sqrt{\frac{\sum_{i=1}^{n}\hat{Y}_i \times REM^2}{\sum_{i=1}^{n}\hat{Y}_i}} \quad 6.3$$

$$Relative\ error\ measurement, REM = \frac{(\hat{Y}_i - \bar{Y})}{\hat{Y}_i} \quad 6.4$$

4. Relative Aggregate Transfer Error (RATE) quantifies the ratio between the transfer WRMSE and the local WRMSE.

$$Relative\ Aggregate\ Transfer\ Error\ (RATE) = \frac{WRMSE_t}{WRMSE_a} \quad 6.5$$

In the above equations 6.1 to 6.5, $\hat{Y}_i$ = predicted dependent variable values produced by transferred model (estimation context) operating on independent variable values in the application context data; $\bar{Y}$ is the mean dependent variable values in application context data; Y is the observed dependent variable values in application context data; R^2 = coefficient of determination of linear regression model fitted to application context data; $WRMSE_t$ and $WRMSE_a$ represent the calculation of WRMSE of transferred model and application context model respectively using application context data; n = number of observations in the linear regression model of the application context.

There are various transfer methods. We used combined transfer estimation and joint context estimation. The model parameters were updated using the combined transfer estimation method, and the formulation for updating through this method is as follows.

$$Combined\ Transfer\ Estimation\ Method, \beta' = \frac{\frac{\beta_t}{\sigma_t^2 + d^2} + \frac{\beta_a}{\sigma_a^2}}{\frac{1}{\sigma_t^2 + d^2} + \frac{1}{\sigma_a^2}} \tag{6.6}$$

Where, β' is the updated model parameter; β_t and β_a are the parameters of the estimation context model and application model, respectively; σ_t and σ_a are corresponding standard deviations in estimation and application contexts; *d* is the difference of parameters of the estimation context model and application context model.

In joint context estimation, the data is pooled. With this pooled data, the model is developed, considering the location as the binary variable. In our case, the data from urban and suburban establishments were pooled, and the models were developed considering the dummy variable.

6.3 Results and Discussion

6.3.1 Preliminary Analysis

The descriptive statistics of FP and the establishment characteristics (GFA and NE) for both urban and suburban areas are presented in Table 6.1. The mean FP values of the establishments vary among the urban and suburban regions. It can be seen that the mean FP of the suburban establishments is more than that of the urban establishments. Also, the means of GFA and NE are higher in the case of suburban establishments. Regarding the classified industries, the average GFA and average NE are the highest in the urban- (900.1 m^2) and suburban- (35) establishments, respectively, and these establishments belong to segment 2. In contrast, the highest average FP is noticed in segment 1 of the suburban establishments.

Table 6.1 Summary statistics

Variables	FP	GFA	NE	Variables	FP	GFA	NE
Unit	Tons/ Week	m^2	No.	Unit	Tons/ Week	m^2	No.
All Establishments							
Urban Establishments				*Suburban Establishment*			
n = 327				*n = 105*			
Minimum	0.6	53.0	1	Minimum	1.9	65.3	3
Average	19.6	749.8	26	Average	26.8	848	31
Maximum	173.1	4333.5	200	Maximum	136.5	4599.4	185
CV	1.1	0.9	1.1	CV	1.0	0.9	1.0
Establishments Classified under Segment 1							
Urban Establishments				*Suburban Establishment*			
n = 77				*n = 24*			
Minimum	1.4	122.4	4	Minimum	8.7	260.5	4
Average	36.4	779.4	20	Average	52.6	865.9	26
Maximum	123.1	3888.5	76	Maximum	136.5	4599.4	83
CV	0.7	0.7	0.7	CV	0.6	1.0	0.7
Establishments Classified under Segment 2							
Urban Establishments				*Suburban Establishment*			
n = 70				*n = 24*			
Minimum	1.4	73.0	3	Minimum	2.7	65.3	5
Average	23.8	900.1	32	Average	29.2	863.8	35
Maximum	173.1	4333.5	180	Maximum	109.6	2660.3	160
CV	1.1	1.0	1.0	CV	1.0	0.9	1.0
Establishments Classified under Segment 3							
Urban Establishments				*Suburban Establishment*			
n = 180				*n = 57*			
Minimum	0.6	53.0	1	Minimum	1.9	159.2	3
Average	10.7	678.7	27	Average	14.9	833.9	32
Maximum	94.2	4333.2	200	Maximum	44.2	3681.3	185
CV	1.0	0.9	1.2	CV	0.8	0.9	1.0

Pearson coefficients and the scatter plots indicate a strong linear association between FP and business variables (GFA and NE) for urban and suburban establishments. Pearson correlation coefficient values for FP explanatory variables vary between 0.767 and 0.849, which indicates a strong linear relationship. The scatter plots for urban and suburban establishments are shown in Figure 6.1. The scatter plots between weekly FP and explanatory variables (GFA and NE) for urban and suburban establishments exhibit a robust positive correlation. These findings suggested that a linear model is appropriate to estimate the weekly FP.

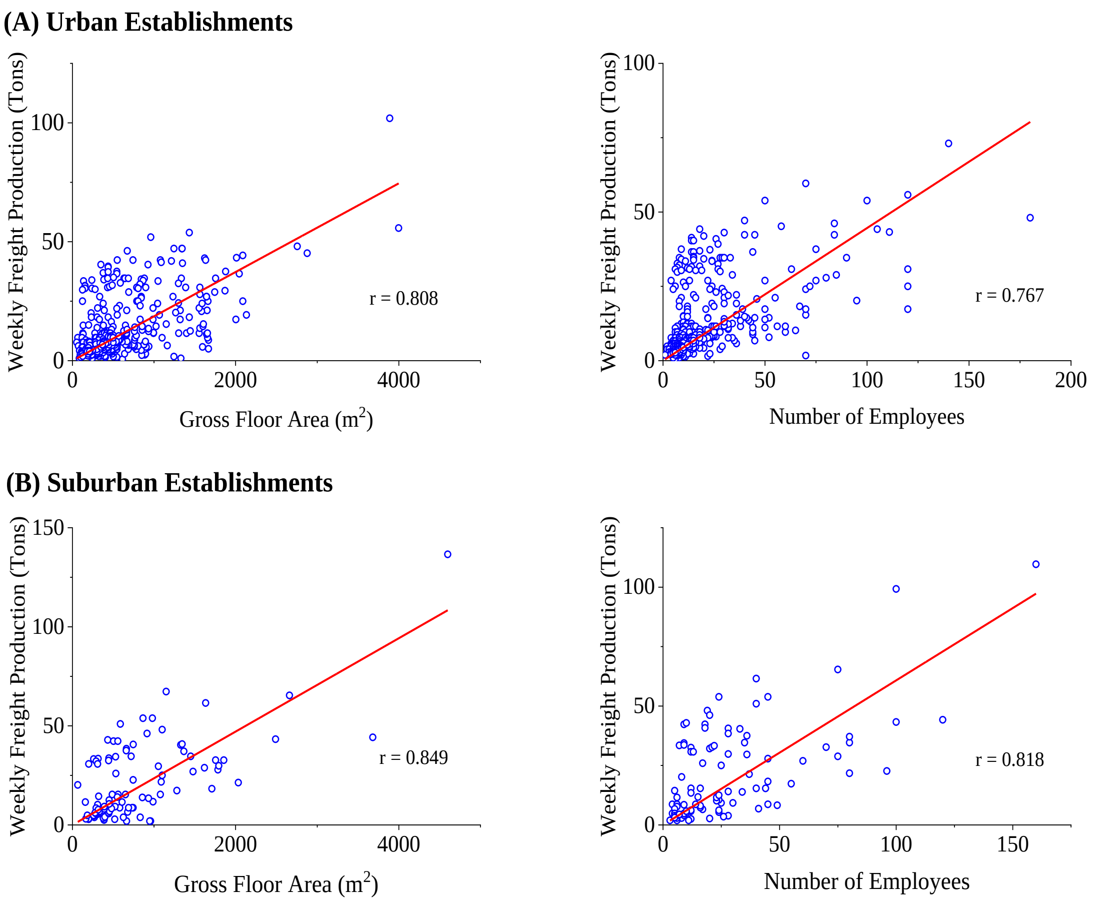


Figure 6.1 Scatter plots for FP models of urban and suburban establishments

6.3.2 Estimation of Freight Production Models

The model estimation results are given in Table 6.2 summarises the FP models for urban and suburban establishments, and Table 6.3, which summarises the FP models for different industrial segments. These tables constitute the sample size (*n*), coefficient of determination (R^2), standard error (*SE*), standard deviation (*SD*), and *t*-statistics. All the developed FP models of urban and suburban establishments are significant at 99.9%, and the R^2 value lies between 0.6 and 0.7. On the other hand, FP models for the three industrial segments show that all the parameters are significant at a 99.9% confidence interval, and goodness of fit (R^2) varies between 0.7 and 0.9.

Table 6.2 FP models based on geographical location

FP Models for Urban and Suburban Establishments												
	Area-based						Employment-based					
	n	*GFA* (β_1)	R^2	*SE*	*SD*	*t-stat*	*n*	*NE* (β_2)	R^2	*SE*	*SD*	*t-stat*
Urban	304	1.87	0.655	12.4	13.9	23.9***	302	0.45	0.588	13.2	13.4	20.7***
Suburban	97	2.36	0.721	15.9	20.5	15.7***	97	0.61	0.670	17.2	20.1	14.0***
Note: *** indicates significant at 99.9%; GFA in 100 m^2												

Table 6.3 FP models for '*a posteriori*' segmentation of establishments

Area-based FP models for establishments in 3 different segments												
Segment	FP models for urban establishments						FP models for suburban establishments					
	n	*GFA* (β_1)	R^2	*SE*	*SD*	*t-stat*	*n*	*GFA* (β_1)	R^2	*SE*	*SD*	*t-stat*
1	71	3.52	0.765	18.7	19.2	15.1***	23	4.11	0.746	30.7	31.7	8.0***
2	68	1.89	0.656	15.7	16.5	11.3***	23	2.85	0.786	17.1	25.5	9.0***
3	171	1.31	0.762	5.8	7.5	23.3***	52	1.53	0.849	6.8	11.3	16.9***
Employment-based FP models for establishments in 3 different segments												
Segment	FP models for urban establishments						FP models for suburban establishments					
	n	*NE* (β_2)	R^2	*SE*	*SD*	*t-stat*	*n*	*NE* (β_2)	R^2	*SE*	*SD*	*t-stat*
1	72	1.80	0.803	17.2	24.2	20.4***	24	1.87	0.907	19.1	32.3	15.0***
2	66	0.58	0.822	10.9	16.1	17.3***	24	0.76	0.874	14.8	29.1	12.6***
3	166	0.32	0.810	5.1	7.1	26.6***	53	0.40	0.841	7.1	11.5	16.6***
Note: *** indicates significant at 99.9%; GFA in 100 m^2												

The FP rates for urban and suburban establishments can be deduced from the FP models presented in Table 6.2. The area-based FP rates for urban and suburban establishments are 1.87 tons per 100 m^2 and 2.36 tons per 100 m^2, respectively. On the other hand, the employment-based FP rates are 0.45 tons per employee and 0.61 tons per employee, respectively. The higher FP rates in suburban regions compared to urban regions can be attributed to factors such as the availability of land with affordable prices and the employment of more people for lesser wages. This explanation is consistent with the theories of production functions in neoclassical economics, which postulate that various inputs, including land, employment, and capital, influence the output (quantity of products). This finding aligns with a previous study conducted by Pani et al. (2018).

Table 6.3 presents the FP models for different industrial segments. The area-based FP rates for urban establishments range from 1.31 tons per 100 m^2 (segment 3) to 3.52 tons per

100 m^2 (segment 1). In contrast, the employment-based FP rates in these establishments vary from 0.32 tons per employee (segment 3) to 1.80 tons per employee (segment 1). For suburban establishments, the area-based FP rates range from 1.53 tons per 100 m^2 to 4.11 tons per 100 m^2, and the employment-based FP rates range from 0.4 tons per employee to 1.87 tons per employee. When comparing industrial establishments regardless of geographical location, higher FP rates are observed in low-value density industries (segment 1), including ISIC 16: Wood, wood products, furniture, and fixtures, and ISIC 24-25: Basic metal, alloy, and metal products. The likely reason behind these higher FP rates is that these industries save more on overhead charges related to handling, packing, and transporting freight. Investing these savings in production leads to higher output (FP rates). However, the developed models suggest that suburban establishments' FP rates are higher than urban FP rates. Additionally, based on the standard error, it is evident that the urban models are more suitable for predicting FP than the suburban models.

6.3.3 Transferability Assessment

The FP models of urban and suburban establishments within Kerala are assessed for transferability. The assessment results of naïve and updated area-based FP models are presented in Table 6.4; employment-based FP models are presented in Table 6.5. The tables compare the performance of naïve and updated models to evaluate the improvement in transferability achieved through parameter updating. The findings indicate that the updated models outperform the naïve models in all cases. This outcome aligns with previous studies, including Sikder and Pinjari (2013), which also showed the superiority of updated models. Given the superior performance of the updated models, most of the discussion in this study focuses on these improved models.

Table 6.4 Transferability assessment of area-based FP models

Urban (Estimation Context) to Suburban (Application Context)								
	Naïve Models				Updated Models			
	TR^2	TI	WRMSE	RATE	TR^2	TI	WRMSE	RATE
All Establishments	0.450	0.625	1.326	1.331	0.414	0.625	1.196	1.200
A Posteriori Segmentation								
Segment 1	*-0.380*	*-0.509*	1.199	1.252	*-0.416*	*-0.509*	1.125	1.175
Segment 2	0.529	0.673	1.532	1.650	0.498	0.673	1.281	1.379
Segment 3	0.338	0.398	0.708	1.247	0.296	0.398	0.655	1.153
Suburban (Estimation Context) to Urban (Application Context)								
	Naïve Models				Updated Models			
	TR^2	TI	WRMSE	RATE	TR^2	TI	WRMSE	RATE
All Establishments	0.034	0.053	0.934	0.781	0.280	0.053	1.090	0.912
A Posteriori Segmentation								
Segment 1	*-0.455*	*-0.594*	0.750	0.830	*-0.225*	*-0.594*	0.856	0.946
Segment 2	*-0.817*	*-1.245*	0.945	0.653	*-0.058*	*-1.245*	1.234	0.853
Segment 3	0.023	0.030	0.725	0.689	0.205	0.030	0.807	0.766

Table 6.5 Transferability assessment of employment-based FP models

Urban (Estimation Context) to Suburban (Application Context)								
	Naïve Models				Updated Models			
	TR^2	TI	WRMSE	RATE	TR^2	TI	WRMSE	RATE
All Establishments	0.388	0.580	1.751	1.461	0.357	0.580	1.536	1.282
A Posteriori Segmentation								
Segment 1	*-0.082*	*-0.090*	0.503	1.059	*-0.103*	*-0.090*	0.492	1.037
Segment 2	0.377	0.510	0.918	1.477	0.327	0.510	0.825	1.326
Segment 3	0.271	0.322	0.963	1.407	0.221	0.322	0.866	1.265
Suburban (Estimation Context) to Urban (Application Context)								
	Naïve Models				Updated Models			
	TR^2	TI	WRMSE	RATE	TR^2	TI	WRMSE	RATE
All Establishments	*-0.362*	*-0.615*	1.079	0.704	-0.025	*-0.615*	1.354	0.883
A Posteriori Segmentation								
Segment 1	*-0.054*	*-0.068*	0.575	0.953	*-0.013*	*-0.068*	0.592	0.982
Segment 2	*-0.799*	*-0.972*	0.644	0.745	*-0.210*	*-0.972*	0.791	0.915
Segment 3	*-0.782*	*-0.964*	0.686	0.755	*-0.311*	*-0.964*	0.828	0.912

When transferability assessment is carried out from urban to suburban, it is evident that most of the area-based urban FP models can be transferred to suburban contexts. Similar observations are seen when employment-based urban FP models are transferred to the suburban context. Only in the case of 'segment 1' the urban model is not transferable to the suburban context. TR^2 and TI in the case of segment 1 are negative, suggesting that the transferred models perform worse than locally estimated models. This negative value is achieved when the transferred model predicts behaviour contrary to that observed (Wilmot, 1995). When the reverse assessment of transferability is done (suburban to urban), the TI value of updated models ranges from -1.245 to 0.053 (area-based and employment-based). This range of TI shows that even though the urban models are transferable to the suburban context, suburban models are not transferable to the urban. It is evident that transferability is not symmetric between two regions within the state, and this finding is consistent with a previous study (Nowrouzian & Srinivasan, 2013). Also, it is noticed that the area-based models are more transferable compared to employment-based models. The possible reason is that compared to area, employment is not best representing freight activities due to possible variations in employment among various industries and the replacement of human resources with automation. In addition, it is noticed that some industrial segment models are not transferable. The possible reason for this is sample size used in developing these models is too small. The small data sets give poor transferability results (Sikder et al., 2014).

In this study, WRMSE and RATE are used to assess the aggregate-level prediction of the transferred model. On updating the coefficients of models, it is noticed that RATE values have improved for updated models. In general, RATE values of updated come closer to 1. The RATE value close to 1 suggests that the aggregate prediction error of the transferred model in local data equals the error of the locally estimated model. When area-based updated models in the urban context are transferred to the suburban, RATE values range between 1.153 and 1.379 for area-based urban updated models in the suburban context. These values indicate that the area-based urban model prediction error in the suburban context lies between 15.3% and 37.9%. When the employment-based models are transferred from urban to suburban, the error ranges between 3.7% and 28.2%. The transferability assessment results show that the area-based urban models are more transferable than employment-based urban models. Also, when *'a posteriori'* segmentation FP models are compared for transferability, it is noticed that all urban models are transferable except low-value density industry (segment 1) models.

For a better perspective of the transferability assessment results, TI and RATE values are considered. As TI is dependent on TR^2 and RATE is dependent on WRMSE, it is appropriate to consider these two metrics. The absolute transfer errors are calculated from RATE values, and absolute transfer error is calculated as a percentage of the absolute value of (1 – RATE). If the RATE equals 1, the absolute transfer error is zero. If the error is zero, then the transferability is perfect. Considering TI and absolute transfer error values, a graph is plotted among them to visualize the transferability. Figure 6.2 represents the graph to visualise the transferability assessment of region (urban and suburban) models; Figure 6.3 visualises the transferability assessment of industrial models. In the graphs, only the assessment metrics of updated models are considered as transferability results in these models are better than the naïve model. The summary of the transferability assessment results is given in Figure 6.4.

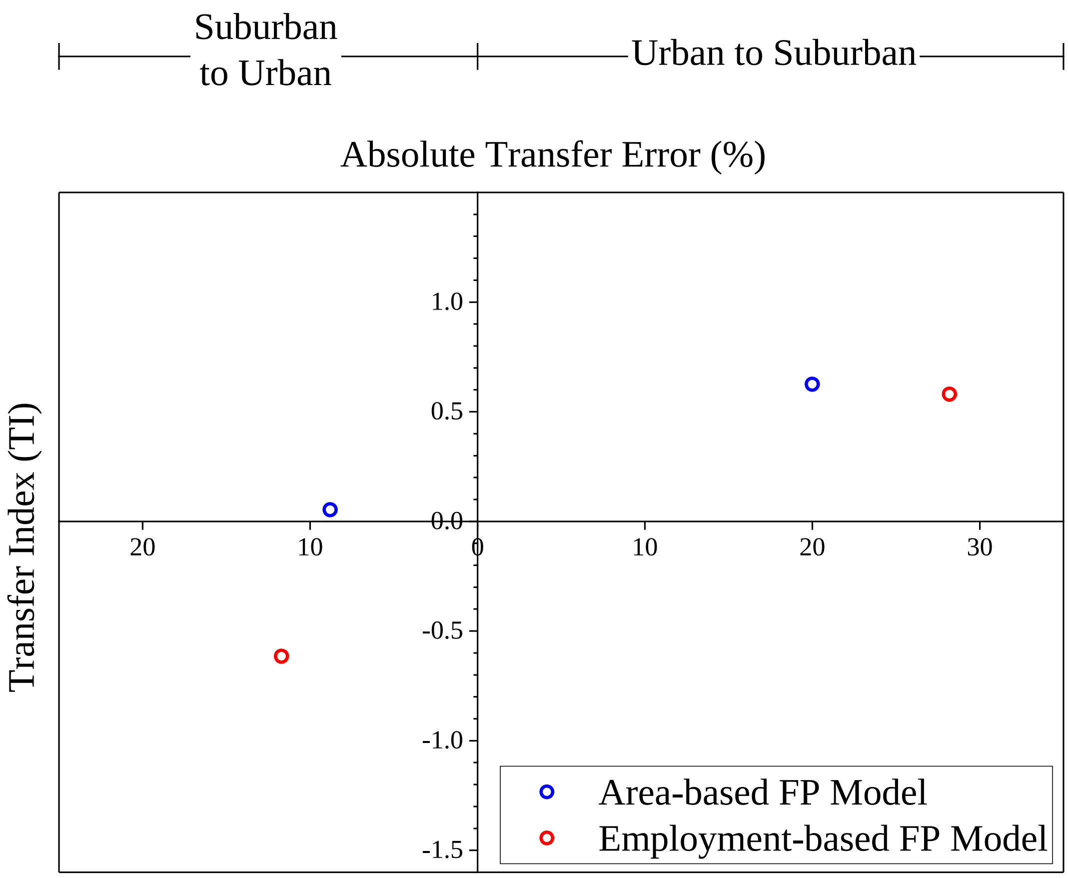
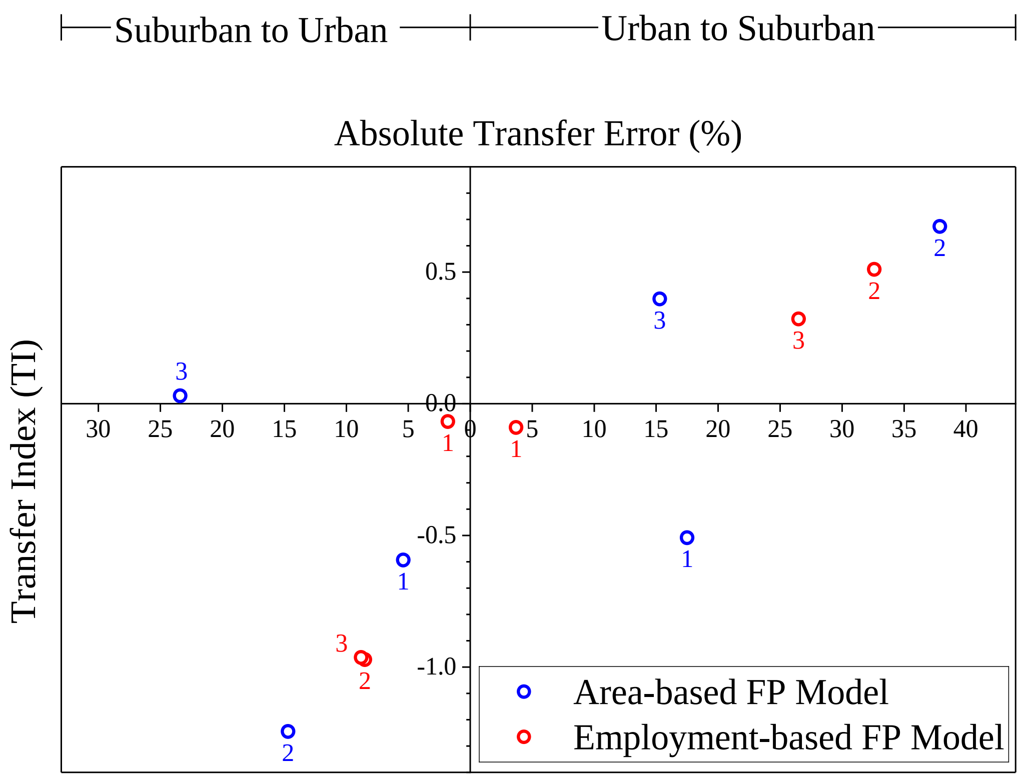


Figure 6.2 Transferability assessment of urban and suburban FP models



Note: 1, 2, 3 represents segments 1, 2 and 3 respectively

Figure 6.3 Transferability assessment of a posteriori segmentation FP models

URBAN TO SUBURBAN TRANSFERABILITY

Area-based FP models

- More transferable than employment-based FP models.
- Except segment 1 industries FP models, all other segment industries FP models are transferable.

Employment-based FP models

- Except segment 1 industries FP models, all other segment industries FP models are transferable.

SUBURBAN TO URBAN TRANSFERABILITY

Area-based FP models

- No FP model is transferable

Employment-based FP models

- No FP model is transferable

Segment 1 industries (Low value density industries)
ISIC 16: Wood, wood products, furniture, and fixtures
ISIC 24-25: Basic metal, alloy, metal products

Segment 2 industries (Medium value density industries)
ISIC 11: Beverages
ISIC 13: Textile mills
ISIC 17-18: Paper, paper products, and printing
ISIC 22: Plastic and rubber products
ISIC 29-30: Transportation equipment

Segment 3 industries (High value density industries)
ISIC 10: Food products
ISIC 14: Wearing apparel
ISIC 20-21: Basic chemicals, chemical products and pharmaceuticals
ISIC 23: Non-metallic mineral products
ISIC 26-28: Machinery and equipment
ISIC 32: Other manufacturing industries

Figure 6.4 Summary of transferability assessment

The transferability assessment was also done using the joint-context estimation method. First, single variable base models (area-based and employment-based) are estimated using the data from the urban and suburban regions. Next, for each selected region, a dummy variable for that region interacts with the variable in the base models to form 'difference variables'. From t-statistics results on these 'difference variables', the following observations are made (i) the area-based base model is transferable to urban and suburban contexts, (ii) the employment-based base model is transferable only to the suburban context. From the t-test, we can determine whether the model is transferable but not the extent of transferability. These base models are further assessed using TI to find the extent of transferability in different regional contexts and industrial segments. In addition, the TI values obtained from combined transfer estimation and joint context estimation are compared to understand if there is any improvement in transferability after pooling the data. Table 6.6 represents the TI values obtained from the combined transfer and joint context estimations. On comparing the base models and updated models, the extent of transferability of the area-based base model (excluding industrial segments) to the urban context has improved. In all other cases, the extent of transferability of updated models using the combined transfer estimation technique is greater. The geographical dissimilarities of urban and suburban regions in terms of population density may be a possible reason for no improvement in TI values on the transfer of the base model. The transferability assessment results show that updated models are better transferable than naïve transfer.

6.4 Summary

This study contributes to existing knowledge of freight demand generation in India by establishing statistically significant relations between FP and business size variables. For this reason, we used data from 432 establishments in Kerala, India, obtained through an EBFS. The establishments were categorized into urban and suburban based on the geographical location to analyse the freight demand model transferability assessment. In addition, '*a posteriori'* segmentation is used to interpret the influence of homogeneous industries. OLS regression method was used to develop single variable models without intercept. The models indicate that the FP rates in suburban regions are more significant than that of urban, irrespective of '*a posteriori'* segmentation. Among the '*a posteriori'* segments, irrespective of urban or suburban, higher FP rates are noticed in Segment 1 (ISIC 16: Wood, wood products, furniture, and fixtures).

Table 6.6 TI for combined transfer and joint context estimations

Transfer Method	Area-based FP Models			Employment-based FP Models		
	All Establishments					
		Transferred to			Transferred to	
	Transferred from	Urban	Suburban	Transferred from	Urban	Suburban
Combined Transfer Estimation^	Urban^	NA	0.625	Urban^	NA	0.580
	Suburban^	0.053	NA	Suburban^	-0.615	NA
Joint Context Estimation	Base Model	0.163	0.454	Base Model	-0.164	0.473
	Segment 1					
		Transferred to			Transferred to	
Transfer Method	Transferred from	Urban	Suburban	Transferred from	Urban	Suburban
Combined Transfer Estimation	Urban^	NA	-0.509	Urban^	NA	-0.090
	Suburban^	-0.594	NA	Suburban^	-0.068	NA
Joint Context Estimation	Base Model	-0.154	-0.512	Base Model	-0.5	-0.667
	Segment 2					
		Transferred to			Transferred to	
Transfer Method	Transferred from	Urban	Suburban	Transferred from	Urban	Suburban
Combined Transfer Estimation	Urban^	NA	0.673	Urban^	NA	0.510
	Suburban^	-1.245	NA	Suburban^	-0.972	NA
Joint Context Estimation	Base Model	-0.215	0.610	Base Model	0.035	0.476
	Segment 3					
		Transferred to			Transferred to	
Transfer Method	Transferred from	Urban	Suburban	Transferred from	Urban	Suburban
Combined Transfer Estimation	Urban^	NA	0.398	Urban^	NA	0.322
	Suburban^	0.030	NA	Suburban^	-0.782	NA
Joint Context Estimation	Base Model	-2.130	-1.200	Base Model	-2.964	-0.893

Note: NA = Not Applicable, ^ represents updated models

The financial burden on Metropolitan Planning Organizations (MPOs) in freight data collection urges for spatial transferability of freight demand models. Transferability assessment is carried out using urban and suburban models to analyse the model performance and the direction of transferability. Transferability assessment performance is performed using TR^2, TI, WRMSE and RATE. The models are directly transferred initially; later, the naïve models are transferred by updating their coefficients using the combined transfer estimation technique. The transferability assessment is done for naïve and updated models, and it can be seen that updated models offer better transferability performance. It is also noticed that most urban models are transferable in the suburban context. However, only a few models are transferable when the suburban models are transferred to the urban context. These statements about transferability between urban and suburban conclude that transferability is asymmetric. The exclusion of spatial factors like population density, road density, road intersection, and the distance of establishment from the seaport and city centre in the models can be a possible reason for the asymmetric transferability between the urban and suburban regions. Also, the suburban models are not transferable because the sample size used for developing these models is small. Finally, among the urban models, area-based models are more transferable than employment-based models. The medium and high-value density industry (segments 2 and 3) urban models are transferable to the suburban context. Among them, medium value density industry (segment 2) models are more transferable. Only the area-based model of high value density industry (segment 3) is transferable in suburban models. A joint context estimation technique is also used to check whether or not the pooled data models give better transferability results. We did not notice higher TI values than the combined transfer estimation technique in this case. In general, due to the rapid growth of freight traffic and a shortage of freight data in suburban regions, the spatial transferability of urban models to suburban contexts can overcome the freight demand model availability to a certain extent in suburban regions.

Several inferences have emerged with significant implications for planning and policymaking. The first finding is that the degree of spatial transferability widely varies among different industrial classes and depends on the following factors: (i) business variables employed for measuring FP, (ii) type of community, and (iii) commodity value density of industry sectors. The first factor suggests that the area-based models show better transferability than the employment-based models. The second factor suggests that transferability is possible from an urban community to a suburban one, not vice versa. The agencies looking forward to transferring FP models may do well if the transfer is done from regions with high to low population densities. The last factor suggests that the transferability of models is better in medium and high-value industries. In this study, the transferability of the FP models opens the possibility for transportation agencies, planners and practitioners to identify the direction and the extent of transferability. The research findings are expected to help save financial resources for data collection exercises and develop a freight model system within budgetary constraints. However, there are certain limitations which pave the way for further research.

The findings may not be generalised as the data pertains to only one state in India. However, the methodology is generic and can be utilised to investigate the freight model transferability direction in cities in other Indian states. This research's findings are helpful for the planning agencies in Kerala or a similar coastal state in India or elsewhere. The research results will assist them in data collection programs and subsequently help develop comprehensive mobility plans. This research is limited to intrastate model transferability assessment and requires further investigation on interstate model transferability. It may also be noted that the small sample size is the limitation of this study. Models built with small data sets produce poor or moderate transferability findings, suggesting the significance of using sufficient samples in building models. The generalisation of this study requires data with a larger sample size from several cities across different states in India. Also, it is essential to identify the factors causing the transferability to be asymmetric and to know the reasons for the transferability of only some industrial segment models.

Chapter 7: Effect of Sample Size on Spatial Transferability

7.1 General

The spatial transferability of a model refers to the technique of employing a model developed in one geographic location to predict outcomes in other geographic locations. In other words, it is the extent to which a model can be applied to new locations outside of the area where the model was developed (Holguín-Veras et al., 2013; Pani, Sahu, & Bhat, 2021; Sikder et al., 2013). This technique is helpful for decision-making and planning since it enables us to extrapolate the model's predictions to other locations. The freight demand models play an essential role in analysing the commodity flow behaviour and anticipating the changes in freight demand for transportation planning. Several freight surveys need to be conducted, each requiring considerable time and money to gather the data needed to estimate these models. Establishment-based freight survey (EBFS) is one such survey, which involves the collection of data at the establishment level - the cost per completed response was around $198 in Calgary in 2000 (Hunt et al., 2006), $185 in Edmonton during 2002 (Hunt et al., 2006), and $500 in Paris during 2012 (Toilier et al., 2016). The freight data collection incurs not only high costs but includes a massive investment in terms of time. Pani and Sahu (2019b) reported that an interviewer, on average, receives 2.2 responses per day in EBFS. The response rate in EBFS is meagre due to the establishment's employees' deterrence towards data sharing (Pani & Sahu, 2019b, 2022). In the USA (Samimi et al., 2013), the response rate for the web-based freight survey was 7%. In the Netherlands (Iding et al., 2002), for EBFS conducted in hybrid mode, the response rate was 15%. The response rates for EBFS are meagre, and responses improve only when the EBFS is conducted through face-to-face interviews; however, the resource requirement (time and cost) is very high (Pani & Sahu, 2019b).

Assessing the spatial transferability of the freight demand models is an efficient solution to manage the constrained survey resources and escalation in demand for the freight data. In the absence of spatially transferable models available to transportation planning practitioners in India, there are high chances of them relying on locally estimated models based on smaller sample sizes. The underfit models with smaller samples may offer inaccurate estimates - leading to erroneous decisions in the planning process of a city or region. In addition, the smaller samples can also lead to unreliable transferability results. It is imperative to understand the extent of transferability of the existing models with varied sample sizes to avoid such errors/mistakes. Especially there is a need for research on the transferability of freight demand models in developing economies like India and many other countries in the South Asian region, where budgetary constraints limit the size of freight data collection programs. This limitation subsequently leads to a dearth of freight demand models, and spatial transferability is a rational approach to deal with this limitation. The model transferability success will depend on the model's predictive ability in the application context with a reasonable sample size. Therefore, this study aims to determine: 1) a reliable modelling approach for estimating freight production (FP) on spatial transfer and 2) the effects of the sample size variations on the model transferability. To the best of our knowledge, this research is unique to freight transportation planning. The study uses the Indian EBFS dataset (the only available dataset) to examine these objectives and recommends future best practices.

7.2 Methodology

The freight data was collected from the cities in India – Calicut, Cochin, Jaipur and Hyderabad. Calicut and Cochin are the coastal cities in Kerala, the southernmost state of India. Calicut has a long-standing history as an important trading centre. It is a city with a well-developed commercial infrastructure known for its active trade in commodities and 24,586 establishments. The city's population density is 5,169 people per km^2 as per 2011 census. On the other hand, Cochin emerged as a significant commercial and industrial hub. It is home to the Cochin Port, one of India's busiest seaports, and has a strong presence in industries. The population density of this city is 6,324 people per km^2. The number of establishments located across the city is 46,042. Unlike these cities, Jaipur and Hyderabad are landlocked cities. Jaipur is the capital city of Rajasthan State, with a population density of 6,531 persons per km^2. The number of establishments located in this city is 2,00,876. It has a diverse manufacturing sector. This city has industrial areas and special economic zones that facilitate industrial growth, attracting investments and generating employment. Among all these study cities, Hyderabad has more industrial activity. It is one of the Indian cities with the highest gross domestic product. Its population density is 10,477 people per km^2. It is the hub for many manufacturing sectors, especially the biotechnology and pharmaceutical sectors. The number of establishments in this city is 97,733. A well-developed transportation system connects the industrial zones in different parts of the city. The population data mentioned above are as per the census 2011 (Office of the Registrar General & Census Commissioner, 2011), and the number of establishments data is as per government records (Udyam) in December 2022 (Government of India, 2022).

The establishment-based freight survey, or EBFS, was a face-to-face interview to collect the freight data. These data were collected from shippers, also known as establishments, in the study cities. The questions in the survey covered topics related to the fundamental information of the establishments, such as the geocoordinates of the establishment, its gross floor area (GFA) in m^2, and the number of employees working there (NE). In addition, it included freight operations, weekly tonnage produced (also known as freight production, or FP in tons), and products manufactured. Based on products manufactured information, the establishments were classified into different industrial classes based on the guidelines of the International Standard Industrial Classification (ISIC) (Department of Economic and Social Affairs, 2008). We considered ISIC 10: Food and food products, ISIC 16: Wood and wood products, ISIC 17-18: Paper and paper products, ISIC 22: Rubber and rubber products, and ISIC 24-25: Basic metal, alloy and metal products. The data from the cities of Calicut and Cochin were collected in 2013. In Jaipur, the data were collected from February 2017 to April 2017. The data from Hyderabad were collected from December 2020 to April 2021. For more details on the data collected, the readers can refer to Pani and Sahu (Pani & Sahu, 2019b). Since the freight data collection process is time-intensive, collecting data from different cities in the same time period is challenging. Our survey duration was more since the data was collected from various Indian cities. However, this study assumes that the freight data undergoes no drastic changes during this time period. This assumption holds particularly for disaggregated data, where short time periods are unlikely to cause substantial differences in establishment characteristics. Therefore, it is reasonable to assume that the freight data remains relatively stable and consistent throughout the survey duration, enabling meaningful analysis and reliable modelling results.

After the data collection program, the research methodology can be broken down into three main steps. First, various modelling approaches were utilised to develop freight generation (FG) models. Second, the models developed in the previous stage were evaluated for their potential to be transferred using direct transfer, also known as a naive transfer. Third, the effect of sample size was examined. The following subsections provide a deeper explanation of each stage.

7.2.1 Model Estimation

FP was estimated using a variety of modelling approaches, such as ordinary least squares (OLS), robust regression (RR), and multiple classification analysis (MCA). Among these methods, OLS and RR can be used to model FP for most of the industrial classes because preliminary analysis (which includes Pearson correlation coefficients and scatter plots) had shown a strong correlation between FP and business size variables – employment (NE) and gross floor area (GFA). Using these specific methods, OLS and RR, single variable FP models were developed for different industrial classes in the study cities. These models included the explanatory variables NE (employment-based FP models) and GFA (area-based FP models) of establishments. Only the single variable models are presented in this study because the models developed with multiple variables were not statistically significant. In the past, freight studies demonstrated the significance of single variable freight generation models. The NCFRP Research Report 37 models were estimated using only employment as an explanatory variable (Holguín-Veras et al., 2016). These models were kept as simple as possible to facilitate their use in data-constrained application environments.

FP models were estimated using the OLS regression technique, and the intercept was considered zero in these models. The literature provided the logical justification for the no intercept model - there is no possibility of economic activities when no employees are employed, or no establishment is constructed (i.e., zero employees and zero area) (Pani et al., 2018; Sahu & Pani, 2020). The model structures adopted for modelling FP are given in 7.1 and 7.2. Unlike OLS, RR models resist errors caused by heteroscedasticity and outliers. Different RR estimation methods eliminate heteroscedasticity and outliers. The methods are M estimation, R estimation, and L estimation. We used M estimation for determining FP. Even within RR models, the intercept was regarded as zero. RR models have a structure similar to that of OLS models.

$$Employment - based\ FP\ Model:\ FP_i = \beta_1 * NE_i + \varepsilon_i \tag{7.1}$$

$$Area - based\ FP\ Model:\ FP_i = \beta_2 * GFA_i + \varepsilon_i \tag{7.2}$$

Where FP_i is the weekly freight produced by the i^{th} establishment (in tons); ε_i is the stochastic error term assumed to be normally distributed in OLS and nonnormally distributed in RR.

OLS regression method gives the linear relationship between the explanatory variables and the response variable. In OLS, the unknown parameters are obtained by minimizing the residual sum of squares (RSS).

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{7.3}$$

Where y_i is the observed value and $\hat{y}_i$ is the predicted value. The M method was used as given by Mendenhall and Soncich (Mendenhall & Sincich, 2012). In this method, the parameters are estimated by minimising the following quantity.

$$\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{7.4}$$

MCA is a frequently employed straightforward method that presents the average FP rates in tabular format and does not follow any functional form (Alho & Silva, 2017; Balla et al., 2023; Guevara & Thomas, 2007; Sahu & Pani, 2020). This technique is comparable to OLS, but dummy variables are employed. The explanatory variables NE and GFA were categorical variables for the MCA. In NE, there are four categories: 0-15, 15-30, 30-45, and >45, and in GFA, there are four: 0-400 m^2, 400-800 m^2, 800-1200 m^2, and >1200 m^2. The interval widths of 15 employees and 400 m^2 of gross floor area were determined based on the performance of the models with different interval widths. A similar method of determining the interval width was adopted by Sahu and Pani (2022) in estimating freight tonnage. After determining the interval width, the FP averages for each cross-classified category were calculated. The structural form of the MCA is as follows.

$$FP_i = \beta_0 + \sum_{m=1}^{3} \delta_m (NE_m)_i + \sum_{k=1}^{4} \mu_k (I_k)_i + \varepsilon_i \tag{7.5}$$

$$FP_i = \beta_0 + \sum_{p=1}^{3} \delta_p (GFA_p)_i + \sum_{k=1}^{4} \mu_k (I_k)_i + \varepsilon_i \tag{7.6}$$

Where FP_i is the freight produced for i^{th} establishment predicted using the m^{th} level of employment and p^{th} level of area. NE_m is a dummy variable defined for NE level ∀ m, m = *1 to 3* (NE ≥ 45 is taken as reference). GFA_p is a dummy variable defined for GFA level ∀ p, p = *1 to 3* (GFA ≥ 1200 m^2 is taken as reference). I_k is the dummy variable for I_k∀ $k = 1 \text{ to } 4$ (ISIC 24-25: Basic metal, alloy and metal products is taken as reference)

7.2.2 Spatial Transferability Assessment

Naïve transfer is an approach in which a model from one region (known as estimation context) is directly applied to another region (known as application context). This transferability approach was assessed using the relative mean absolute error (relative MAE), which measures the transfer error relative to the error of the application context model. Relative MAE is always positive, and the upper bound is infinity. If this metric is less than 1, it indicates that the prediction ability of the transferred model (estimation context model) is better than the application context model.

$$Mean\ absolute\ error\ (MAE) = \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{n} \qquad 7.7$$

$$Relative\ mean\ absolute\ error\ (Relative\ MAE) = \frac{MAE_{est}}{MAE_{app}} \qquad 7.8$$

Where $\hat{y}_i$ is the predicted value for the i^{th} establishment; y_i is the value of the response variable for i^{th} establishment; n is the number of observations, MAE_{est} represents the MAE value of the estimation context model when transferred to application context data, MAE_{app} represents the MAE value of the application context model.

7.2.3 Effect of Sample Size

The previous studies (Karasmaa, 2007; Santoso & Tsunokawa, 2005; Sikder et al., 2014) revealed that the sample size plays a crucial role in understanding the extent of transferability. This study examined the influence of the sample size on transferability by considering different sample sizes of local data (application context data). The local data was divided into three different sizes in every application context, and the sample sizes considered were one-third, two-thirds, and the total size of application context data. The sample size was varied by resampling the data using bootstrapping.

7.3 Results and Discussion

7.3.1 Descriptive Statistics

The descriptive statistics of various industrial classes for the data received from each study city are presented in Table 7.1. Compared to other study cities, Jaipur has a higher average weekly FP for ISIC 10: Food and food products. It can be noticed that the average weekly FP for ISIC 16: Wood and wood products in Calicut is relatively high. This city also has significant weekly FP averages in the industrial class ISIC 17-18: Paper, paper products and printing. Cochin has a high average FP in ISIC 22: Rubber and plastic products. For Calicut in ISIC 24-25: Basic metal, alloy and metal products, the average weekly FP is higher than in other study cities.

7.3.2 Freight Production Models

Table 7.2, a summary of the FP models developed using OLS and RR, can be found. Some of the models do not show statistical significance. In contrast, most models show statistical significance of more than 90%. It is to be noted that the statistically insignificant models can be assessed for their ability to be spatially transferred (Sikder et al., 2013). It has been observed that the NE and GFA positively influence the quantity of product produced (in tons) each week. It has also been noticed that the FP rates (weekly tonnage per employee or weekly tonnage per 100 m^2) for ISIC 17-18: Paper, paper products, and printing are significantly higher. It is noticed that after this industrial class, ISIC 16: Wood and wood products has demonstrated a slightly higher FP rate compared to other industrial classes. However, the industrial models – ISIC 16 and 17-18 are developed with a smaller sample. Therefore, it is crucial to approach conclusions about the FP rates of these industries with caution. More detailed comparisons of FP rates across various cities are compared and illustrated in Figure 7.1. Only the models developed with a reasonable sample size (minimum of 20) were considered in this figure. In addition, only the FP rates of models with better prediction ability (i.e., with lower MAE values) were considered among OLS and RR. This figure shows that the area-based FP rate of Calicut (2.962 tons/100 m^2) is the highest in ISIC 10. Furthermore, within this industrial class, the employment-based FP rate is the highest in Calicut, amounting to 0.928 tons per employee. In addition, it is noticed that Hyderabad leads with the FP rates of 2.769 tons/100 m^2 and 0.665 tons/employee in ISIC 24-25.

Table 7.1 Descriptive statistics of variables

ISIC 24-25: Basic metal, alloy and metal products (n=166)			
	Weekly FP	NE	GFA
Hyderabad (n=118)			
Minimum	0.13	2.00	26.76
Maximum	125.88	85.00	5759.99
Average	26.45	19.19	692.74
Standard Deviation	30.91	16.67	773.15
Jaipur (n=25)			
Minimum	1.20	4.00	53.00
Maximum	120.00	75.00	1562.20
Average	24.36	24.80	567.80
Standard Deviation	30.72	22.59	440.87
Calicut (n=14)			
Minimum	1.35	4.00	131.90
Maximum	123.08	76.00	1944.90
Average	43.43	27.14	836.90
Standard Deviation	36.91	21.59	592.61
Cochin (n=9)			
Minimum	4.81	8.00	430.90
Maximum	53.85	50.00	1643.80
Average	32.76	21.00	803.10
Standard Deviation	13.85	12.78	373.90

ISIC 22: Rubber and plastic products (n=38)			
	Weekly FP	NE	GFA
Hyderabad (n=25)			
Minimum	0.26	4.00	104.20
Maximum	83.11	120.00	72843.40
Average	17.81	27.60	3772.20
Standard Deviation	23.40	25.76	14429.61
Jaipur (n=7)			
Minimum	4.80	8.00	480.20
Maximum	48.60	170.00	1869.90
Average	26.40	45.00	1104.90
Standard Deviation	17.66	56.93	550.49
Calicut (n=3)			
Minimum	3.85	3.00	73.00
Maximum	9.62	18.00	290.00
Average	6.54	11.33	182.00
Standard Deviation	2.90	7.64	108.50
Cochin (n=3)			
Minimum	9.62	7.00	498.80
Maximum	40.38	33.00	1327.50
Average	28.21	24.33	902.70
Standard Deviation	16.35	15.01	414.75

ISIC 16: Wood and wood products (n=41)			
	Weekly FP	NE	GFA
Calicut (n=15)			
Minimum	6.73	6.00	122.40
Maximum	93.65	35.00	1844.80
Average	38.45	18.87	622.80
Standard Deviation	27.07	8.70	471.43
Hyderabad (n=11)			
Minimum	3.12	5.00	146.50
Maximum	64.66	84.00	780.40
Average	34.51	41.60	422.70
Standard Deviation	25.39	29.62	262.39
Cochin (n=8)			
Minimum	5.96	7.00	446.50
Maximum	65.38	46.00	1583.20
Average	34.69	17.50	925.10
Standard Deviation	18.01	13.21	440.49
Jaipur (n=7)			
Minimum	2.40	4.00	123.90
Maximum	25.80	67.00	1500.00
Average	13.54	19.14	691.60
Standard Deviation	9.50	22.63	625.32

ISIC 10: Food and food products (n=77)			
	Weekly FP	NE	GFA
Calicut (n=25)			
Minimum	1.92	4.00	53.00
Maximum	94.23	100.00	1662.10
Average	14.37	17.20	409.90
Standard Deviation	19.67	22.32	503.52
Cochin (n=21)			
Minimum	1.73	1.00	282.50
Maximum	29.62	52.00	2003.40
Average	7.34	13.48	744.40
Standard Deviation	8.20	13.46	390.91
Jaipur (n=20)			
Minimum	3.00	5.00	102.20
Maximum	63.60	84.00	1662.10
Average	19.41	19.45	380.90
Standard Deviation	17.56	21.01	402.81
Hyderabad (n=11)			
Minimum	0.52	5.00	199.20
Maximum	49.62	60.00	1254.20
Average	19.04	29.45	501.90
Standard Deviation	17.09	19.22	320.03

ISIC 17-18: Paper, paper products, and printing (n=29)			
	Weekly FP	NE	GFA
Hyderabad (n=15)			
Minimum	0.91	9.00	240.00
Maximum	46.20	128.00	1437.00
Average	18.95	37.07	666.80
Standard Deviation	15.45	30.92	297.64
Jaipur (n=7)			
Minimum	2.40	6.00	96.00
Maximum	96.00	32.00	4800.00
Average	24.00	13.86	1005.80
Standard Deviation	33.20	9.32	1683.61
Calicut (n=4)			
Minimum	25.00	6.00	102.20
Maximum	36.54	14.00	200.00
Average	31.44	10.75	140.70
Standard Deviation	4.90	3.40	42.02
Cochin (n=3)			
Minimum	5.77	5.00	376.70
Maximum	50.96	40.00	588.30
Average	26.92	16.67	472.10
Standard Deviation	22.73	20.21	107.32

Note: (i) ISIC stands for International Standard Industrial Classification; n refers to sample size, (ii) FP stands for freight production (in tons), (iii) NE and GFA refer to the number of employees and gross floor area (in m^2), respectively

Table 7.2 Summary of OLS and RR models

Employment-based Freight Production Models			Area-based Freight Production Models			Employment-based Freight Production Models			Area-based Freight Production Models		
Calicut						**Cochin**					
ISIC 10: Food products (n=25)						**ISIC 10: Food products (n=21)**					
Term	OLS	RR	Term	OLS	RR	Term	OLS	RR	Term	OLS	RR
NE	0.788***	0.928***	GFA	2.962***	1.949**	NE	0.497***	0.334***	GFA	0.998***	0.615**
R^2	0.831	0.968	R^2	0.624	0.723	R^2	0.744	0.834	R^2	0.592	0.735
MAE	6.074	5.792	MAE	8.492	8.730	MAE	3.957	3.882	MAE	4.960	4.549
ISIC 16: Wood and wood products (n=15)						**ISIC 16: Wood and wood products (n=8)**					
Term	OLS	RR	Term	OLS	RR	Term	OLS	RR	Term	OLS	RR
NE	1.857***	2.206#	GFA	5.068***	4.852***	NE	1.366*	2.617***	GFA	3.271**	3.468*
R^2	0.680	0.748	R^2	0.707	0.709	R^2	0.576	0.916	R^2	0.738	0.732
MAE	24.361	22.938	MAE	20.313	20.173	MAE	23.224	20.574	MAE	18.355	17.773
ISIC 17-18: Paper, paper products and printing (n=4)						**ISIC 17-18: Paper, paper products and printing (n=3)**					
Term	OLS	RR	Term	OLS	RR	Term	OLS	RR	Term	OLS	RR
NE	2.815**	2.802***	GFA	20.785*	20.714**	NE	1.326*	1.311	GFA	6.021	6.070
R^2	0.978	0.975	R^2	0.906	0.893	R^2	0.904	0.913	R^2	0.781	0.754
MAE	4.120	4.091	MAE	7.975	7.960	MAE	6.780	6.579	MAE	12.766	12.681
ISIC 22: Rubber and plastic products (n=3)						**ISIC 22: Rubber and plastic products (n=3)**					
Term	OLS	RR	Term	OLS	RR	Term	OLS	RR	Term	OLS	RR
NE	0.493#	0.492#	GFA	3.209#	3.198*	NE	1.142**	1.142**	GFA	3.189*	3.185**
R^2	0.839	0.812	R^2	0.871	0.855	R^2	0.993	0.992	R^2	0.971	0.966
MAE	2.768	2.768	MAE	2.047	2.054	MAE	2.463	2.462	MAE	4.914	4.900
ISIC 24-25: Basic metal, alloy and metal products (n=14)						**ISIC 24-25: Basic metal, alloy and metal products (n=9)**					
Term	OLS	RR	Term	OLS	RR	Term	OLS	RR	Term	OLS	RR
NE	1.588***	1.587***	GFA	4.956***	4.902***	NE	1.189**	1.237*	GFA	3.337**	3.363**
R^2	0.935	0.931	R^2	0.800	0.773	R^2	0.666	0.659	R^2	0.689	0.656
MAE	11.188	11.188	MAE	21.486	21.453	MAE	16.359	16.256	MAE	17.244	17.236
Jaipur						**Hyderabad**					
ISIC 10: Food products (n=20)						**ISIC 10: Food products (n=11)**					
Term	OLS	RR	Term	OLS	RR	Term	OLS	RR	Term	OLS	RR
NE	0.655***	0.506***	GFA	2.780**	1.760**	NE	0.614***	0.993***	GFA	3.123**	2.946**
R^2	0.511	0.912	R^2	0.345	0.613	R^2	0.723	0.999	R^2	0.536	0.503
MAE	11.345	11.329	MAE	14.016	13.972	MAE	10.33	10.192	MAE	14.280	13.966
ISIC 16: Wood and wood products (n=7)						**ISIC 16: Wood and wood products (n=11)**					
Term	OLS	RR	Term	OLS	RR	Term	OLS	RR	Term	OLS	RR
NE	0.365#	0.328**	GFA	1.377*	1.337**	NE	0.612#	0.767***	GFA	6.387#	8.315***
R^2	0.413	0.555	R^2	0.592	0.718	R^2	0.534	0.999	R^2	0.559	0.999
MAE	8.849	8.595	MAE	6.705	6.637	MAE	21.270	18.445	MAE	21.839	18.513
ISIC 17-18: Paper, paper products and printing (n=7)						**ISIC 17-18: Paper, paper products and printing (n=15)**					
Term	OLS	RR	Term	OLS	RR	Term	OLS	RR	Term	OLS	RR
NE	1.400	0.654***	GFA	2.021***	2.013***	NE	0.371**	0.358***	GFA	2.896***	2.477
R^2	0.343	0.975	R^2	0.924	0.936	R^2	0.535	0.934	R^2	0.760	0.855
MAE	21.739	16.629	MAE	7.667	7.611	MAE	9.373	9.224	MAE	8.015	8.527
ISIC 22: Rubber and plastic products (n=7)						**ISIC 22: Rubber and plastic products (n=25)**					
Term	OLS	RR	Term	OLS	RR	Term	OLS	RR	Term	OLS	RR
NE	0.274	0.264#	GFA	2.250**	2.163**	NE	0.499***	0.297***	GFA	0.040	0.433*
R^2	0.374	0.335	R^2	0.777	0.76	R^2	0.413	0.724	R^2	0.200	0.397
MAE	18.807	18.772	MAE	12.316	11.992	MAE	14.148	13.136	MAE	17.446	26.434
ISIC 24-25: Basic metal, alloy and metal products (n=25)						**ISIC 24-25: Basic metal, alloy and metal products (n=118)**					
Term	OLS	RR	Term	OLS	RR	Term	OLS	RR	Term	OLS	RR
NE	0.691**	0.296***	GFA	3.494***	1.589***	NE	0.953***	0.665**	GFA	2.769***	4.823***
R^2	0.352	0.821	R^2	0.414	0.713	R^2	0.355	0.394	R^2	0.500	0.882
MAE	18.770	17.887	MAE	18.631	17.472	MAE	22.661	22.254	MAE	19.248	21.067

Note: (i) GFA represents the gross floor area of an establishment in 100 m^2, (ii) NE represents the number of employees working, (iii) R^2 is the coefficient of determination, (iv) MAE means mean absolute error, (v) ***p<0.001, **p<0.01, *p<0.05, #p<0.1, and no symbol specified for coefficient – statistically insignificant, (vi) All the above models are single variable models without intercept, (vii) ISIC stands for International Standard Industrial Classification, (viii) OLS, and RR stand for Ordinary Least Squares, and Robust Regression, respectively, (ix) n represents the number of observations used in modelling (both OLS and RR)

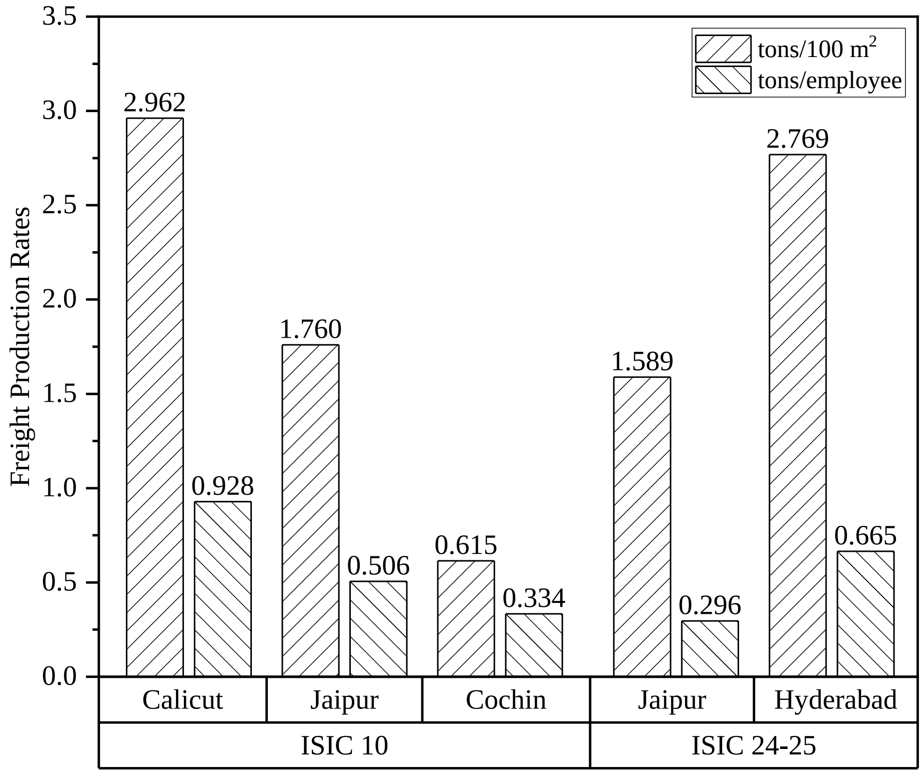


Figure 7.1 Comparison of FP rates across various industries

The MCA results are presented in Table 7.3 with interval widths of 15 employees and 400 m^2 of gross floor area. The methodology section elaborates on the rationale behind selecting these specific interval widths. All these models are statistically significant at 90%. The coefficient of determination (R^2) ranges from 0.236 to 0.750. In specific cases, it is observed that the FP rates increase with the increase in NE and GFA. But, for some cases, there is a non-linear pattern in FP rates. This non-linear pattern is observed in establishments with higher employment (NE ≥30). The possible reasons for this pattern are as follows – (i) MCA is a non-parametric modelling methodology, which does not presume any data distribution, and (ii) the smaller sample size in the case of some industrial classes. For some industrial classes, the MCA approach's predictive ability (lower MAE values) is better than the OLS and RR models.

Table 7.3 Summary of Multiple Classification Analysis (MCA)

Calicut (n=61)											
	Weekly Freight Production for Different Industrial Classes (in tons)						Weekly Freight Production for Different Industrial Classes (in tons)				
Employment Levels	ISIC 10	ISIC 16	ISIC 17-18	ISIC 22	ISIC 24-25	Area levels (in m^2)	ISIC 10	ISIC 16	ISIC 17-18	ISIC 22	ISIC 24-25
0-15	5.677	27.750	31.442	2.382	22.686	0-400	6.720	28.165	31.442	6.539	25.028
15-30	18.151	40.224	43.916	14.856	35.160	400-800	17.374	38.819	42.096	17.193	35.682
30-45	36.030	58.103	61.795	32.735	53.039	800-1200	9.410	30.855	34.132	9.229	27.718
≥45	80.438	102.511	106.203	77.143	97.447	≥1200	49.196	70.641	73.918	49.015	67.504
R^2 = 0.656, Adjusted R^2 = 0.610, MAE = 11.363						R^2 = 0.506, Adjusted R^2 = 0.461, MAE = 14.103					
Cochin (n=44)											
	Weekly Freight Production for Different Industrial Classes (in tons)						Weekly Freight Production for Different Industrial Classes (in tons)				
Employment Levels	ISIC 10	ISIC 16	ISIC 17-18	ISIC 22	ISIC 24-25	Area levels (in m^2)	ISIC 10	ISIC 16	ISIC 17-18	ISIC 22	ISIC 24-25
0-15	4.031	27.668	18.774	11.908	22.553	0-400	2.372	27.112	24.671	19.976	26.772
15-30	21.873	45.510	36.616	29.750	40.395	400-800	5.750	30.490	28.049	23.354	30.150
30-45	28.478	52.115	43.221	36.355	47.000	800-1200	10.202	34.943	32.502	27.807	34.602
≥45	6.650	30.287	21.393	14.527	25.172	≥1200	15.855	40.596	38.155	33.459	40.255
R^2 = 0.750, Adjusted R^2 = 0.702, MAE = 6.301						R^2 = 0.548, Adjusted R^2 = 0.481, MAE = 8.400					
Jaipur (n=66)											
	Weekly Freight Production for Different Industrial Classes (in tons)						Weekly Freight Production for Different Industrial Classes (in tons)				
Employment Levels	ISIC 10	ISIC 16	ISIC 17-18	ISIC 22	ISIC 24-25	Area levels (in m^2)	ISIC 10	ISIC 16	ISIC 17-18	ISIC 22	ISIC 24-25
0-15	16.041	9.093	23.275	23.460	20.369	0-400	13.108	3.971	12.377	0.988	13.953
15-30	22.486	15.538	29.720	29.905	26.814	400-800	42.622	33.485	41.891	30.502	43.467
30-45	14.666	7.718	21.900	22.085	18.994	800-1200	26.306	17.168	25.574	14.186	27.150
≥45	34.302	27.354	41.536	41.721	38.630	≥1200	35.443	26.306	34.712	23.323	36.287
R^2 = 0.288, Adjusted R^2 = 0.201, MAE = 16.956						R^2 = 0.236, Adjusted R^2 = 0.189, MAE = 15.316					
Hyderabad (n=180)											
	Weekly Freight Production for Different Industrial Classes (in tons)						Weekly Freight Production for Different Industrial Classes (in tons)				
Employment Levels	ISIC 10	ISIC 16	ISIC 17-18	ISIC 22	ISIC 24-25	Area levels (in m^2)	ISIC 10	ISIC 16	ISIC 17-18	ISIC 22	ISIC 24-25
0-15	10.959	22.846	9.916	9.937	21.338	0-400	14.478	33.262	11.718	7.886	18.221
15-30	17.258	29.145	16.215	16.236	27.637	400-800	17.606	36.390	14.846	11.014	21.349
30-45	28.461	40.348	27.418	27.439	38.840	800-1200	25.937	44.721	23.177	19.345	29.680
≥45	22.626	34.513	21.583	21.604	33.005	≥1200	52.158	70.942	49.398	45.566	55.901
R^2 = 0.266, Adjusted R^2 = 0.227, MAE = 21.031						R^2 = 0.240, Adjusted R^2 = 0.208, MAE = 18.215					

Note: (i) ISIC stands for International Standard Industrial Classification, (ii) n represents the number of observations used in modelling, (iii) All models are statistically significant at 90%, (iv) R^2 means coefficient of determination, (v) MAE means mean absolute error

7.3.3 Transferability Assessment of Freight Production Models

The models were transferred across different cities and assessed for transferability using the metrics MAE and relative MAE. From the transferability assessment results, the best transferable modelling methodologies for each industrial class are represented in Table 7.4. This table presents the best transferable modelling methodology for different cases (different combinations of application context and estimation context) based on relative MAE values. For example, let us transfer Cochin's employment-based FP model in ISIC 10 to Calicut. In this case, the relative MAE of OLS is lesser compared to that of other modelling methodologies. Therefore, the OLS model may be best transferable in this case. On a closer look at the best methods (in terms of spatial transferability) in the table, it is observed that OLS is preferred in the case of ISIC 10: Food and food products. Also, the preferable modelling methodology for ISIC 24-25: Basic metal, alloy and metal products is OLS. For all the other industrial classes, ISIC 16: Wood and wood products, ISIC 17-18: Paper, paper products and printing and ISIC 22: Rubber and plastic products, MCA is preferred. If we look at the details of the sample sizes, it is noticed that there are many cases in which the sample size of the estimation context is smaller than the application context. Under such circumstances, it is advised to exercise caution when selecting the methodology. Subsequently, a more comprehensive investigation is carried out to gain a deeper understanding of the effect of sample size on transferability.

7.3.4 Examination of Sample Size Effect

We have initially summarised the best transfer cases to understand sample size's effect on transferability. Later, we assessed the transferability by varying the sample size and recommended certain transfer cases. Table 7.5 summarises the transfer cases with the best estimation contexts and transferable modelling methodology. For example, let us transfer Hyderabad's employment-based model in ISIC 10 to Calicut. Here, as estimation context, when Hyderabad model is transferred to Calicut, it gives relatively more minor values of relative MAE than other cities (Cochin and Jaipur). Also, the best transferable model among different modelling methodologies is RR. Further discussion on the table is as follows.

Employment-based FP models: On a closer look at the summary of employment-based FP models in Table 7.5, it is clear that for ISIC 10: Food and food products, Hyderabad FP model gives better results on transfer. Also, it is to be noted that RR is the preferred modelling methodology for transfer in this industrial class. Considering ISIC 16: Wood and wood products, Hyderabad models are better transferable; MCA is the preferred methodology for transfer. Considering ISIC 17: Paper and paper products, models developed for Jaipur are more transferable, and MCA is the methodology of choice for transfer. Hyderabad FP models produce superior results when it comes to the transfer of models in ISIC 22: Rubber and plastic products. It is important to note that OLS and MCA are the most accurate methodologies for transfer in this particular industrial class. Similar observations are noticed in ISIC 24-25: Basic metal, alloy and metal products. Hyderabad models are better transferable in this industrial class, and MCA is preferred for transfer.

Area-based FP models: From Table 7.5, considering ISIC 10: Food and food products, models developed in Jaipur are better transferable; both OLS and RR are preferred for transfer. Considering ISIC 16: Wood and wood products, it is noticed that the transferability of Cochin models is more compared to other city models. It is to be noted that MCA is preferred for transfer in this industrial class. In ISIC 17-18: Paper and paper products, MCA is the preferred modelling methodology for transfer. Considering ISIC 22: Rubber and plastic products, it is clear that Cochin city model is the best transferable; RR is the best choice for transfer. For ISIC 24-25: Basic metal, alloy and metal products, Cochin models are better transferable; OLS and MCA are preferred for transfer.

From the above discussion about the transferability of employment-based and area-based models, it is noticed that Hyderabad employment-based FP models are better transferable as per count, and the preferred methodology for transfer is MCA. In the case of area-based FP models, Cochin models are transferable as per the count, and MCA is preferred. However, in all these cases, the sample size is not considered. Considering the sample size, it is to be noted that certain transfer cases in Table 7.5 are highlighted. These highlighted cases are cases where the sample size of the estimation context is more than that of the application context. Considering only the highlighted cases, Hyderabad is a better estimation context for employment-based FP models. Based on count, 8 out of 10 cases indicate that Hyderabad models are better transferable. 6 out of 10 cases show that MCA is the best modelling method in terms of spatial transferability. In the highlighted cases of area-based FP models, Calicut is the best estimation context (4 out of 7 cases), and MCA is preferred for transfer. From the discussion (all the transfer cases, which includes highlighted transfer cases in Table 7.5), Hyderabad employment-based models are better for spatial transferability. MCA is the better transferable modelling methodology for employment-based models. Calicut area-based models are better transferable than other city models; however, no specific modelling methodology is preferred for transfer.

The key findings from Table 7.4 and Table 7.5 are as follows – (i) OLS and MCA are better transferable compared to others on not considering the sample size, (ii) MCA is better transferable when the sample size is considered, and (iii) Hyderabad employment-based and Calicut area-based FP models are better transferable when the sample size is considered. In the past, some studies investigated on importance of sample size for transferability assessment. Sikder et al. (2014) invested in the transferability of choice models, and it was found that the degree of transferability increased with the sample size. Karasmaa (2007) evaluated the transferability of travel demand models and found that models developed using larger sample sizes yield more dependable outcomes in terms of transferability. Based on the literature, it can be concluded that (i) MCA is the preferred modelling methodology for transferability of both employment-based and area-based FP models, (ii) Hyderabad as estimation context is better in transferability of employment-based FP models, and (iii) Calicut as estimation context is better in transferability of area-based FP models. Also, one of the findings from transferability assessment results is that the transferability is asymmetric, i.e., the transfer direction may be from one region to another but not vice versa. Similar findings were found in some of the transferability studies. Sikder et al. (2014) developed choice models for Alameda City and Sonoma City. These models were assessed for transferability, and it was found that Alameda's model was transferable to Sonoma but not vice versa. A similar observation was seen in the study by Pani et al. (2021), where FP models were assessed for transferability. Also, Balla et al. (2021) proved that the transferability of FP models is asymmetric. The models were transferable from urban to suburban but not vice versa.

Table 7.4 Summary of best transferable modelling approaches

Estimation Context	Application Context	Best Transferable Modelling Methodology: Employment-based FP Model	Best Transferable Modelling Methodology: Area-based FP Model
ISIC 10: Food and food products			
Cochin (n=21)	Calicut (n=25)	OLS	MCA
Jaipur (n=20)	Calicut (n=25)	OLS	OLS
Hyderabad (n=11)	Calicut (n=25)	RR	RR
Calicut (n=25)	Cochin (n=21)	OLS	MCA
Jaipur (n=20)	Cochin (n=21)	RR	RR
Hyderabad (n=11)	Cochin (n=21)	OLS	MCA
Calicut (n=25)	Jaipur (n=20)	OLS	RR
Cochin (n=21)	Jaipur (n=20)	OLS	OLS
Hyderabad (n=11)	Jaipur (n=20)	OLS	MCA
Calicut (n=25)	Hyderabad (n=11)	OLS	OLS
Cochin (n=21)	Hyderabad (n=11)	OLS	OLS
Jaipur (n=20)	Hyderabad (n=11)	OLS	OLS
Most repeated modelling methodology		OLS (83.3%)	OLS (41.7%)
ISIC 16: Wood and wood products			
Cochin (n=8)	Calicut (n=15)	RR	RR
Jaipur (n=7)	Calicut (n=15)	OLS	MCA
Hyderabad (n=11)	Calicut (n=15)	OLS	MCA
Calicut (n=15)	Cochin (n=8)	OLS	RR
Jaipur (n=7)	Cochin (n=8)	OLS	OLS
Hyderabad (n=11)	Cochin (n=8)	OLS	OLS
Calicut (n=15)	Jaipur (n=7)	RR	MCA
Cochin (n=8)	Jaipur (n=7)	MCA	MCA
Hyderabad (n=11)	Jaipur (n=7)	MCA	MCA
Calicut (n=15)	Hyderabad (n=11)	MCA	OLS
Cochin (n=8)	Hyderabad (n=11)	MCA	MCA
Jaipur (n=7)	Hyderabad (n=11)	MCA	MCA
Most repeated modelling methodology		MCA, OLS (41.7% each)	MCA (58.3%)
ISIC 17-18: Paper, paper products and printing			
Cochin (n=3)	Calicut (n=4)	MCA	MCA
Jaipur (n=7)	Calicut (n=4)	MCA	MCA
Hyderabad (n=15)	Calicut (n=4)	MCA	MCA
Calicut (n=4)	Cochin (n=3)	MCA	MCA
Jaipur (n=7)	Cochin (n=3)	OLS	OLS
Hyderabad (n=15)	Cochin (n=3)	MCA	MCA
Calicut (n=4)	Jaipur (n=7)	MCA	MCA
Cochin (n=3)	Jaipur (n=7)	OLS	MCA
Hyderabad (n=15)	Jaipur (n=7)	OLS	MCA
Calicut (n=4)	Hyderabad (n=15)	MCA	MCA
Cochin (n=3)	Hyderabad (n=15)	MCA	MCA
Jaipur (n=7)	Hyderabad (n=15)	MCA	RR
Most repeated modelling methodology		MCA (75.0%)	MCA (83.3%)
ISIC 22: Rubber and plastic products			
Cochin (n=3)	Calicut (n=3)	MCA	RR
Jaipur (n=7)	Calicut (n=3)	OLS	OLS
Hyderabad (n=25)	Calicut (n=3)	OLS	OLS
Calicut (n=3)	Cochin (n=3)	MCA	RR
Jaipur (n=7)	Cochin (n=3)	MCA	OLS
Hyderabad (n=25)	Cochin (n=3)	MCA	MCA
Calicut (n=3)	Jaipur (n=7)	OLS	OLS
Cochin (n=3)	Jaipur (n=7)	MCA	MCA
Hyderabad (n=25)	Jaipur (n=7)	RR	MCA
Calicut (n=3)	Hyderabad (n=25)	OLS	MCA
Cochin (n=3)	Hyderabad (n=25)	MCA	MCA
Jaipur (n=7)	Hyderabad (n=25)	OLS	MCA
Most repeated modelling methodology		MCA (50.0%)	MCA (50.0%)
ISIC 24-25: Basic metal, alloy and metal products			
Cochin (n=9)	Calicut (n=14)	RR	RR
Jaipur (n=25)	Calicut (n=14)	MCA	OLS
Hyderabad (n=118)	Calicut (n=14)	MCA	OLS
Calicut (n=14)	Cochin (n=9)	OLS	RR
Jaipur (n=25)	Cochin (n=9)	OLS	OLS
Hyderabad (n=118)	Cochin (n=9)	OLS	OLS
Calicut (n=14)	Jaipur (n=25)	MCA	OLS
Cochin (n=9)	Jaipur (n=25)	MCA	OLS
Hyderabad (n=118)	Jaipur (n=25)	RR	OLS
Calicut (n=14)	Hyderabad (n=118)	OLS	RR
Cochin (n=9)	Hyderabad (n=118)	OLS	RR
Jaipur (n=25)	Hyderabad (n=118)	OLS	RR
Most repeated modelling methodology		OLS (50.0%)	OLS (58.3%)

Note: (i) OLS, RR, and MCA stand for Ordinary Least Squares Regression, Robust Regression, and Multiple Classification Analysis, respectively, (ii) In most repeated modelling methodology, the percentage in parenthesis represents the frequency of repetition of that particular methodology in the industrial class), (iii) FP means freight production.

Table 7.5 Summary of best transfer cases

Employment-based FP Models

Estimation Context	*Application Context*	*Best Transferable Modelling Methodology*	*MAE*	*Relative MAE*
ISIC 10: Food and food products				
Hyderabad (n=11)	Calicut (n=25)	RR	6.151	1.062
Jaipur (n=20)	Cochin (n=21)	RR	3.992	1.028
Hyderabad (n=11)	Jaipur (n=20)	OLS	11.118	0.980
Calicut (n=25)	Hyderabad (n=11)	RR	10.216	1.002
ISIC 16: Wood and wood products				
Cochin (n=8)	Calicut (n=15)	MCA	15.32	1.348
Hyderabad (n=11)	Cochin (n=8)	MCA	10.231	1.624
Hyderabad (n= 11)	Jaipur (n=7)	OLS	10.797	1.220
Jaipur (n=7)	Hyderabad (n=11)	MCA	22.408	1.065
ISIC 17-18: Paper, paper products, and printing				
Cochin (n=3)	Calicut (n=4)	MCA	15.32	1.348
Jaipur (n=7)	Cochin (n=3)	OLS	7.763	1.145
Hyderabad (n=15)	Jaipur (n=7)	MCA	17.995	1.061
Jaipur (n=7)	Hyderabad (n=15)	MCA	16.955	1.838
ISIC 22: Rubber and plastic products				
Hyderabad (n=25)	Calicut (n=3)	OLS	2.773	1.002
Hyderabad (n=25)	Cochin (n=3)	MCA	10.231	1.624
Hyderabad (n=25)	Jaipur (n=7)	MCA	17.995	1.061
Jaipur (n=7)	Hyderabad (n=25)	OLS	13.323	0.942
ISIC 24-25: Basic metal, alloy and metal products				
Cochin (n=9)	Calicut (n=14)	RR	14.846	1.327
Hyderabad (n=118)	Cochin (n=9)	MCA	10.231	1.624
Hyderabad (n=118)	Jaipur (n=25)	MCA	17.995	1.061
Jaipur (n=25)	Hyderabad (n=118)	OLS	22.219	0.981

Area-based FP Models

Estimation Context	*Application Context*	*Best Transferable Modelling Methodology*	*MAE*	*Relative MAE*
ISIC 10: Food and food products				
Jaipur (n=20)	Calicut (n=25)	OLS	8.414	0.991
Jaipur (n=20)	Cochin (n=21)	RR	8.115	1.784
Calicut (n=25)	Jaipur (n=20)	RR	13.951	0.998
Jaipur (n=20)	Hyderabad (n=11)	OLS	13.674	0.958
ISIC 16: Wood and wood products				
Cochin (n=8)	Calicut (n=15)	MCA	16.149	1.145
Calicut (n=15)	Cochin (n=8)	MCA	12.288	1.463
Cochin (n=8)	Jaipur (n=7)	OLS	15.291	2.281
Jaipur (n=7)	Hyderabad (n=11)	MCA	20.446	1.122
ISIC 17-18: Paper, paper products, and printing				
Cochin (n=3)	Calicut (n=4)	MCA	16.149	1.145
Calicut (n=4)	Cochin (n=3)	MCA	12.288	1.463
Hyderabad (n=15)	Jaipur (n=7)	RR	10.669	1.402
Jaipur (n=7)	Hyderabad (n=15)	OLS	9.29	1.159
ISIC 22: Rubber and plastic products				
Cochin (n=3)	Calicut (n=3)	OLS	2.059	1.006
Calicut (n=3)	Cochin (n=3)	RR	4.941	1.008
Cochin (n=3)	Jaipur (n=7)	RR	16.222	1.353
Jaipur (n=7)	Hyderabad (n=25)	MCA	20.446	1.122
ISIC 24-25: Basic metal, alloy and metal products				
Cochin (n=9)	Calicut (n=14)	MCA	16.149	1.145
Calicut (n=14)	Cochin (n=9)	MCA	12.288	1.463
Cochin (n=9)	Jaipur (n=25)	OLS	18.292	0.982
Cochin (n=9)	Hyderabad (n=118)	OLS	19.651	1.021

Note: (i) In highlighted transfer cases, the sample size of the estimation context is larger than the application context, (ii) n refers to sample size, (iii) MAE stands for mean absolute error, (iv) OLS, RR, and MCA stand for Ordinary Least Squares Regression, Robust Regression, and Multiple Classification Analysis, respectively

Further investigation was done to recommend transfer cases from Table 7.5. This investigation was done by considering different sizes (one-third, two-thirds, and total) of the local data (application context data). The MAE values were calculated for different proportions of sample sizes, and these varied samples were resampled using bootstrapping. Bootstrapping is preferred because it is a robust method that can handle small sample sizes without making assumptions. After computing MAE values through bootstrapping, the coefficient of variance was calculated for these values to examine the error dispersion when transferring the estimation context model across different application context data variations.

The transfer cases from Table 7.5 are recommended based on two criteria. First, the sample size of the estimation context is greater than 20. The single variable models with a minimum sample size 20 give relatively reliable results. Second, the value of the relative MAE is closer to one. If the relative MAE value is closer to 1, the performance of estimation and application contexts' models are almost similar. Table 7.6 presents recommended cases based on these criteria. A closer look at the table shows that Hyderabad is better as an estimation context for transferring the employment-based FP model. Also, MCA is better transferable in some of these recommended cases. The coefficient of variance values ranges from 0.0004 to 0.0016. This range indicates that the error dispersion of transferability is negligible with the sample size variation.

Table 7.6 Transferability assessment of recommended cases

Estimation Context	Application Context	Industrial Class	Best Transferable Modelling Methodology	MAE Values for Different Proportions of Application Context's Sample Size			Coefficient of Variance	Relative MAE
				1/3rd	2/3rd	Total		
Employment-based Freight Production Models								
Calicut (n=25)	Hyderabad (n=11)	ISIC 10	RR	10.182	10.206	10.176	0.0016	1.002
Hyderabad (n=25)	Calicut (n=3)	ISIC 22	OLS	2.773	2.775	2.774	0.0004	1.002
Hyderabad (n=25)	Jaipur (n=7)	ISIC 22	MCA	18.033	18.007	17.997	0.0010	1.061
Hyderabad (n=118)	Jaipur (n=25)	ISIC 24-25	MCA	18.033	18.007	17.997	0.0010	1.061
Area-based Freight Production Models								
Calicut (n=25)	Jaipur (n=20)	ISIC 10	RR	13.945	13.969	13.955	0.0009	0.998
Jaipur (n=20)	Hyderabad (n=11)	ISIC 10	OLS	14.081	14.088	14.067	0.0008	0.958

Note: OLS, RR, and MCA stand for Ordinary Least Squares Regression, Robust Regression, and Multiple Classification Analysis, respectively

7.4 Summary

In this study, the degree of spatial transferability of freight production (FP) models produced using a variety of modelling methodologies was investigated, and it was determined which models are more transferable. In addition, the effect of transferability due to sample size was examined to get reliable transferability results. Using an establishment-based freight survey (EBFS), the data needed to develop these FP models were gathered from establishments in various cities in India, including Calicut, Cochin, Jaipur, and Hyderabad. A set of single variable FP models was developed for each city under investigation. Explanatory variables in these models included the gross floor area of the establishment and the total number of employees working there. The ordinary least squares (OLS) regression and robust regression (RR), and multiple classification analysis (MCA), were utilised in the development of these FP models.

The FP models built for one region (estimation context) were used in another region (application context), and the idea of this transfer is referred to as spatial transferability. Metrics such as mean absolute error (MAE) and relative mean absolute error (relative MAE) were utilised to evaluate the FP models in terms of their potential for spatial transferability. It is important to note that OLS and MCA are the modelling approaches recommended for spatial transferability. On the other hand, considering the effect of sample size on the results, the transferability among the study cities in India has produced some intriguing implications. MCA is the better approach for spatial transferability. The transfer error of the Hyderabad employment-based FP model is negligible to many other study cities across most industrial classes. In the case of area-based FP models, Calicut is better as an estimation context. Further investigation on transferability considering the sample size variation, MCA is still the better method for transferability. The variation in the sample size has not influenced the extent of transferability of MCA models.

The freight demand models (which include FP models) are developed using freight data, which is very challenging to collect through freight surveys like EBFS. This survey requires a large number of resources in terms of human resources, money and time. Collecting data in the regions or cities that cannot invest in the survey may be impossible. In such regions or cities, we can refer recommended transfer cases of this study. With this approach, regions or cities lacking survey resources can have solid freight demand estimates, which will help plan the freight infrastructure, including terminals, warehouses, and special corridors for freight movement. In addition, freight demands can be helpful for policymakers in easy governance. With the recommended transfer cases from the study, we can implement the preexisting policies in geographically similar areas. In addition, industrialists can borrow or transfer the logistic strategies to regions or cities with similar demographics.

This study acknowledges certain limitations. One of the assumptions made in this study is the temporal stability of the freight data. However, it is essential to recognize that changes in the establishments' economic activities can occur during the survey period and lead to minor errors in freight demand estimates. In addition, this study assumes that the models can be directly transferred and recommends transferable modelling methodologies. However, updating the parameters of the models using different transfer methods can enhance the transferability and discover better transferable modelling methodologies. Also, the study assumes model transfer based solely on statistical evidence without considering geographical demographics associated with freight activities. Apart from the acknowledged limitations, there are several potential future directions to expand the scope of this study. The transferability between Calicut and Cochin has been more significant in certain transfer cases. This extent of transferability may be due to the following reason: there is a possibility that there are similarities in the patterns of freight data between these two cities because they are both located in the same Indian state - Kerala. Additional research and analysis are required to understand better why the models established in specific locations are more transferable to some regions. There is a possibility that one of the reasons for increased transferability is due to the geographical similarities between these places. Compiling a more comprehensive data set from other geographic areas and urban centres is necessary to investigate these parallels. In addition, it has been observed that MCA models are more easily transferable when directly transferred (naïve transfer). It is necessary to do additional research to get a better understanding of the reasons why this modelling methodology has good transferability. In addition, the transfer method used in the study is naïve transfer. The other transfer methods are not discussed since these methods for the parametric models (OLS and RR) and the non-parametric model (MCA) are different. Also, we used only the application-based approach of transferability in the study. In this approach, the models are developed for one specific region and later transferred to another. The other transferability approach is estimation-based, which is not used in this study. In this approach, the data is pooled and assessed for transferability based on the statistical significance of the difference parameters. This transferability approach can give us more details on the transferability of models. One crucial question that can broaden the scope of the study is whether these findings can be generalised to cities or regions beyond the ones examined in the study. Identifying geographically similar regions with similar economic indicators or activities is essential to address this question.

Part IV: Measurement

Chapter 8: Framework for Reducing Freight Survey Resources Using Spatial Transferability

8.1 General

Prior research has examined into the assessment of the direction and extent of transferability in freight data. However, these findings have often remained confined to their original contexts and have not been extrapolated to newer regions or areas lacking comprehensive freight data. This critical knowledge gap presents significant challenges when it comes to estimating freight demand in developing nations. In these nations, residents and government entities are increasingly concerned about the potential adverse impacts stemming from freight establishments. Therefore, it becomes imperative to gain a deep understanding of the transferability of freight data to these novel settings. Such knowledge can prove invaluable for policymakers and urban planners, as it empowers them to proactively address and mitigate the externalities associated with freight operations, plan ecologically sustainable infrastructure to support long-term growth, and effectively manage the movement of goods within these regions (Mohapatra et al., 2021; Pani et al., 2018). Another noteworthy shortcoming in prior research is the lack of exploration regarding how transferability findings can be applied to estimate the optimal sample size when faced with limited survey resources. By addressing this research gap, we can equip investors, planners, and researchers with valuable insights, enabling them to streamline processes, reduce costs, and save time when conducting freight surveys. This, in turn, enhances the efficiency and effectiveness of data collection efforts in the realm of freight transportation research.

This study aims to (i) investigate the direction and extent of transferability using various transfer methods, (ii) identify geographically similar regions based on different geographical characteristics, and (iii) determine the sample size requirements for geographically similar regions using an approach guided by transferability findings. By addressing these objectives, this study aims to contribute to the comprehension of transferability in freight demand estimation and provide practical guidance for cost-effective survey planning and resource allocation.

8.2 Methodology

This study conducted a freight survey, Establishment-based Freight Survey (EBFS), in North Kerala (Kozhikode, Kannur, Malappuram, Kochi), Central Kerala (Kottayam, Palakkad, Thrissur), Jaipur, and Hyderabad using face-to-face interviews. Primary data includes the establishment's gross floor area (GFA) and number of employees (NE). The reliability of the employment data was verified using Commercial Tax Departments and local websites. Additionally, Google Earth's measurement tools verify the building's gross floor. Table 8.1 shows EBFS freight data descriptive statistics. In addition to freight data, publicly available data on population density (people/sq. km), number of establishments, land value (INR/sq. ft), road density (km/sq. km), and seaport proximity (km) is used to compare Indian regions. All data sources are shippers. Shippers are classified as “C – Manufacturing” by the International Standard Industrial Classification.

Table 8.1 Summary of primary data used in modelling

	Number of Employees (NE)	Gross Floor Area (GFA) in m^2	Weekly Freight Production (FP) in tons	Weekly Freight Attraction (FA) in tons
		Hyderabad region (HYD), n=349		
Minimum	2.00	72.46	0.13	0.10
Maximum	134.00	14864.49	155.84	155.84
Average	23.82	925.61	20.60	17.69
S.D.	23.13	1683.36	28.45	25.34
		Jaipur region (JAI), n=162		
Minimum	3.00	53.00	1.20	0.20
Maximum	250.00	4800.00	264.00	140.40
Average	27.69	645.60	25.91	15.94
S.D.	35.77	680.45	31.75	25.95
		North Kerala region (NK), n=202		
Minimum	3.00	53.00	0.58	0.73
Maximum	180.00	4599.40	136.54	104.20
Average	21.35	644.80	20.95	14.81
S.D.	22.44	607.00	23.38	14.00
		Central Kerala region (CK), n=230		
Minimum	1.00	65.32	0.96	1.15
Maximum	200.00	4333.50	173.08	67.31
Average	32.90	886.86	21.69	14.45
S.D.	33.81	736.25	22.48	11.26

Note: S.D. stands for the standard deviation

The entire methodology is divided into three parts – (i) model estimation, (ii) model application, and (iii) sample size determination. The model estimation included the estimation of FG models using the RR method. The developed models were applied (model application) to other regions, and this application is called spatial transferability. The transferability assessment was done using different transfer methods. After this assessment, the geographically similar regions to the study regions were clustered using the K-means clustering algorithm (KMCA) and multidimensionality scaling (MDS). The sample size was determined by the transferability assessment results and information on geographically similar regions. This discussion on the methodology is presented in Figure 8.1 Framework to determine required sample size for new regions. More details about the framework are discussed in the following subsections.

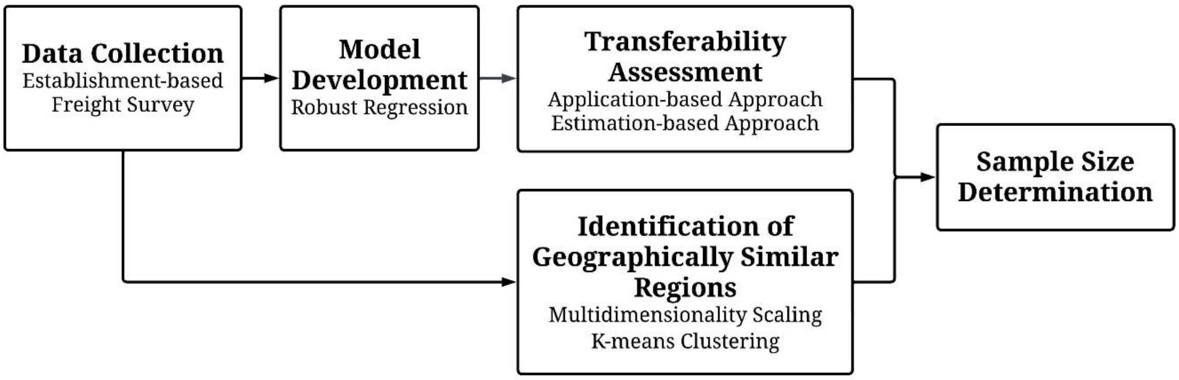


Figure 8.1 Framework to determine required sample size for new regions

8.2.1 Model Estimation

Single variable FG models and multivariable FG models were estimated using the RR method. The structure of the model is as follows.

(i) Single variable FP models

$$Area - based\ FP\ Model:\ FP_i = \beta_0 + \beta_1 \times GFA_i + \varepsilon_i \qquad 8.1$$

$$Employment - based\ FP\ Model:\ FP_i = \beta_0 + \beta_2 \times NE_i + \varepsilon_i \qquad 8.2$$

(ii) Single variable FA models

$$Area - based\ FA\ Model:\ FA_i = \beta_0 + \beta_1 \times GFA_i + \varepsilon_i \qquad 8.3$$

$$Employment - based\ FA\ Model:\ FA_i = \beta_0 + \beta_2 \times NE_i + \varepsilon_i \qquad 8.4$$

(iii) Multivariable models

$$FP\ Model:\ FP_i = \beta_0 + \beta_1 \times GFA_i + \beta_2 \times NE_i + \varepsilon_i \qquad 8.5$$

$$FA\ Model:\ FA_i = \beta_0 + \beta_1 \times GFA_i + \beta_2 \times NE_i + \varepsilon_i \qquad 8.6$$

Where FP_i is the weekly freight produced by the i^{th} establishment (in tons); FA_i is the weekly freight attracted by the i^{th} establishment (in tons); β_0 is the intercept; β_1, β_2 are the coefficients of the explanatory variables GFA and NE, respectively; ε_i is the stochastic error term assumed to be non-normally distributed in RR.

In this study, the RR method is employed instead of OLS due to the presence of heteroscedasticity and influential observations in the data. Additionally, RR is preferred because it does not assume a normal error distribution, which is essential given the observed data characteristics. Heteroscedasticity in the study data was identified through scatter plots and statistical tests like the Breusch-Pagan test.

8.2.2 Model Application

In previous studies, a wide variety of methods and metrics were utilised in order to conduct an analysis of the transferability of model parameters. In this study, an application-based approach as well as an estimation-based approach were utilised to evaluate the developed model's potential for transferability to a variety of study regions. Application-based methods and estimation-based methods were the two broad categories into which Bowman et al. (2014) divided the transferability assessment approaches. Using data from a single region as the estimation context, the application-based approach involves estimating model parameters, which are then applied to data from a different region (application context). Researchers can learn more about the model's general transferability by assessing the model's predictive power in this different region. This method, however, does not specify which specific parameters can be transferred.

The joint-context estimation method, also known as the estimation-based method (Karasmaa, 2007), is more thorough. To estimate a single model that takes into account potential variations between the two contexts, it pools data of estimation and application contexts. By estimating difference parameters that reflect the variations between the contexts, this model estimate is obtained. We can get a sense of whether the parameter estimates differ significantly between the estimation and application contexts by performing simple t-tests on these difference parameters. Notably, this methodology tests the transferability of each model parameter to provide a thorough understanding of the extent of transferability (Bowman et al., 2014).

Application-based Approach

The transferability assessment used mean absolute error (MAE) and relative MAE. The relative MAE measures the relative error of the transferred model (estimation context model) to the application context model, where both models are applied in application context data. The formulae for these metrics are as follows.

$$Mean\ absolute\ error\ (MAE) = \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{n} \qquad 8.7$$

$$Relative\ mean\ absolute\ error\ (Relative\ MAE) = \frac{MAE_{est}}{MAE_{app}} \qquad 8.8$$

Where $\hat{y}_i$ is the predicted value when the estimation context model is operated in the application context for the i^{th} establishment, y_i is the value of the response variable for i^{th} establishment in the local data, and n is the number of observations in the local data, MAE_{est} represents the MAE value of the estimation context model when transferred to the application context, and MAE_{app} represents the MAE value of the application context model.

The assessment was done using naïve transfer and model update transfer. Naïve transfer is the transfer method in which the model is directly transferred, whereas, in updated model transfer, the model's intercept is updated using the following equation.

$$\beta_0' = \bar{Y}_a - \sum_{i=1}^{n} \beta_{e_i} \bar{X}_{a_i} \qquad 8.9$$

Where β_0' is the locally estimated constant/intercept in the estimation context model, $\bar{Y}_a$ is the mean of the observed values in the application context, $\bar{X}_a$ is the average explanatory variable in the application context, β_e is the coefficient of X (NE or GFA) in the estimation context model.

Estimation-based Approach

The models were built by pooling the data from the two different regions – estimation and application contexts, to determine if models developed for one region can be transferred to another. The binary variables representing each region were incorporated into the models to account for locational differences. The models were estimated using the fixed effects, including the binary variables and the interaction terms, including significant variables such as GFA and

NE. The difference parameters were estimated to account for the estimation and application contexts' differences. On these difference methods, a t-test was conducted to determine whether the models could be transferred. If this parameter is statistically insignificant, there is no distinction between the two regions.

8.2.3 Sample Size Determination

The results of the transferability assessment reveal the comparative performance of transferred models and transfer methods. However, this evaluation is limited to the study regions. Can we interpolate the transferability assessment results to the other Indian regions? An analysis was conducted to identify geographically comparable regions to the study regions to answer this research question. The geographical similarities were quantified using MDS. MDS aims to convert similarities between multiple variables into distances represented in low-dimensional ordination space (Cox & Cox, 2000). The distances between the study and new regions (regions other than study regions) in this ordination space represented their similarities. In effect, shorter distances in ordination space indicate a stronger relationship between the two observations. Cox and Cox (2000) provide additional details regarding the application of MDS. This study used secondary data collected from Indian cities with a population of one million or more to determine the geographical similarities between these cities/regions and the study regions. Population density (people/sq. km), number of establishments, land value (INR/sq. ft.), road density (km/sq. km), and seaport proximity (km) were considered as determining factors (INR stands for Indian Rupee). Using MDS, geographical regions comparable to the regions of the study were identified.

Following MDS-based quantification, the Indian regions were clustered to distinguish between similar regions. Utilising the K-means clustering algorithm (KMCA), the clustering was carried out. Everitt (1980) provides additional details regarding the application of KMCA. KMCA optimises the clustering of similar regions iteratively by reassigning each region to its nearest centroid. KMCA and MDS were used to differentiate highly similar regions, and this combination was employed in a recent study (Paea & Baird, 2018).

The sample size for new regions (which are geographically similar to study regions) is determined using the transferability results as follows – (i) Understanding the direction of transferability and the best transfer method: From transferability results, a decision on which estimation context model is more transferable to which application context can be made. Also, the comparison of naïve transfer and updated model transfer assessment results can give us a better transfer method. From the above inputs, we can know the transfer direction and the best method in that direction, (ii) Determination of sample size based on transfer method: If naïve transfer shows better transferability from estimation to application contexts, then there is no need for data collection in the new regions similar to the application context. If the updated model transfer is better, the sample size required for the new regions equals the sample size used to update the estimation context model.

8.3 Results and Discussion

The results are discussed briefly in this section. First, it includes model estimation (FG modelling). Second, it includes model application (application-based and estimation-based approaches). Third, it includes measurement (sample size determination)

8.3.1 FG Modelling

The summary of the FG models developed using RR is presented in Table 8.2. This summary shows that the correlation between the FP and the explanatory variables is positive; FA and the explanatory variables are positively correlated. Therefore, it indicates that FG (which includes FP and FA) is positively correlated with the gross floor area of the establishment and positively correlated with the number of employees working there.

8.3.2 Application-based Approach

All the single variable FG models are naïve transferred (transferred directly without updating the parameters) across various regions. The naïve transfer assessment summary is tabulated in Table 8.3. The relative MAE values represent how best an estimation context model is comparable to the application context model. On a closer look at the results, it is evident that in some naïve transfer cases, the relative MAE values are less than one, which indicates that in an application context, the prediction ability of the estimation context model is better than its own model (application context model/local model). However, there are some cases in which the estimation context models are not performing well upon transfer.

Table 8.2 Summary of FG models

Single Variable Freight Generation Models						
Area-based freight production models						
	n	Intercept	GFA		R^2	MAE
Hyderabad	349	2.531***	1.019***		0.798	14.501
Jaipur	162	6.364**	1.440***		0.398	16.462
North Kerala	202	5.505***	1.251***		0.278	12.742
Central Kerala	230	6.204*	1.108***		0.292	13.130
Employment-based freight production models						
	n	Intercept	NE		R^2	MAE
Hyderabad	349	-	0.338***		0.490	16.555
Jaipur	162	8.403**	0.197***		0.228	17.278
North Kerala	202	5.861***	0.244***		0.379	13.337
Central Kerala	230	-	0.452***		0.708	13.469
Area-based freight attraction models						
	n	Intercept	GFA		R^2	MAE
Hyderabad	348	2.967***	0.838***		0.648	12.701
Jaipur	152	-	1.083***		0.527	12.047
North Kerala	146	3.491*	1.378***		0.478	7.849
Central Kerala	162	4.624***	1.116***		0.560	5.898
Employment-based freight attraction models						
	n	Intercept	NE		R^2	MAE
Hyderabad	348	-	0.247***		0.346	14.794
Jaipur	152	-	0.327***		0.584	10.708
North Kerala	146	-	0.439***		0.607	9.534
Central Kerala	162	7.778***	0.170***		0.404	6.585
Multivariable Freight Generation Models						
Freight production models						
	n	Intercept	GFA	NE	R^2	MAE
North Kerala	202	4.144***	0.643*	0.149***	0.449	12.712
Central Kerala	230	3.805*	0.836***	0.142*	0.421	12.630
Freight attraction models						
	n	Intercept	GFA	NE	R^2	MAE
Jaipur	162	-	0.594**	0.176()	0.582	10.789
Central Kerala	230	4.190**	0.827***	0.073*	0.607	5.776

Note: (1) ***p<0.001, **p<0.01, *p<0.05, ()p<0.1. (2) '-' indicates that the intercept is zero. (3) some regional models are not statistically significant, so they are not represented in this table.

Table 8.3 Summary of naïve transfer of FG models

Relative Mean Absolute Error (MAE) Values

Area-based Freight Production Models		Application Context			
		HYD	JAI	NK	CK
Estimation Context	HYD	1.00	1.06	1.10	1.04
	JAI	1.14	1.00	**1.02**	0.99
	NK	1.07	**0.99**	1.00	**0.97**
	CK	**1.05**	1.02	1.05	1.00

Area-based Freight Attraction Models		Application Context			
		HYD	JAI	NK	CK
Estimation Context	HYD	1.00	0.92	1.05	1.12
	JAI	**1.02**	1.00	1.19	1.18
	NK	1.12	0.94	1.00	**1.02**
	CK	1.06	**0.91**	**1.00**	1.00

Employment-based Freight Production Models		Application Context			
		HYD	JAI	NK	CK
Estimation Context	HYD	1.00	1.06	1.10	1.04
	JAI	1.03	1.00	**1.02**	0.99
	NK	1.01	**0.99**	1.00	**0.97**
	CK	**1.00**	1.02	1.05	1.00

Employment-based Freight Attraction Models		Application Context			
		HYD	JAI	NK	CK
Estimation Context	HYD	1.00	1.07	1.08	1.27
	JAI	**0.98**	1.00	1.01	**1.21**
	NK	**0.98**	**0.95**	1.00	1.33
	CK	0.99	**0.95**	**0.89**	1.00

Mean Absolute Error (MAE) Values

Freight Production Models		Application Context			
		HYD	JAI	NK	CK
Estimation Context	HYD	na	na	na	na
	JAI	na	na	na	na
	NK	15.084	16.530	12.712	**12.650**
	CK	**14.862**	**16.348**	**12.546**	12.630

Freight Attraction Models		Application Context			
		HYD	JAI	NK	CK
Estimation Context	HYD	na	na	na	na
	JAI	**13.069**	10.789	8.898	**6.429**
	NK	na	na	na	na
	CK	13.521	**10.499**	**7.804**	5.776

Note: (i) na stands for not applicable. (ii) The bolded values represent the best values among all the other values in the application context. (iii) HYD, JAI, NK and CK represent Hyderabad, Jaipur, North Kerala, and Central Kerala.

Area-based FP models: In these models, it is seen that the error of transferability of models to Central Kerala is lesser compared to other regions — the relative MAE values in this application context range between 0.97 to 1.04. After Central Kerala, the FG models are more transferable to Jaipur, and the relative MAE values range from 0.99 to 1.06. The models are least transferable to Hyderabad.

Employment-based FP models: The models are more transferable to Central Kerala; relative MAE ranges from 0.97 to 1.04. Also, there is a slightly lower transferability to Hyderabad, and relative MAE ranges between 1.00 to 1.03. The least transferability is noticed in North Kerala.

Area-based FA models: The models have shown better transferability in Jaipur, and relative MAE values of this region are between 0.91 and 0.94. After Jaipur, the transferability is better in Hyderabad, and the relative MAE ranges from 1.02 to 1.12. The least transferability is noticed in both Central Kerala and North Kerala.

Employment-based FA models: The more transferability of these models is seen in North Kerala with a relative MAE between 0.89 and 1.08. Also, Hyderabad has shown better transferability with a relative MAE between 0.98 and 0.99. The transferability of these models is the least in Central Kerala.

Multivariable FG models: In the case of these multivariable FG models, the metric used is MAE only. The reason for using MAE and not relative MAE is that some of these models are not statistically significant. Due to the unavailability of statistically significant models for all regions, we could not make a relative comparison. The multivariable FP models are transferable more across North and Central Kerala regions. Also, the transferability across these regions is higher in multivariable FA models.

Transferability Assessment of Updated Models

The models were updated and assessed for transferability. The assessment results are presented in Table 8.4, and the discussion follows.

Area-based FP models: It is noted that the transferability of these models in Central Kerala as application context is better, and relative MAE values are from 1.08 to 1.10, followed by North Kerala, where the relative MAE values range from 1.12 to 1.16. The least transferability of these models is noticed in Jaipur, which has relative MAE values between 1.16 and 1.18.

Employment-based FP models: It is seen that the models are more transferable to Central Kerala with relative MAE in the range of 1.06 and 1.10. After this region, better transferability is noticed in Jaipur, which has relative MAE values from 1.11 to 1.12. As an application context, Hyderabad has shown the least transferability, with a relative MAE between 1.14 and 1.18.

Table 8.4 Summary of updated model transfer of FG models

Relative Mean Absolute Error (MAE) Values

Area-based Freight Production Models	Application Context				Area-based Freight Attraction Models	Application Context			
Estimation Context	HYD	JAI	NK	CK	Estimation Context	HYD	JAI	NK	CK
HYD	1.00	1.18	1.16	1.10	HYD	1.00	1.07	1.09	1.07
JAI	**1.16**	1.00	**1.12**	**1.08**	JAI	**1.13**	1.00	1.08	1.03
NK	**1.16**	**1.16**	1.00	1.09	NK	1.16	**1.04**	1.00	**1.02**
CK	**1.16**	1.17	1.15	1.00	CK	**1.13**	1.05	**1.07**	1.00

Employment-based Freight Production Models	Application Context				Employment-based Freight Attraction Models	Application Context			
Estimation Context	HYD	JAI	NK	CK	Estimation Context	HYD	JAI	NK	CK
HYD	1.00	1.12	1.15	**1.06**	HYD	1.00	1.05	**0.96**	**1.07**
JAI	1.18	1.00	1.18	1.10	JAI	1.10	1.00	0.97	1.14
NK	1.16	1.14	1.00	1.08	NK	**1.08**	**0.94**	1.00	1.32
CK	**1.14**	**1.11**	**1.14**	1.00	CK	1.12	1.12	**0.96**	1.00

Mean Absolute Error (MAE) Values

Freight Production Models	Application Context				Freight Attraction Models	Application Context			
Estimation Context	HYD	JAI	NK	CK	Estimation Context	HYD	JAI	NK	CK
HYD	na	na	na	na	HYD	na	na	na	na
JAI	na	na	na	na	JAI	14.701	**10.789**	8.721	**6.403**
NK	17.100	19.191	12.712	**14.179**	NK	na	na	na	na
CK	**16.779**	**19.000**	**14.668**	12.630	CK	**14.426**	12.123	**8.523**	5.776

Note: (i) na stands for not applicable. (ii) The bolded values represent the best values among all the other values in the application context. (iii) HYD, JAI, NK and CK represent Hyderabad, Jaipur, North Kerala, and Central Kerala.

Area-based FA models: It is discerned that Central Kerala as an application context has given the least transferability errors, and the transfer error is between 1.02 and 1.07. After Central Kerala, it is Jaipur, which has relative MAE values between 1.04 and 1.07. The higher transfer error is observed in Hyderabad with MAE between 1.13 and 1.16.

Employment-based FA models: The greater extent of transferability is noticed in North Kerala, where relative MAE values range from 0.96 to 0.97, followed by Jaipur, which has relative MAE values from 0.94 to 1.05. The highest error of transferability is in Central Kerala, with a relative MAE from 1.07 to 1.32.

Multivariable FG models: The intercept of multivariable FG models is also updated, and the MAE values of these models' transferability are summarised in Table 8.4. The reason for using only MAE values is discussed before. The transferability of FP models is more in North Kerala and Central Kerala.

8.3.3 Estimation-based Approach

A summary of transferable models based on an estimation-based approach is presented in Table 8.5. The models are developed by pooling data from two different regions and considering the regions as binary variables. The model is considered transferable if the difference parameter is not statistically significant. The results indicate that single variable FP models can be transferred across various regions, except for between Hyderabad and North Kerala. Multivariable FP models are transferable between Jaipur and North Kerala, as well as North Kerala and Central Kerala. However, this pattern of transferability is not observed in FA models, where only a limited number of single variable FP models are considered transferable, and no multivariable FA models are transferable.

Table 8.5 Summary of transferable models through estimation-based approach

Region	Freight Production			Freight Attraction		
	Area-based	Employment-based	Multivariable	Area-based	Employment-based	Multivariable
HYD and JAI	✓	✓		✓		
HYD and NK				✓	✓	
HYD and CK	✓	✓				
JAI and NK	✓	✓	✓			
JAI and CK	✓	✓				
NK and CK	✓	✓	✓	✓		

✓ indicated that the model is transferable

8.3.4 Summary of Spatial Transferability of Freight Generation Models

The application-based approach provides insights into the extent of transferability. On the other hand, the estimation-based approach focuses on the direction of transferability. Table 8.6 represents a comprehensive overview of the results of both approaches. This table showcases only the best transferability cases for different transfer methods (naïve transfer and updated model transfer) based on the application-based approach while summarizing the transferability status of models based on the estimation-based approach.

The summarised transferability results have shown the best possible way of transferability in an application context. Let us take the example of North Kerala region. The employment-based FA model of Hyderabad is transferable to this region, and the best transfer method is the updated model transfer. From Table 8.6, it is clear that the naïve transfer is a better approach. The transferability of the area-based FP models is possible in all application contexts. Among these, there are two cases – (i) transfer from North Kerala to Jaipur and (ii) transfer from North Kerala to Central Kerala, where the estimation context model (transferred model) showed better prediction ability for application context data than the application context model itself (i.e., a relative MAE of lower than 1). A similar observation is seen in the case of the transferability of employment-based FP models. Suppose we have a closer look at FA models. In that case, it is seen that the transferability of the FA model of Jaipur is better in Hyderabad. In all the other cases of transferability of area-based FA models, the transferability of models is not possible as per estimation-based, even though the relative MAE values as per application-based approach are a bit lower. There is transferability of Hyderabad employment-based FA model to North Kerala; for this case, the best transfer method is updated model transfer. On considering the transferability of multivariable models, it is seen that the transferability of North Kerala FP model is possible in Central Kerala, where naïve transfer is a better transfer method.

The transferability observed between the North Kerala and Central Kerala regions can be attributed to their shared coastal location and geographical proximity as part of the same state. Furthermore, the transferability between Jaipur and Hyderabad is evident due to the absence of coastal access in both regions, which leads to the likelihood of similar transportation network systems. The transferability across North Kerala, Hyderabad, and Jaipur can be attributed to their comparable geographical positions. Some cities in North Kerala are landlocked regions. These landlocked regions also exhibit distinct economic focuses, placing greater emphasis on industries like manufacturing, agriculture, tourism, or service sectors. The transferability is also noticed across the Kerala regions and Jaipur; the contributing factor is their economic status.

Table 8.6 Summary of transferability assessment results

Naïve Transfer				Updated Model Transfer			
Single Variable Models							
Application Context	Estimation Context	Relative MAE	T	Application Context	Estimation Context	Relative MAE	T
Area-based Freight Production Models				Area-based Freight Production Models			
HYD	**CK**	**1.05**	**Yes**	HYD	JAI/CK	1.16	Yes
JAI	**NK**	**0.99**	**Yes**	JAI	NK	1.16	Yes
NK	**JAI**	**1.02**	**Yes**	NK	JAI	1.12	Yes
CK	**NK**	**0.97**	**Yes**	CK	JAI	1.08	Yes
Employment-based Freight Production Models				Employment-based Freight Production Models			
HYD	**CK**	**1.00**	**Yes**	HYD	CK	1.14	Yes
JAI	**NK**	**0.99**	**Yes**	JAI	CK	1.11	Yes
NK	**JAI**	**1.02**	**Yes**	NK	CK	1.14	Yes
CK	**NK**	**0.97**	**Yes**	CK	HYD	1.06	Yes
Area-based Freight Attraction Models				Area-based Freight Attraction Models			
HYD	**JAI**	**1.02**	**Yes**	HYD	JAI	1.13	Yes
JAI	CK	0.92	No	JAI	NK	1.04	No
NK	CK	1.00	No	NK	CK	1.07	No
CK	NK	1.02	No	CK	NK	1.02	No
Employment-based Freight Attraction Models				Employment-based Freight Attraction Models			
HYD	**JAI**	**0.98**	**Yes**	HYD	NK	1.08	Yes
JAI	NK/CK	0.95	No	JAI	NK	0.94	No
NK	CK	0.89	No	**NK**	**HYD**	**0.96**	**Yes**
CK	JAI	1.21	No	CK	HYD	1.07	No
Multivariable Models							
Naïve Transfer				Updated Model Transfer			
Application Context	Estimation Context	MAE	T	Application Context	Estimation Context	MAE	T
Freight Production Models				Freight Production Models			
HYD	CK	14.862	No	HYD	CK	16.779	No
JAI	CK	16.348	No	JAI	CK	19.000	No
NK	CK	12.546	No	NK	CK	14.668	No
CK	**NK**	**12.650**	**Yes**	CK	NK	14.179	Yes
Freight Attraction Models				Freight Attraction Models			
HYD	JAI	13.069	No	HYD	CK	14.426	No
JAI	CK	10.499	No	JAI	CK	12.123	No
NK	CK	7.804	No	NK	CK	8.523	No
CK	JAI	6.429	No	CK	JAI	6.403	No

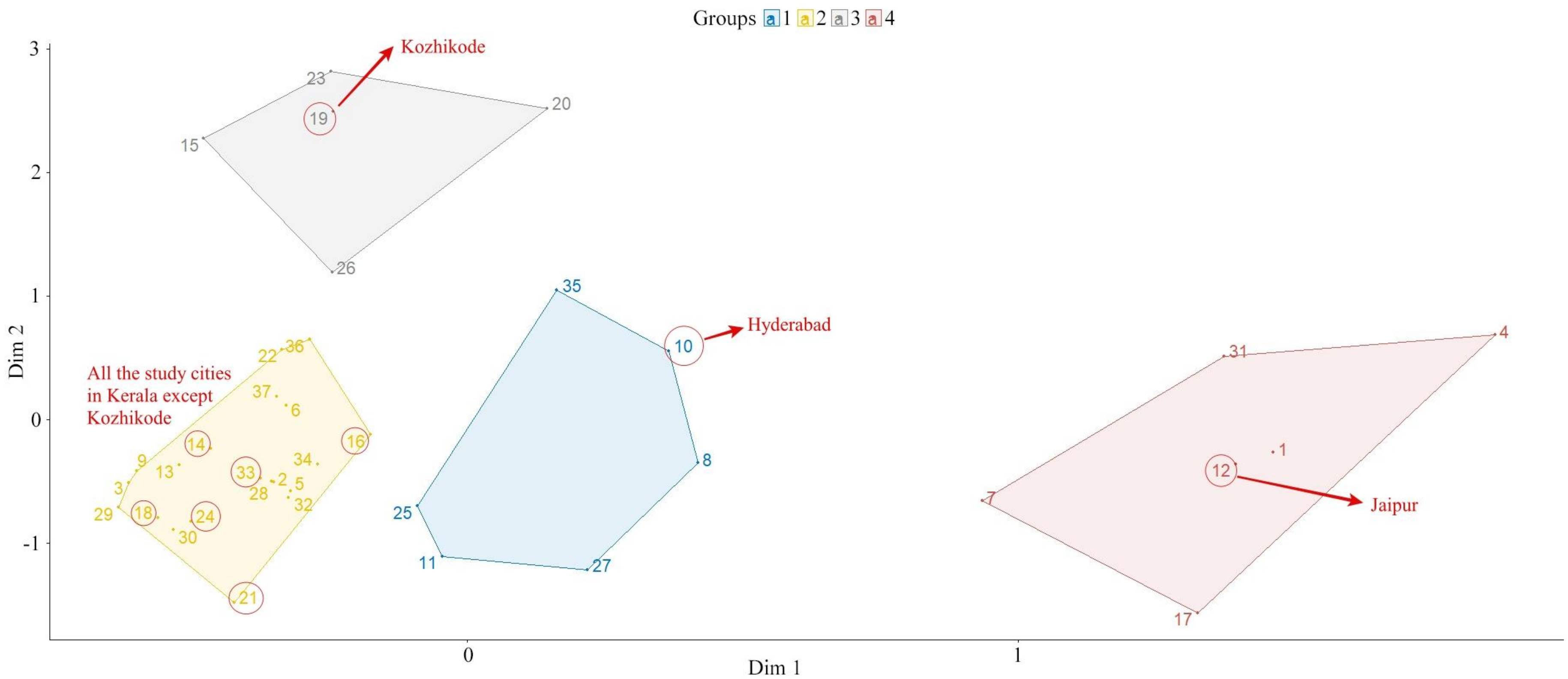
Note: The bolded cases represent the best transferable models based on results from application-based and estimation-based approaches; MAE stands for Mean Absolute Error; 'T' represents whether the model is transferable – Yes or No

8.3.5 Exploring Geographically Similar Regions

The limitation of many freight studies which assessed spatial transferability is that they could not answer the question – Are these spatial transferability results applicable to other regions (which are not study regions)? This study explores this question to some extent by understanding the geographical similarities of Indian cities/regions with a million-plus population. The determinants of freight traffic, such as population density (people/sq. km), number of establishments, land value (INR/sq. ft), and road density (km/sq. km), are considered for knowing the geographical similarities. All this data about the regions are collected from publicly available data sources. A two-dimensional Multidimensionality Scaling (MDS) solution is provided. MDS results show that many regions are very closely placed in the plot. In order to sort out into better groups, the MDS distance matrix is used for the K-means clustering algorithm. Figure 8.2 shows the map that can help delineate clusters with similar geographical characteristics. The results show that Hyderabad and Jaipur belong to different clusters. All the Kerala regions except Kozhikode belong to one cluster. It can be assumed that all the regions belonging to one cluster have similar geographical characteristics. Hence, it can be said that the models transferable to a study city/region can also be transferred to other regions in the same cluster. For example, models transferable to Hyderabad as an application context are also transferable to Coimbatore, Indore, Patna, Rajkot, and Vadodara regions.

8.3.6 Sample Size Determination

The details about the determination of sample size for the new regions are explained in Figure 8.3. This flowchart starts with the model estimation. After estimation, the transferability assessment is done using both application-based and estimation-based approaches. The sample size is determined based on which transfer method (naïve transfer or updated model transfer) is better. If naïve transfer gives good results, then there is no need for surveying the new region. Else, the minimum sample requirement is 40% of the application context sample size. Here the question is – why 40%? The application context data was divided into 20%, 40%, 60%, 80%, and 100%. With these proportions, the transferred model was assessed for transferability. It was observed that there was no variation in transfer errors. Therefore, a 20% of sample size is the requirement. However, we are recommending a bit higher proportion which is 40%.



1. Ahmedabad	2. Amritsar	3. Aurangabad	4. Bengaluru	5. Bhopal	6. Bhubaneswar	7. Chennai	8. Coimbatore	9. Guwahati	**10. Hyderabad**
11. Indore	**12. Jaipur**	13. Jammu	**14. Kannur**	15. Kanpur	**16. Kochi**	17. Kolkata	**18. Kottayam**	**19. Kozhikode**	20. Lucknow
21. Malappuram	22. Mangalore	23. New Delhi	**24. Palakkad**	25. Patna	26. Raipur	27. Rajkot	28. Ranchi	29. Shimla	30. Srinagar
31. Surat	32. Thiruvananthapuram	**33. Thrissur**	34. Tiruchirappalli	35. Vadodara	36. Varanasi	37. Visakhapatnam			

Note: (1) Some points in cluster 2 could not be labelled due to space constraints. However, all the cities unavailable in other clusters are in cluster 2. (2) All the encircled and labelled regions represent the study regions. In addition, these regions are bolded in the legend.

Figure 8.2 Map of MDS and KMCA

8.3.7 Examination of Transferability Strategy

Pani and Sahu (2019b) presented results about the average number of resources consumed for conducting EBFS in India. The following are the results – (i) the average cost per completed response per interviewer is USD 1.61 (INR 133.3), (ii) the average number of completed responses per interviewer is 2.2, and (iii) the average number of interviewers in the survey team per day is 6. Therefore, the average number of responses collected per day is 13.2. The minimum sample size requirement is as per the standard statistical formula, which is given as Eq. (10) (Cochran, 1977; Thompson, 2012).

$$n = \left[\frac{\frac{z^2 \times p(1-p)}{e^2}}{1 + \left(\frac{z^2 \times p(1-p)}{e^2 N} \right)} \right] \quad 8.10$$

Where n is the minimum sample size required; N is the population; z is the z score (z = 1.96 if we consider a confidence level of 95%); p is the proportion of the population, which is taken as 0.5 by default; e is the margin of error which is by default considered as 0.05 or 5%.

Let us take the example of the transferability of the employment-based FA model of Hyderabad to North Kerala. North Kerala Region and other cities form cluster 2 as per Figure 8.2. The sample size requirement as per the conventional approach and transferability approach (proposed framework) are presented in Table 8.7. On comparing the conventional and transferability approaches of sample sizes for this example, it is clear that the transferability approach proposes a size of 59. In contrast, the conventional approach proposes a sample size between 367 and 381. This minimisation of sample size can reduce the cost in the range of INR 41,056 and INR 42,922. This approach can reduce the time of data collection to 5 days. This comparison between the conventional and transferability approaches is illustrated in Figure 8.4 (minimum sample size requirements), Figure 8.5 (survey durations), and Figure 8.6 (survey costs).

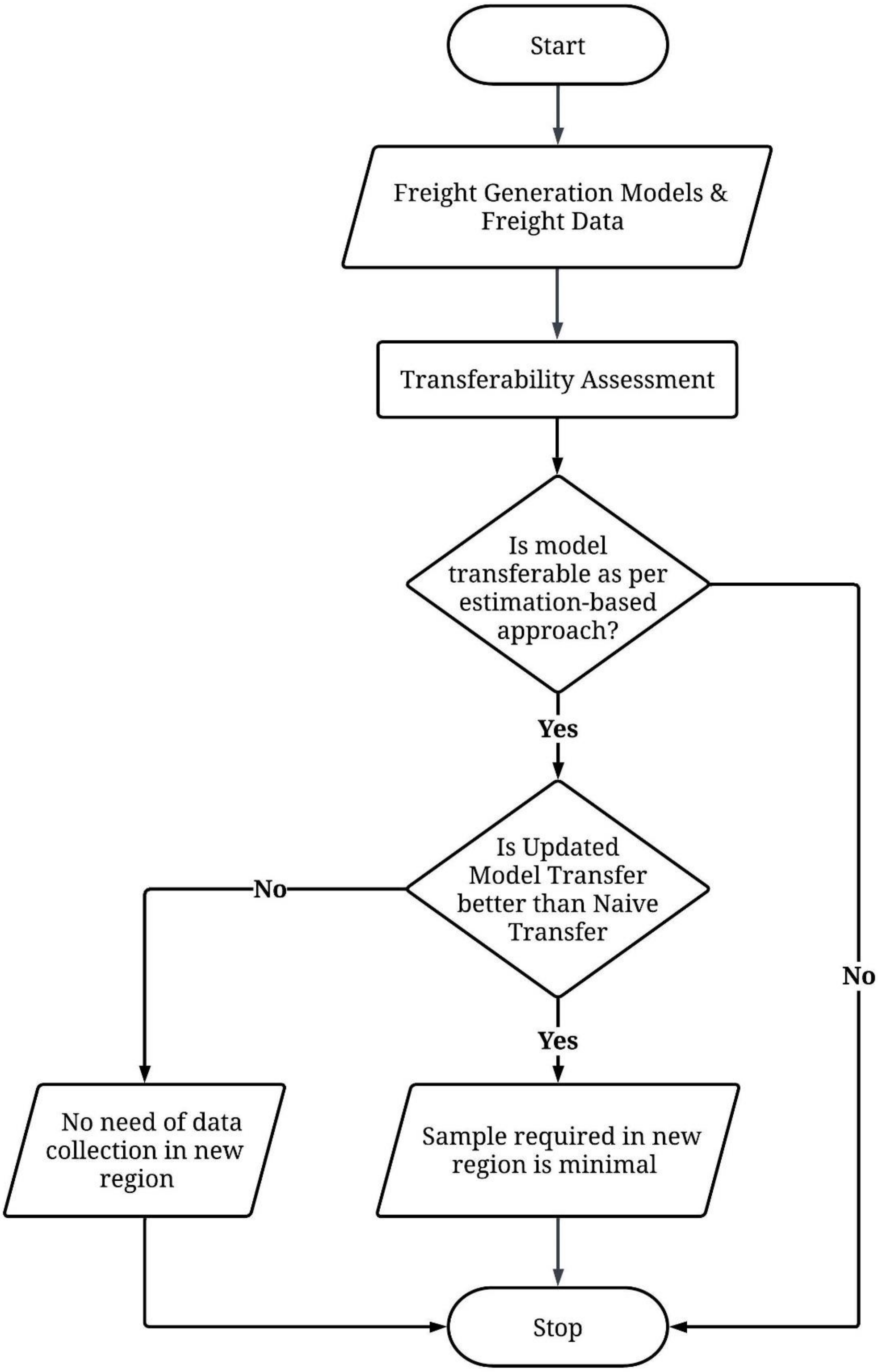


Figure 8.3 Flowchart to determine sample size for new regions

Table 8.7 Comparison between conventional and transferability approaches

	Conventional Approach	Transferability Approach		Conventional Approach	Transferability Approach
Amritsar			**Ranchi**		
Sample Size	380	59	Sample Size	380	59
Time Required for Data Collection (in days)	29	5	Time Required for Data Collection (in days)	29	5
Cost of Data Collection (in INR)	50,654	7,865	Cost of Data Collection (in INR)	50,654	7,865
Aurangabad			**Shimla**		
Sample Size	367	59	Sample Size	367	59
Time Required for Data Collection (in days)	28	5	Time Required for Data Collection (in days)	28	5
Cost of Data Collection (in INR)	48,921	7,865	Cost of Data Collection (in INR)	48,921	7,865
Bhopal			**Srinagar**		
Sample Size	381	59	Sample Size	377	59
Time Required for Data Collection (in days)	29	5	Time Required for Data Collection (in days)	29	5
Cost of Data Collection (in INR)	50,787	7,865	Cost of Data Collection (in INR)	50,254	7,865
Bhubneshwar			**Thiruvananthapuram**		
Sample Size	380	59	Sample Size	381	59
Time Required for Data Collection (in days)	29	5	Time Required for Data Collection (in days)	29	5
Cost of Data Collection (in INR)	50,654	7,865	Cost of Data Collection (in INR)	50,787	7,865
Guwahati			**Tiruchirappalli**		
Sample Size	370	59	Sample Size	381	59
Time Required for Data Collection (in days)	29	5	Time Required for Data Collection (in days)	29	5
Cost of Data Collection (in INR)	49,321	7,865	Cost of Data Collection (in INR)	50,787	7,865
Jammu			**Varanasi**		
Sample Size	376	59	Sample Size	380	59
Time Required for Data Collection (in days)	29	5	Time Required for Data Collection (in days)	29	5
Cost of Data Collection (in INR)	50,121	7,865	Cost of Data Collection (in INR)	50,654	7,865
Mangalore			**Visakhapatnam**		
Sample Size	380	59	Sample Size	380	59
Time Required for Data Collection (in days)	29	5	Time Required for Data Collection (in days)	29	5
Cost of Data Collection (in INR)	50,654	7,865	Cost of Data Collection (in INR)	50,654	7,865

Note: (i) INR stands for Indian Rupees. (ii) Here employment-based FA model of Hyderabad is transferred to North Kerala. (iii) Cost per response is INR 133.3 (Pani & Sahu, 2019b)

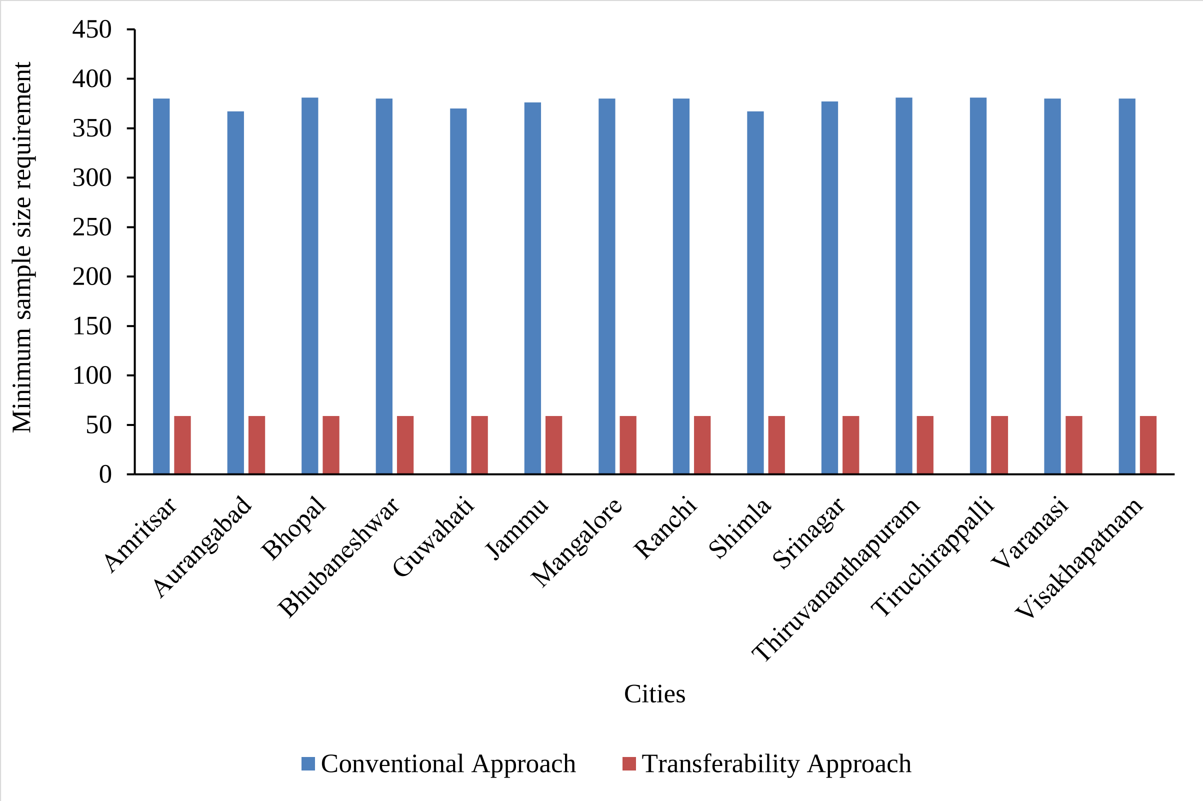


Figure 8.4 Comparison of minimum sample size requirements

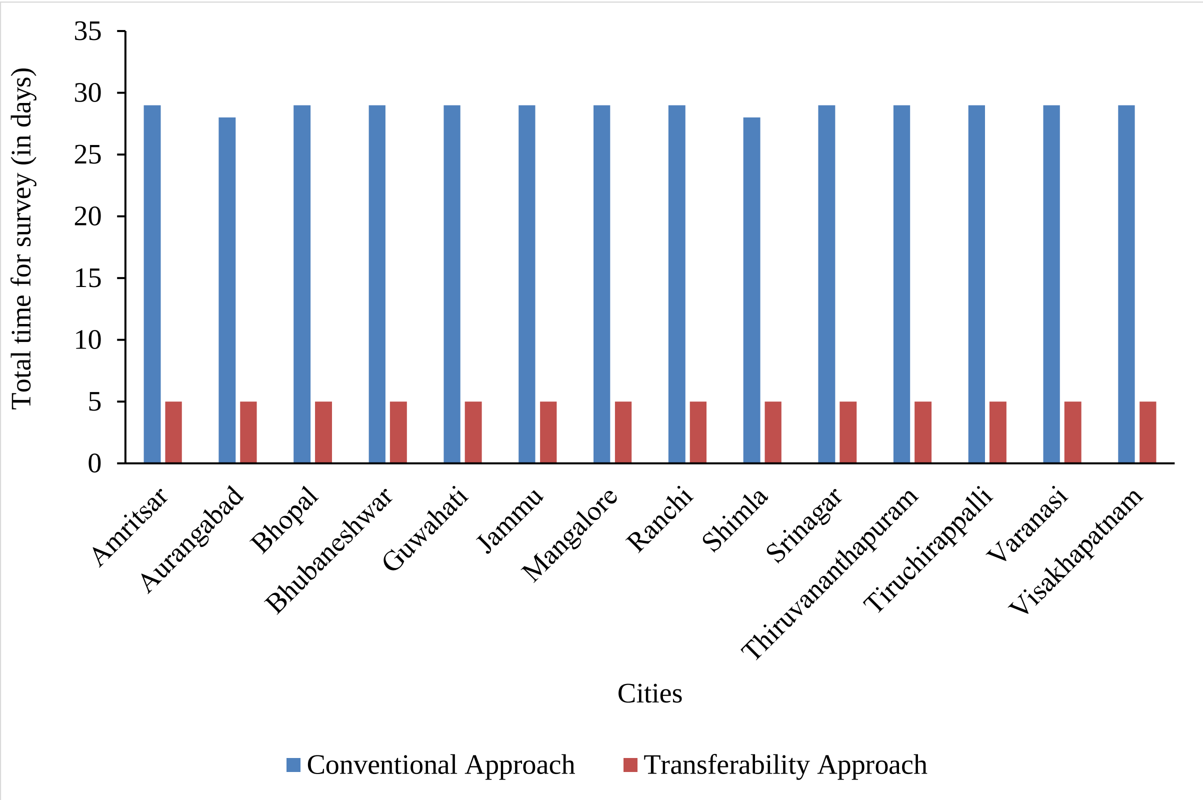


Figure 8.5 Comparison of survey durations

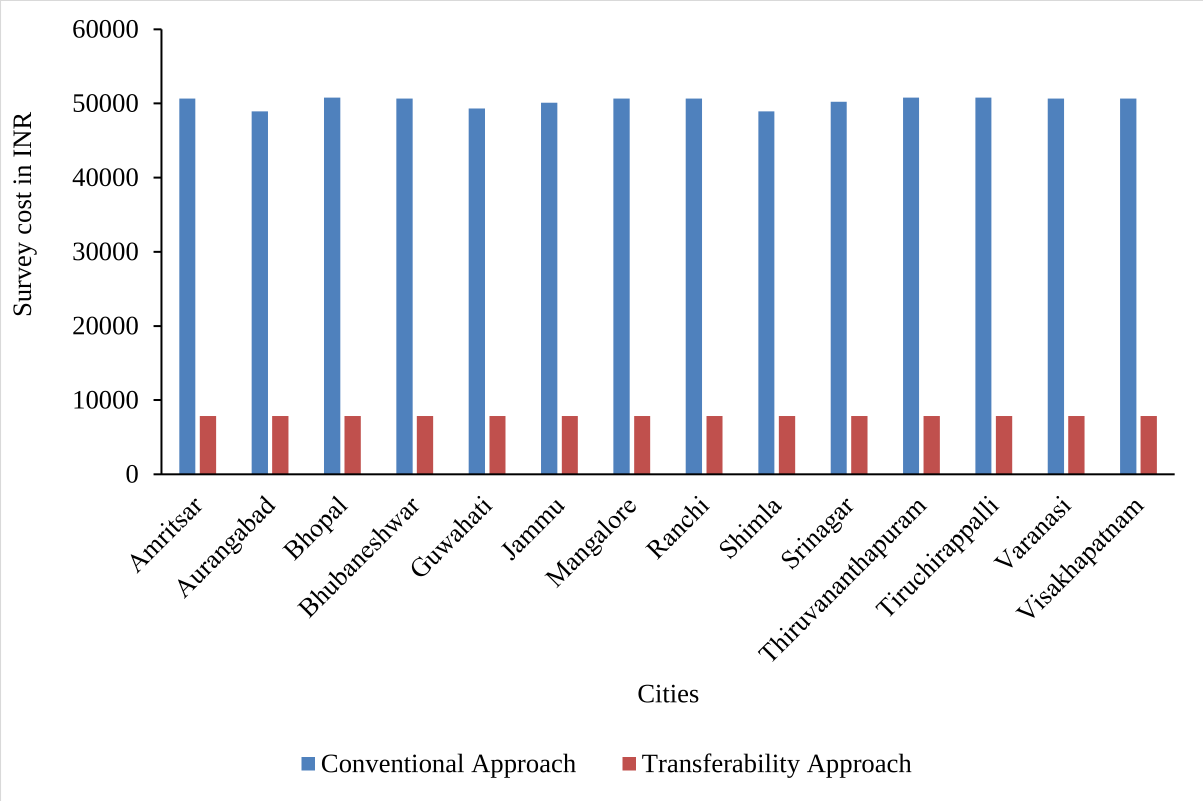


Figure 8.6 Comparison of survey costs

8.4 Research Implications

Freight data collection is crucial in various aspects, including cost analysis, time management, and financial considerations. Collecting disaggregate freight data for regions with little to no data is a tedious task which is resource and time intensive. This resource limitation brings us to the subsequent implication of the present study - the minimisation of resources required for establishment-level freight data collection via spatial transferability of the FG models. Instead of collecting data and generating FG models for a region with insufficient freight data at a disaggregate level, transferring an already developed FG model for a geographically similar region to the region of interest can save time and funds spent on data collection. The geographical transfer of the FG models simplifies the process of FG modelling for a new region with reasonable accuracy. Furthermore, the policymakers and stakeholders can borrow the strategies already under implementation in a region with topographic similarity. Effective policy and decision-making regarding infrastructural development and planning of trucking terminals, freight lanes and corridors, and freight consolidation centres enable easy governance. Similarly, industrialists can take the establishments in a geographically similar region as case studies to plan the resource and capacity allocation and design efficient and responsive supply chains.

The freight information gathered can assist in formulating policies concerning infrastructure investment, economic development, environmental sustainability, and transportation pricing strategies. Accurate and timely freight data enables policymakers to identify areas that require additional investments to facilitate the efficient movement of goods. Government agencies can invest in data collection technologies and methodologies to ensure effective freight data collection at a national level. These investments may involve implementing automatic vehicle location systems, electronic toll collection systems, and remote sensing technologies. By adopting these technologies, agencies can gather more precise and up-to-date freight data, leading to improved decision-making processes. In order to minimize the effort required for data collection, particularly in regions experiencing rapid industrial growth, the proposed methodology in this study suggests utilizing existing models that have already been established.

8.5 Summary

This study attempts to develop a framework to determine the reduced sample size of a region using the transferability approach. For this reason, 943 disaggregate observations (data from an establishment is an observation) were collected across various study regions in India, namely Hyderabad, North Kerala region (cities - Malappuram, Kannur, Kozhikode, and Kochi), and Central Kerala region (cities - Kottayam, Thrissur, and Palakkad) and Jaipur. A set of single variable freight generation (which includes freight production (FP) and freight attraction (FA)) models and multivariable freight generation models are built using this data with business size variables – gross floor area of an establishment and number of employees working in an establishment as explanatory variables. Models are assessed for their spatial transferability across the study regions by two approaches – (i) Application-based approach and (ii) Estimation-based approach. In the application-based approach, the transfer methods used are naïve transfer and updated model transfer. After this approach, an estimation-based approach is used, and in this approach, the data from two regions is pooled. The models are developed with pooled data, considering regions as binary variables. From the results of both approaches, it is observed that single variable FP models are more transferable. In the case of single variable FA models, the transferability is limited to only a certain number of regions.

However, the multivariable FP and FA models did not have a reasonable extent of transferability.

Following modelling and assessing transferability, the regions geographically similar to the study regions are identified. The identification process involves considering population density, number of establishments, land value, road density, and proximity to seaports. To accomplish this, the Multidimensionality Scaling yields a similarity matrix, which is then utilised as input for the K-means clustering algorithm. As a result, four distinct clusters are formed. Hyderabad, Jaipur, Kozhikode, and all other study regions of Kerala belong to separate clusters. These clusters also contain new regions that exhibit geographic similarity to the study regions and share similar transferability characteristics. By obtaining comprehensive information on the transferability of these new regions, it becomes straightforward to calculate the required sample size for them. This study introduces a novel framework that utilises a transferability approach to determine the sample size.

The methodology described in this research can save money and time for planning organisations when gathering survey data. This novel method can reduce the money spent on data collection in the states lacking freight data due to resource constraints. In these regions, the resources can be minimised to an unimaginable extent - many regions or cities in a state can be surveyed with the budget allocated to a city or region. Suppose a state has ample funding for the survey. In that case, more data can be collected from that state, increasing the accuracy of the results obtained from the transferability approach. Also, this method can make the planning agencies switch to other survey modes, such as web-based and hybrid surveys, instead of face-to-face interviews. This switch can reduce the number of enumerators required for surveying. In addition, the time needed to complete the survey can be reduced. Planning organisations can collect a more significant amount of data within the stipulated time. If they spend more time, they receive more observations. More observations can update the models and transferability findings with a greater magnitude. Finally, this novel framework can save time, reduce cost per response, and minimise the human resources requirement in freight surveys.

When calculating the freight demand in the new region, the government agencies can significantly improve their accuracy by combining the little data set they have acquired from this region with an existing model of the freight demand market. These estimations can be beneficial in strategically planning the geographical location of freight distribution centres, truck terminals, and consolidation warehouses. In addition, the authorities can develop freight lanes or corridors with complete confidence. In addition, the amount of congestion generated by the movements of commercial vehicles can be forecasted using the transferred demand models. Policymakers can decide on the required data and the funding necessary for collecting those data. Also, they can make decisions on restricting truck travel if needed. Industrialists who want to establish a new manufacturing or assembling unit in a region can determine the amount of land that will be necessary, as well as the number of workers that will need to be hired, using only a limited amount of information on freight flows. Additionally, decisions on the types of trucks and carriers can be made at the firm level in a more appropriate way for logistics. In addition, the logistics manager will be better able to provide an accurate estimate of the cost of transportation.

Nevertheless, the findings of this study are limited because determinants considered to identify the geographical similarities may be restricted. Further similarities should be classified for more accuracy. The sample size calculation with a small amount of data from various regions can create constraints regarding transferability accuracy. In order to overcome these limitations, more data need to be gathered on the characteristics of the regions, and more freight data needs to be collected from various regions. However, the methodology proposed in this study is generic. It can be utilised to investigate the freight model transferability direction and sample size for regions inside and outside India.

Part V: Conclusions

Chapter 9: Conclusions and Recommendations

9.1 Specific Conclusions

This thesis commenced with a primary research goal aimed at developing cost-efficient freight demand models with disaggregate freight data using the practice of spatial transferability. The establishment of this research was built upon three precise research inquiries connected to this objective: (i) model estimation (Which modelling methodology accurately estimates FG at industrial, regional and state levels? Which establishment typology increases the accuracy of FG models?), (ii) model application (How effective are urban models using various transfer methods when utilised in suburban regions? How do various modelling methodologies perform in the context of spatial transferability? What is the influence of sample size on spatial transferability of the FG models?), and (iii) measurement (How do we minimise the efforts of freight data collection for the regions that are constrained in terms of resources?). The summary of findings and insights drawn from the result interpretations of each research question are presented in subsequent sections.

9.1.1 Model Estimation

Which modelling methodology accurately estimates FG at industrial, regional and state levels?

Both parametric and non-parametric modelling techniques were utilised when developing a set of freight generation (FG) models (which includes freight production (FP) and freight attraction (FA)). These models take into account gross floor area (GFA) and number of employees (NE). The ordinary least squares (OLS) regression, weighted least squares (WLS) regression, robust regression (RR), and seemingly unrelated regression (SUR) are all examples of parametric modelling approaches. Multiple classification analysis (MCA) and support vector regression (SVR) are examples of non-parametric modelling approaches. This study's most critical findings and contributions include the following:

- Based on the model estimation findings, establishments in suburban areas exhibit higher FP and FA rates. These elevated rates can be credited to two potential factors:

(i) the availability of inexpensive land in suburban regions, facilitating the establishment of more extensive facilities, and (ii) the employment of more workers willing to accept lower wages due to the lower cost of living in these areas.

- The increased FG rates, correlated with greater GFA and NE, align with the principles of production theory, which suggests that output quantity is influenced by factors like land, employment, and capital. Moreover, the higher average FP and FA in establishments located in suburban areas could be a consequence of heavily congested urban roads with insufficient width.

- The models developed at various levels were validated using internal and external validation techniques. The validation results show that SVR models, a non-parametric modelling approach, perform better than other approaches when it comes to modelling at the state, regional, and level of specific industrial segments.

- In the case of suburban models, MCA, a non-parametric approach, displays a higher degree of accuracy in FG prediction. RR (a parametric modelling methodology) exhibit superior prediction ability in forecasting FG in specific industrial segments.

- Overall, comparing results and their interpretation point to the superiority of non-parametric models in FG prediction, with RR being the only parametric approach offering performance comparable to that of the non-parametric models.

- Nevertheless, SVR is the method of choice for modelling FG when the establishments' freight data distribution is unknown. This choice of methodology is because SVR allows for more accurate predictions in the presence of data with unknown distribution.

Which establishment typology increases the accuracy of FG models?

FP models were developed for various latent classes (establishment typology), which were statistically derived. Different variables were considered for the typology, and the following are the conclusions.

- The developed class models have shown better prediction ability than the regional models.

- A comparison of the employment-based and area-based FP models shows that the former models have a higher prediction ability.

- The class models for classes developed with conditional latent class analysis have shown more accuracy compared to the class models for models derived through unconditional latent class analysis.

- Class models for classes conditioned with industry type have shown better accuracy compared to the class models conditioned with business size variables – gross floor area and employment.

9.1.2 Model Application

How effective are urban models using various transfer methods when utilised in suburban regions?

Initially, the FP models developed for both urban and suburban regions were directly transferred. Subsequently, the naïve models were transferred by updating their coefficients using the combined transfer estimation technique. The conclusions of this study are as follows.

- The assessment of transferability was conducted for both naïve and updated models, revealing that the updated models exhibit superior transferability performance.

- The findings indicate that most urban models can be transferred effectively to the suburban context. However, the transferability of suburban models to the urban context is limited, with only a few models being successfully transferable. Thus, the conclusion drawn is that transferability is asymmetric in this scenario.

- Among the urban models, area-based models demonstrate higher transferability than employment-based models.

- Specifically, medium and high-value density industry (segments 2 and 3) urban models exhibit transferability to the suburban context, with medium value density industry (segment 2) models being particularly more transferable. Only the area-based model of high-value density industry (segment 3) shows transferability in suburban models.

- As an alternative approach, a joint context estimation technique was employed to examine whether pooled data models offer improved transferability results. However, it was observed that the combined transfer estimation technique outperformed the joint context estimation technique regarding transferability.

How do various modelling methodologies perform in the context of spatial transferability?

The FP models were assessed for their spatial transferability. The industry class models were assessed for transferability across various Indian cities, and the following conclusions are derived from the results.

- The parametric modelling methodology, OLS, showed a better transferability.
- Also, MCA is the recommended modelling approach for achieving spatial transferability.
- In most transfer cases, where MCA is the modelling methodology, the transferability between Hyderabad and Jaipur is more.

What is the influence of sample size on spatial transferability of the FG models?

Considering the influence of the sample size on spatial transferability, the following are the key conclusions.

- The employment-based FP model in Hyderabad exhibits negligible transfer error when applied to many other study cities across various industrial classes.
- Concerning area-based FP models, Kozhikode (also known as Calicut) proves to be a more suitable estimation context.
- Further examination of transferability while considering variations in sample size reaffirms MCA's superiority in achieving transferability.
- It is also observed that OLS can also be considered to be a relatively better modelling approach in terms of the extent of transferability.

9.1.3 Measurement

How to minimise the efforts of freight data collection for the regions which are constrained in terms of resources?

FG models were formulated with business size variables – NE and GFA, as predictor variables. The models were evaluated for spatial transferability across various Indian regions using application-based and estimation-based approaches. A novel approach MMM was proposed and the following are the conclusions of the study.

- The transferability assessment of both approaches (application-based and estimation-based approaches) indicates that single variable FP models are transferable to all study regions whereas, the single variable FA models are transferable to only specific regions.
- No significant improvement was observed in the transferability of multivariable models (including FP and FA) compared to single variable models.
- The minimum sample size required for the various cities was determined based on transferability and geographical characteristics. The minimum sample size requirement based on MMM approach is observed to be 80% less than the conventional approach of sample size calculation.
- The reduction in the minimum sample size requirement has also reduced the requirement of the survey resources in terms of time and money by 80%.
- In cases where the regions are geographically similar, the pre-existing model of the study region can be directly used. Hence, the data collection in these regions can be avoided.

Chapter 10: Specific Contributions

The thesis makes several significant contributions to the field. These contributions are outlined as follows:

- Novel Freight Generation Estimation: The study introduces innovative methods for estimating freight generation, which is particularly crucial in India, where conventional commodity flow surveys are lacking. By providing reliable estimates of freight generation, the thesis addresses a critical gap in understanding transportation logistics in the country.

- Strategic Planning Insights: The developed freight generation models offer valuable insights into strategic planning for logistics facilities such as truck terminals, warehouses, and freight consolidation centres beyond urban boundaries. These insights enable policymakers and planners to make informed decisions regarding infrastructure development and establishment policies, thereby enhancing overall logistics efficiency.

- Informative Decision-Making Tools: Policymakers and planners can utilise the models presented in the study to make data-driven decisions on investments and operational strategies specific to freight transportation. This research enables more effective planning at various levels, including state, regional, and corridor planning, contributing to improved transportation networks and capacity management.

- Policy Implications: Accurate freight data collection is essential for designing and implementing effective policies on economic development, infrastructure investment, and environmental sustainability. The thesis underscores the importance of financing advanced data-gathering technologies to enhance freight data collection processes, ultimately improving decision-making in transportation planning.

- Guidelines Development: The thesis contributes to the development of authoritative guidelines related to donor context identification and transferable model parameters in freight generation modelling. This provides a framework for future researchers and practitioners to follow, ensuring greater consistency and reliability in modelling approaches.

- Optimised Resource Allocation: Selecting appropriate modelling approaches allows for optimized resource allocation in transportation planning. By strategically identifying optimal locations for logistics facilities and strategically clustering industries, the models presented in the thesis facilitate efficient resource allocation and promote sustainable economic growth.

- Enhanced Data Collection Strategies: The spatial transferability of freight generation models proposed in the thesis minimizes resource requirements for data collection, particularly in regions with limited data availability. This approach streamlines the modelling process and optimizes resource allocation, contributing to more efficient freight data collection practices.

- Stimulating Further Research: The research stimulates further investigations and discussions on freight generation modelling in India and other developing economies. The thesis lays the groundwork for potential advancements in freight logistics practices and policies globally by addressing this area of study.

Chapter 11: Future Scope of Work

This research has certain limitations like any other research. Initially, the transferability assessment was conducted across four specific regions, potentially limiting the applicability of the findings to broader geographical contexts. Furthermore, due to data constraints, only five factors—seaport proximity, land value, population density, number of establishments, and road density—were utilized to categorize regions with geographical similarities. It is plausible that other factors related to economic activity and policy measures, which may influence freight movements, were not considered. Additionally, the study did not examine the temporal stability of the transferable models. Moreover, further investigations are warranted to validate the transferability outcomes and establish a reliable framework for benefit transfer in freight transportation planning. The proposed model transfer framework relies on established industrial attributes, providing transportation agencies, planners, and practitioners a means to identify factors beneficial to model transfer. However, it is essential to empirically verify whether these findings hold true across different nations and diverse geographical contexts. In addition, future studies should explore the proposed framework across various regions, aggregation levels, and industrial characteristics. Furthermore, this study did not address the impact of harvest seasons, commodities, business cycles, weather conditions, natural disasters, and economic fluctuations on model parameters, presenting an intriguing avenue for further inquiry. Lastly, the study models, developed from shippers' data, did not capture the sequence of freight movements involving urban consolidation centres, warehouses, hubs, and other freight infrastructures.

Insights for future research endeavours involve conducting freight surveys in a wider array of regions to deepen our comprehension of demand estimation and broaden the scope of transferability. It is imperative to explore machine learning algorithms like transfer learning to scrutinise the accuracy of transferability, although necessitating a larger dataset for robust analysis. Hence, expanding freight surveys across diverse cities would facilitate the extrapolation of research findings. Moreover, delving further into measurement analysis could entail considering additional variables to pinpoint geographically analogous regions. Collecting data spanning various timeframes is vital for grasping temporal fluctuations in freight patterns. Furthermore, broadening the analysis to encompass a broader range of factors influencing freight movements, including harvest seasons, commodities, business cycles, weather conditions, natural disasters, and economic fluctuations, would yield a more comprehensive understanding of freight transportation dynamics. Additionally, gathering freight data directly from carriers and warehouses would yield insights into the intricacies of truck movements.

Furthermore, weather-related disruptions such as monsoons, floods, or cyclones can potentially disrupt supply chains and alter freight flow patterns. To mitigate such risks, models may integrate historical weather data and predictive analytics to anticipate disruptions and their ramifications on freight movements. Understanding the details of one-way freight transport and hinterland logistics is paramount, particularly in countries like India, characterised by diverse topographies and regional inequalities. Models designed for such contexts should consider factors such as trade imbalances, infrastructure availability, and regional preferences for transportation modes. Lastly, the impact of exceptional events like pandemics on truck movements can be comprehensively captured if sufficient data is available before and after these occurrences.

References

Agyemang-Duah, K., & Hall, F. L. (1997). Spatial transferability of an ordered response model of trip generation. *Transportation Research Part A: Policy and Practice*, *31*(5), 389–402. https://doi.org/10.1016/S0965-8564(96)00035-3

Alho, A. R., & Silva, J. de A. e. (2014). Freight-Trip Generation Model - Predicting Urban Freight Weekly Parking Demand from Retail Establishment Characteristics. *Transportation Research Record*, *2411*(1), 45–54. https://doi.org/10.3141/2411-06

Alho, A. R., & Silva, J. de A. e. (2017). Modeling retail establishments' freight trip generation: a comparison of methodologies to predict total weekly deliveries. *Transportation*, *44*(5), 1195–1212. https://doi.org/10.1007/s11116-016-9700-z

Allen, J., Browne, M., & Cherrett, T. (2012). Survey Techniques in Urban Freight Transport Studies. *Transport Reviews*, *32*(3), 287–311. https://doi.org/10.1080/01441647.2012.665949

Arentze, T., Hofman, F., Van Mourik, H., & Timmermans, H. (2002). Spatial transferability of the Albatross model system: Empirical evidence from two case studies. *Transportation Research Record*, *1805*, 1–7. https://doi.org/10.3141/1805-01

Atherton, T., & Ben-Akiva, M. (1976). Transferability and updating of disaggregate travel demand models. *Transportation Research Record, 610*, 12–18.

Balla, B. S., & Sahu, P. K. (2023). Assessing regional transferability and updating of freight generation models to reduce sample size requirements in national freight data collection program. *Transportation Research Part A*, *175*(May), 103780. https://doi.org/10.1016/j.tra.2023.103780

Balla, B. S., Sahu, P. K., & Pani, A. (2021). Are Freight Production Models Transferable between Urban and Suburban Areas? Guiding Model Transfer in Geographically Sprawling Indian Cities. *Journal of The Institution of Engineers (India): Series A*, *102*(3), 643–656. https://doi.org/10.1007/s40030-021-00556-7

Balla, B. S., Sahu, P. K., Pani, A., Sharma, S., & Majumdar, B. B. (2023). Comparison of Parametric and Non-Parametric Methods for Modeling Establishment-Level Freight

Generation. *Transportation Research Record: Journal of the Transportation Research Board*, *2677*(2), 154–172. https://doi.org/10.1177/03611981221116369

Bastida, C., & Holguín-Veras, J. (2009). Freight generation models: Comparative analysis of regression models and multiple classification analysis. *Transportation Research Record*, *2097*, 51–61. https://doi.org/10.3141/2097-07

Ben-Akiva, M., & Bolduc, D. (1987). Approaches to model transferability and updating: the combined transfer estimator. *Transportation Research Record*, *1139*, 1–7.

Bowman, J. L., Bradley, M., Castiglione, J., & Yoder, S. L. (2014). Making advanced travel forecasting models affordable through model transferability. *Transportation Research Board 93rd Annual Meeting*.

Bowman, John L, Bradley, & Mark. (2017). Testing Spatial Transferability of Activity-Based Travel Forecasting Models. *Transportation Research Record: Journal of the Transportation Research Board*, *2669*, 62–71. http://dx.doi.org/10.3141/2669-07

Campbell, S., Holguín-Veras, J., Ramirez-Rios, D. G., González-Calderón, C., Kalahasthi, L., & Wojtowicz, J. (2018). Freight and service parking needs and the role of demand management. *European Transport Research Review*, *10*(2). https://doi.org/10.1186/s12544-018-0309-5

Cantillo, V., Jaller, M., & Holguín-Veras, J. (2014). The Colombian Strategic Freight Transport Model Based on Product Analysis. *PROMET - Traffic&Transportation*, *26*(6), 487–496. https://doi.org/10.7307/ptt.v26i6.1460

Chandra, A., Pani, A., Sahu, P. K., Majumdar, B. B., & Sharma, S. (2022). Identifying large freight traffic generators and investigating the impacts on travel pattern: A decision tree approach for last-mile delivery management. *Research in Transportation Business and Management*, *43*(October 2020), 100695. https://doi.org/10.1016/j.rtbm.2021.100695

Chandra, A., Pani, A., Sahu, P. K., & Sharma, S. (2023). Integrating Commodity-Based and Trip-Based Approaches of Freight Demand Modelling Using Trip Length Distributions. *Journal of The Institution of Engineers (India): Series A*, *104*(2), 417–434. https://doi.org/10.1007/s40030-023-00712-1

Chang, L. Y. (2005). Analysis of freeway accident frequencies: Negative binomial regression versus artificial neural network. *Safety Science*, *43*(8), 541–557. https://doi.org/10.1016/j.ssci.2005.04.004

Cochran, W. G. (1977). *Sampling Techniques* (Third). John Willey and Sons, Inc.

Comi, A., Donnelly, R., & Russo, F. (2013). Urban Freight Models. In L. Tavasszy & G. de Jong (Eds.), *Modelling Freight Transport* (pp. 163–200). Elsevier Inc. https://doi.org/10.1016/B978-0-12-410400-6.00008-2

Cox, T. F., & Cox, M. A. A. (2000). *Monographs on Statistics and Applied Probability 88 - Multidimensional Scaling*. Chapman & Hall/CRC.

de Oliveira, L. K., França, J. G. da C. B., Nascimento, C. de O. L., de Oliveira, I. K., Meira, L. H., & Rabay, L. (2021). Evaluating problems and measures for a sustainable urban freight transport in Brazilian historical cities. *Sustainable Cities and Society, 69*. https://doi.org/10.1016/j.scs.2021.102806

Department of Economic and Social Affairs. (2008). International Standard Industrial Classification of All Economic Activities (ISIC), Rev.4. In *Statistical Papers*. https://doi.org/10.4337/9781781955659.00009

Everitt, B. (1980). Cluster Analysis. In *Quality and Quantity* (pp. 75–100). Springer. https://doi.org/https://link.springer.com/article/10.1007/BF00154794

Fu, H., Wilmot, C. G., & Baker, E. J. (2006). Sequential logit dynamic travel demand model and its transferability. *Transportation Research Record*, *1977*, 17–26. https://doi.org/10.3141/1977-05

Galbraith, R. A., & Hensher, D. A. (1982). Intra-Metropolitan Transferability of Mode Choice Models. *Journal of Transport Economics and Policy*, *16*, 7–29.

Garrido, R. A., & Mahmassani, H. S. (2000). Forecasting freight transportation demand with the space-time multinomial probit model. *Transportation Research Part B: Methodological*, *34*(5), 403–418. https://doi.org/10.1016/S0191-2615(99)00032-6

Giuliano, G., Gordon, P., Pan, Q., Park, J. Y., & Wang, L. L. (2010). Estimating freight flows for metropolitan area highway networks using secondary data sources. *Networks and Spatial Economics*, *10*(1), 73–91. https://doi.org/10.1007/s11067-007-9024-9

Gonzalez-Calderon, C. A., Holguín-Veras, J., Amaya, J., Sánchez-Díaz, I., & Sarmiento, I. (2021). Generalized noortman and van es' empty trips model. *Transportation Research Part A: Policy and Practice*, *145*(February), 260–268. https://doi.org/10.1016/j.tra.2021.01.005

Gonzalez-Feliu, J., & Peris-Pla, C. (2017). Impacts of retailing attractiveness on freight and shopping trip attraction rates. *Research in Transportation Business and Management*, *24*(July), 49–58. https://doi.org/10.1016/j.rtbm.2017.07.004

Gonzalez-Feliu, J., & Sánchez-Díaz, I. (2019). The influence of aggregation level and category construction on estimation quality for freight trip generation models. *Transportation Research Part E: Logistics and Transportation Review*, *121*(July 2018), 134–148. https://doi.org/10.1016/j.tre.2018.07.007

Government of India. (2022). *Ministry of Micro, Small & Medium Enterprises*. https://dashboard.msme.gov.in/udyam_dist_wise.aspx?stid=33

Guevara, C. A., & Thomas, A. (2007). Multiple classification analysis in trip production models. *Transport Policy*, *14*(6), 514–522. https://doi.org/10.1016/j.tranpol.2007.08.001

Günay, G., Ergün, G., & Gökaşar, I. (2016). Conditional Freight Trip Generation modelling. *Journal of Transport Geography*, *54*, 102–111. https://doi.org/10.1016/j.jtrangeo.2016.05.013

Ha, D.-H., & Combes, F. (2015). Building a model of freight generation with a commodity flow survey. In U. Clausen, H. Friedrich, C. Thaller, & C. Geiger (Eds.), *Commercial Transport, Proceedings of the 2nd Interdiciplinary Conference on Production Logistics and Traffic* (pp. 23–37). Springer, Cham. https://doi.org/10.1007/978-3-319-21266-1

Hensher, D. A., & Ton, T. T. (2000). A comparison of the predictive potential of artificial neural networks and nested logit models for commuter mode choice. *Transportation Research Part E: Logistics and Transportation Review*, *36*(3), 155–172. https://doi.org/10.1016/S1366-5545(99)00030-7

Holguín-Veras, J., Jaller, M., Sánchez-Díaz, I., Campbell, S., & Lawson, C. T. (2014). Freight Generation and Freight Trip Generation Models. In L. Tavasszy & G. de Jong (Eds.), *Modelling Freight Transport* (pp. 43–63). Elsevier Inc. https://doi.org/10.1016/B978-0-12-410400-6.00003-3

Holguín-Veras, J., Jaller, M., Sanchez-Diaz, I., Wojtowicz, J., Campbell, S., Levinson, H., Lawson, C., Powers, E. L., & Tavasszy, L. (2012a). *NCFRP 19: Freight Trip Generation and Land Use: Final Report*.

Holguín-Veras, J., Jaller, M., Sanchez-Diaz, I., Wojtowicz, J., Campbell, S., Levinson, H., Lawson, C., Powers, E. L., & Tavasszy, L. (2012b). NCFRP Research Report 19: Freight

Trip Generation and Land Use. In *Freight Trip Generation and Land Use*. https://doi.org/10.17226/23437

Holguín-Veras, J., Lawson, C., Wang, C., Jaller, M., González-Calderón, C., Campbell, S., Kalahashti, L., Wojtowicz, J., & Ramirez, D. (2016). *NCFRP Research Report 37: Using Commodity Flow Survey Microdata and Other Establishment Data to Estimate the Generation of Freight, Freight Trips, and Service Trips: Guidebook*. The National Academies Press. https://doi.org/10.17226/24602

Holguín-Veras, J., López-Genao, Y., & Salam, A. (2002). Truck-trip generation at container terminals: Results from a nationwide survey. *Transportation Research Record*, *1790*, 89–96. https://doi.org/10.3141/1790-11

Holguín-Veras, J., Sánchez-Díaz, I., Lawson, C., Jaller, M., Campbell, S., Levinson, H., & Shin, H. S. (2013). Transferability of freight trip generation models. *Transportation Research Record*, *2379*, 1–8. https://doi.org/10.3141/2379-01

Hunt, J. D., Stefan, K. J., & Brownlee, A. T. (2006). Establishment-based survey of urban commercial vehicle movements in Alberta, Canada survey design, implementation and results. *Transportation Research Record*, *1957*, 75–83. https://doi.org/10.3141/1957-11

Iding, M. H. ., Meester, W. J., & Tavasszy, L. A. (2002). Freight trip generation by firms. *42nd European Congress of the Regional Science Association*, 14. http://www.raumplanung.tu-dortmund.de/rwp/ersa2002/cd-rom/papers/453.pdf

Jaller, M., Sánchez-Díaz, I., & Holguín-Veras, J. (2015). Identifying freight intermediaries: Implications for modeling of freight trip generation. *Transportation Research Record*, *2478*, 48–56. https://doi.org/10.3141/2478-06

Johnston, R. J. (2007). Choice experiments, site similarity and benefits transfer. *Environmental and Resource Economics*, *38*(3), 331–351. https://doi.org/10.1007/s10640-006-9073-4

Kamboj, P., Malyan, A., Kaur, H., Jain, H., & Chaturvedi, V. (2022). *India Transport Energy Outlook*.

Karasmaa, N. (2007). Evaluation of transfer methods for spatial travel demand models. *Transportation Research Part A: Policy and Practice*, *41*(5), 411–427. https://doi.org/10.1016/j.tra.2006.09.009

Kawamoto, E. (2003). Transferability of standardized regression model applied to person-based trip generation. *Transportation Planning and Technology*, *26*(4), 331–359.

https://doi.org/10.1080/03081060310001635896

Kemperman, A., & Timmermans, H. (2009). Influences of Built Environment on Walking and Cycling by Latent Segments of Aging Population. *Transportation Research Record*, *2134*(1), 1–9. https://doi.org/10.3141/2134-01

Klodzinski, J., & Al-Deek, H. M. (2003). Transferability of an Intermodal Freight Transportation Forecasting Model to Major Florida Seaports. *Transportation Research Record*, *1820*, 36–45. https://doi.org/10.3141/1820-05

Koppelman, F. S., & Wilmot, C. G. (1982). Transferability analysis of disaggregate choice models. *Transportation Research Record*, *895*, 18–24. https://doi.org/10.1016/S0065-2113(08)60255-2

Kriegel, H. P., Kröger, P., Sander, J., & Zimek, A. (2011). Density-based clustering. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *1*(3), 231–240. https://doi.org/10.1002/widm.30

Kulpa, T. (2014). Freight Truck Trip Generation Modelling at Regional Level. *Procedia - Social and Behavioral Sciences*, *111*, 197–202. https://doi.org/10.1016/j.sbspro.2014.01.052

Kuo, C.-W., & Tang, M.-L. (2011). Survey and empirical evaluation of nonhomogeneous arrival process models with taxi data. *Journal of Advanced Transportation*, *47*(June 2010), 512–525. https://doi.org/10.1002/atr

Kuzmyak, R. (2008). NCHRP Synthesis 384: Forecasting Metropolitan Commercial and Freight Travel. In *National Academies of Sciences, Engineering, and Medicine*. The National Academies Press. https://doi.org/10.17226/22069

Lawson, C. T., Holguín-Veras, J., Sánchez-Díaz, I., Jaller, M., Campbell, S., & Powers, E. L. (2012). Estimated generation of freight trips based on land use. *Transportation Research Record*, *2269*, 65–72. https://doi.org/10.3141/2269-08

Lindsey, C., Mahmassani, H. S., Mullarkey, M., Nash, T., & Rothberg, S. (2014). Industrial space demand and freight transportation activity: Exploring the connection. *Journal of Transport Geography*, *37*, 93–101. https://doi.org/10.1016/j.jtrangeo.2014.04.009

McArthur, D. P., Kleppe, G., Thorsen, I., & Ubøe, J. (2011). The spatial transferability of parameters in a gravity model of commuting flows. *Journal of Transport Geography*, *19*(4), 596–605. https://doi.org/10.1016/j.jtrangeo.2010.06.014

Mccabe, S., Kwan, H., & Roorda, M. J. (2013). Comparing GPS and Non-GPS Survey Methods for Collecting Urban Goods and Service Movements. *International Journal of Transport Economics*, *40*(2).

Mendenhall, W., & Sincich, T. (2012). *A Second Course in Statistics: Regression Analysis* (7th ed.). Prentice Hall.

Miodonski, D., & Kawamura, K. (2012). Effects of Built Environment on Freight Consumption. *Procedia - Social and Behavioral Sciences*, *39*, 74–88. https://doi.org/10.1016/j.sbspro.2012.03.092

Mohapatra, S. S., Pani, A., & Sahu, P. K. (2021). Examining the Impacts of Logistics Sprawl on Freight Transportation in Indian Cities: Implications for Planning and Sustainable Development. *Journal of Urban Planning and Development*, *147*(4), 04021050. https://doi.org/10.1061/(asce)up.1943-5444.0000745

Mostafa, M. M. (2004). Forecasting the Suez Canal traffic: A neural network analysis. *Maritime Policy and Management*, *31*(2), 139–156. https://doi.org/10.1080/0308883032000174463

National Transport Development Policy Committee. (2014). *India Transport Report, Vol. 1: Executive Summary*.

Novak, D. C., Hodgdon, C., Guo, F., & Aultman-Hall, L. (2011). Nationwide Freight Generation Models: A Spatial Regression Approach. *Networks and Spatial Economics*, *11*(1), 23–41. https://doi.org/10.1007/s11067-008-9079-2

Nowrouzian, R., & Srinivasan, S. (2013). Modeling the effect of land use on person miles traveled by using geographically weighted regression. *Transportation Research Record*, *2397*, 108–116. https://doi.org/10.3141/2397-13

Office of the Registrar General & Census Commissioner. (2011). *Population finder*. Ministry of Home Affaris, GoI. https://censusindia.gov.in/census.website/data/population-finder

Oka, H., Hagino, Y., Kenmochi, T., Tani, R., Nishi, R., Endo, K., & Fukuda, D. (2019). Predicting travel pattern changes of freight trucks in the Tokyo Metropolitan area based on the latest large-scale urban freight survey and route choice modeling. *Transportation Research Part E: Logistics and Transportation Review*, *129*(October 2017), 305–324. https://doi.org/10.1016/j.tre.2017.12.011

Ortúzar, J. de D., & Willumsen, L. G. (2011). *Modelling Transport* (4th ed.). Wiley.

https://doi.org/10.1002/9781119993308

Paea, S., & Baird, R. (2018). Information Architecture (IA): Using Multidimensional Scaling (MDS) and K-Means Clustering Algorithm for Analysis of Card Sorting Data. *Journal of Usability Studies*, *13*(3), 138–157.

Pani, A., Bhat, F. A., & Sahu, P. K. (2020). Effects of Business Age and Size on Freight Demand: Decomposition Analysis of Indian Establishments. *Transportation Research Record*, *2674*(2), 112–126. https://doi.org/10.1177/0361198120902432

Pani, A., Mishra, S., & Sahu, P. (2022). Developing multi-vehicle freight trip generation models quantifying the relationship between logistics outsourcing and insourcing decisions. *Transportation Research Part E: Logistics and Transportation Review*, *159*(August 2020), 102632. https://doi.org/10.1016/j.tre.2022.102632

Pani, A., & Sahu, P. K. (2019a). Comparative Assessment of Industrial Classification Systems for Modeling Freight Production and Freight Trip Production. *Transportation Research Record*, *2673*(3), 210–224. https://doi.org/10.1177/0361198119834300

Pani, A., & Sahu, P. K. (2019b). Planning, designing and conducting establishment-based freight surveys: A synthesis of the literature, case-study examples and recommendations for best practices in future surveys. *Transport Policy*, *78*(April), 58–75. https://doi.org/10.1016/j.tranpol.2019.04.006

Pani, A., & Sahu, P. K. (2022). Modelling non-response in establishment-based freight surveys: A sampling tool for statewide freight data collection in middle-income countries. *Transport Policy*, *124*(August), 128–138. https://doi.org/10.1016/j.tranpol.2019.10.011

Pani, A., Sahu, P. K., & Bhat, F. A. (2021). Assessing the Spatial Transferability of Freight (Trip) Generation Models across and within States of India: Empirical Evidence and Implications for Benefit Transfer. *Networks and Spatial Economics*, *21*(2), 465–493. https://doi.org/10.1007/s11067-021-09530-z

Pani, A., Sahu, P. K., & Holguín-Veras, J. (2021). Examining the determinants of freight transport emissions using a fleet segmentation approach. *Transportation Research Part D: Transport and Environment*, *92*(February). https://doi.org/10.1016/j.trd.2021.102726

Pani, A., Sahu, P. K., & Majumdar, B. B. (2019). Expenditure-based segmentation of freight travel markets: Identifying the determinants of freight transport expenditure for developing marketing strategies. *Research in Transportation Business and Management*, *33*(October 2019), 100437. https://doi.org/10.1016/j.rtbm.2020.100437

Pani, A., Sahu, P. K., Patil, G. R., & Sarkar, A. K. (2018). Modelling urban freight generation: A case study of seven cities in Kerala, India. *Transport Policy*, *69*(May 2018), 49–64. https://doi.org/10.1016/j.tranpol.2018.05.013

Patil, G. R., & Sahu, P. K. (2017). Simultaneous dynamic demand estimation models for major seaports in India. *Transportation Letters*, *9*(3), 141–151. https://doi.org/10.1080/19427867.2016.1203582

Puente-Mejia, B., Palacios-Argüello, L., Suárez-Núñez, C., & Gonzalez-Feliu, J. (2020). Freight trip generation modeling and data collection processes in Latin American cities. Modeling framework for Quito and generalization issues. *Transportation Research Part A: Policy and Practice*, *132*, 226–241. https://doi.org/10.1016/j.tra.2019.10.013

Rose, G., & Koppelman, F. S. (1984). Transferability of disaggregate trip generation models. *Proceedings of 9th International Symposium on Transportation and Traffic Theory*, 471–491.

Sahu, P. K., Chandra, A., & Pani, A. (2022). Examining the Impacts of Highway Proximity on Freight Flow Patterns of Establishments. *Transportation Research Record*. https://doi.org/10.1177/03611981221133081

Sahu, P. K., Chandra, A., Pani, A., Majumdar, B. B., & Bandhu, B. (2020). Designing freight traffic analysis zones for metropolitan areas : identification of optimal scale for macro-level freight travel analysis. *Transportation Planning and Technology*, *0*(0), 1–18. https://doi.org/10.1080/03081060.2020.1780711

Sahu, P. K., & Pani, A. (2020). Freight generation and geographical effects: modelling freight needs of establishments in developing economies and analyzing their geographical disparities. *Transportation*, *47*(6), 2873–2902. https://doi.org/10.1007/s11116-019-09995-5

Samimi, A., Mohammadian, A., & Kawamura, K. (2013). A nationwide web-based freight data collection. *Canadian Journal of Civil Engineering*, *40*(2), 114–120. https://doi.org/10.1139/cjce-2011-0555

Sánchez-Díaz, I. (2017). Modeling urban freight generation: A study of commercial establishments' freight needs. *Transportation Research Part A: Policy and Practice*, 3–17. https://doi.org/10.1016/j.tra.2016.06.035

Sánchez-Díaz, I. (2018). Potential of Implementing Urban Freight Strategies in the Accommodation and Food Services Sector. *Transportation Research Record*, *2672*(9), 194–203. https://doi.org/10.1177/0361198118796926

Sánchez-Díaz, I., Holguín-Veras, J., & Wang, X. (2016). An exploratory analysis of spatial effects on freight trip attraction. *Transportation*, *43*(1), 177–196. https://doi.org/10.1007/s11116-014-9570-1

Santoso, D. S., & Tsunokawa, K. (2005). Spatial transferability and updating analysis of mode choice models in developing countries. *Transportation Planning and Technology*, *28*(5), 341–358. https://doi.org/10.1080/03081060500319694

Schiffer, R. G. (2012). *NCFRP Report 735: Long-Distance and Rural Travel Transferable Parameters for Statewide Travel Forecasting Models*. The National Academies Press. https://doi.org/10.17226/22661

Sikder, S., Augustin, B., Pinjari, A. R., & Eluru, N. (2014). Spatial transferability of tour-based time-of-day choice models: Empirical assessment. *Transportation Research Record*, *2429*(2429), 99–109. https://doi.org/10.3141/2429-11

Sikder, S., & Pinjari, A. R. (2013). Spatial transferability of person-level daily activity generation and time use models. *Transportation Research Record*, *2343*, 95–104. https://doi.org/10.3141/2343-12

Sikder, S., Pinjari, A. R., Srinivasan, S., & Nowrouzian, R. (2013). Spatial transferability of travel forecasting models: a review and synthesis. *International Journal of Advances in Engineering Sciences and Applied Mathematics*, *5*(2–3), 104–128. https://doi.org/10.1007/s12572-013-0090-6

Tang, L., Xiong, C., & Zhang, L. (2018). Spatial Transferability of Neural Network Models in Travel Demand Modeling. *Journal of Computing in Civil Engineering*, *32*(3), 04018010. https://doi.org/10.1061/(asce)cp.1943-5487.0000752

Taniguchi, E., Thompson, R. G., & Yamada, T. (2016). New Opportunities and Challenges for City Logistics. *Transportation Research Procedia*, *12*(June 2015), 5–13. https://doi.org/10.1016/j.trpro.2016.02.004

The World Bank Group. (2022). *World Bank Open Data*. https://data.worldbank.org/

Thompson, S. K. (2012). *Sampling*. John Wiley & Sons.

Tian, G., Park, K., & Ewing, R. (2019). Trip and parking generation rates for different housing types: Effects of compact development. *Urban Studies*, *56*(8), 1554–1575. https://doi.org/10.1177/0042098018770075

Toilier, F., Gardrat, M., Routhier, J. L., & Bonnafous, A. (2018). Freight transport modelling in urban areas: The French case of the FRETURB model. *Case Studies on Transport Policy*, *6*(4), 753–764. https://doi.org/10.1016/j.cstp.2018.09.009

Toilier, F., Serouge, M., Routhier, J. L., Patier, D., & Gardrat, M. (2016). How can Urban Goods Movements be Surveyed in a Megacity? the Case of the Paris Region. *Transportation Research Procedia*, *12*(June 2015), 570–583. https://doi.org/10.1016/j.trpro.2016.02.012

Wafa, Z., Bhat, C. R., Pendyala, R. M., & Garikapati, V. M. (2015). Latent-segmentation-based approach to investigating spatial transferability of activity-travel models. *Transportation Research Record*, *2493*, 136–144. https://doi.org/10.3141/2493-15

Wilmot, C. G. (1995). Evidence on transferability of trip-generation models. *Journal of Transportation Engineering*, *121*(5), 405–410. https://doi.org/10.1061/(ASCE)0733-947X(1995)121:5(405)

Xie, C., Lu, J., & Parkany, E. (2003). Work Travel Mode Choice Modeling with Data Mining: Decision Trees and Neural Networks. *Transportation Research Record*, *1854*, 50–61. https://doi.org/10.3141/1854-06

Zhang, Y., & Xie, Y. (2008). Travel mode choice modeling with support vector machines. *Transportation Research Record*, *2076*, 141–150. https://doi.org/10.3141/2076-16

List of Publications

Refereed Journal Publications

1. **Balla, B. S.**, Sahu, P. **(2023)**. *Assessing Regional Transferability and Updating of Freight Generation Models to Reduce Sample Size Requirements in National Freight Data Collection Program*. **Transportation Research Part A: Policy and Practice**, 175 (103780). https://doi.org/10.1016/j.tra.2023.103780

2. **Balla, B. S.**, Sahu, P. **(2023)**. *Examining the Effects of Local Sample Sizes on Spatial Transferability of Freight Production Models*. **Transportation Research Record**. https://doi.org/10.1177/03611981231197649

3. **Balla, B. S.**, Sahu, P., Pani, A., Sharma, S., & Majumdar, B. B. **(2023)**. Comparison of Parametric and Non-Parametric Methods for Modeling Establishment-Level Freight Generation. **Transportation Research Record**, 2677(2), 154–172. https://doi.org/10.1177/03611981221116369

4. **Balla, B. S.**, Sahu, P., & Pani, A. **(2021)**. *Are Freight Production Models Transferable between Urban and Suburban Areas? Guiding Model Transfer in Geographically Sprawling Indian Cities*. **Journal of The Institution of Engineers (India): Series A**, 102(3), 643–656. https://doi.org/10.1007/s40030-021-00556-7

Conference Presentations

1. **Balla, B. S.**, Sahu, P. K., and Pani, A. (2024). *Framework for Reducing Freight Survey Resources Through the Use of Spatial Transferability*. 103rd Annual Meeting, Transportation Research Board, Washington DC, USA, January.

2. **Balla, B. S.**, and Sahu, P. K. (2023). *Comparison of Correction Methods in Spatial Transferability of Freight Production Models*. 7th Conference of the Transportation Research Group of India, Surat, India, December.

3. **Balla, B. S.**, and Sahu, P. K. (2023). *Transferability Approach for Freight Data Measurement and Model Development in Resource-Constrained Regions*. 7th Conference of the Transportation Research Group of India, Surat, India, December.

4. **Balla, B. S.**, Sahu, P. K., and Pani, A. (2023). *Enhancing Spatial Transferability of Freight Generation Models using an Adaptive Neural Network to Optimise Freight Survey Resources*. 5th VREF conference on Urban Freight, Gothenburg, Sweden, October.

5. **Balla, B. S. (2023)**. *Optimisation of Resources in Pre-planning Stage of Freight Data Collection using Model Transferability Approach*. Presented in Freight and Marine Young Members Council (YMC-FM) session at **102nd Annual Meeting, Transportation Research Board**, Washington DC, USA, January.

6. **Balla, B. S.**, Pani, A., and Sahu, P. **(2022)**. *Enhancing the Transferability Accuracy of Urban Freight Demand Models using a Novel Establishment Typology*. **15th Urban Mobility India Conference & Expo**, Kochi, India, November.

7. **Balla, B. S.**, Pani, A., and Sahu, P. **(2021)**. *Spatial Transferability Assessment of Establishment Level Freight Attraction Models: A Case Study of Indian Cities*. **6th Conference of the Transportation Research Group of India**, Tiruchirappalli, India, December.

8. **Balla, B. S.**, Sahu, P., Pani, A., & Arvind, K. **(2021)**. *Spatial Transferability Assessment of Freight Production Models: Guiding Model Transfer between Urban and Suburban Areas in Geographically Sprawling Indian Cities*. **100th Annual Meeting, Transportation Research Board**, Washington DC, USA, January.

Brief Biography of the Candidate

Mr. Balla Bhavani Shankar is a PhD student at Birla Institute of Technology and Science Pilani (BITS Pilani), Hyderabad Campus, focusing on Freight Transportation Planning. His research in the field of freight can be encapsulated by the three interconnected aspects known as the "MMM": "measurement" of establishment-level freight data, "modelling" freight demand, and "model transfer" concerning spatial aspects. In essence, his primary focus is on developing cost-efficient freight demand models by utilising data collected from various Indian cities, leveraging the concept of spatial transferability. This approach allows for the efficient use of previously estimated model parameters in new application contexts, thereby minimising the need for extensive data collection efforts. He published his doctoral research in Transportation Research Part A: Policy and Practice and Transportation Research Record. He co-authored a paper with Prof. Prasanta Kumar Sahu, “Assessing regional transferability and updating of freight generation models to reduce sample size requirements in national freight data collection program", published in Transportation Research Part A: Policy and Practice. In this paper, a novel transferability-based framework was proposed to minimise the efforts of freight data collection. However, the applications of this framework are not limited to freight transportation planning; it has multidisciplinary applications.

Mr. Balla Bhavani Shankar
Research Scholar
Department of Civil Engineering
BITS Pilani, Hyderabad Campus
www.bsballa.in

Brief Biography of the Supervisor

Prof. Prasanta Kumar Sahu is an Associate Professor in the Department of Civil Engineering, Birla Institute of Technology and Science Pilani (BITS Pilani) in Hyderabad, India. He holds a Ph.D. in Transportation Systems Engineering (TSE) from Indian Institute of Technology Bombay (IIT Bombay) and a Master's in TSE from Indian Institute of Technology Kanpur (IIT Kanpur). He has specialised in the application of rigorous quantitative methods to the study of freight travel demand and land-use planning. He authored or co-authored more than 85 refereed journal publications in the last eight years. The key underlying research interest in these publications has been related to the quantification of disaggregate-level freight demand from establishments and ports. He also focuses on green mobility planning, quality of life, and demand management schemes such as congestion pricing and transport emissions. He is an Adjunct Professor at University of Manitoba, Canada. He is a visiting faculty at University of Regina, Regina, Canada and Cardiff University, Cardiff, Wales, UK. He holds many prestigious positions – Handling Editor, Transportation Research Record (TRR), Editorial Board Member – Transportation in Developing Economies (TiDE) and Standing Committee Member and Sub-editor (AT015) - Freight Transportation Planning and Logistics, Transportation Research Board (TRB), USA. Currently, he is the Joint Principal Investigator for Technologies for Urban Transit to Enhance Mobility and Safe Accessibility (TUTEM), a project sponsored by the Asian Development Bank. In addition, he is leading an associate research center (ARC) of Center of Excellence on Sustainable Urban Freight Systems (CoE-SUFS) at BITS Pilani. Through ARC, he is working towards fostering sustainability efforts in urban freight transport planning and operations.

Prof. Prasanta K. Sahu
Associate Professor
Department of Civil Engineering
BITS Pilani, Hyderabad Campus
www.prasantsahu.com