**Strategic Decisions in Agent-Based Freight Transportation Models:**

**Methods and Data**

BY

MONIQUE ANNE STINSON
B.S., University of Texas at Austin, 2000
M.S., University of Texas at Austin, 2002

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Civil Engineering
in the Graduate College of the
University of Illinois at Chicago, 2020

Chicago, Illinois

Defense Committee:

Abolfazl (Kouros) Mohammadian, Chair and Advisor
Jane Lin
Bo Zou
Kazuya Kawamura, Urban Planning and Policy
Anthony Pagano, Center for Supply Chain Management and Logistics

## Acknowledgements

Author's unpublished research prepared at another institution, Chapter 3: Stinson, M. *A Proposed Agent-Based Freight Modeling Framework for SimMobility's Mid-Term and Long-Term,* working document (2016/09/05). In Chapter 3, I incorporate most of this working paper, which I conceived of and wrote while studying at the Massachusetts Institute of Technology under the direction of Professor Moshe Ben-Akiva. I am grateful to Professor Ben-Akiva, Professor Vittorio Marzano, Dr. Carlos Azevedo, Dr. Diem-Trinh Le and Dr. Andre Ahlo for their feedback on this document.

Published conference paper, Chapter 3.3: M. Stinson, J. Auld, and A. (Kouros) Mohammadian, "A large-scale, agent-based simulation of metropolitan freight movements with passenger and freight market interactions," *Procedia Computer Science*, vol. 170, pp. 771–778, 2020, doi: 10.1016/j.procs.2020.03.157. © 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/). Individual contributions to the paper as follows: study conception and design: M. Stinson; methodology development: M. Stinson, K. Mohammadian; draft manuscript preparation: M. Stinson; review: all authors.

Conference presentation, Chapter 3: M. Stinson, J. Auld, and A. (Kouros) Mohammadian, A. Rousseau, "A Behavioral Framework To Model The Movement Of Commodity Flows Through Supply Chains", Presented at the 3rd VREF Conference on Urban Freight, October 17-19, 2018, Gothenburg, Sweden. Individual contributions: study conception and design: M. Stinson; methodology development: M. Stinson, K. Mohammadian; draft manuscript preparation: M. Stinson; review: all authors.

Published conference paper: Chapter 4 M. Stinson and A. (Kouros) Mohammadian, "Modeling Firm Transportation Strategy using Big Text Data," to be presented at and published in the proceedings of the Forum on Integrated and Sustainable Transportation System (FISTS), Delft, The Netherlands, Nov. 3-5, 2020. Individual contributions: study conception and design: M. Stinson; data collection: M. Stinson; analysis and interpretation of results: M. Stinson, K. Mohammadian; draft manuscript preparation: M. Stinson; review: all authors.

Conference presentation, Chapter 6:  M. Stinson, Jane Lin, and A. (Kouros) Mohammadian, "Linking Business Strategy To Transportation Energy Consumption And Emissions Estimates: Effects Of Automobile Assembly Location Strategies", presented at the Transportation Research Board Annual Meeting, January 13–17, 2019, Washington, D.C.  Individual contributions: study conception and design: M. Stinson, J. Lin, K. Mohammadian; data collection: M. Stinson, J. Lin; analysis and interpretation of results: M. Stinson, J. Lin, K. Mohammadian; draft manuscript preparation: M. Stinson; review: all authors.

Financial support

The author is grateful to the University of Illinois at Chicago for waiving tuition during her PhD studies there. The author is grateful to the Massachusetts Institute of Technology for waiving tuition, and to the Federal Highway Administration for funding a research assistantship, during her PhD studies there. The author would also like to thank Argonne National Laboratory for assistance with educational expenses, and more broadly for supporting the POLARIS modeling framework that is integrated with parts of this research.

Personal acknowledgements

Any successful PhD graduate owes an unmeasurable debt to her advisor in the PhD journey. I hold the deepest gratitude for my advisor, Professor Kouros Mohammadian, whose vast knowledge of transportation behavior and methods enriched my research. Kouros also taught me invaluable lessons in leadership and mentoring. I am grateful to my initial PhD advisor, Professor Moshe Ben-Akiva of the Massachusetts Institute of Technology, who helped me connect my theoretical notions to technical methodology development. If super-commuting had been more bearable, I surely would have stayed. Finally, I thank each member of my dissertation committee for generously sharing expertise and stimulating improvements in this work.

In some ways, this thesis was largely inspired by two excellent colleagues, Mike Fischer (now retired) and Dan Beagan of Cambridge Systematics. Through working with Mike and Dan, freight quickly became my top research interest. I am sincerely thankful for their expertise and camaraderie. I am grateful for additional mentorship from Chandra Bhat, Hani Mahmassani, Pat Mokhtarian, Ann Schlenker and Aymeric Rousseau, and earlier Kara Kockelman, Ellen Rathje, Susan Handy and Wilma Slyoff. I learned a lot from innumerable other talented colleagues and friends throughout my career.

I am particularly thankful to Kermit Wies, formerly of the Chicago Metropolitan Agency for Planning, Vidya Mysore and Birat Pandey of the Federal Highway Administration, Vladimir Livshits of the Maricopa Association of Governments, and David Anderson, Prasad Gupte, Erin Boyd and Heather Croteau of the US Department of Energy, for funding my research in advanced, agent-based freight modeling.

Numerous individuals were personally critical to my PhD journey. My parents prioritized education and gifted me with a profound work ethic. With these lessons, I reached for my goal. I am deeply grateful to Cindy, Rob, Jean, Kelley, Scott, Laurie, Pixie, Skip, Rachel and Danny for graciously sharing wisdom, encouragement, and support throughout. Friends at Pilgrim Lutheran Church and School provided a spiritual home and community, plus many volunteer opportunities that gave me balance. Isis and Frances, you make my life a blessing every day. I am grateful for support and understanding from Jenny. Allison and my Monday dinner group helped me regain focus. Accountability check-ins from Laurie pulled me out of a pandemic slump. Bilge, Katie and Eunice provided kindness and aid. My MIT classmates looked out for me – Mazen, Yinjin, Isabel, Haizheng, Na, Rounaq, Ajinkya, Adam, Adam, Mike, and many others – my only regret is that I did not have more time to spend as a friend. I am thankful to the support and kindness of my UIC colleagues, including Ramin, Ali, Ehsan, Noosheen, Amir Bodahar, Homa, MJ, Ahmad, Diana and Alicia – you have been wonderful classmates.

MAS

# Table of Contents

List of Tables

List of Figures

## Summary

Freight transportation has significant impacts on transportation-related energy consumption and emissions. It also plays a vital economic role by enabling trade, bringing goods to consumers and providing jobs. Because of the significance of goods movement, it is important to have quantitative tools that allow government agencies and researchers to study the impacts of freight transportation. Such tools should also be sensitive to the impacts of policies, infrastructure, supply chain trends, and economic scenarios, so that they can be used for scenario analysis as government agencies identify ways to maintain and improve the transportation system.

However, modeling freight transportation has many challenges. The business environment is complex, with significant heterogeneity among actors and their behaviors. Differences in industry sector, commodities, logistics needs, and scale of operations are just a few important dimensions that characterize an agent in this setting. Since freight crosses geographic and political boundaries, models likewise should consider national and global trade in their representation of both trade networks and transportation networks. Finally, mathematical models of freight transportation in the travel demand forecasting domain have had considerably less attention than passenger models. As a result, the ideal modeling system is not yet realized and existing frameworks, despite their achievements, have room for improvement. For example, existing frameworks use unaffiliated establishments instead of firms and lack behavioral models of fleet and distribution center ownership, which are fundamental to goods transport.

The core of this thesis is an innovative, theoretical foundation for freight transportation forecasting that employs the concept of firm strategy to inform agent behavior and unify upstream and downstream model components. This work implements several major steps toward meeting this vision. The thesis adopts the perspective that modeling the behavior of individual firms and establishments is necessary for adequately addressing the underlying heterogeneity and complexity of goods movement. The new modeling structure is developed for an agent-based modeling context, which is well suited to modeling the preferences and activities of individual agents. Agent partnerships, which are foundational to supply chains, are also represented intuitively in an agent-based approach.

The freight system architecture is designed to realistically capture key elements of firms and firm operations as they relate to transportation. Firm strategy is integral to the modeling system design since, in reality, strategy informs firm decision-making and aligns firm decisions to be consistent with its overall goals. But in addition to the conceptual architecture development, it is important to address practical questions regarding how to mathematically model strategy and its effects, and where to obtain data on firm strategy. Currently, there are no known quantitative data sources of firm strategy, and no known methods to estimate firm strategies and strategic behavior. The impact of strategy on transportation outcomes, particularly energy and emissions, is also not known. This thesis makes several major contributions in these areas.

Acknowledging the lack of data on firm strategy, the thesis develops two novel algorithms that generate strategy data. Each algorithm creates strategy measurements from natural language textual sources. The algorithms integrate a mix of older and newer methods, including the text mining bag-of-words method, Principal Components Analysis and the recent word2vec method from the Natural Language Processing field. My algorithms expand on these methods and create latent attitudinal (strategy) measurement data in a completely new way. These algorithms present an alternative to the standard method of using attitudinal surveys to collect this kind of data, and therefore have potential applications in passenger modeling as well.

The mathematical approach to modeling strategy is likewise an innovative contribution. First, I develop a theoretical model that illustrates the linkages between (latent) strategy, observed strategic decisions, and exogenous variables. Next, I formulate a mathematical system that operationalizes this conceptual model, integrating latent variable models into a Seemingly Unrelated Regression with unrestricted covariance. I then apply the model in a proof of concept to analyze firm choices in private fleet ownership and distribution center operations. My model jointly generates several types of useful outcomes—binary, continuous, and censored—and provides an elegant framework for modeling multiple discrete-continuous decisions and choice set generation parameters. These mathematical aspects play an important role in linking high-level strategy decisions to subsequent decisions in downstream areas of the model.

Lastly, the thesis demonstrates the link between strategy and two key transportation outcomes, energy and emissions, thereby addressing sustainability impacts of firm strategy. A multimodal, agent-based tool is developed and applied to in a case study of automobile manufacturers, examining baseline impacts and the impact of a strategic shift in production location.

Extensive proof of concept analyses demonstrate important results in several under-researched areas. Strategies of Fortune 500 firms in freight-intensive sectors are analyzed. Their strategic decisions regarding fleet and distribution center use are modeled. The decisions are shown to impact subsequent decisions, including the geographic structure of national, firm-level distribution centers and regional distribution center location choice decisions. The sustainability analysis demonstrates unique results for the automobile manufacturing industry.

While the modeling architecture and tools are developed for the freight context, they are expected to provide valuable examples for passenger modeling as well. For instance, activity-based passenger models, like freight, can benefit from the proposed modeling structure, which improves behavioral consistency between the agent's higher-level and lower-level decisions. The mathematical formulation that jointly links overarching, strategic (latent) preferences to multiple decisions will also be valuable to both passenger and freight domains. The novel setup for modeling choice set parameters and multiple discrete-continuous decisions is likewise expected to be useful regardless of domain. Finally, the novel methods for attitudinal data development offer many advantages over existing methods for collecting such data.

# 1 Introduction

Freight transportation, or goods movement, is an essential service that brings goods where they are needed. Freight transportation is also central to the economy and creates jobs as discussed in Chicago Metropolitan Agency for Planning (CMAP) (2012). However, goods movement generates notable negative impacts as well. According to the US Department of Transportation (2017), an estimated $27B in time and fuel is wasted annually in the US due to truck congestion while the International Energy Agency (2017) estimates that the ocean shipping industry emitted 866 million metric tons of carbon dioxide (CO2) equivalents in 2014, which is comparable to CO2 emissions for the entire county of Japan (1,142 MMT) or Germany (730 MMT) in 2015 (Union of Concerned Scientists, 2020). These factors motivate the need for policy analysis tools that can be applied to understand baseline impacts of goods movement, potential future impacts, and the anticipated success of policy interventions. Metropolitan planning organizations (MPOs) rely on travel demand forecasting models to estimate these impacts by quantifying commodity flow volumes, origins and destinations, and transportation paths adopted in various scenarios.

However, many factors make it challenging to evaluate freight impacts with transportation forecasting models. Transportation is rapidly changing today as new technologies emerge, including electrified vehicles, crowd-shipping, parcel lockers, and more. Freight transportation is especially affected by large-scale environmental factors such as tariffs and pandemic-induced supply and demand shocks. With the rise in e-commerce, distribution is inching ever closer to the end consumer. Moreover, the business population varies enormously, demonstrating substantial heterogeneity in observable ways: the types of goods produced, revenue, geography of operations, supply chain decisions, vehicle fleet and distribution channel decisions, and so on. Businesses rely more than ever on access to the latest information to make rapid, data-driven adjustments to everything from production to delivery, but this reliance differs depending on the location of the business in the supply chain.

Like individual persons, businesses also exhibit a variety of preferences, or strategies, that are unobservable and more challenging to quantify. To facilitate discussion throughout the remainder of this thesis, the following terms are defined:

- **Strategy**: a latent (unobservable) objective that (a) is adopted by an individual company, (b) lends consistency to company's decision-making processes, (c) can be represented mathematically as a latent variable, and (d) informs outcomes that are observable to and measurable by analysts
- **Strategic decision**: a manifest (observable) outcome of a decision that an individual company makes based on strategy and other factors

Strategy is critical for firm operations because it lends consistency to decisions ranging from major to minor, including basic but critically important decisions regarding which markets to operate in and whether to outsource transportation. As a result, strategy has a very real connection to goods movement.

For all of these reasons mentioned above, the ideal policy analysis tool will be capable of addressing both the complex business environment and business heterogeneity in observable and latent attributes. Agent-based models meet these requirements and are widely used in passenger transportation analysis. Briefly, the typical setup synthesizes a population of agents, assigns attributes to each agent, models agent activities for a 24-hour period, then simulates their collective movements throughout the transportation network over the day. The freight forecasting community began using these methods in the mid-2000s, with implementations to date in numerous regions worldwide. Despite this progress, existing modeling frameworks have several critical gaps (Shabani et al., 2018), demonstrating that the freight transportation community currently is far from realizing an ideal framework.

## 1.1 Significance of the Problem

The goal of travel demand forecasting is to generate credible estimates of transportation impacts. To achieve this goal, forecasting models must mathematically represent transportation activity with a sufficient level of realism. For example, the population of agent-based passenger transportation models is generally represented using both household agents and person agents. This representation acknowledges the fact that some decisions, such as time of departure from work, can be highly individualized whereas other decisions can have significant influence from other household members. Household location choice and vehicle ownership are two decisions that are typically modeled at the household level with this rationale. Moreover, based on my experience, certain decisions are universally regarded as essential to transportation models in the US. One example of a fundamental decision is vehicle ownership since, for instance, a person with a car will likely drive much more than a person without a car. It is unlikely that any agent-based passenger model that ignored households, household fleet decisions, or workplace locations would be considered credible.

However, by and large, extant approaches in behavioral-based freight modeling do not account for major, real-world features of businesses. For example, no existing framework models firms (related establishments), fleet ownership, and distribution center ownership. As a result, the validity of any model output that is, in reality, informed by these missing features is questionable. Specifically, these omissions likely compromise the validity of the following model output. Origin-destination flows are likely inaccurate because, for example, a firm may ship deliveries to all of its regional establishments from the same distribution center. Without acknowledging this relationship, such deliveries may appear to come from other, unrelated locations. In another example, a firm that owns a fleet will generate truck tours that serve its own establishments. In contrast, truck tours that are made by carriers are more likely to travel between unrelated establishments. Fleet management may be driven by different priorities, with in-house fleets, for instance, prioritizing customer service and reliability while carriers may prioritize cost savings. As a result, the operational characteristics (average daily miles traveled, fuel efficiency, and so on) may differ significantly between private fleets and for-hire carrier fleets. These are just a few examples of potential discrepancies between reality and simulation that can introduce major credibility concerns into the validity of freight transportation models and the estimates that they generate.

Moreover, none of the existing agent-based freight frameworks model firm strategy and its impact on decision making. This is a critical gap, as the literature demonstrates that strategy is an important determinant of firm behavior. In fact, strategy is a major driver of strategic behavior, which is characterized by major resource allocation decisions. For the business context, examples of such decisions are whether or not to own a private fleet and whether or not to invest in distribution centers. While these are binary (yes/no) decisions, each of these has continuous aspects as well. For instance, if distribution centers are a worthwhile investment, how much floor space should be operated? Details of fixed investments, such as location choice for a distribution center, also can be considered a strategic decision. Ultimately, firm strategy is a unifying feature that ties together major firm decisions. Omitting firm strategy causes agent decisions to be disjointed in simulation, whereas in reality they are aligned. As a result, the simulation may introduce firm actions that are incongruous. For instance, a firm that prioritizes customer satisfaction may operate large warehouses to ensure that customers never experience a shortage, whereas a firm that prioritizes lean supply chain operations may operate much smaller warehouses.

The fundamental issues in current freight models are serious enough to warrant a major revision of the modeling paradigm. Such an overhaul requires attention to the following details. First, the entire freight modeling system should be reconstructed to account for key details such as firms and strategies. Major transportation decisions, such as whether to own or outsource fleet and distribution center operations, should be simulated for the firm. Other high-impact phenomena in recent years, such as the rise of the information age and e-commerce, should be accounted for. Second, a theoretical framework should be designed and implemented to enable firms to behave consistently according to their strategies. Of course, there needs to be more than a theory: the mathematical features of a strategy model should be designed and worked out, accompanied by solution methods. There is also a question of data, as companies may not openly share information about their strategies. Data on company strategy needs to be found, and mathematical constructs for deploying them in freight forecasting models need to be developed.

Ultimately, empirical evidence regarding the real-world impact of strategy should be prepared to “make the case” for including it in transportation models. Good empirical examples will demonstrate a clear linkage between firm strategy and its impact on transportation decisions or externalities such as energy consumption and emissions.

## 1.2 Statement of Purpose

The central thesis of this work is that strategy unifies decisions that individual actors make, and that this consistency can be achieved in agent-based models through carefully constructed framework design. Strategy should be integrated into the model stream in a way that permits, or even enforces, consistency between high-level, long-term decisions and any number of other long-term, mid-term, or possibly short-term downstream decisions. This is a simple yet compelling idea that, quite remarkably, has not been previously introduced into the transportation modeling paradigm.

The purpose of this study is to develop theoretical constructs, mathematical methods, data sources, and proof of concept applications using this thesis as a guiding vision. Specifically, the purpose of Chapter 3 is to develop an entirely new conceptual design for an innovative, agent-based freight modeling framework. The chapter presents a theoretical framework that models strategy and structures it as a mechanism that enforces consistency among decisions that the agent makes elsewhere in the framework. In the process, this chapter also aims to improve upon the state of the art in agent-based freight modeling in multiple other ways. It develops novel theoretical approaches to include the effect of information on agent decisions, to operationalize the push-pull boundary, and to model interactions between freight agents and households. It also aims to remedy fundamental issues that are common among existing freight models.

Chapter 4 and Chapter 5 form the core methodological chapters of this thesis. Their objectives are to develop methods for creating attitudinal (strategy) measurement data and for jointly modeling strategy and strategic decisions, respectively. This thesis treats strategy as a latent attitude that can be measured with indicators, thereby expressing qualitative notions with quantitative measurements. The purpose of Chapter 4 is to develop methods for creating strategy measurements using natural language textual data sources. The desired output is a dataset of measurements than can subsequently be input to an agent-based freight forecasting system, or in a stand-alone behavioral model. The purpose of Chapter 5 is to first develop a theoretical framework for the joint modeling of latent strategies and manifest strategic decisions, then to develop a mathematical approach, including a solution method, that estimates the parameters that are specified in the system.

The objective of Chapter 6 is twofold. First, the chapter presents a methodology for estimating the global energy and emissions impacts that are attributable to a change in firm strategy. Second, this chapter applies this method in order to conclusively demonstrate the significance of this work, showing the impacts of company strategy on transportation energy and emissions in a real-world case study.

## 1.3 Theoretical Basis of Study

Figure 1 illustrates the structure of this thesis and how each of eight key theoretical areas, marked A through H, constitute its foundation. A: Xu et al. (2003) and Caplice (2006) present a theoretical framework for freight modeling consisting of three layers. Strategic or economic decisions are in the highest layer, the middle layer includes tactical decisions or logistics (distribution-related) decisions, while the lowest layer is operational or transport decisions. B: Shortly afterwards, the Aggregate-Disaggregate-Aggregate (ADA) is formulated (de Jong & Ben-Akiva, 2007 and Ben-Akiva and de Jong, 2008) and implemented in Sweden and Norway. The ADA model is the first disaggregate implementation to convert commodity flows to shipments that move through supply chains. Samimi et al. (2013) and Urban et al. (2012) are the earliest efforts to implement the ADA in the US and extend it for the US context. However, as the literature review discusses, while the ADA framework is a ground-breaking innovation with many strengths, it also has several important areas that warrant improvement or replacement.

C: Leveraging the three-layered concept and the early ADA efforts, Stinson (2016) develops a new conceptual construct to improve upon the ADA implementation. Major improvements include using firms instead of establishments, operationalizing the push-pull boundary, and including the effect of information. The resulting framework forms the basis of Chapter 3. The integration of the features in A through C is shown more clearly in this thesis, and a proof of concept including many of the proposed features is demonstrated. But critically, this thesis extends the work of Stinson (2016) by including a strategy model and demonstrating the features that operationalize it as a unifying link with other agent decisions.

D: Earlier theories in strategic behavior help establish the basis for this work. Shapiro (1989) and Teece (2019) provide the foundational theoretical notions of (a) what firm strategy is and (b) how it relates pragmatically to consistency in observable decisions that the firm makes. Their theoretical presentations illustrate that firm strategy is a guiding policy or set of policies that add consistency to other firm decisions, such as asset investments. In other words, these theories show that latent strategies inform manifest strategic decisions, which is a central tenet of this thesis. However, these works are entirely theoretical and do not establish a modeling architecture or provide mathematical or data details regarding implementation. Choo and Mokhtarian (2018) and Ben-Akiva (2010), on the other hand, show that higher-level, advance planning or strategies inform actions for the individual in the transportation context specifically, but do not show how to apply this viewpoint to a set of decisions that an agent makes. These works are foundational to the entire thesis, but especially to Chapter 3 and Chapter 5.

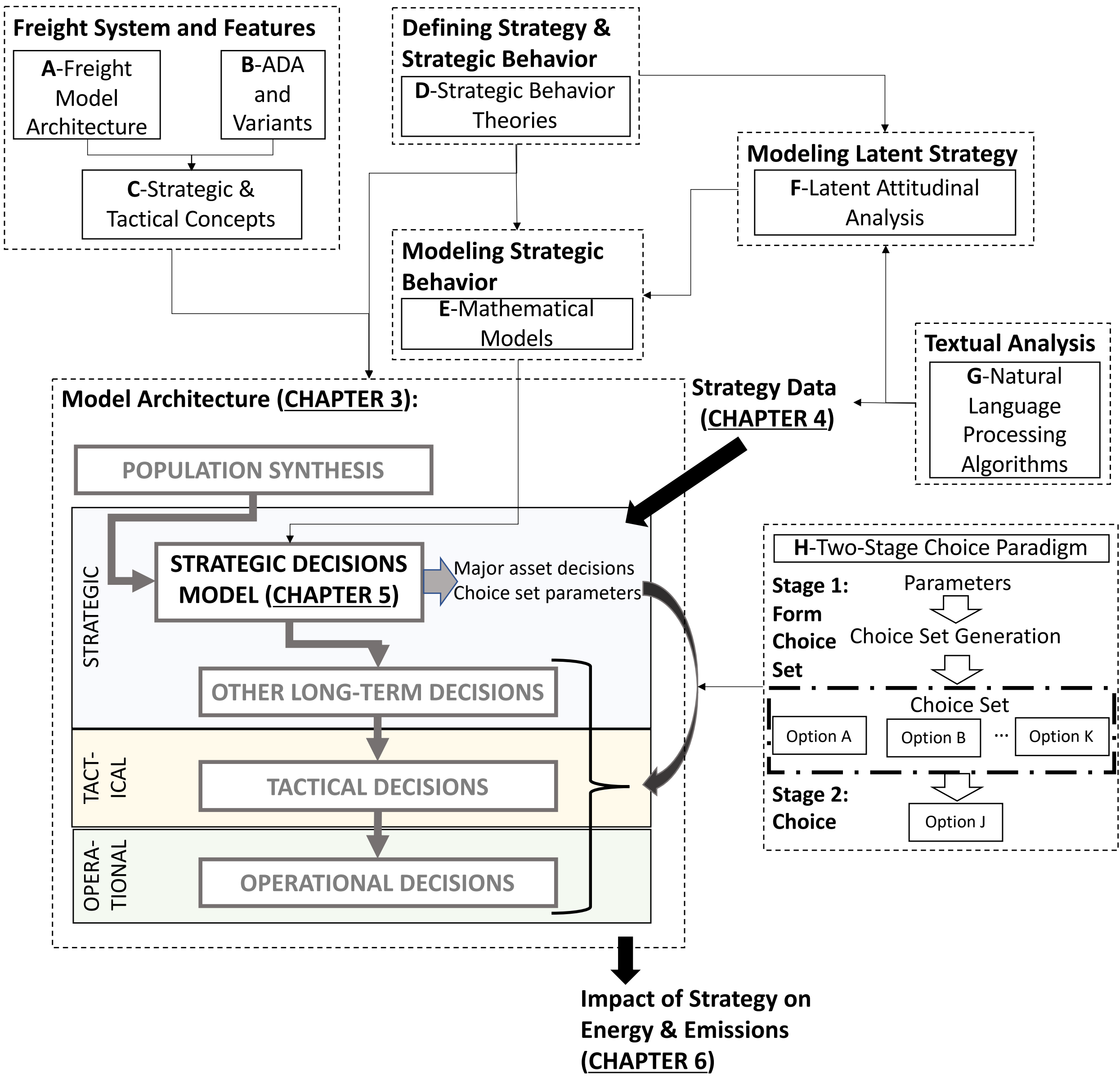



**Figure 1. Theoretical basis of this study.**

E: Mathematical and data development aspects are critical for developing a source of strategy data and estimating models of strategy and strategic decisions in Chapter 5. Fang (2008) shows how to develop models that are extremely useful for modeling strategic decisions, which often have a yes/no decision accompanied by one or more ordinal or continuous decisions that is conditional on a “yes” decision. The system is implemented as a Seemingly Unrelated Regression (SUR), with which correlations between decisions can be modeled readily. The latter feature is desirable for modeling strategic decisions, which are often related. The system is solved using Gibbs Sampling, originally developed in Albert and Chib (1993). However, the work does not include latent variables, which this thesis uses as the mechanism for representing strategy. Daziano (2015) shows how to include latent variables in choice models using Gibbs Sampling, but does not demonstrate this for SUR models. This thesis leverages these works to develop a new mathematical model and solution method that has many advantages for analyzing strategy and strategic decisions.

F: Building on the underlying theories of Shapiro (1989) and Teece (2019), this thesis assumes that strategies are unobservable and can be treated as latent variables. Bollen (1989) summarizes the rich history of development in latent variable modeling, which emerged primarily from psychology and economics. Ben-Akiva et al. (2002) and Walker (2001) summarize the breakthrough application of latent variables in transportation behavioral modeling.

G: Mikolov et al. (2013) develops a novel method called word2vec, which is based on recent developments in Natural Language Processing (NLP), to represent words as vectors in some N-dimensional space. Each dimension of the space represents a unique concept. The authors show how the results naturally embed information that provides a way to measure differences in words. The algorithm generates a single vector for each word, which represents it average contextual use. The algorithm, along with Principal Component Analysis (PCA) and text mining, are key inputs to the algorithms that Chapter 4 proposes.

H: Finally, a powerful innovation in my strategy modeling framework is its utilization as a new method for developing choice set generation parameters. The two-stage choice process is first described in Manski (1977). The theoretical framework is designed with the notion that most individuals (or companies) cannot recognize the universal choice set in many instances – for example, when choosing a residential or factory location. So, the theory goes, the individual or company creates a “short list” based on parameters (e.g., minimum building size or proximity to an airport). Although this theory has existed for over five decades, it remains a challenging and active area of research.

## 1.4 Research Questions and Hypotheses

This thesis investigates the following research questions and poses a hypothesis for each:

**Question**: Assuming strategies are important, how can they be integrated in an agent-based framework to lend consistency to other agent decisions?

*Hypothesis*: A strategy model can be placed in the agent-based modeling architecture immediately after population synthesis. Outcomes from this model can then feed into downstream components. Chapter 3 and Chapter 5 present the conceptual modeling architectures that operationalize this.

**Question**: How can strategy be measured?

*Hypothesis*: Strategy can be measured using text-based sources of attitudinal data. Attitudinal surveys are one approach to measure strategy. As an alternative, Chapter 4 develops two innovative, natural language-based approaches for this purpose.

**Question**: How can natural language sources be used to quantify firm strategy or strategic decisions?

*Hypotheses*: Firms use certain words (a) more or less frequently (on a relative basis), or (b) in different ways, depending on their strategies or strategic decisions. Chapter 4 and Chapter 5 use visualization examination, statistical tests, factor analysis, and mathematical choice models to test hypotheses (a) and (b).

**Question**: What is the relationship between strategy, a latent or unseen preference of the firm, and manifest or observable strategic decisions that firms make?

*Hypothesis*: Strategy influences strategic decisions. The theoretical basis discussed above determines the directionality. The statistical strength of the relationship is tested in Chapter 5.

**Question**: What, if any, issues related to jointness or simultaneity in decision making are important for modeling strategy?

*Hypothesis*: Firms can jointly determine a strategy set and a set of strategic decisions to adopt. Chapter 5 provides a theoretical framework for this hypothesis and performs empirical tests using real-world data.

**Question**: How can the agent-based architecture of this thesis model choice set parameter generation?

*Hypothesis*: Choice set parameter generation can be treated as a strategic decision, as Chapter 3 and Chapter 5 explore.

**Question**: Do strategy and strategic decisions impact two critical end outcomes of transportation: transportation energy use and emissions (TECE)?

*Hypothesis*: Yes, Chapter 6 shows conclusively in a real-world case study that strategy can have substantial impacts on TECE. Chapter 6 also develops a method to measure TECE for global, multimodal flows.

**Question**: How does strategy impact strategic decisions regarding logistics control (private fleet ownership and distribution center control)?

*Hypothesis*: Strategy influences logistics control decisions, although not as much as other factors such as industry sector.

**Question**: How does a firm make decisions regarding fleet ownership and outsourcing, and third-party logistics (3PL) versus in-house use for distribution?

*Hypothesis*: These can be treated as strategic decisions. The factors that inform these decisions are investigated using the modeling system that is tested in the Chapter 5 proof of concept. Specifically, I hypothesize that several factors including revenue and industry sector, as well as firm strategies, inform these strategic decisions.

**Question**: How can national distribution structure, which is a Multiple Discrete-Continuous decision, and regional distribution center locations, which uses the two-stage choice construct, be modeled for a firm?

*Hypothesis*: The firm and these attributes can be modeled using a carefully constructed system architecture, as Chapter 3 demonstrates. A real-world application is developed in the Chapter 5 proof of concept. The formulation treats national distribution structure as a Multiple Discrete-Continuous decision and regional distribution center location choice as the second stage of the two-stage choice construct.

**Question**: How can complex but pivotal features of the freight landscape, including the push-pull boundary, the effect of information, e-commerce, and interactions between consumers and freight agents be modeled?

*Hypothesis*: These features can be integrated in an agent-based framework using the theoretical architecture presented in Chapter 3.

## 1.5 Significance of the Study

Chapter 3 introduces two powerful concepts, those of strategy and strategic decisions, into the highly practical context of agent-based modeling and presents an innovative conceptual architecture to enable the integration of these notions. The objective of this chapter is to design a realistic, agent-based modeling framework that can be used to evaluate a variety of goods movement and related activities. Chapter 3 makes numerous, novel contributions to the state of research and practice while meeting this objective:

- Firm strategy adoption is treated explicitly as a decision to be modeled, with consequences for downstream activities. This unifies agent decisions throughout the framework and mirrors the strategic behavior of real-world business entities.
- This framework explicitly models the effect of information on production, procurement and inventory decisions. As such, this framework demonstrates how the famed push-pull boundary

can be modeled in a full-scale, agent-based framework. Additionally, this device permits supply chain innovations including e-commerce and Just In Time policies to be modeled.

- E-commerce demand from households is integrated with parcel delivery supply.

The model design permits other interactions between businesses and households, accommodating on-demand delivery and crowdshipping. Although one other framework offers this capability (Sakai et al., 2020), this is an emerging area with room for research contributions moving forward. Other unique proposed features include modeling carrier services, interactions between trade and carrier agents, and distribution center structure. The carrier aspects have received some attention in a limited modeling exercise with no strategy behavior (e.g., in Liedtke et al., 2015) while distribution center analysis receives some attention using optimization models, but due to computational challenges these are mainly limited to a single industry.

The main objective of Chapter 4 is to develop methods for measuring individual- or company-specific attitudes quantitatively using large-scale, passive data sources. A secondary objective is to address a major data gap in the domain of freight transportation modeling in particular. By meeting these objectives, the chapter makes the following contributions to research:

- Two novel methods to measure attitudes using natural language-based text are developed.
- Both of the methods offer the unique capability of permitting "natural" or unscaled measurements of attitudes.
- A longstanding, text-based source of freight establishment data is recognized and its potential contribution to the pool of freight data sources is proposed and explored.
- A new source of freight attitudinal data, which is readily available and can be collected passively, is developed using the methods of this study, thereby addressing a major gap in freight data.
- An automated attitudinal data development engine (ADDE), which extracts, compiles, and prepares attitudinal data is developed.
- The new methods and data sources are deployed in a proof of concept application to explore a potential freight modeling application, using attitudinal data to detect guiding strategies among Fortune 500 companies in freight-intensive sectors.

The main objective of Chapter 5 is to develop a theoretical framework accompanied by a methodology to model agent strategies. In meeting this objective, this study makes the following contributions to the state of the art in transportation demand modeling:

- A theoretical framework for the joint modeling of strategies and strategic decisions is developed.
- A methodology to jointly estimate both strategies and strategic decisions is developed, incorporating both observable and attitudinal input data.
- The proof of concept demonstrates the first real-world, behavioral modeling application of attitudinal measurements that are developed using natural language processing (NLP) methods (Chapter 4).
- A new methodology to generate consideration set parameters is proposed, developed and implemented.
- Consideration set parameter selection are treated as a strategic decision, and their adoption is modeled jointly with other strategic decisions.
- The methodology is demonstrated in an agent-based freight modeling context.

Chapter 6 is the final core chapter in this thesis. Although it is last in this respect, it provides an important illustration of the potential impact of strategy on transportation energy consumption and emissions (TECE), which accrues from transporting shipments. The analysis is completed in an agent-based framework, which uses a broader geographic resolution than the framework of Chapter 3. The case study in Chapter 6 focuses on firms in the auto manufacturing industry, or Original Equipment Manufacturers (OEMs), and on automobiles sold in the contiguous US. The case study quantifies the TECE that is associated with a strategic shift in production location from a US plant to China. As such, the main contributions of this chapter are:

- Developing an agent-based model of the OEM industry, replete with a set of global, multimodal energy use and emissions rates related to their commodity shipments; and
- Applying the model to analyze the TECE associated with a change in strategy for one of the OEMs.

## 1.6 Definition of Terms

In addition to the terms *strategy* and *strategic decision*, which are defined earlier, the following terms and definitions are used throughout this work:

**Establishment**: "a single physical location where one predominant activity occurs" (definitions from US Census Bureau (2020a) and Sadeghi et al. (2016)

**Company** or **firm**: a collection of establishments that are part of the same business but operate across two or more different sites (e.g., a chain of retail stores); an unaffiliated establishment is technically also a firm

**Transport**: the carriage of goods between locations by motorized conveyances including trucks, airplanes, and so on.

**Storage**: a location or area where goods are held; or, the activity of holding goods

**Distribution**: the transmission of a good from its origin to its destination

**General-purpose (GP) Storage**: any storage that is used for internal company purposes (e.g., holding inventory), including warehouses, yards, tanks, or other

**Transshipment**: the transfer of goods from one vehicle to another

**Transshipment hubs**: facilities where transshipping occurs, such as intermodal yards, truck terminals, crossdocks, railroad classification yards, airport cargo warehouses, and seaport docks.

**Private fleet**: a truck fleet that is owned by the company that uses them. In contrast, companies without a private fleet must rely on a carrier, or a for-hire trucking firm, to transport their goods.

**DC** or **distribution center**: a DC is a facility through which a good is routed on its way to a customer. This thesis also assumes that in real-world data, based on visual inspection of aerial images for roughly 100 properties:

- Any property that is labeled with a "Distribution" use is considered to be a DC
- All "Refrigeration/Cold Storage" ("Light Distribution", Warehouse") properties that are 20,000 (100,000, 150,000) SF or larger are considered to be DCs

**DC control**: a company that owns or leases its own space in one or more DCs has "DC control"

**Freight-intensive**: shipping and/or receiving notable volumes of cargo – for example, firms in the manufacturing sector are freight-intensive while telecommunications firms are not

**Word vector** or **word embedding**: a real-valued vector in some N-dimensional space that represents the location of the word.

# 2 Literature Review

## 2.1 Historical Trends in Freight Transportation Modeling

Historically, freight transportation modeling techniques have followed a similar trajectory as passenger modeling. The first few decades in freight modeling focused on three-step and four-step frameworks. These early models were aggregate, analyzing total tons or trips between zones (e.g., Demetsky, 1974 and Beagan et al., 2007), sometimes with mode choice as in (Abdelwahab and Sargious, 1992). They did not account for logistics activities, such as transloading, that often occur on the shipment journey. In the 1990s, models that address logistics activities began to emerge. Early examples include the SMILE model in Tavasszy et al. (1998), the GoodTrip model in Boerkamps and van Binsbergen (1999), and the EUNET model in Jin et al. (2005). These models advance a theme of major importance—logistics analysis for shipment paths—but have natural limitations due to their aggregate nature.

Aggregate models have many advantages and are considered for use in this thesis. Aggregate models are useful for analyzing many policy or strategic decisions of system authorities (Chow et al., 2010). As Gonzalez-Feliu (2019) notes, aggregate model development has relatively low resource requirements, but still can be used to evaluate high-level features of freight flows, including the flow of commodities or trucks between zones. They are most suited for an environment that has simple transportation options and is static, with no major changes in the economy, vehicle technologies, highway infrastructure, and so on. Unfortunately, these models are poorly suited to capturing heterogeneity among individual actors unless they are enriched to the point where they represent behaviorally homogenous groups of agents (Thaller et al., 2016)). Nevertheless, even this enhanced aggregate approach weakens when changing agent preferences disturb the homogeneity of the grouping system due to, for example, changes in the operating environment. As a result, using a fully agent-based approach is a more straightforward and flexible option.

To mitigate issues related to aggregate modeling, agent-based models of freight transportation began to emerge in the 2000s. This review discusses several of them, highlighting their key features and limitations, to present the context for understanding the innovations provided by the current work.

One direction in agent-based freight research is modeling the activities of individual trucks, as Hunt and Stefan (2007) shows, in metropolitan areas such as Calgary, Alberta in Canada. The second direction, which is more relevant to this thesis, focuses on modeling the activities of individual businesses. The most widely used framework is based on the Aggregate-Disaggregate-Disaggregate (ADA) framework (de Jong and Ben-Akiva, 2007). In this framework, aggregate commodity flows are input to the model system then disaggregated to shipper-receiver pairs using a process that incorporates economic trade partnership information. The disaggregate shipments are then aggregated to zone-to-zone flows of goods or trucks for traffic simulation. The groundbreaking 2010 Freight Activity Microsimulation Estimator (FAME) (Samimi et al., 2013) is the first to adapt the ADA framework to the US context and includes a novel data collection effort. Subsequently, the 2011 Chicago Metropolitan Agency for Planning (CMAP) Mesoscale model (Cambridge Systematics, Inc., 2011; Urban et al., 2012) develops a variant with global supply chains, the wholesale sector, and models for establishment and transload locations. The code base and features of the CMAP Mesoscale model are incorporated and extended in 2012 to include truck touring and a distribution channel model, in an FHWA Broad Agency Announcement (BAA) study (Resource Systems Group et al., 2012). The resulting integrated model saw widespread adoption around the US (Shabani et al., 2018). Additional efforts, including the 2014 Maricopa Association of Governments (MAG) SHRP2-C20 model (Maricopa Association of Governments, 2018), extend the ADA-based framework to integrate truck touring and successfully develop a novel trade partnership module.

While the ADA variants provide a behavioral basis for modeling the flow of shipments through supply chain, they have two fundamental issues. One issue is the lack of firms. Although summary documentation (e.g., Shabani et al., 2018) specifies the agent as a “firm”, in reality, in all of the ADA variants described here, each agent is an unaffiliated establishment. A second issue is that these variants ignore key agent behaviors that are fundamental to freight transportation outcomes. Namely, they do not explicitly model logistics assets of the firm, namely fleets and distribution centers. I suspect that these gaps cause issues in important model functions, including estimation of origin-destination flows, distribution paths of goods, and truck tour development. Some evidence of this is found in tests (Wisconsin Department of Transportation, 2017), although more empirical studies should be conducted to demonstrate the range and severity of any issues.

Various other establishment-centered, agent-based freight models have also been proposed but have seen limited development beyond the conceptual design stage. Nevertheless, these have innovative and useful conceptual features that inform this work. Among these, the Transportation And Production Agent-based Simulator (TAPAS) model (Holmgren et al., 2012) and framework in Roorda et al. (2010) (further specified in FREMIS of Cavalcante and Roorda, 2013) focus on decisions and interactions between agents, for example through the use of contracts, as a major driver of freight activity, in addition to many other useful and realistic features. The latter also discusses sensitivity to emerging trends in policy, business and technology, and it includes transportation and logistics activities.

Currently, an operational, integrated agent-based passenger and freight models is the SimMobility model (Adnan et al., 2016 and Sakai et al., 2020). The latter discusses the freight component and its foundations in (Alho et al., 2017). This model is similar to the proposed framework in that it includes both behavioral-based and dynamic traffic assignment features along with a method for modeling crowd-shipping. However, although fleets are simulated based on an observed distribution, it does not model firms, fleets, strategy, global trade partnerships, and other features that are addressed in this thesis.

Two other research streams are worth mentioning as they relate to some key themes in this study. First, optimization models are sometimes used in transportation freight models (e.g., Friedrich, 2010). However, optimization is difficult to implement with reasonable computation times for multiple sectors and large geographic regions, so this work for the time being is limited to individual industry sectors (food retail, in this case). Second, another optimization-based study, (Liedtke et al., 2015) models shipper decisions including lot size, carrier decisions including fleet size and mix, and shipper-carrier interactions. However, as a proof of concept, the study uses two prototypical carrier companies rather than a full population of carrier agents. Nevertheless, these studies present insightful implementations and motivate the argument that optimization methods or related heuristics are valuable tools for agent-based freight models.

Reviews of the state of the art and state of the practice in freight modeling are found in de Jong et al. (2013), Chow et al. (2010), and Thaller et al. (2016). These summaries establish that while progress has been made, more research is required to transform these analysis systems into more realistic models of freight activity. This thesis directly addresses many gaps noted in de Jong et al. (2013), including use of firms, modeling spatial decisions such as warehouse structure, and others.

## 2.2 Strategic Firm Behavior

The definition of strategy that this thesis adopts is informed by perspectives from the business, economics, and transportation domains. The 1980 classic work (Porter, 1980) defines strategy as the "...broad formula for how a business is going to compete, what its goals should be, and what policies will be needed to carry out those goals" and the "...combination of the ends (goals) for which the firm is striving and the means (policies) by which it is seeking to get there… The essence of formulating competitive strategy is relating a company to its environment." Mintzberg (1987) and Rumelt (2011) further explain how business strategies promote consistency in objectives across company actions. Strategy can also be viewed from a Value Chain perspective as guiding company functions in areas where they can (and can't) gain an advantage by differentiate themselves from the competition. For instance, companies may focus on transportation delivery to provide improved speed of delivery to customers (Rodrigue, 2020).

Two key theoretical studies from economics literature inform this thesis. In particular, they motivate the connection of strategy to strategic decisions, consisting chiefly of major investments such as asset decisions. They also discuss at the conceptual level how strategy, when implemented in a consistent way by a firm, serves to align the purpose of various investments so that these actions help achieve the broader firm goals. First, Shapiro (1989) discusses business strategy from a philosophical and historical perspective, noting that the firm's "strategic decisions, which involve long-lasting commitments" determine its "tactical decisions, which are short-term responses to the current environment". The study suggests a range of variables that qualify as strategic variables, including investment in distribution or other physical facilities. Second, the theory of firm capability (Teece, 2019) differentiates between 'ordinary' capabilities and dynamic capabilities, with the former being enduring investments and the latter being characterized by flexibility in adapting to the environment. The work aligns with Kahneman and Tversky (1979) in seeking alternative explanations for observed firm behavior, which, the study suggests, appears to be largely inconsistent with fundamental economic maxims such as cost minimization. Similar to transportation models that incorporate attitudes, it also seeks to better explain variation between firms using information that can be difficult to quantify.

Quantitative analysis involving business strategy is a major area within the field of business management (see literature examples in Strategic Management Journal, among others). Game theoretic models of strategic behavior are the most common approach to modeling firm actions, however, these models are severely problematic for scenario analysis in futuristic settings (Shapiro, 1989). Due to these issues, this thesis instead uses a combination of econometric and psychometric analysis as the foundation of the strategic behavior models developed in Chapter 5. Additionally, some of the business literature presents applications that examine firm strategy in a freight transportation context. For example, the impact of Just-In-Time, lean production, and other strategies on transportation cost are analyzed (Ehrler et al., 2018). While informative, such a model only examines one outcome, cost, and does not evaluate the inter-relationship of cost to other outcomes. In contrast, this thesis proposes to model and operationalize strategies to inform a number of decisions throughout the agent-based modeling structure.

## 2.3 Strategy in Transportation Models

The concept of strategy is addressed in transportation studies[1] using mathematical models that are designed to generate either insights or practical, applicable results. The following treatments of strategy analysis are consistent with the definition that this thesis outlines in Section 0.

A series of efforts (see Choo and Mokhtarian, 2012 and Choo and Mokhtarian, 2008) examines “strategy bundle” adoption by individuals in their travel pursuits. The relevance of the “strategy bundle” approach is its emphasis on understanding the core, high-level goals of the individual and how these goals drive the specific, actionable mobility decisions that the individual subsequently makes. This perspective is distinctly different from hybrid choice models, which are discussed shortly, where strategies or attitudes are modeled jointly with

[1] The transportation literature also uses the term “strategy” in the context of system governance or control (Kraus et al., 2010), survey responses (Crastes dit Sourd et al., 2018) and the process of making decisions (Hensher, 2014), but these uses are not relevant to this work.

one of the lower-level, actionable decisions. However, these studies aim to inform policy decisions, and do not establish an operational framework that connects strategy bundles to the lower-level decisions.

Another relevant transportation-based strategy study is the development of a theoretical framework in which individuals make plans, or strategies, that drive an observable action (Ben-Akiva, 2010). Like the work just discussed, this study focuses on the passenger context, but has relevance for freight also. Building on the notion that individual actions are preceded by a planning stage, a conceptual framework for transforming a plan into an action is illustrated. The framework is operationalized using a Hidden Markov Model. The plan can be a goal, intention, or choice criterion. The framework is designed to produce one choice outcome. This thesis extends the notion of planning to action for a single decision to a full set of strategies and strategic decisions that are operationalized in an agent-based modeling system.

In contrast, concepts of strategy surprisingly are a major gap in extant agent-based freight models (e.g., Auld and Mohammadian, 2012; Maricopa Association of Governments, 2018; Alho et al., 2017; de Bok and Tavasszy, 2018; and Federal Highway Administration, 2020). Given the underling business focus of agent-based freight models, and the critical impacts of strategy on business actions, it seems intuitive that these models should already account for business strategy. Further, since agent-based modeling techniques support a rich characterization of agents and decisions, incorporating strategy would not be an insurmountable task. This thesis addresses this gap, which is first identified by the author (Chapter 3 and Stinson et al., 2018), and expands the initial model implementation documented in (Stinson et al., 2020) and illustrated in (Stinson et al., 2019).

## 2.4 Latent Variable Measurement with Attitudinal Data

Theoretical and computational developments in recent decades have enabled behavioral models, including those used in transportation, to include unobservable (or latent) attitudinal factors in addition to observable factors (e.g., Ben-Akiva & Boccara, 1995; Ben-Akiva et al., 2002; Kamargianni & Polydoropoulou, 2013; Daziano, 2015; and Cambridge Systematics et al., 2007). The main advantage of adding such factors is that they provide an additional level of behavioral realism, which ideally will support the development of more robust models and, ultimately, better predictions about the future—but see caveats (Vij and Walker, 2016).

However, despite their intuitive appeal, the development of attitudinal data sources and even the very measurement of attitudes is known to be cumbersome and fraught with issues (Ben-Akiva et al., 2002 and Vij and Walker, 2016 summarize challenges). The standard process for gathering attitudinal information consists of asking a series of questions about the respondent's preferences. Answers are typically provided using a Likert or similar pre-determined scale – e.g., with choices ranging from "Strongly agree" to "Strongly disagree". The main issue with this technique it its intrinsic subjectivity, both in terms of the questions asked and the answers provided. There are no clear rules regarding what questions will accurately and precisely elicit the desired information from the respondent. Both the content of questions and their wording are highly dependent on the experience and judgment of the questionnaire developer. An answer of "Neutral" to one person may be equivalent to "Somewhat agree" for another. The highest and lowest categories of the pre-imposed scale (e.g., one to seven) may not capture the full range of attitudes and preferences, for example, an "off-the-charts" passion for environmental sustainability. Using discrete, unit steps between each response category raises issues in analysis.

Finally, surveys are expensive to conduct and often suffer from low response rates, which hinders the potential to collect a random sample (Ben-Akiva and Lerman, 1985, describes how random sampling delivers the best statistical properties). So many topics are important – household or company attributes, daily transportation patterns, information about the available fleet – that transportation-related surveys are, in general, already quite long. Adding attitudinal questions exacerbates this issue. Moreover, due to their subjective nature, many attitudinal questions are normally asked in order to help ensure that the respondent's true attitudes are extracted, potentially fatiguing respondents and engendering poor-quality responses.

Attitudinal surveys, while challenging for passenger traveler understanding, are essentially a complete gap in freight transportation surveys of companies. One reason for this is that the freight domain is challenged by more urgent gaps (Transportation Research Board, 2003). Another reason is that businesses are busy and often unwilling to share information. As a result, with some exceptions (Jin and Shams, 2016 and Ben-Akiva et al., 2013, to name a few), transportation activity surveys of companies are uncommon, hence many agencies choose to develop freight transportation models without attitudinal features (e.g., Maricopa Association of Governments, 2018; de Jong et al., 2013). In spite of this gap, it is reasonable to assume that attitudes are just as integral to company decisions as they are to individual decisions.

The key innovation of Chapter 4 is proposing and developing new methods to use large-scale text data to generate attitudinal data that can be used in lieu of traditional attitudinal survey sources. The idea to use text data for attitudinal analysis is not entirely new to this thesis. Collins et al. (2013), Ghiassi et al. (2013) and others have used Twitter to study attitudes towards transportation and other areas. However, sentiment bias is a known issue with Twitter data (Barbosa & Feng, 2010). Furthermore, the user base of Twitter is far from a random sample of the population. Text-based analysis of companies and freight is starting to emerge as well. One freight-related application uses NLP to learn the type of commodity that is shipped by a company (Moscardi, 2019). A longitudinal sample of annual company reports utilizes topic modeling to detect changes in company strategy over time (Menon et al., 2018). While insightful, however, these studies do not generate the kind of statistically sound, attitudinal data that are required as inputs to factor analysis, hybrid choice models, and so on.

## 2.5 Distributed Representation Learning and Code Bases for PCA and word2vec

The methods of this study utilize both Principal Component Analysis (PCA) and Natural Language Processing (NLP). PCA has a long and well-documented history of development and application, as one synopsis shows (Jolliffe and Cadima, 2016). The field of NLP is less mature but has already advanced far by leveraging techniques from cognitive science, computational linguistics and computer science. This study relies particularly on methods that quantify symbolic data using distributional representations, with each feature of the distribution representing a unique concept (Barlow, 1972 and Hinton, 1986). The implication of these notions for NLP is that each word can be treated as a mix of N orthogonal concepts that are represented in a corresponding N-dimensional vector space (N is chosen by the analyst) (Bengio et al., 2003; Mikolov, Chen, et al. 2013; and Pennington et al., 2014). Figure 2 shows how several words may be represented as vectors in a two-dimensional construct, with one dimension having a living/non-living interpretation and the other a mobile/immobile interpretation.

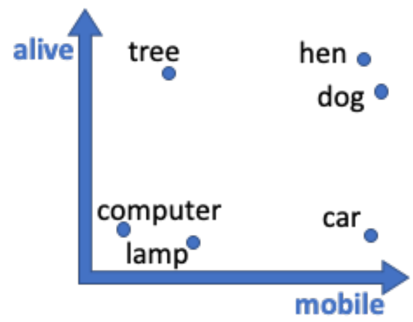



**Figure 2. Example two-dimensional conceptual construct.**

Neural networks are intrinsic to computational performance in these algorithms. They also provide an elegant alignment of each neuron in the hidden layer with one concept. Features of word vectors in this space are learned using information from word contexts.

The algorithms that are designed in this study build upon two existing code bases. The study uses the NLP word2vec code, which is made publicly available by (Google Code Archive, 2013) with Python accessible modules (Rehurek, 2020). Python Scikit-learn modules are used for PCA (Pedregosa et al., 2011). To the author's knowledge, this research is the first effort to propose and develop methods that adapt these algorithms and combine them in such a way as to generate attitudinal measurement data.

The next section describes the innovations of this work and how these innovations contribute to the state-of-the art in attitudinal analysis and freight data development.

## 2.6 Methods for Modeling Joint Decisions with Latent Variables

Moving beyond the conceptual motivation to include strategy in transportation models, determining suitable mathematical approaches to model strategy and its transportation impacts is a new area of research. This thesis adopts earlier thinking around "strategy bundles" (Choo and Mokhtarian, 2012), in which each (latent) strategy comprises multiple aspects, and the notion that multiple strategies can inform decisions simultaneously (Kamargianni and Polydoropoulou, 2013). Moreover, this work combines these unobservable aspects with the idea that multiple strategies inform multiple outcomes simultaneously. Conceptually and quantitatively, an emerging stream of research into "mobility bundle" or "lifestyle" choices relates to this study. In one recent example, model several long-term, medium-term, and short-term choices (location, vehicle ownership, and activity and travel choices, respectively) simultaneously with attitudes and lifestyle preferences (Paleti et al., 2013). Bhat, Astroza et al. (2016) further motivate this work by pointing out that when multiple unobserved factors jointly influence multiple outcomes, greater efficiency in coefficient estimates are obtained when the modeling system appropriately addresses this jointness. Together, these and other studies inform the structure of the methods that are developed in this study and demonstrate examples that bridge the conceptual and quantitative aspects.

This work models unobserved strategies following the foundations of attitudinal analysis as discussed in Bollen (1989), Kaplan (2009) and others. The above studies regarding individual attitudes and strategies likewise adopt this foundation, which comprises factor analysis and structural equation modeling, as do various studies of company strategy and attitudes in the business domain (e.g., Lai et al., 2002). The most widely used framework, known as the Hybrid Choice Model (HCM), has both a latent component and one observed choice component (Walker, 2001 and Ben-Akiva et al., 2002). Latent variable estimates (scores) are predicted by exogenous variables, while choice outcomes are predicted by both exogenous variables and the latent variable scores. This study adopts this general structure, although with a number of enhancements and refinements to extend it for this context, in particular developing the capability to jointly model a range of continuous and truncated variables.

The rest of this discussion reviews existing methods to model joint decisions, attitudes and their impacts on decisions, and choice set parameters. Various mathematical models have been developed and applied to problems involving joint decision making. Fang (2008) develops a Seemingly Unrelated Regression (SUR) construct to study mixed discrete-continuous (MDC) questions involving the number of vehicles of different classes that are owned by a household and the vehicle-miles traveled (VMT) using each class.

The model in Fang (2008) offers several appealing features. First, it represents VMT using a truncated variable (Tobin, 1958; Woodridge, 2010). This elegant construct allows one variable to capture two choice dimensions: one discrete and the other continuous. This is especially useful to major, high-level decisions that involve a yes/no aspect as well as "how much" aspect. Amore and Murtinu (2019) discusses examples where this is relevant to high-level, strategic decisions in the business context. Using an ordinal model with a continuous latent variable corresponding to the number of vehicles, the entire system of outcomes is represented as a set of continuous variables, which generally lend themselves to more straightforward and faster computations than discrete variables. The approach in Fang (2008) also readily permits unrestricted covariance between the error terms of each pair of outcomes, which is another extremely desirable property. The mathematical foundations of this approach are well established and equivalent to the SUR model. Finally, the approach readily permits additional categories of a variable to be added – for example, two types of vehicles (passenger cars and trucks) are examined in Fang (2008). Due to its similarities to the Multiple Discrete Continuous Extreme Value model (MDCEV) (Bhat, 2008), an MDCEV application is implemented using the same data. Similar results are obtained for each application.

However, while elegant, the model in Fang (2008) does not address attitudinal dimensions. Furthermore, its application is designed using long-term, locational decisions as exogenous inputs, focusing only on predicting medium-term, downstream decisions (vehicle ownership and use).

Numerous recent works have developed powerful methods to jointly model multiple latent attitudes and outcomes of different types (nominal, continuous, etc.). The development of the Generalized Heterogeneous Data Model (GHDM) in Bhat (2015a) enables a significant step forward in this respect. Bhat, Pinjari et al. (2016) and Lavieri et al. (2017) demonstrate the value in applying this framework using a number of decisions ranging from long-term to short-term in nature (location choice, vehicle ownership, mode choice, and so on) jointly with multiple latent variable constructs.

However, these demonstrations illustrate a disadvantage with the current GHDM setup, namely related to the treatment joint decisions that are associated with one discrete outcome. For example, Fang's use of the SUR permits the modeling of three decisions for each vehicle class: (1) own one or more vehicles of that class (yes/no); if "yes", then (2) how many vehicles to own and (3) how many miles to drive them. To generate a similar decision structure using the GHDM, it appears that that a new instance of the GHDM must be established for a particular combination of activities. For example, the application in Bhat, Pinjari et al. (2016) presupposes that the household has members who commute to work, implying that the estimated model may not be valid for households that are making other choices within the application (such as residential location) but that have no workers. It is not immediately clear how the GHDM handles discrete or continuous choices for which some individuals may choose to consume nothing. Presumably, a new set of parameters needs to be estimated for each such case, with one set representing individuals with non-zero consumption and another set representing individuals with zero consumption.

## 2.7 Methods to Model Choice Set Generation Parameters

In addition to mitigating this drawback, the approach proposed here offers a key innovation over existing works. Namely, this study proposes to treat choice set generation parameters as a set of strategic decisions that are made jointly with other types of strategic decisions. As put forth in Manski (1977), decision making often involves two stages (at least conceptually, although they may overlap in reality). First, in choice set generation, an individual determines a set of choices from which to choose one option. Second, the individual selects an option from this set. There is general agreement that non-compensatory processes, such as adhering to some maximum threshold for price, play a role in choice set generation for such decisions (e.g., Ben-Akiva and Boccara, 1995 and Castro et al., 2011). As such, ignoring the choice set generation process (by imposing a single, compensatory choice model), will generally bias the estimates of the true model parameters by conflating the estimates with factors that are not really considered in the second, compensatory decision stage.

For these reasons, choice set generation is widely used in transportation and land use models, particularly in preparation for modeling spatial or other decisions for which the universal choice set is vast (e.g., in passenger travel demand modeling, housing location choice models such as Kaplan et al., 2012 and Rashidi et al., 2012). Key factors in the choice set generation process for location-related choices may include both property requirements (e.g., maximum price) and transportation accessibility requirements (e.g., distance to the nearest passenger rail station). This study evaluates freight-related accessibility from a similar perspective and uses pre-existing software Geographic Information Systems (GIS) packages, in particular the geospatial "sf" package in R (Pebesma, 2018), to measure distances.

Moreover, the method proposed here permits the joint modeling of choice set generation parameters that are associated with multiple types of choices – for example, minimum fleet size and maximum warehouse size. This extends modeling efforts to date that have developed choice set parameters for one decision only (for instance, see Swait and Ben-Akiva, 1986 and Bhat, 2015b).

The process of solving such a rich modeling system is a challenge. Estimation using classical, full information methods involves computing multidimensional integrals to maximize the log likelihood function, which is computationally challenging in applications with multiple latent variables. As a result, researchers rely widely on Maximum Simulated Likelihood (MSL) (Train, 2003), in which the integrals are approximated. Even so, MSL remains computationally challenging for problem instances with substantial interdependency between alternatives and multiple latent variables, which is the type of application that is addressed in this thesis.

This study turns to Hierarchical Bayes techniques to estimate model parameters. The benefits of Bayesian approach, and Gibbs sampling in particular, are discussed in previous studies (Fang, 2008; Daziano, 2015; and Train, 2003, to name a few) and summarized here. A Bayesian approach with Gibbs sampling and variable augmentation avoids the use of multidimensional integrals. This reduces computational burden and eliminates the need for approximations, thereby providing exact inference with finite samples. Moreover, Bayesian estimators are statistically at least as good asymptotically as classical estimators. With a good, informative prior, Bayesian estimators can be superior to classical estimators. Ultimately, Bayesian methods offer a powerful foundation for developing detailed statistical models with large amounts of interdependency in outcomes and multiple latent variables.

Using Gibbs sampling with data augmentation (Albert and Chib, 1993; see also Geweke, 2005), this work extends earlier, Tobit-based implementations (e.g., Fang, 2008; Cowles, et al. 1996) by incorporating attitudes as an additional dimension. McCulloch and Rossi (1994) and Fang (2008) inform the methods used in this study for drawing from a truncated distribution conditional on other distributions. Daziano (2015) also uses Gibbs sampling in the behavioral modeling context, but with a discrete choice kernel rather than a SUR.

The Maximum Approximate Composite Marginal Likelihood (MACML) method (Bhat, 2011) may also work well for this application. Using MACML or other solution methods to solve the modeling system in this application is a potential extension to this work.

## 2.8 Data Fusion and Real-World Analyses of Firm Strategy

This thesis presents a novel use and fusion of numerous existing sources in Chapter 4 and Chapter 5. Fortune 500 magazine (Fortune) regularly compiles and makes available salient statistics on the largest companies in the US and internationally. The FleetSeek database has comprehensive coverage of practically all commercial truck fleets in the US (FleetSeek, 2017 as analyzed in Mele, 2017, and posted online in BigMackTrucks.com, 2017). The CoStar real estate database contains information for commercial properties throughout the US (CoStar, 2020). Data on population density, rail-truck intermodal yards, and major water ports are obtained from the US Bureau of Transportation Statistics (BTS) (2020), the US Census Bureau (2020b), and the FHWA Freight Analysis Framework (FAF) (Oak Ridge National Laboratory (for FHWA Freight Management and Operations, 2017). Nearly all of these sources are publicly available—only CoStar is propriety. Such a comprehensive analysis of these companies and their logistics behaviors using these various data sources has not been undertaken previously.

The proof of concept results in Chapter 3 and Chapter 6 also integrate a number of real-world data sources. The data foundation for the former follows Urban et al. (2012) and Stinson, Enam et al. (2019), and is summarized briefly within the chapter. Likewise, sources used in the latter are documented within that chapter.

The behavior-based results that are generated in each proof of concept are themselves an important contribution to the domain of agent-based freight modeling. Although a few stand-alone models of freight fleet ownership have been developed (e.g., Rashidi and Roorda, 2018, and Sillaparcharn, 2007), by and large vehicle ownership models have been confined to the passenger transportation domain (Anowar et al., 2014). Moreover, existing agent-based freight models including those in Sakai et al. (2020) and Outwater et al. (2013) do not recognize affiliations between distribution centers and their operators. In general, they also do not account for private fleet vs. for-hire transport decisions, which may be an issue for the accuracy of origin-destination flows and the characteristics of truck tours. While de Bok and Tavasszy (2018) develops an insightful simulation of urban goods transport of carrier activity and warehousing, it has limited application for private fleet and privately-operated distribution. The framework proposed in Federal Highway Administration (2020) acknowledges the importance of fleet and distribution channel decisions and integrates some real-world data, but fleet and distribution decisions are not integrated. This study improves upon these works by acknowledging the integrated nature of fleet and distribution decisions, by devising methods to model these decisions, and by developing real-world models and parameters that are immediately useable in agent-based modeling platforms.

Estimating the impact of company strategy on its global energy and emissions footprint has no precedent in the literature. More generally, related works have quantified the negative impacts of freight transportation nationally (e.g., Olmer et al., 2017 quantifies global shipping impacts), but not in the context of strategy impacts analysis.

# 3 A Behavioral Framework to Model the Movement of Goods through Supply Chains

Part of this chapter was previously published as Stinson, M., J. Auld, and A. (Kouros) Mohammadian, "A large-scale, agent-based simulation of metropolitan freight movements with passenger and freight market interactions," *Procedia Computer Science*, vol. 170, pp. 771–778, 2020, doi: 10.1016/j.procs.2020.03.157. © 2020 The Authors.

## 3.1 Introduction

This chapter presents a novel, agent-based framework for simulating the movement of goods through the transportation system. Like earlier works, the model architecture uses a three-layered conceptual construct of decision-making and outcomes (Figure 3). The three-layered construct illustrates the activities and decisions of agents, and how various elements interact, in a systematic fashion. The construct also provides a systematic and intuitive way to translate this conceptual understanding into a computational framework. There is continuity between and among the layers, as decisions from one layer naturally feed into downstream decisions.

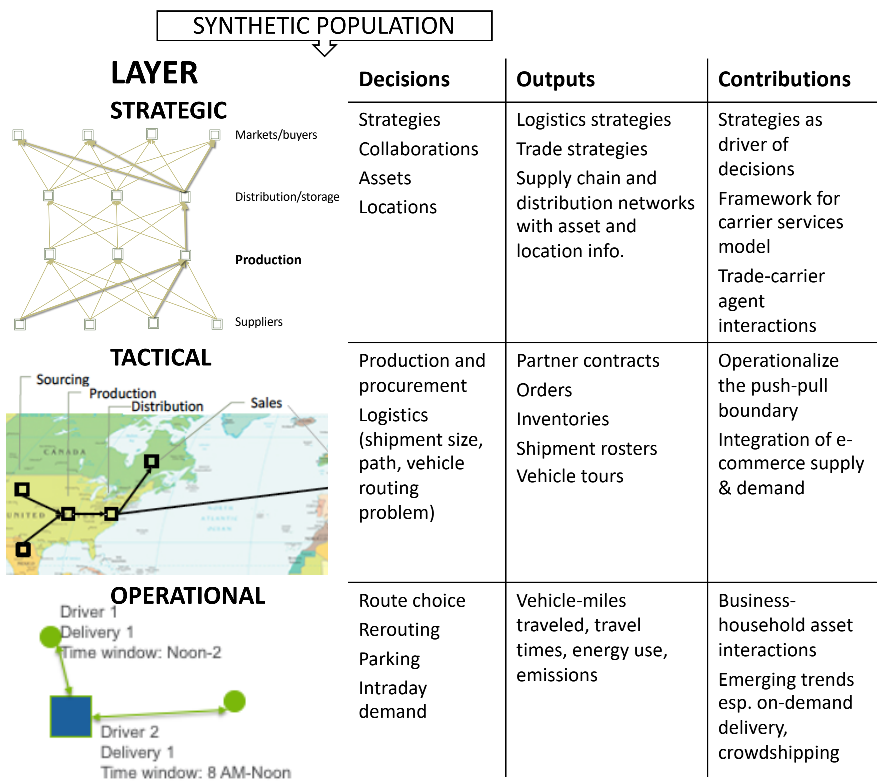



**Figure 3. Model overview.**

The layers are distinguished by the types of decisions and actions that are featured. Generally speaking, decisions are categorized based primarily on their temporal qualities, namely duration of impact and frequency. For example, in general, long-term decisions are infrequently made and have a relatively long-lasting impact. Characteristic outcomes of each decision also informs its categorization:

- Long-term (LT): decisions that drive origin-destination flows of goods
- Mid-term (MT): decisions that result in delivery rosters and goods storage
- Short-term (ST): operational decisions that generate traffic

The model simulates the flow of shipments by modeling the following agent behaviors and decisions:

- Strategy adoption and strategic decisions;
- Collaborations, which create the underlying demand for goods movement;
- Asset choices including both facilities and fleets;
- Trade volumes, including production, procurement, and inventory levels;
- (Carriers only) Transport and logistics service offerings;
- Shipping decisions, including shipment size, frequency, and path choices;
- Vehicle routing and touring decision; and
- En-route decisions including routing, parking, and on-demand delivery adjustments.

Since the model design is based on underlying business behaviors rather than specific market segments, the framework can be applied across all types of geographies: urban, national, and international levels. It can also be scaled to different temporal resolutions ranging from minutes to years. However, for computational and calibration reasons, this work discusses the most recommended application with a detailed metropolitan area focus that has national and global trade and distribution ties (discussed in Section 3.2.1), and with traffic simulation for a single, 24-hour period. Furthermore, the model supports any number of industry classes and transportation modes. Finally, the design of the model facilitates scenario analysis under varying political, socioeconomic, technological or other trends. For example, trends in the economy, government regulation, population growth, new vehicle technologies, e-commerce, sharing economy, and infrastructure can be evaluated for their impacts on goods movement.

The remainder of this chapter is organized as follows. First, the objectives and contributions are documented. Second, key features of the framework – its "building blocks" – are described. The behavioral basis of the model and features that are unique to this framework are explained further. Next, the integration of all features into a single modeling system is presented. Agent decisions are discussed in more detail in this section. Finally, a proof of concept is presented.

## 3.2 Key Features

### *3.2.1 Geography*

Although the model is built with greatest detail for a specified metropolitan region, the entire modeling system is global. Throughout the model, the location of agents is extremely detailed (parcel level) in the central metropolitan region. In keeping with Urban et al. (2012), locations of agents in other US and international areas are tracked using the FAF zone system (Oak Ridge National Laboratory (for FHWA Freight Management and Operations), 2017).

Simulated trade partnerships can be global, national or regional in nature (Pane A in Figure 4). In the partnership stage, transport and logistics paths likewise are modeled for these three scales. The model keeps track of path attributes including travel time on each mode when trade agents are choosing paths. In addition, as Pane B illustrates, domestic transshipment locations are specified as part of the path choice process. Foreign transshipment points are represented by a single location. The network formation models do not decide the exact route (e.g., where does the truck turn), but instead factors the great circle distance based on the average mode speed to estimate travel time between each pair of consecutive transshipment points (see Stinson et al., 2017 for more detail).

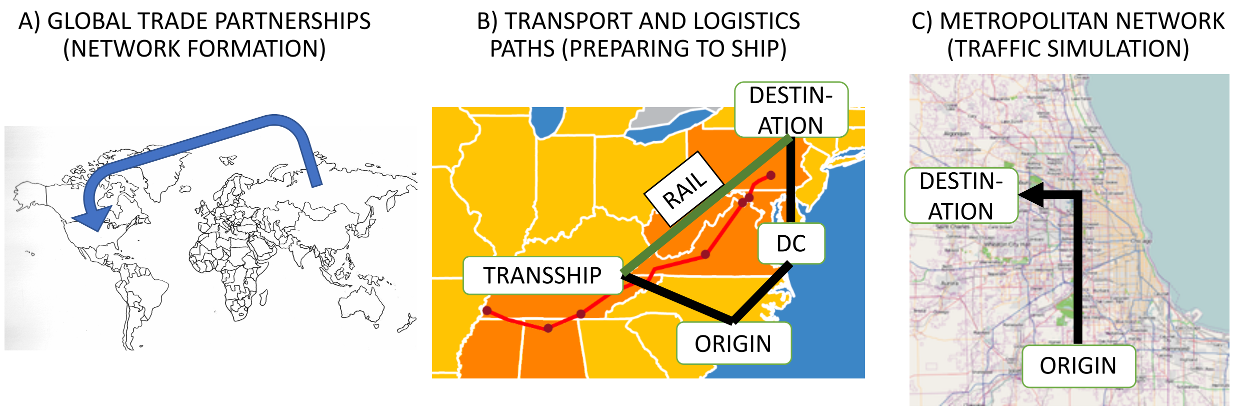





**Figure 4. Geographic scale - examples.**

For the short-term traffic simulation (Pane C), origins and destinations are already established. Each vehicle agent chooses a route. The geographic resolution is extremely detailed in this stage, using a parcel-level geography for all regional trips and a detailed highway network.

### *3.2.2 Agent typology*

This framework uses two fundamental types of agents: the establishment and the firm (or company; see Figure 5). It currently stops short of collections of firms, which BLS denotes as enterprises, but may be extended to this in the future.

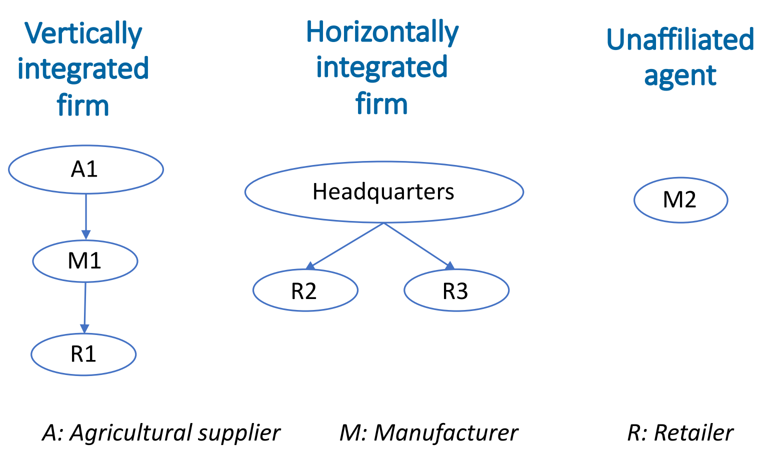



**Figure 5. Illustration: Firms and unaffiliated establishments.**

Firms can be horizontally integrated, vertically integrated, or both. Integration shapes the flows of goods and information across agents. A vertically integrated firm generates at least some of its own inputs and thus includes establishments in two or more industries. A horizontally integrated firm operates in one industry and has two or more establishments in that industry. For example, a horizontally integrated firm may operate a regional chain of retail stores. While vertical integration in trade has fallen out of favor in recent decades (Hall and Braitwaite, 2001), it is still used (e.g., "Phillips 66" (2020) notes that Phillips 66 owns both refineries and gas stations). Therefore, both are modeled in this framework. Vertical integration of logistics agents is evident with well-known, integrated service providers (e.g., FedEx) or can be organized by a freight forwarder with each party—ocean carrier, port terminal, and inland carrier—controlling its own segment (Meerseman and Van de Voorde, 2001). Horizontal integration of logistics agents provides greater economies of scale, e.g., in vehicle fleets (Meerseman and Van de Voorde, 2001).

Agents are characterized at the establishment level based on which activities are conducted at the site, in other words, the functional role(s) of the establishment. For this purpose, this framework defines trade as the physical exchange of goods between one establishment and another. As such, trade can occur between either unrelated establishments or affiliated establishments (for intrafirm movements). The shipper is the sending establishment and the receiver is the recipient. The goods are owned by the shipper until they reach the receiver.

Three categories are used to classify individual establishments: trade agents, logistics agents, and rule-makers. Classes are based on the following hierarchy.

Trade agents supply or buy goods (or do both). Any establishment that engages in trade is designated as a trade agent. Trade agents have three potential additional roles: goods production, goods consumption, and logistics functions. When the trade agent conducts its own transport, it is said to have a private fleet. A trade agent that conducts its own distribution is said to have DC control. This framework assumes that the trade agent uses its own logistics services exclusively. Likewise, such services are assumed to be available exclusively to the affiliated trade agent. Logistics activities, if conducted by the trade entity, can be accommodated at the same site as trade activity (for example, a factory with an adjacent parking garage for its private truck fleet) or at an off-site location. The latter is referred to as an ancillary logistics agent, meaning an establishment that serves logistics function(s) and is part of a firm. Otherwise, logistics activities are associated with a unique establishment.

For trade agents, it is necessary to elaborate on distinctions between supplier and buyer activities. Suppliers provide goods to entities that purchase the goods. There are two main types of suppliers. Producer agents make goods through a production process such as harvesting crops or baking. Agriculture, mining, and manufacturing industries are prominent in goods production. Agents that produce goods usually also consume goods that become part of the product or output (for example, seeds grow into crops).

Wholesale agents, also called distributors, are the second type of supplier in this framework. Agents in the wholesale sector purchase goods from producers or other wholesalers and sell the goods to other agents. Wholesale agents do not produce goods, but as Urban et al. (2012) notes, they buy and sell goods, and are important enough to constitute nearly 50 percent of the establishment sample in the US Commodity Flow Survey (US Census Bureau, 2020a).

Buyers purchase goods from a supplier. There are two main types of buyers. Consumer agents use or absorb goods. For example, a manufacturer that uses steel to make beams is a consumer. End consumers use goods but either produce services or have no economic output. For example, automobile shops purchase auto parts and use the parts to fix cars, which is classified as a service, while individuals in the population purchase food for personal consumption with no economic output. All wholesale agents are classified as buyers as well as suppliers.

The second type of agent is the logistics agent. This agent transports or stores and distributes goods for trade agents, but does not trade goods. Two types of logistics agents are defined. A carrier is defined as a logistics agent that owns and operates a fleet. It may also own and operate one or more storage facilities. Truck carriers may specialize in conveying parcel, less-than-truckload, full truckload, or intermodal container shipments. The framework defines third-party logistics (3PL) firms as a logistics agent that owns one or more distribution centers.

The third kind of agent is the rule-maker, which is an authority that makes decisions that govern the operating environments of trade and logistics agents. Rule-makers include government agencies, which are unaffiliated with trade or logistics agents. They also include transshipment hub operators (THOs), which may be operated by a public authority (e.g., a port authority) or a private party (e.g, Walmart or a 3PL). The THO is assumed to be responsible for determining fees, storage allocations, and the like. Storage may occur in staging areas where while goods await their next pickup.

The model uses a fourth kind of agent: a truck driver. At this time, drivers simply appear for a trip and disappear when the trip ends. The driver has route choice behavior but makes no other decisions.

### *3.2.3 Strategy*

A unique and powerful innovation of this framework is its emphasis on integrating agent strategies as a key factor in the agent's decision-making processes. Strategy models are introduced early in the model architecture – immediately following population synthesis – since they provide a mechanism to guide and unify agent decision-making from the long-term to the short-term horizons.

This work focuses on trade and logistics strategies because they have a major impact on origin-destination patterns and transport decisions, which are both pivotal to simulating freight transportation. Proof of concept models of strategy for trade agents in the logistics context are developed in Chapter 5.

In addition to logistics strategies, trade agents will develop trade strategies. The planned extension to trade strategies will focus on identifying factors that shape origin-destination flows. The focus is on evaluating strategic decisions regarding which sales markets and procurement markets are targeted by trade agents, with markets defined based on geography, trading partnership characteristics, and qualitative dimensions that capture the impact of human factors in the decision process. Disposition towards risk is of particular interest, as Kahneman and Tversky (1979) explored and documented its effects for individual persons. Risk aversion can manifest as, for example, searching out domestic versus international trade partnerships in emerging markets.

Sales (procurement) strategies relate to the volume of products that sellers (buyers) target for trading across a mix of markets. These are planned to be high-level in nature – for example, the percentage of production that is targeted for international versus domestic sales. Relationship mechanisms, e.g., type of contract as discussed below, will be assessed as part of this.

Logistics agents, like trade agents, must invest in assets without perfect knowledge of demand. Logistics agents plan for anticipated freight demand with two types of strategic decisions. First, they select assets, including an operational base, facilities and fleet, that will enable them to serve anticipated demand in a cost-effective way. As part of this, logistics agents decide whether to offer transport, storage, or both. Second, they establish service parameters to accommodate the needs of trade agents. The following service decisions are planned to be included in the model:

- Goods: all types of goods vs. specific commodities
- Geographic scope: urban, intercity/same-day, national/multi-day
- Operational base(s) locations
- Relationship and exclusivity preferences: amount of service dedicated to specific movements (e.g., port, a single business, etc.)
- Carriage offerings: full truckload, less-than-truckload, parcel (specialty offerings including refrigeration may be added)

### *3.2.4 Assets*

Asset ownership and use decisions are key behaviors that are modeled in this framework. The framework includes fixed assets and mobile assets. Fixed assets are production sites (e.g., factories), storage sites, transshipment sites, and other immobile facilities. Production sites are initially represented using three broad classes (farmland, mining, and manufacturing); the exact classes will be refined as model development continues. Mobile assets, including size and mix of truck and van fleets, are modeled. While transport by non-highway modes is modeled, non-highway conveyances are not currently modeled. In other words, capacity on non-highway modes is assumed to be unrestricted.

Storage is space that is used to hold goods. Three types of storage are modeled. First, storage occurs at logistic agent or THO sites while goods await processing. Second, storage as part of distribution is modeled. Third, GP storage is modeled. The second and third types are defined in Section 0. Further, to maintain consistency with Chapters 4 and 5, DCs are assumed to fulfill certain minimum size requirements.

### *3.2.5 Relationships and supply chains*

In the goods movement context, relationships can be formed between trade agents, trade and logistics agents, trade agents and THOs, or logistics agents and THOs. This subsection focuses on cooperative relationships, which exhibit different degrees of formality varying from casual to binding:

- "Spot market" relationships are the least formal and least enduring partnerships. The spot market is where short-term partnerships are formed, e.g., a short-term agreement to transport a single load from Business A to Business B. The relationship is based on cost and the provider's ability to meet other requirements, such as supply volumes, delivery time windows or storage space minima.
- Contractual relationships guarantee certain elements of the partnership – for example, that a certain level of goods will be purchased at regular intervals by a receiving establishment. This framework specifies volume guarantees (for trade contracts) and level of service guarantees (for logistics contracts) in contractual relationships.
- Binding relationships involve an enduring legal mechanism that effectively mandates partnership functions. Relationships among establishments that belong to the same firm are treated as binding and exclusive for all activities for which the firm has a function. For example, if the firm owns a private fleet, then that fleet is assumed to handle all of the deliveries for the trade agents of that firm.

Other types of relationships occur in real-world scenarios. These include supply chain management, in which various actors in the supply chain share information on customer demand to better control costs up and down the chain (Heaver, 2001). Alliances, e.g., between major ocean carriers, enable participants to achieve the benefits that are associated with scale of operations effects as Meerseman and Van de Voorde (2001) suggests. Finally, competitive relationships also influence agent behavior. These are not currently included in the framework but may be in the future.

Trade relationships are formed between agents that trade goods with one another. In general, relationships are many-to-many: a given trade establishment typically has multiple suppliers and multiple buyers. A complex web of trading activity results (Figure 6).

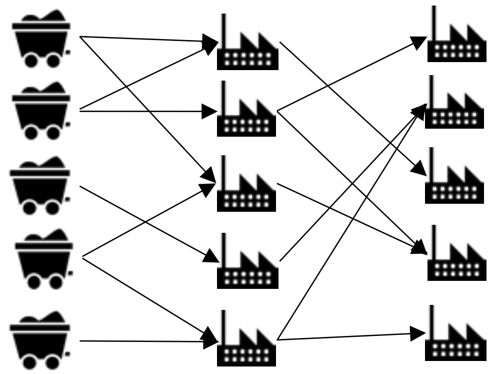


**Figure 6. A web of trading activity**

A supply chain is a cross-section of all trading and logistics activity that is specific to one product that is produced by a specific company. Figure 7 shows an example of a supply chain for a trade agent that manufactures retail products. The Council of Supply Chain Management Professionals (CSCMP) (2013) defines the term “supply chain” as follows:

> 1) Starting with unprocessed raw materials and ending with the final customer using the finished goods, the supply chain links many companies together; 2) the material and informational interchanges in the logistical process stretching from acquisition of raw materials to delivery of finished products to the end user. All vendors, service providers and customers are links in the supply chain.

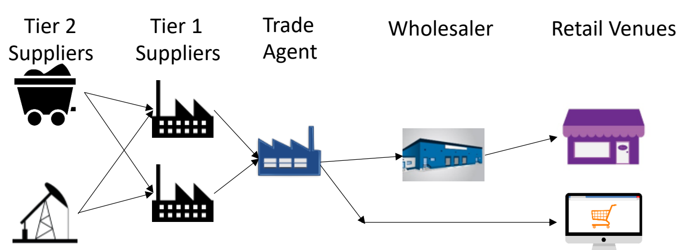



**Figure 7. A supply chain with trade agents shown**

Decisions in the framework are made both by individual agents and by groups of agents. When agents interact, the group must prioritize the needs and objectives of its members in a way that benefits the entire group. Initially, agent needs and subsequently decision making are prioritized in the following hierarchy:

- the end consumer
- the principal firm establishment
- the purchasing establishment
- the shipping establishment
- the prime logistics agent (if multiple logistics agents are used)
- supporting logistics agents

This scheme is used for prioritizing both trade and logistics decisions. This prioritization scheme may be revised based on data from an establishment survey.

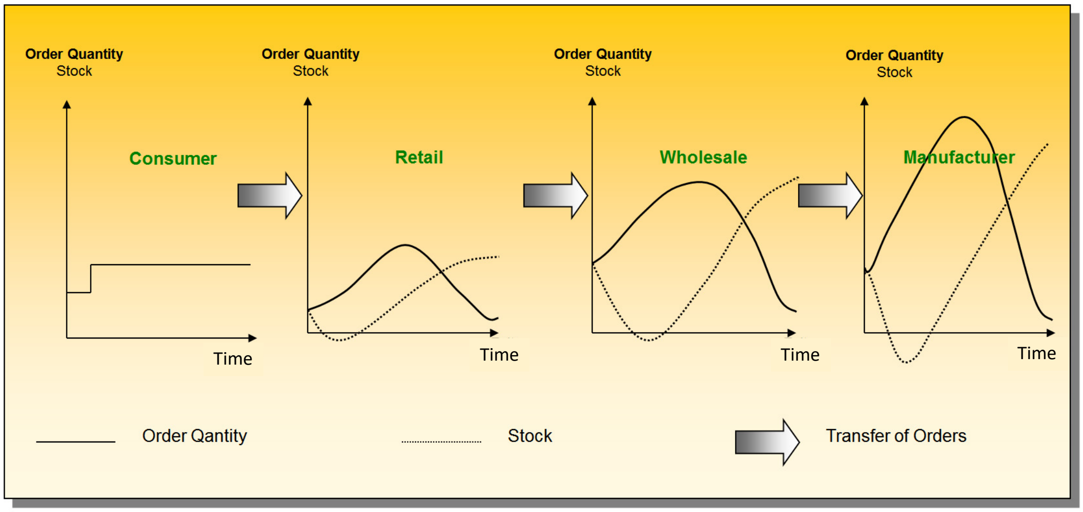
The remainder of this discussion is presented from the supply chain perspective since supply chains are the dominant business structure in today's economy. However, the phenomena described below generally apply to vertically and horizontally integrated firms as well. A major difference is that the activities of integrated firms do not suffer from the lack of information that is inherent to supply chains. These shared and differing aspects of business operations are important features of this modeling framework.

### *3.2.6 The effects of information*

Visibility in the supply chain describes the available level of transparency of other members of the supply chain. More information, especially information on demand, typically translates into greater efficiency of goods production and goods movement.

#### 3.2.6.1 Order Quantity and the Bullwhip Effect

Lack of information regarding the needs and actions of other supply chain members generates uncertainty that can be amplified from one member to the next, creating a "bullwhip effect" (Forrester, 1961) (Figure 8). This is particularly true when demand at the endpoint is unknown. The results of this compounded uncertainty include locally optimal production and a buildup of buffer inventory to unhealthy levels, leading to financially harmful effects such as huge discounts on unwanted inventory.




*Creative common license: Grap-Own Work (2010)*

**Figure 8. The bullwhip effect**

At the other extreme, a given trade agent may have perfect information on its customer's demand. In this case, the agent can utilize the "just-in-time" (JIT) delivery method, which emerged in recent decades as a cost-effective method of arranging the flow of goods (Norwich University Online, 2017). In JIT systems, goods are produced (1) in the right way (meeting specifications exactly) and (2) in the necessary quantity; then they are delivered (3) to the place where they are used (4) at exactly the right time. One objective of JIT is to reduce safety inventory to zero levels, thus minimizing waste and achieving cost savings. This method requires high quality goods (ideally with no defects), highly accurate knowledge of demand and extremely reliable transport and logistics systems. Geographic concentration of suppliers often is associated with JIT since it facilitates more efficient transport of goods from multiple suppliers (Estall, 1985).

When trade establishments face uncertainty in demand, they use an ordering method with built-in buffers to account for known and unknown variability in demand.  This method is the classic economic order quantity (EOQ) formula, which was developed about 100 years ago (Harris, 1915) and remains widely used today (Waters, 2001).

#### 3.2.6.2 The Push-Pull Boundary

Ultimately, it is helpful to characterize demand uncertainty by visualizing where it can occur in the supply chain. In fact, the flow of goods can be generated in one of two ways:

- “Push” system: The demand for goods downstream is forecast by upstream suppliers who then produce goods based on the forecast demand. True demand by downstream consumers (especially the end customer) is not known with certainty.
- “Pull” system: The demand for goods downstream directly generates orders upstream. As such, a high level of certainty is embedded in this process.

Pull systems are preferable since greater certainty in demand translates into lower inventory needs and decreased waste in production. However, supply chains typically fall short of this ideal due to disparities in visibility, or information, among supply chain members. As a result, the provision of supply typically becomes uncoupled from demand at some point upstream. Referred to as the “decision point”, this is where production becomes driven by supplier forecasts rather than actual consumer demand (Hoekstra and Romme, 1992). The decision point can be located anywhere upstream of sales in the supply chain, thus forming the “push-pull boundary” as shown in Figure 9 (Rich, 1995).

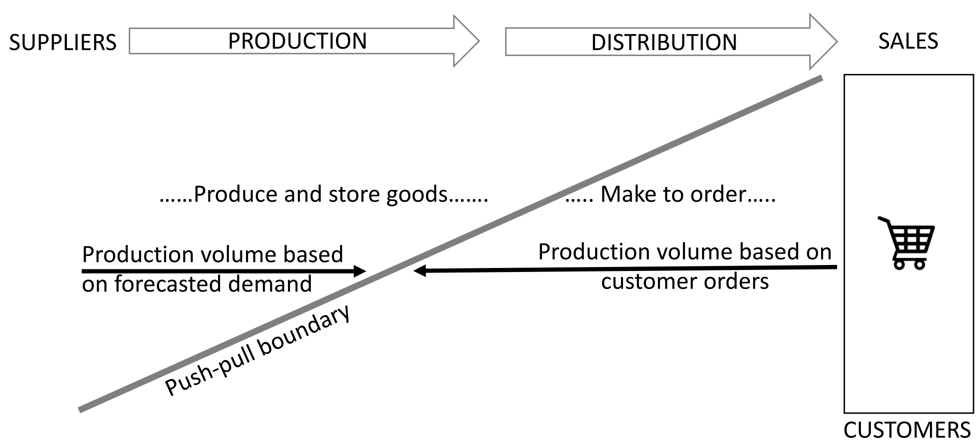



**Figure 9. Potential locations of the push/pull decision point.**

Push/pull behavior and the decision point have important analytical implications for freight demand models:

- Clearly, orders for goods sometimes are precisely in line with end consumer demand while in other cases orders are generated by forecasting processes that include buffer stocks to address uncertainty in demand.

- Suppliers near to the end consumer are more likely than upstream suppliers to use demand-driven (pull) orders.
- Improved supply chain coordination, especially the sharing of information on customer demand, can reduce costs and improve efficiency of upstream suppliers who are enabled with "forward visibility" in this process.

In other words, demand forecasting methodology differs depending on (a) the location of the supplier in the supply chain and (2) the level of coordination among members in the supply chain. For example, an upstream supplier in an uncoordinated supply chain may rely on time series demand forecasting and use the economic order quantity (EOQ) formula to establish the size and frequency of its orders. Suppliers in coordinated supply chains or those close to the end consumer will need less buffer inventory and can utilize a cost-effective JIT delivery system.

#### 3.2.6.3 E-commerce

In many ways, e-commerce is the penultimate example of end consumer visibility since it makes information sharing between consumers and producers so easy and fast. Consumers can use an online system to express their product specifications to a manufacturer or online retailer (or e-tailer), at which point this supplier can fill the customer order. As such, e-commerce reduced overall reliance of supply chains on intermediary trade agents (especially wholesalers and brick-and-mortar retailers), thereby shortening the length of the supply chain while simultaneously modifying its configuration (Figure 10).

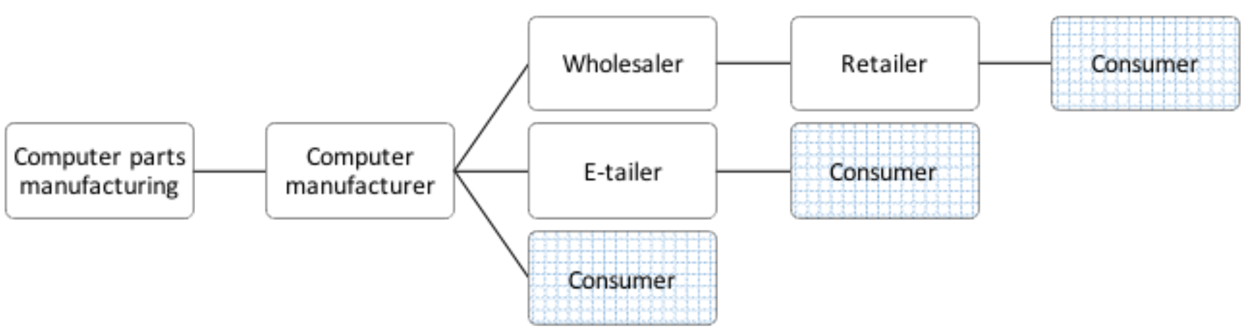



**Figure 10. Supply chain reconfigurations due to e-commerce.**

E-commerce also impacts the delivery system. Orders from online retailers are typically delivered using postal or parcel services. In-store pickup is sometimes offered by establishments that have both online and a brick-and-mortar stores. The e-tailer may stock goods or use “drop shipping”, where the e-tailer transfers orders directly to the manufacturer, a wholesaler, or another retailer who then fills the customer order (“Drop Shipping,” 2019).

The proof of concept describes implementation results from an e-commerce analysis.

### *3.2.7 Integrated framework*

Figure 11 illustrates the components that operationalize the various building blocks and shows their integration. An “X” denotes elements that have been implemented so far. The proof of concept in this chapter and Chapter 5 are the key references that document the implementation to date.

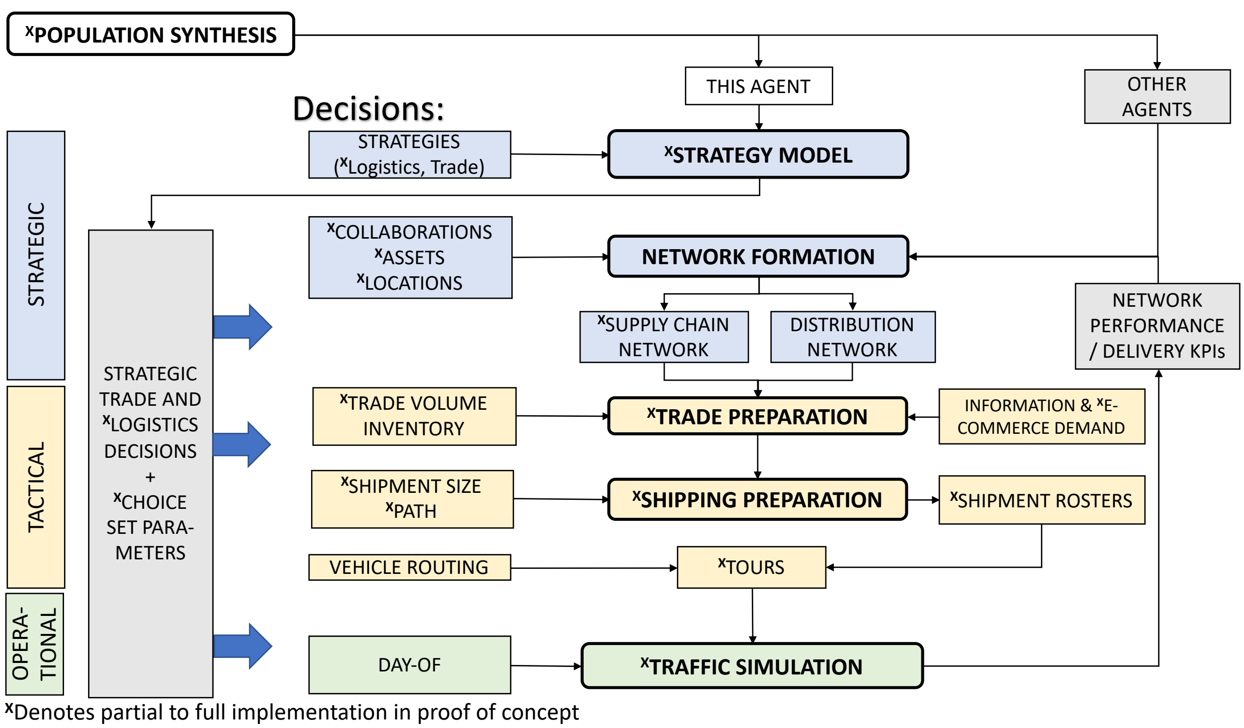



**Figure 11. Overview of model components.**

The first model component addresses the task of generating a synthetic population of agents. This population will be used in simulation runs of the model. The synthetic population is intended to be realistic but it is not an exact replica of the business establishment population. The characteristics of establishment agents in the synthetic population of firms includes:

- Primary industry;
- Location (region), including accessibility to transportation facilities;
- Size measures: Revenue, employment;
- Assets and their characteristics, especially fleet size and mix, types of facilities at site (manufacturing, storage, parking), and facility characteristics (floor area); and
- Firm membership of establishments.

Next, model agents make strategic decisions regarding trade and/or logistics. The resulting strategies inform their remaining long-term, medium-term, and short-term activities. The discussion of strategies and strategic decisions is covered extensively in Section 3.2.3 and in Chapter 5. This model focuses on sourcing, sales, and transport strategies since these types of strategies have a major impact on origin-destination patterns and transport decisions, which are both pivotal to simulating freight transportation. The next several subsections describes each component in greater detail. The proof of concept discussion in Section 3.3 describes the methodology and data sources that are employed in the initial implementation. Otherwise, methodology for most stages is to be determined but is expected to rely on statistical models, including discrete choice models, and heuristics.

### *3.2.8 Network formation models*

This stage first simulates supply chains, distribution networks, and firm assets, then simulates location decisions for fixed assets. Supply chain formation and distribution network formation are performed in the same stage as asset selection because of partnership needs are driven to some extent by availability of intrafirm assets. For example, trade firms that own vehicle fleets will not seek a transport agent partnership.

Supply chain network formation is initialized by simulating partnerships between trade agents (Figure 12). Relationship type (e.g., spot market versus contractual) will be modeled as part of this. Next, each logistic entity forms a physical collection of facilities and routes, which collectively is referred to as its distribution network. Figure 13 outlines the analysis steps that are planned for the distribution network simulation. Carrier agents, whether affiliated with a purely carrier firm or a private trade firm, will choose capacity and levels of service (e.g., geography, price). Finally, this stage simulates asset decisions include fixed trade assets (examples: factories, retail stores); facilities dedicated to transport and logistics (examples: warehouses; intermodal yards); and vehicle fleets.

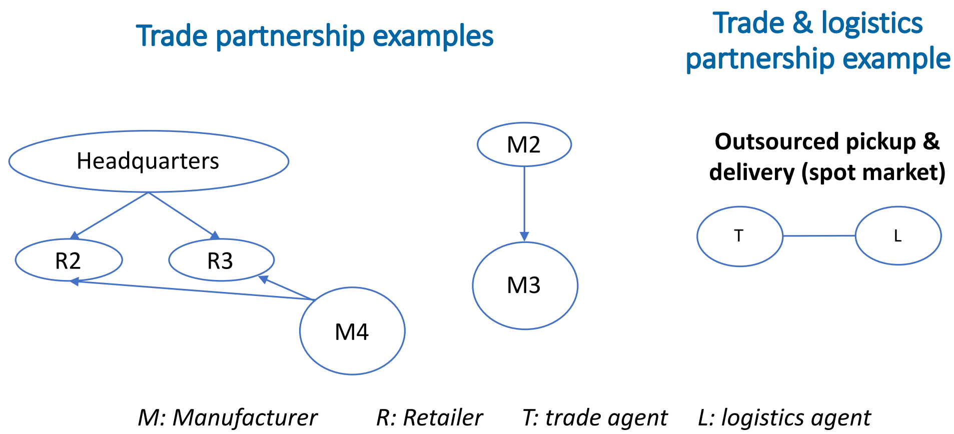



**Figure 12. Examples of supply chain collaboration.**

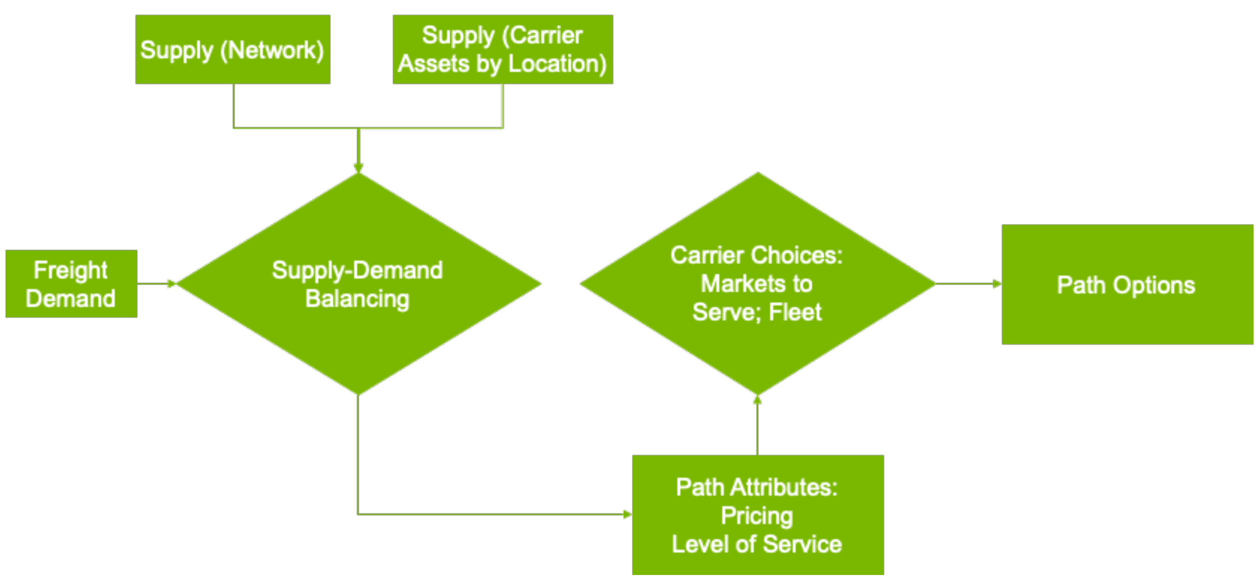



**Figure 13. Development of path options by logistics agents.**

The traffic simulation requires a detailed location to be identified for each vehicle trip. Detailed locations are also needed to assess the characteristics of potential transport and logistics path options, which are evaluated in the next component. To support these needs, a *Parcel Location Algorithm* assigns each establishment to a specific parcel of land. This process will be extended to assign each vehicle to a specific parcel as its base of operations or home station.

### *3.2.9 Mid-term models: trade and shipping preparation*

The trade preparation models simulate the steps that trade agents take as they prepare to trade goods with one another. First, each trade agent forecasts demand for its products, then sets its output levels in accordance with the forecast. The demand forecast has two components: (a) predicted demand and (b) realized orders. As described earlier, access to information and proximity to the customer will be used to determine the extent to which each establishment relies solely on forecasts versus known orders. Next, the buyer determines its annual volume requirement for each input that it uses. The demand forecast informs the resulting procurement decisions. Each trade establishment determines the volumes of inputs to procure in order to support the forecasted demand and inventory (safety stock) needs. Finally, demand for goods is realized when buyers, after determining their input needs, place orders that will meet their needs. The output of this step is a list of inbound orders, outbound orders, production volumes and inventory volumes for each establishment. Currently, the supply of e-commerce goods is assumed to use perfect information on demand from households (discussed more in the proof of concept and in Stinson, Enam, et al., 2019).

Figure 14 shows the flow of the trade preparation module and conveys how the push-pull boundary is operationalized. At some point, every seller receives an order, or information on realized demand, from the buyer. However, depending on information and visibility, sellers receive this information either before or after making production and procurement decisions. The information or customer proximity mechanism is labeled “Visibility” in the figure. Sellers with higher visibility will have greater weighting of actual demand, while those with less visibility will have greater weighting on forecasted demand. Through this mechanism, the so-called bullwhip or Forrester effect (Forrester, 1961 and Chen and Lee, 2017) will be assessed. This will be operationalized by first evaluating demand among end consumers, then evaluating forecasted demand for each upstream layer of the supply chain, using the visibility model at each step.

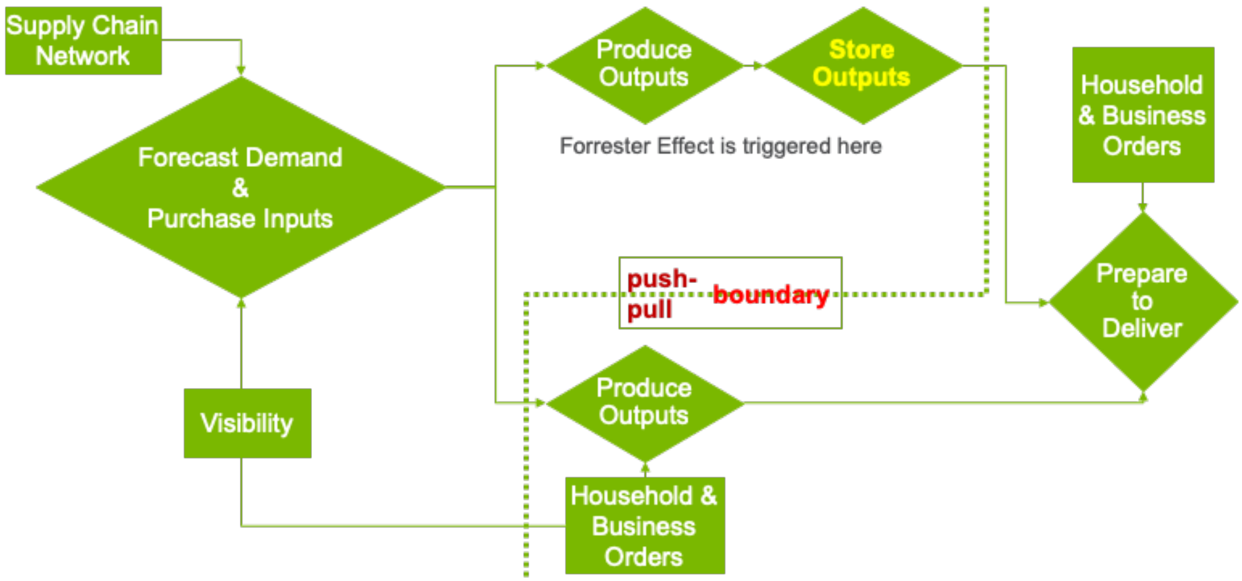



**Figure 14. Trade preparation and operationalization of the push-pull boundary.**

The shipping preparation models generate shipments and the high-level transport and logistics path of each shipment. Annualized volumes from the previous stage are converted into shipments using the EOQ or JIT formulation, depending on information effects as described earlier. EOQ is used for all shipments in the initial implementation. Each resulting shipment is a customer order that must be delivered. This process simultaneously generates shipment size and frequency.

Figure 15 illustrates the steps that convert customer orders into freight tour itineraries. First, for each customer order, the shipper and/or receiver selects a transport and logistics path with attributes that meet shipment requirements. This process uses high-level path attributes: mode(s) and number and location of transshipment points. Next, shipments are assigned to vehicles and vehicles to tours. The sequence of stops along the tour by time of day is planned.

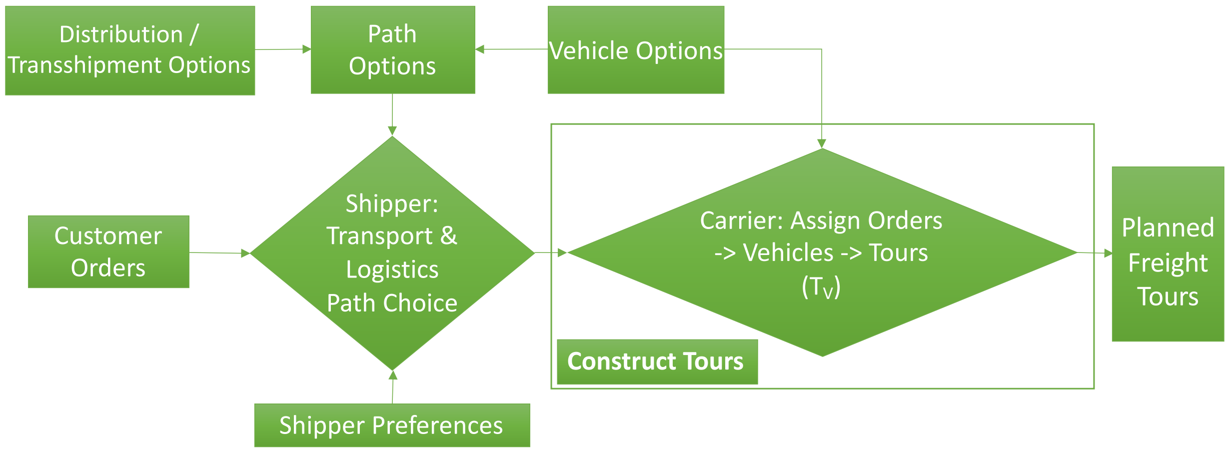



**Figure 15. Shipping preparation models.**

### *3.2.10 Short term: multimodal traffic simulation*

The actual exchange of goods, or shipping, occurs in the Short-Term layer. This layer involves the simulation of one full day of activities and trips in the selected region. This study leverages the existing dynamic traffic assignment (DTA) model that is contained in the open-source POLARIS software (Auld et al., 2016; also see Auld and Mohammadian, 2012). This freight framework is being implemented for the Chicago metropolitan region, which already contains a full-scale implementation of activities and trips for passenger agents in the region. As a result, this implementation permits a comprehensive, simultaneous assessment of both passenger and freight traffic.

To supplement the proof of concept implementation, the next steps are to develop a mechanism by which agents adjust their plans based on performance of the transportation system. Furthermore, Figure 16 shows the planned design features that will permit the analysis of freight movements that are driven by emerging trends. In particular, the design features include the ability to model express or on-demand deliveries as well as expanding the pool of commercial vehicle operators and assets to support crowd-sourced shipping, or crowdshipping (Punel & Stathopoulos, 2018). These extensions will leverage the dynamic nature of the existing DTA and the fact that both freight and passenger populations are modeled simultaneously.

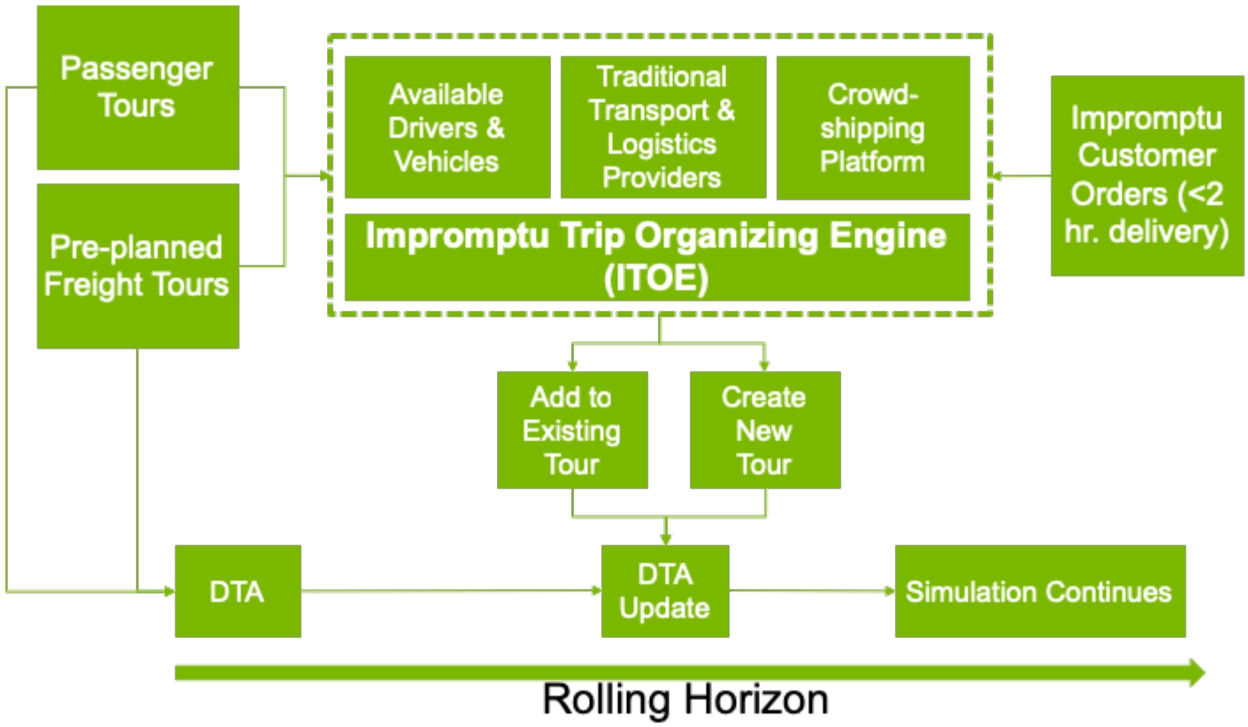



**Figure 16. Planned extensions for on-demand delivery and crowd-shipping.**

## 3.3 Proof of Concept

This section contains the entire content of the paper that is mentioned at the beginning of the chapter.

### *3.3.1 Abstract*

This study summarizes the first stage in the implementation of an agent-based freight modeling system that has a global representation of agents and detailed modeling of a large-scale transportation network. The model is used to evaluate the transportation and energy impacts of goods movement across urban and national scales. The framework is implemented within POLARIS, a C++-based Planning and Operations Language for Agent-based Regional Integrated Simulation, which consists of an activity-based modeling (ABM) and dynamic traffic assignment (DTA) system that has robust features for passenger travel. This platform provides a tool to model interactions among consumers, producers, and the transportation system. The main objective of this initial implementation is to implement a freight model within POLARIS following an agent-based paradigm with behavioral and simulation methods. This paper presents the initial framework and illustrates the

application of the model. Building upon earlier works, a parcel location assignment algorithm for business establishments in the population is documented, along with a method for estimating establishment production and consumption volumes. In addition to population generation, other features of the model include push-pull supply chains, multimodal path choice, choice of transportation logistics node, and dynamic traffic assignment. A module with e-commerce supply and demand was also developed to analyze the effects of e-commerce delivery on last-mile energy use and congestion.

### *3.3.2 Introduction and background*

Transportation demand models are powerful tools for evaluating current and future demands on transportation system. They feature prominently in energy and emissions analysis due to transportation's effects in these areas. Including freight modes in transportation models is critical for accurate assessment of the transportation impacts.

Nomenclature

| | |
|---|---|
| ABM | agent-based model |
| DTA | dynamic traffic assignment |
| VMT | vehicle-miles travelled |

For instance, freight trucks comprise about 10 percent of VMT in the US and consume about 30 percent of transportation energy (US Department of Transportation, 2017). If freight is left out of transportation demand models, then a large part of VMT, energy and emissions cannot be sufficiently accounted for.

However, modeling freight transportation has many challenges, which are mainly due to the large variety of agents that are involved with freight production, consumption, or carriage, as well as the wide array of options that are available to these businesses in forging business partnerships, choosing modes of transport, and other decisions. Further, they operate in a global environment, which complicates the decision-making process and expands the number and nature of agents that are relevant to local agents.

Aggregate models, which are designed mainly for estimating total flows or trips (Beagan et al., 2007 and Abdelwahab and Sargious, 1992) at the zonal level, often focus on major decision factors such as distance while ignoring more nuanced factors. For instance, early models generally did not consider logistics activities, such as transloading, that often occur on the shipment journey. Subsequently, models such as Tavasszy et al. (1998), Boerkamps and van Binsbergen (1999) and Jin et al. (2005) began to include logistics in shipment paths but still had some limitations due to the use of aggregate techniques.

To address these complexities in the freight environment, a detailed freight demand forecasting framework that utilizes integrated ABM with DTA was proposed by the authors in Stinson, Auld et al. (2019). In the full framework, variety among businesses is handled by using individual agents. Interactions are handled by modeling behavioral and economic preferences of agents. Agent response to potential partnerships and their environments can be handled by modeling their behavioral preferences – in this model, strategic choices are explicitly made, which have major impact on transportation externalities (M. Stinson et al., 2019). Responses to traffic conditions are evaluated using DTA. Design features to address major gaps in extant agent-based freight models, which focus on both trucks as agents (Hunt and Stefan, 2007) and on businesses as agents (examples include de Jong and Ben-Akiva, 2007; Samimi et al., 2013; Urban et al., 2012; Hong et al., 2017; Stinson et al., 2017; and Alho et al., 2017) were planned in detail. These earlier efforts are limited in the way they handle agent interactions, especially shipper-carrier relationships and firm partnerships. Further, the full proposed model introduces several important key features, including modeling of strategic decisions, the operationalizing of the push-pull boundary, and key emerging trends. Other frameworks are relevant but have not been fully implemented: Holmgren et al. (2012) and Roorda et al. (2010) elaborate on agent decisions and interactions, such as contractual relationships, and to some extent covers sensitivities to emerging trends in policy, business and technology.

Due to the number of features to be modeled, it is not possible to develop all features immediately. Therefore, an incremental development approach is used wherein major elements from each stage of the model are implemented, in some cases with placeholder model formulations and parameters, in order to operationalize the model. This document, then, describes the initial implementation of the model. The initial implementation is centered on the Chicago metropolitan region in the Midwestern US. Thus, the model contains the most detail for this region while also including agents and transportation network features that are external to the region.

The rest of the paper focuses on this initial implementation. First, an overview of the model architecture is presented and its key conceptual and operational features are summarized. Second, each component of the initial model is presented in more detail. Test results are shown for various components in order to demonstrate the application. Finally, a summary and listing of next steps is presented.

### *3.3.3 Approach: model overview*

Three tiers serve as a schematic guide for the full framework (Figure 17, left) (Fischer et al., 2005) in order to frame an understanding of agents, decisions, and the operational structure of the model for implementation in a computational framework. In the strategic layer, model agents make long-term (3 mo.-1 yr.) decisions regarding business-to-business collaborations, trade more generally, and transportation and logistics capacity. They form strategies that guide their long-term as well as short-term activities. Our model focuses on sourcing, sales, and transport strategies since these types of strategies have a major impact on origin-destination patterns and transport decisions, which are both pivotal to simulating freight transportation. In the medium-term (1 day-3 mo.) tactical layer, agents engage with their collaborators to arrange specific trade activities—in particular, the virtual exchange of shipments and planning of associated transportation and distribution activities. Practical plans such as driver schedules emerge here. The operational layer simulates the resulting physical flows of vehicular traffic. It also models short term (<1 day) decisions of agents such as route changes, parking decisions and fulfillment of express (1-2 hour or same-day) delivery demand. An e-commerce delivery module, which is coupled with a household-level e-commerce module from the POLARIS passenger component, currently feeds into the DTA also. The right pane of the figure demonstrates the structure of the initial implementation and illustrates which features are used in the initial version.

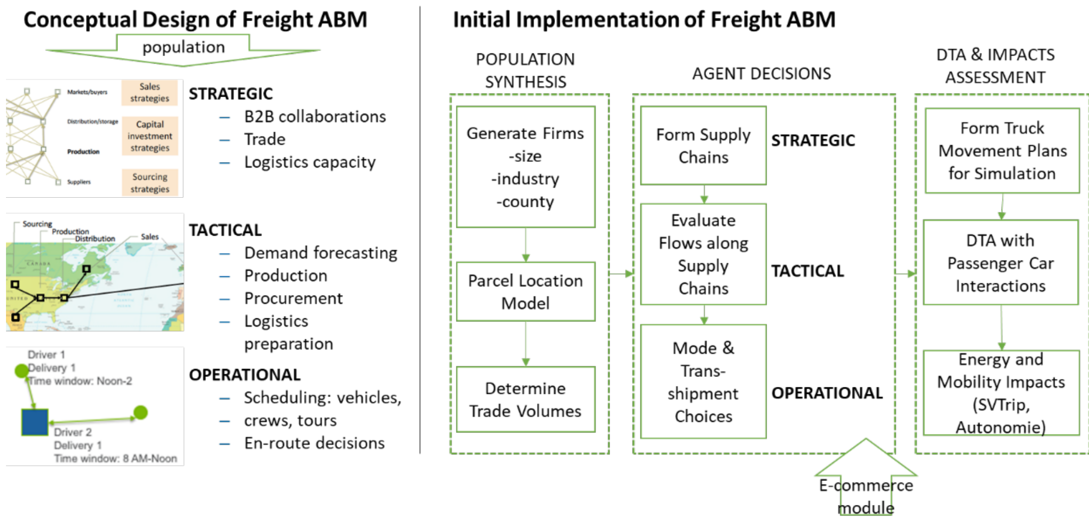



**Figure 17. (left) Conceptual features; (right) Initial implementation.**

The level of geographic detail and network resolution in each layer differs as shown in Table 1 below. Locations of establishments outside of the region are currently at a coarse geographic resolution between US counties and regions of the world. Thus, the framework incorporates the ability to model a wide span of geographies. The initial e-commerce module includes last-mile delivery trips, i.e., trips from depots to homes and businesses.

**Table 1. Geographic level of detail for each layer of the model.**

| Layer | Establishment location in Chicago Region | Network |
|---|---|---|
| Strategic | County | Sketch representation of major links and nodes |
| Tactical | Parcel | Sketch representation of major links and nodes |
| Operational | Parcel | All links and nodes |

### *3.3.4 Implementation and results*

This section presents the implementation of the model and initial results at each stage.

#### **3.3.4.1 Population synthesis**

Currently, model agents include business establishments that engage in trading goods. They may produce goods, consume goods, or both. The agent population is generated by enumerating the number of establishments in each US county based on industry category and size (number of employees). The input data for this is the US Census Bureau County Business Patterns (CBP) (2020c) data. Fig. 2 shows how the input data are transformed into a set of model agents.

| NAICS code | County | 1-5 employees | 6-10 employees | 11-20 employees | |
|---|---|---|---|---|---|
| … | | | | | |
| 313110 (Fiber, yarn and thread mills) | DuPage | 2 | 0 | 1 | … |
| … | | | | | |

| NAICS code | County | Employees |
|---|---|---|
| … | | |
| 313110 (Fiber, yarn and thread mills) | DuPage | 1 to 5 |
| 313110 (Fiber, yarn and thread mills) | DuPage | 1 to 5 |
| 313110 (Fiber, yarn and thread mills) | DuPage | 11 to 20 |
| … | | |

**Figure 18. Input: marginal population totals; output: synthetic population.**

Next, each agent is assigned to a specific location for simulation purposes. This is needed because counties are quite large (on the order of one hundred square miles). Let i denote an index for the set of industries I, z denote an index for the set of zones Z, pz denote an index for the set of parcels in zone Z, and e denote an index for the set of establishments E. The parcel location algorithm pseduo-code is as follows:

For every i in I:

For every z in Z:

Compute $S_{iz}$ = Sum(Employment$_{i,z}$)

Determine ranking $R_z$ (order from highest to lowest) based on $S_{iz}$

For every e in E:

Form set of candidate zones, $Z_c$ as follows:

For firms with 5,000+ employees, $Z_c$ = zones with highest $R_z$

…

For firms with 5 employees, $Z_c$ = all zones

Use Monte Carlo draws to assign each e in E to a z in Z

Finally, after assigning each establishment to a zone, its exact parcel location within the zone is simulated using a Monte Carlo draw from the set of all parcels with a commercial land use in that zone. Figure 19 shows the synthetic agents in the Chicago region at their simulated locations.

Trade volumes, or production and consumption volumes, are then assessed for each establishment. To do so, first a set of rates is computed for each industry. The input data for the rates are the CBP data and the US Department of Transportation Federal Highway Administration Freight Analysis Framework (FAF) data, which has the dollar value of freight flows that are produced and consumed in various areas of the country. The FAF data are in terms of commodity flows, therefore assumptions were used to create a crosswalk between commodity type and industry type that produces or consumes each type of commodity. First, total employment by industry is summarized using the CBP data. Next, total production and attraction values are divided by employment in each industry to estimate the value produced and consumed by each employee in each industry for a given commodity type. Finally, using these rates, production and consumption volumes are simulated for each model agent based on industry type and number of employees.

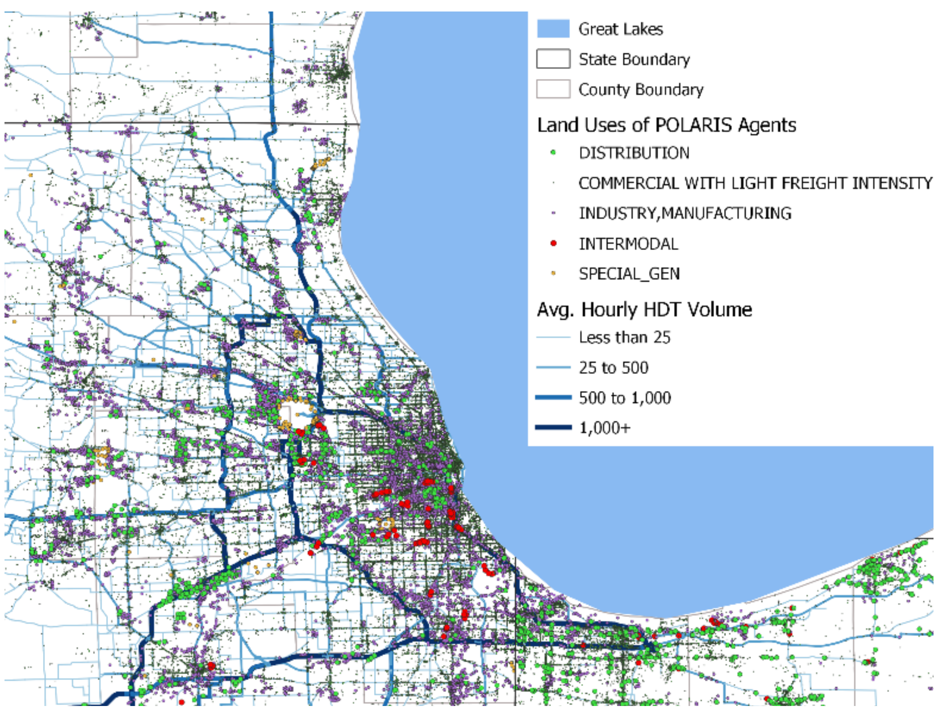



**Figure 19. Synthetic population with parcel locations in the Chicago metropolitan region.**

#### 3.3.4.2 Supply chain formation and flow evaluation along supply chain

Ultimately, a relatively elaborate scheme of collaborations among business is planned for implementation. For now, a supplier selection model is used wherein each buyer selects a supplier to supply an input commodity. First, a candidate set of suppliers is formed by randomly drawing ten potential suppliers based on the suppliers' production industry and simulated production volume. Second, the well-known Multinomial Logit model described in Ben-Akiva and Lerman (1985) is used to select a supplier based on its characteristics as well as a stochastic element. The utility formula is a placeholder as data for specifying a model and estimating its parameters are currently being collected. The utility of supplier $s$ for the initial model is $U_s = \beta' X_s + \epsilon_s$ where $\beta$ is a vector of parameters to be estimated, $X$ is a vector of supplier attributes that includes its size, location (foreign vs. in-region vs. other US-not in region), and distance to the buyer; and $\epsilon_s$ is an error term. Figure 20 illustrates this process.

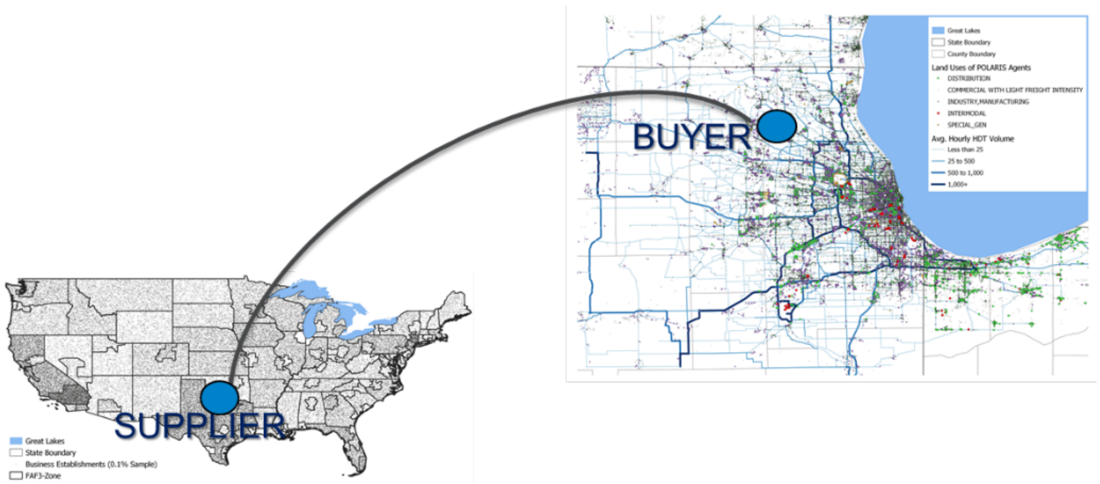



**Figure 20. Supplier selection (supply chain formation).**

At this time, the amount of flow is assumed to be the consumption volume that is needed by the buyer. In the future, the methodology will be revised to account for the production capacity of the seller.

In the remaining downstream areas of the model, the buyer-supplier pair are treated as a single decision maker, as from this point they are effectively working together to furnish the needs of the buyer.

#### 3.3.4.3 Mode and transshipment (path) choices; and DTA

In this step, mode and transhipment path options are chosen for each shipment. The model is multimodal and includes the following modal options: rail carload, rail-truck intermodal, full truckload (FTL), less-than-truckload (LTL), parcel, and air. Transshipment locations are included as part of the multimodal path option. These options include:

- Choice of airport (O'Hare, Rockford, Midway, etc.)
- Choice of trucking terminal / crossdock location (for LTL / FTL shipments)
- Choice of rail terminal (e.g., BNSF Global-I)

The methodology used for path selection is as follows (Caplice, 2016 and Chen, 2014). For supplier-buyer pair and each candidate path, the optimal shipment size and frequency is determined based on commodity characteristics, which is consistent with inventory theory. The first characteristic is discount rate, which covers the cost of physically storing goods (e.g., warehousing leasing rates) as well as perishability factors. The second characteristic is the value of the good, which is normalized as the value density in dollars per pound. Goods are distinguished as Bulk, Intermediate, and Finished goods, with discount rate and value density varying by these distinctions. The optimal shipment size Q*p for path p is calculated as:

$$\boldsymbol{Q_p^*} = \sqrt{\frac{\boldsymbol{2DC_{order}}}{\boldsymbol{C_{holding}}}} \tag{1}$$

where D is annual demand and $C_{order}$ and $C_{holding}$ are the order and holding costs, respectively. Order cost is computed as the value density divided by the square of path transportation cost per unity weight while holding cost is the value density multiplied by the discount rate.

Total relevant path cost, $TRC_p$, using path p is then computed as the sum of total transportation cost, total order cost, total holding cost, and total pipeline inventory cost over the annual demand time horizon, which is one year. Transport cost is the unit cost multiplied by distance and D. Order cost is the unit order cost multiplied by D and divided by $Q_p^*$. Holding cost is the unit holding cost multiplied by $Q_p^*$ and divided by two. Pipeline inventory cost is unit holding cost multiplied by D and the transit time.

Finally, path selection is performed as follows. First, the inverse square root of $TRC_p$ is computed for each path p. Second, the sum of these inverses is computed, and the proportion of the sum is computed for each p. Third, Monte Carlo draws are used to select a path based on these proportions. Although this process is somewhat *ad hoc* and serves as a placeholder for a more carefully specified, data driven model and parameters, it is effective in guiding agents toward selecting modes and transhipment locations that are relatively low cost yet account for tradeoffs in inventory and transport costs.

To illustrate the output, mode shares for agriculture, hunting, fishing and forestry products, which are bulk goods, as estimated by this process are shown in the left pane of Figure 21 below. The right pane of the figure is a bandwidth plot of heavy duty truck (HDT) traffic volumes into, out of and near the Chicago O'Hare International Airport, which illustrates results of the transshipment location selection aspect of the path decision.

For the DTA, each shipment is assigned to one vehicle trip. Fleet selection for each establishment as well as routing algorithms are in progress and will provide a significant enhancement to the current process. For now, Monte Carlo draws are used to simulate the use of truck type and powertrain features. For purposes of brevity, this process is not described in detail here.

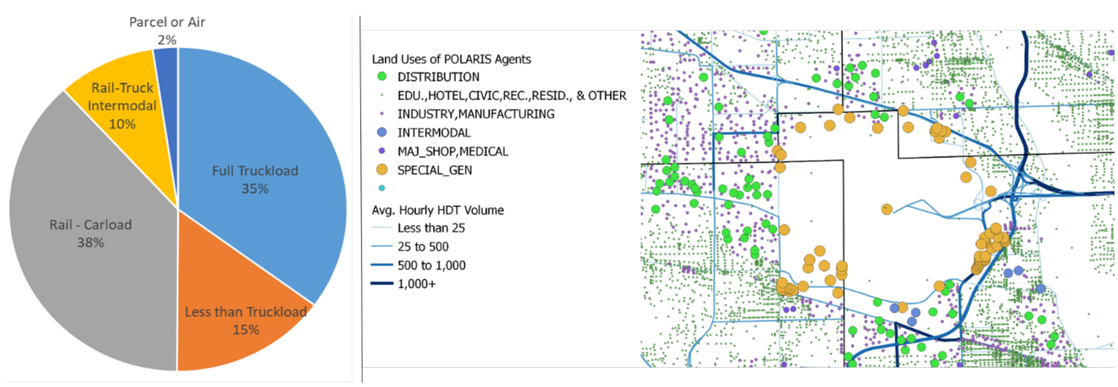



**Figure 21. (left) Mode shares of bulk goods; (right) Truck volumes near O'Hare Airport.**

### 3.3.4.4 E-commerce module

The e-commerce module is fully documented in Stinson, Enam et al. (2019). Briefly, its workings can be summarized as follows. Demand for e-commerce deliveries is generated for each household in the metropolitan region. The demand estimates are provided to an external procedure that uses Python scripting and Geographic Information Systems (GIS) to generate delivery routes for the demand based on observed depot locations and the traveling salesman algorithm. The resulting parcel delivery truck tours are fed into POLARIS. This process is targeted to be fully integrated within POLARIS in the future. Example output from this process is shown below in Figure 22, which illustrates the individual and combined VMT by medium-duty delivery trucks (MDT) and light-duty shopping vehicles (LDV) in the baseline (one delivery per household per week), short-term A scenarios (three deliveries per week) and long-term B/C scenarios (five deliveries per week) with other varying conditions such as high levels of autonomous vehicle penetration in the C scenarios. Full details on the scenario results are available in Stinson, Enam et al. (2019).

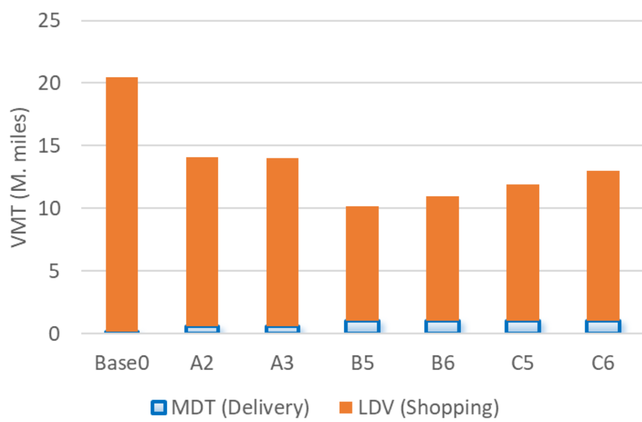



**Figure 22. Estimated VMT associated with last-mile retail activity.**

### *3.3.5 Summary and next steps*

This paper presented the initial development of a freight ABM with DTA. The model structure includes population synthesis as well as agent decision making in the strategic, tactical and operational spheres, and ultimately truck movements that are assigned along with passenger vehicles in a transportation network. Demonstration results for population synthesis, the parcel location algorithm, mode choice, and transshipment location choice have been provided along with an e-commerce application with VMT impacts. Future steps include developing the remainder of the framework according to the proposed outline presented in Stinson, Auld et al. (2019).

# 4 Methods to Measure Attitudes using Natural Language

Parts of this chapter were previously published as Stinson, M. and A. Mohammadian. (2020) "Modeling Firm Transportation Strategy using Big Text Data," proceedings of the IEEE Forum on Integrated and Sustainable Transportation System (FISTS), Delft, The Netherlands, Nov. 3-5, 2020. 

## 4.1 Introduction

The last two decades have witnessed the emergence and widespread adoption of attitudinal factors in behavioral models, which form the foundation of passenger travel behavior modeling. Attitudinal factors, also called latent factors due to their unobservable nature, add value to models that are otherwise quantitative. A mode choice model with only quantitative factors, for example, may measure an individual's sensitivity to travel time while ignoring the individual's preference for modes that are environmentally sustainable. Ideally, behavioral models should consider both quantitative and attitudinal factors, since both types of factors can impact decision making. The development of improved methods in recent years has enabled the inclusion of attitudinal factors in behavioral models to become a more common practice.

Parallel to this trend, a veritable explosion in analytic methods for textual data has occurred in the computer science domain, notably in Natural Language Processing (NLP). NLP, which covers a host of methods that rapidly analyze textual data, holds particular promise for attitudinal analysis applications. However, NLP research and its demonstrations to date focus primarily on real-time applications such as online chatbots and Q&A/search query functions. In other words, applications of NLP mostly involve replacing human functions in situations that previously involved real-time dialogue between two humans.

This study explores the intersection of these two areas and develops attitudinal measurement methods that combine and leverage the beneficial properties of each area. The main idea put forth by this work is that natural language is, in a sense, similar to a survey response; and if we query it in a clever way, we might be able to find the answers that we need to inform model development. In this setup, I treat the NLP-based system as my surveyor, and in doing so, I develop a new way to passively collect attitudinal data that previously could only be collected through elaborate and costly survey efforts.

I propose and build upon two main ideas in this study. First, I propose and test the idea that people or companies will use certain words more or less frequently than others, depending on agent attitudes, strategies, underlying behavioral drivers, etc. Second, setting aside the relative frequency of using various words, I also propose and test the intuitive supposition that individual persons or companies will use certain words differently – for example, in different contexts (with different surrounding words) – depending on what matters to that individual or firm. In this work, I explore these notions, developing quantitative methods to examine both relative frequency of word use and contextual differences in word use. The methods that I develop are based in principle on the foundational work in attitudinal measurement from the psychological, economic, and transportation domains, but in practice shift the measurement methods to algorithms that leverage the most recent computational linguistics paradigms.

In the process, this study addresses a very real data gap in the domain where I conduct most of my research: freight transportation models. Behavioral models that simulate decision-making processes of businesses are limited by a near-complete lack of data in this area. This gap provided the initial motivation for this research. However, this work is relevant to a number of other areas, with potential applications in marketing, psychology, passenger transportation and others. These and other fields can benefit from methods to harvest attitudinal data from large-scale, passive sources.

The remainder of this chapter is structured as follows. The approach and data are discussed first, respectively, followed by a presentation and discussion of the results. The chapter concludes with a summary of findings and suggested extensions.

## 4.2 Approach

This study develops two methods to generate attitudinal measurements using natural language: Bag-of-words (BOW) with simple scaling (SS-BOW) and word2vec with PCA (w2vPCA). in developing these methods, the main challenges are determining (1) what to measure, (2) how to measure it, and (3) what are the implications (if any) for attitudinal analysis. The second and third topics are explored in the next few subsections. In regards to the first question, SS-BOW is based on measuring the relative frequency of certain words among individual companies. In contrast, w2vPCA measures differences in word usage among individual companies.

The reasoning behind these two approaches is as follows. Companies have different goals, and they have different strategies to support their goals. As a result, company reports have varying content. Because of this, one company will use certain words more or less frequently than other companies. Furthermore, a company will use certain words in different contexts than other companies. Ultimately, the frequency and/or in-context usage of certain words in particular are expected to depend on company strategies.

### *4.2.1 Attitudinal measurement methodology: SS-BOW*

This section is an excerpt of the approach from the IEEE paper that is listed in the Acknowledgements section. The paper referred to the algorithm as BOW. This thesis labels it SS-BOW.

The following nomenclature is used. $D$ denotes the set of documents and is indexed by $d$. $W_d$ is the set of words in document $d$ and is indexed by $w_d$. $K$, indexed with $k$, is the set of keywords that is targeted for analysis by the DDE. $K_d$ is the set of keywords present in document $d$ and is indexed by $k_d$. $K_d$ can contain duplicates items. The total count of keyword $k$ in document $d$ is denoted $|k_d|$.

The Simple-Scaled Bag-of-Words (SS-BOW) process can be described as text mining, using a "bag of words" approach, with results subsequently scaled. With this approach, the frequency of words is the variable of interest. Since documents have different sizes and we want to avoid spurious effects that are related to document size, the frequency of each keyword is normalized so that the sum of scaled keyword frequencies is 100 for each document. The normalized value is denoted as $|k_d|^N$ and thus ranges from zero to 100.

The SS-BOW approach uses the following pseudo-code:

For each document $d \in D$:

  Initialize $|k_d|$=0 and $|k_d|^N$=0 for each $k \in K$.

  For each word $w_d \in W_d$:

    For each keyword $k \in K$:

      If $w_d == k$, $|k_d| = |k_d| + 1$

  For each $k \in K$, compute $|k_d|^N$ as $|k_d|$ divided by the sum of $|k_d|$ across $k$.

The resulting set of $|k_d|^N$ are used to model firm strategies in a behavioral modeling construct. For each document $d$, the quantities $|k_d|^N$ sum to 100 due to the normalization. These normalized values are input to the strategies model.

The results of the SS-BOW measurements are analogous to those obtained using ratings scales in questionnaires that ask respondents to “Rate this item on a scale from X1 to X2”. However, a chief difference is that while ratings questions typically constrain a respondent to a certain range (e.g., between 1 and 10), the original text data that are used here does not impose such a constraint. This is because these text data are based on the natural occurrence of key words as they arise in the company’s own discussions of its operations. For example, a scale of one to seven constrains the maximum difference between attitudinal measurements to a factor of seven. The methodology at hand has no such constraint.

### *4.2.2 Attitudinal measurement methodology: w2vPCA*

The w2vPCA method combines two separate methods, the word2vec algorithm and Principal Components Analysis (PCA). This section describes each of these methods and how they are uniquely adapted and combined to generate attitudinal measurement data.

#### 4.2.2.1 The word2vec algorithm

The word2vec algorithm emerged in 2013 as a novel means to conduct quantitative text analysis (Mikolov, Chen, et al., 2013). The algorithm essentially relies on the notion that similar words are used in similar contexts; therefore, similarity and other word relationships can be quantitatively generated by analyzing the contexts of each word. The following description of the word2vec algorithm, and its historical context, follows Lee (2018) and Socher (2017).

Prior to the 2010s, the meaning of a word was typically represented using a “one-hot” vector (Figure 23) in quantitative analysis. The dimensionality of such a vector is based on the number of unique words that are being analyzed. For a given word, the vector contains the value “1” for the dimension that corresponds to the word, and “0” everywhere else. Unfortunately, since a typical vocabulary may exceed 10,000 words, such vectors are extremely high-dimensional and computationally cumbersome. Further, as Figure 23 shows, similarity between two words is not captured since all word vectors are orthogonal.

```
container= [ 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0]
bin        = [ 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0]
```

**Figure 23. Traditional one-hot vectors.**

Despite these issues, the use of a vector inherently is valuable for word representations because it can leverage powerful mathematical applications involving linear algebra. Word2vec also represents words as vectors, but uses a much denser representation using *N* dimensions, with each dimension representing a concept instead of a unique word.

Word2vec achieves this compactness by using context words to generate the *N* values for a given word. It generates similar vectors for words that have similar meanings since such words are commonly used with similar context words. Thus, the algorithm obviates the dimensionality and word similarity issues that are associated with one-hot vectors. Levy et al. (2015) note that a formal proof is desirable, but developing one is beyond the scope of this thesis.

The analyst provides the following inputs to word2vec: (a) the input corpus, (b) the window width of size *m* to identify context, which is the subset of words located in the same sentence and within *m* words from the key word (this analysis uses the entire sentence), (c) dimensionality *N* of the output vector space. This analysis uses the Continuous Bag of Words (CBOW) solution approach with hierarchical softmax sampling in training, which are now described (alternatively skip-gram modeling with negative sampling can be specified). The problem is set up in a two-layer neural network (Figure 24). More explanation follows the figure, with the discussion informed by Mikolov, Sutskever, et al. (2013) and Rong (2014).

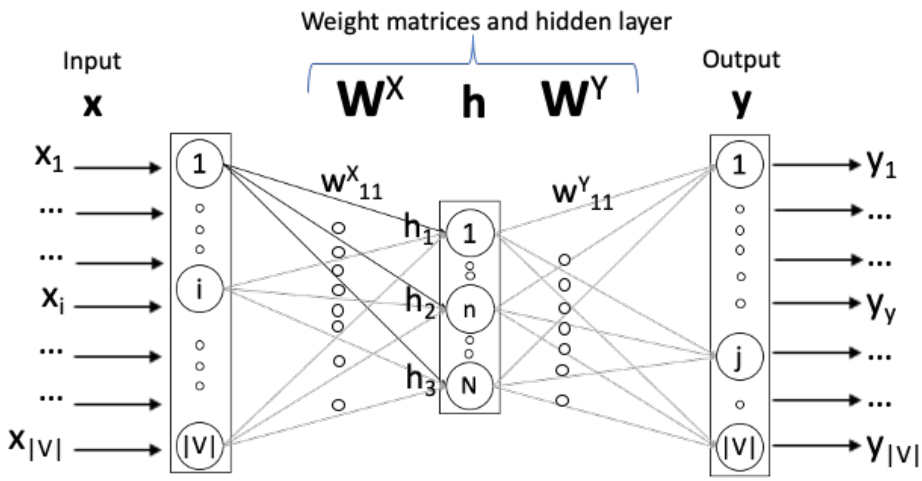



**Figure 24. Two-layer neural network for word2vec algorithm**

Based on a set known context words, the word2vec model predicts the unknown target word. Denote the vocabulary as V and the vocabulary size (number of unique words) as |V|. Let $i, j \in |V|$ denote indices for feature positions in the input and output feature vectors, respectively, let $c \in C$ denote the index of an element of the input context word set, and let $n \in N$ denote the index of an arbitrary element in the hidden layer. The following vectors and matrices are defined:

$\mathrm{x_c}$: "one hot" vector ($|\mathrm{V}|\mathrm{x}1$)

$\mathrm{X} = CBOW$ input vector ($|\mathrm{V}|\mathrm{x}1$), formed using $\{\mathrm{x_c}\}$ (described below)

$x^w, y^w$ = words from V

$\mathrm{y}$ = output feature vector ($|\mathrm{V}|\mathrm{x}1$) of the target word (with elements $\mathrm{y_j}$)

$\mathrm{h}$ = vector representing the hidden layer (Nx1)

$\mathrm{W^X}$ = weight matrix that transforms X into h ($|\mathrm{V}|\mathrm{xN}$)

$\mathrm{W^Y}$ = weight matrix that transforms h into y ($\mathrm{Nx}|\mathrm{V}|$)

The algorithm iterates through the set of target word positions. Therefore, only one input vector and one output vector are used at a time. When only one context word exists, then $X=x_c$. When multiple context words are input, CBOW computes X as:

$$X = \frac{1}{C}(x_1 + x_2 + \cdots + x_C) \tag{2}$$

Next, the hidden layer vector is computed:

$$h = (\mathrm{W}^{\mathrm{X}})'X \tag{3}$$

Using the j[th] row of $(\mathrm{W}^{\mathrm{Y}})'$, each element of y is computed as:

$$\mathrm{y_j} = \mathrm{prob}\left(\mathrm{y_j^w} \middle| \mathrm{x_1^w}, \ldots, \mathrm{x_C^w}\right) = \frac{\exp(\mathrm{u_j})}{\sum_{\mathrm{r}=1}^{|\mathrm{V}|} \exp(\mathrm{u_r})} \tag{4}$$

where the term $\mathrm{u_j}$ is referred to as the score and is computed as:

$$\mathrm{u_j} = (\mathrm{W}^{\mathrm{Y}})'_{\mathrm{j},\cdot} \cdot \mathrm{h} \tag{5}$$

The above expression with exponentiation is called the softmax function, or a log-linear classification model. It produces the posterior distribution of words, which is a multinomial distribution.

The loss function is computed as:

$$\mathrm{E} = -\ln \mathrm{y_j} = -\mathrm{u_j} + \ln \sum_{\mathrm{r}=1}^{|\mathrm{V}|} \exp(\mathrm{u_r}) \tag{6}$$

The typical solution method at this point is Stochastic Gradient Descent (SGD) with modifications to expedite processing. The objective is to minimize the loss function, or equivalently, to maximize the probability that the predicted word is the actual observed word (max $\mathrm{y_j}^*$ = max $\ln(\mathrm{y_j}^*)$ = max(-E)), where the asterisk refers to the actual observed word. In practice, the average loss summed across all target words is typically used (Lee, 2018). More details on SGD are available in Rong (2014) and are widely available online.

#### 4.2.2.2 Principal components analysis (PCA)

This section provides a brief overview of Principal Component Analysis (PCA) for the one-dimensional and two-dimensional cases (Shlens, 2005). PCA is used to compress a multi-dimensional dataset into a smaller number of variables or dimensions while limiting the loss of information contained in the subset of variables.

This is accomplished by analyzing the covariance among variables and measuring which combinations of variables capture the largest amount of variance in the dataset. The resulting new components are independent of one another.

Covariance measures how much two random variables, X and Y, vary together. The sample covariance is computed as:

$$cov(X,Y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

using realizations $x_i, y_i$ and sample means $\bar{x}, \bar{y}$ of the random variables. In PCA, the first step is to calculate the covariance matrix $A$ using the observed data. The eigenvectors and eigenvalues, $v$ and $\lambda$, of this matrix are then computed based on the formula

$$Av = \lambda v \tag{7}$$

The covariance matrix, its eigenvalues and its eigenvectors, and the sample means of each variable are inputs to PCA.

Using these inputs, PCA is performed as follows:

1. Mean transformation: The mean of each variable is subtracted from each value of the variable, for example: $X' = X_i - \bar{X}$;
2. The covariance matrix is calculated using each mean-transformed variable;
3. Eigenvectors and eigenvalues are calculated for the covariance matrix from (2). These eigenvectors are called the "principal components" (PCs);
4. Eigenvectors are sorted in descending order according to their eigenvectors. The eigenvector, or component, with the highest eigenvalue is the 1st principal component.

The resulting component with the highest eigenvalue is the principal component (PC). The one with the second-highest eigenvalue is the 2$^{nd}$ principal component; and so on.

#### 4.2.2.3 Combining word2vec and PCA to analyze attitudes

As noted earlier, the main challenge in using word vectors is identifying a way to learn the different usage of certain words by different companies. Then question, then, is how to go from word vectors to attitudinal measurement?

This next step forms the core methodological contribution of this work. Essentially, the word2vec algorithm constrains each word to have the same vector. This results in an aggregated use of the word across sentences, paragraphs, companies, and so on. The key innovation of this work involves the relaxation of this constraint, which enables the average usage of each keyword to now be specific to individual companies.

Figure 25 shows the first stage in operationalizing this idea. First, the input data are modified, tagging each keyword with a company ID. Second, word2vec is performed on this modified input data. This operation results in different average keyword use measurements, or word embeddings, for each company.

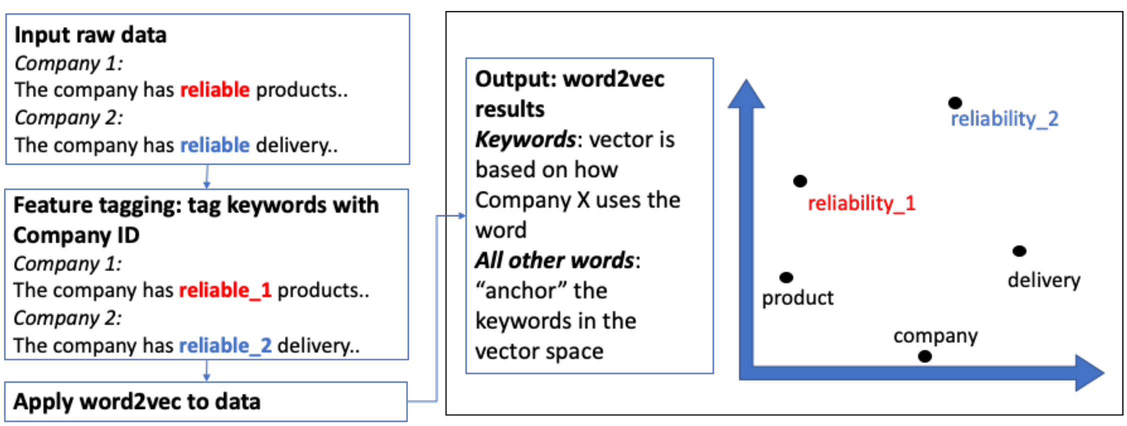



**Figure 25. Key word2vec modification: feature tagging.**

Finally, the question of how to measure differences remains. For this step, I apply Principal Component Analysis (PCA) for each of the tagged keywords and its associated set of company-specific measurements (Figure 26). This process generates measurements based on the company-specific values. Since the word2vec analysis permits any number of dimensions (up to the size of the vocabulary), the resulting word embeddings can be projected from the analysis dimension to any lower dimension. My method uses the first principal component. The appeal of using the first principal component of PCA is that the direction of greatest variance, or difference, among companies is captured. Moreover, the values are not bounded – thus, they provide a natural, unscaled measurement of differences in word usage among companies.

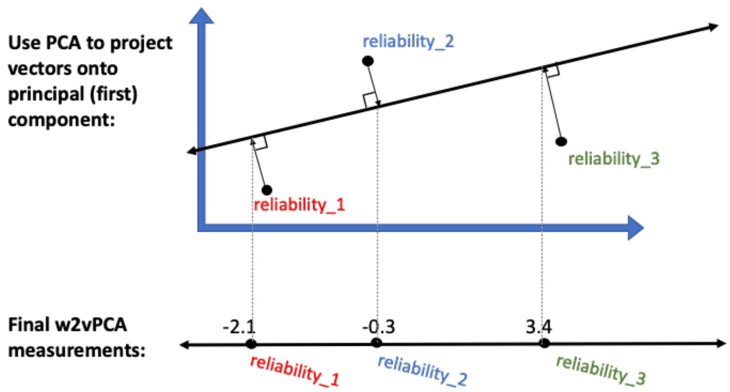



**Figure 26. Measurement process in w2vPCA.**

### *4.2.3 The Attitudinal Data Development Engine (ADDE)*

An Attitudinal Data Development Engine (ADDE) is developed to expedite the processing of large volumes of text, which is both intensive (due to document size) and repetitive. For this particular application, ADDE automatically reads in and processes an annual company report from the United States Securities and Exchange Commission (SEC) EDGAR database (2020) for each company in the study. Python word processing packages are used to support the engine (Richardson, 2020 and NLTK Project, 2020). ADDE can be redirected to read in other sources of text data. ADDE computes and outputs both the SS-BOW and w2vPCA measurements. For the current study, the output of ADDE constitutes the firm strategy data, i.e., the attitudinal measurement for each keyword that is specified by the user.

ADDE iterates through a list of input documents and conducts the following set of processes to input and convert each document into sentences and words, or tokens, for analysis:

1. Read in the report using the "get" command from the "requests" HTTP parsing library
2. Convert the result to text, then to a Python BeautifulSoup (BS) object, then extract the text from the BS object using the BS "get_text" command
3. Split the text, which is a long string, into substrings (each substring is one sentence) using NLTK "sent_tokenize"
4. Clean the text:
    a. Convert all text to lowercase

   b. Eliminate punctuation
   c. Eliminate stopwords (the, and, they, etc.)
5. Split each sentence into tokens, or individual words, using NLTK "word_tokenize"
6. optional: using word stems, rather than full words, can be implemented here
7. optional: convert the output token set into an NLTK text object using the "Text" command to obtain document statistics or perform other operations

In terms of performance, using serial processing, ADDE takes about four hours to create a dictionary of sentences and word tokens for all ~260 companies. A 2015 Macbook Air with 8 GB 1600 MHz DDR3 RAM and 2.2 GHz Dual-Core Intel Core i7 processors was used. Parellelization would improve the processing time with little additional effort.

### *4.2.4 Methodology for proof of concept*

A proof of concept demonstrates the usefulness and performance of the novel attitudinal measurements. The objective of these applications is to detect underlying firm strategies. Raw data sources and the attitudinal measurements are described in the next sections. The proof of concept uses two related methodologies that have evolved primarily in psychology and economics. The methods aim to identify underlying, hidden (or latent) behavioral factors that drive observable (or manifest) actions and decisions. Briefly, this study utilizes first exploratory factor analysis (EFA) then confirmatory factor analysis (CFA, pictured in Figure 27) to detect underlying company strategies using the measurement results that are obtained from the SS-BOW and w2vPCA applications (references on factor analysis include Rosseel, 2012 and Everitt and Dunn, 1991).

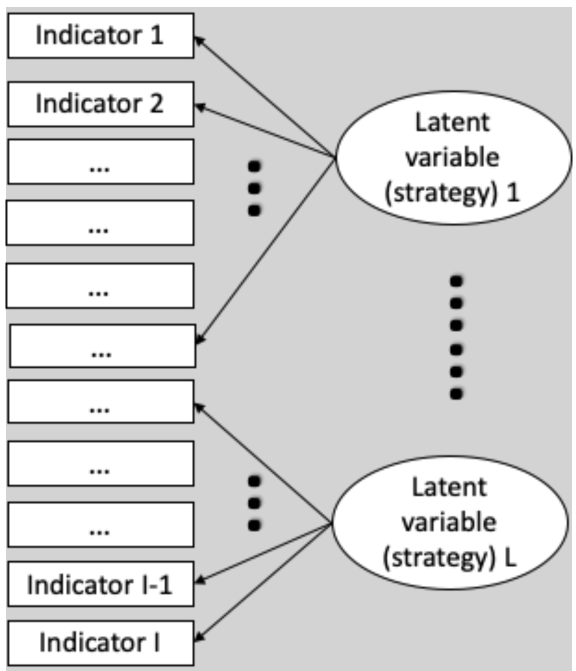
The rest of this description is an excerpt from the IEEE paper that is listed in the Acknowledgements section.

EFA and CFA are often used to examine the underlying and unseen, or latent, relationship among variables that are measured. Examples of measured, or observed variables, include:

- Answers to attitudinal questions—e.g., Likert or opinion-based questions
- Other observed data such as size or revenue
- Latent variables, in contrast, are not observed and must be hypothesized by the analyst.

The methods focus on analyzing the covariance between variables. Higher covariance implies greater similarity. In EFA, the analyst inputs all of the measurement data and interprets the output, which informs her/his hypotheses about latent variables. In CFA, the analyst specifies which measurements belong to which latent factor, then examines the relationship based also on statistical output. Figure 27 shows example relationships theorized *a priori* by the analyst. The CFA results in a set of latent factors which can be also used in Structural Equation Models or Hybrid Choice Models with additional variables.

It is important to note that the arrows in the diagram do not connote causality. This is because the mathematical foundations are built on covariance. Therefore, while the arrows are helpful for thinking about the construct, the results are correctly interpreted as correlations rather than causal relationships.




*Note: Error terms not shown.*

**Figure 27. Theoretical structure of variable relationships (CFA): latent variables represent unobservable company strategies, or emphasis areas.**

## 4.3 Data

The methods are applied to study studies the attitudes and strategies of companies belonging to the year 2017 Fortune 500 list (Fortune). Subsequently, the attitudinal measurements are input to the proof of concept application. Data regarding private fleet ownership and fleet characteristics are obtained from BigMackTrucks.Com (2017), which is a posting of data that FleetOwner magazine in Mele (2017) developed using the FleetSeek (2017) database. The CoStar (2020) real estate database is the source of distribution center data for the study. Company-specific, textual, attitudinal data are obtained from 10-K reports that were filed in 2017 with the Securities and Exchange Commission (SEC) and archived in the SEC's EDGAR system (United States Securities and Exchange Commision (SEC), 2020). A set of attitudinal keywords is specified by the user. This section discusses these data source and their analysis. In addition to these sources, geospatial data are obtained from other sources and processed within the data engine for use in later applications (not discussed here). The data inputs and their processing steps are illustrated in Figure 28, demonstrating the Attitudinal Development Data Engine (ADDE) that is developed in Stinson and Mohammadian (2020).

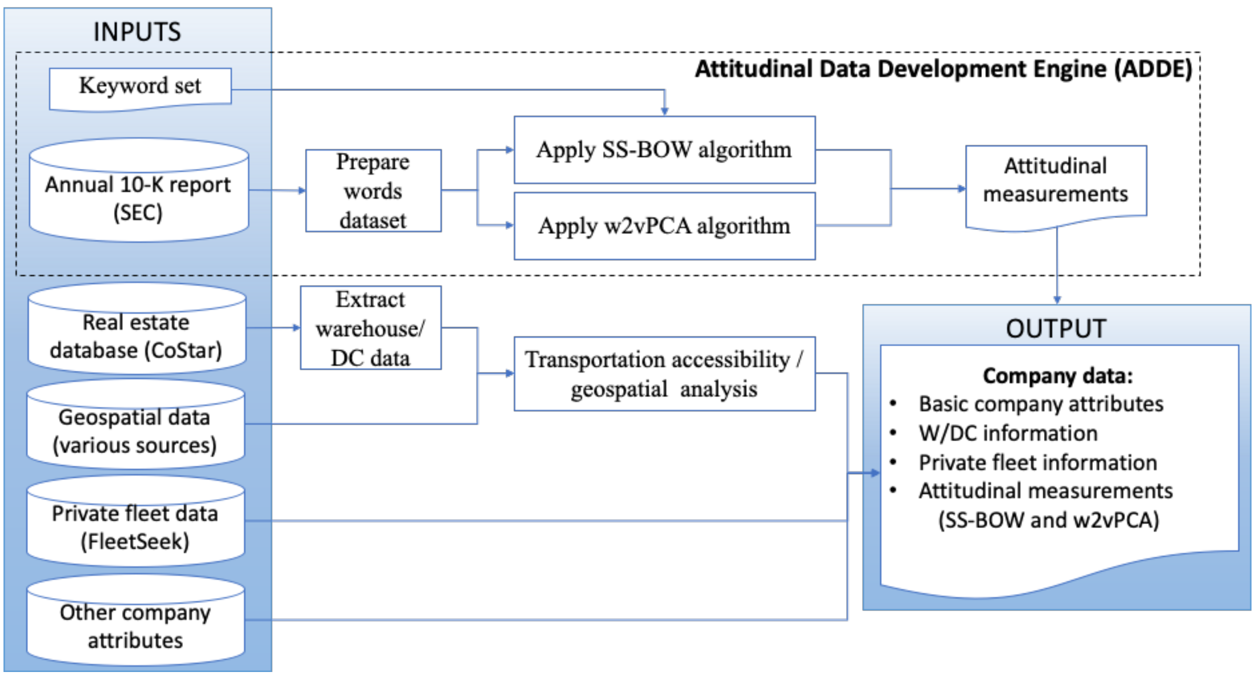



**Figure 28. Data development.**

### *4.3.1 Company attributes and logistics controls*

Fortune 500 companies in sectors that are not freight-intensive (such as telecommunications) are eliminated, leaving a total of 260 companies as the starting figure. An additional 19 companies are excluded from the dataset due to their lack of data (some companies are privately owned and therefore do not file a 10-K report) or sparseness of word usage (using fewer than half of the selected keywords, which causes numerical issues). The latter may introduce bias and should be investigated further, although this is not believed to be an issue here since fewer than 10 companies are eliminated for this reason.

The private fleet data contains information on ownership of tractors, trailers, and total trucks. This study focuses on the binary distinction: whether or not a company owns a private fleet. The other outcomes data are used in Chapter 5.

Distribution center information is obtained by extracting and processing CoStar commercial real estate data on individual properties throughout the US. Information on property use and size (square feet, or SF) is used to identify distribution centers as documented in Section 0.

### *4.3.2 10-K reports and attitudinal keywords*

In general, every publicly owned, US-based company files a 10-K report annually with the US Securities and Exchange Commission (SEC). US laws regarding securities require this and other information disclosures by public companies. The annual report gives "a comprehensive overview of the company's business and financial condition and includes audited financial statements" (United States Securities and Exchange Commission, 2009). The required statements can be quite lengthy and, as such, focusing on certain areas of the document may be advantageous in terms of data cleanliness. For instance, Item 1 of the 10-K report is an overview of the company, including its products and services. However, companies regularly supplement other items with company information (e.g., the number of distribution centers and manufacturing sites that are owned in different regions). So, this study uses the entire 10-K report.

The 10-K reports can be extracted in several formats including *.pdf and *.html. Starting in 2018, SEC began using Inline XBRL as its standard for companies to upload data. For this study, the *.html reports were downloaded from the EDGAR system and processed using ADDE.

Keywords for this analysis are pre-selected by the user for this study as follows. First, a text mining process with journals and industry reports pertaining to shipping and freight transportation was explored. However, this was fairly time-consuming to set up, and was considered lower priority than other tasks in the study. So, a second approach is adopted in which judgment is used to manually select about 30 keywords total for testing. The chosen keywords reflect a range of attitudes and behaviors that are related to shipping and freight transportation. Some keywords are purely attitudinal in nature, while others describe some sort of logistics activity or geographic scale more literally.

The sets of keywords that are tested in this study are categorized as:

- Attitudes: value, quality, reliability, affordability, innovative, diversify, technology, service, convenience, security, efficiency, cost
- Mixed attitudinal and logistics-related: cost, customer, delivery, distribution, efficiency, environmental, fleet, global, growth, logistics, national, provide, quality, reliability, safety, security, service, ship, standard, storage, transportation, value

Several 10-K reports are briefly skimmed to find examples of these keywords in order to determine anecdotally whether the words are used differently by companies. A few example keyword uses in the 10-K documents are as follows:

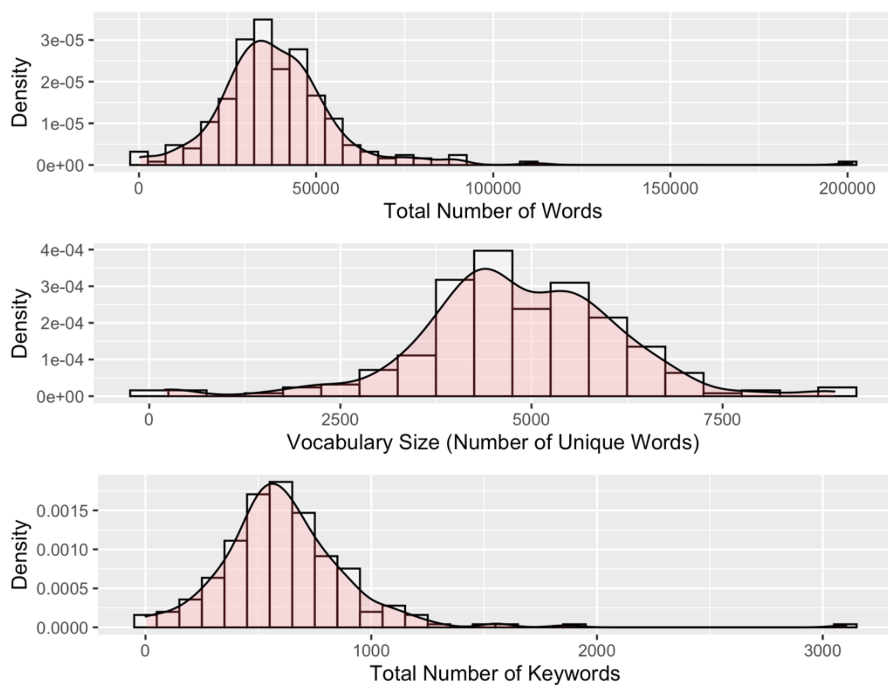
AutoNation (2017) (no fleet, no DCs for vehicles):
"…product quality, affordability and innovation,…"
"… enhance the value of our retail brands…"
"…distribution capabilities of the vehicle manufacturers,…"
"… new vehicles, used vehicles, … automotive repair and maintenance services…"
"…dependent upon the efficient operation of our information systems…"

Aramark (2017) (has both private fleet & DCs)
"…consistency of product, distribution capability, particularly for large multi-location clients…"
"…approximately 374 service locations and distribution centers across North America…"
"…We operate a fleet of service vehicles…"

These examples suggest that, for at least some keyword occurrences, the selected keywords are used differently by companies depending on their logistics controls. However, some keywords are often used in other contexts that provide no insights into the logistics controls questions of this study. For example, the word "value" is used regularly in the 10-K accounting summaries as part of the term "fair market value". However, as virtually all of the companies use this phrase regularly, this study assume that the effect essentially is averaged across firms, and that other instances of the word "value" are the instances that matter for the attitudinal measurement process (which focuses on differences in keyword use between companies). This assumption will be revisited in future studies.

In future research, this process can be improved as follows. First, keywords can be selected based on meta-analysis of industry journals and other literature. Second, the data can be further processed (e.g., using word sense analysis) to reduce the impact of keywords when their use is unrelated to the aims of the study.

Figure 29 shows the distribution of document sizes in terms of both the total number of words per document and the vocabulary size, or the number of unique words. Most of the 10-K reports are under 50,000 words in length and use between 2,500 and 7,500 unique words. The total number of keywords observed in documents generally is under 1,000, with an average of about 500 attitudinal keywords per document. This figure uses the attitudinal keyword set.




*The lower pane is based on the attitudinal keyword set.*

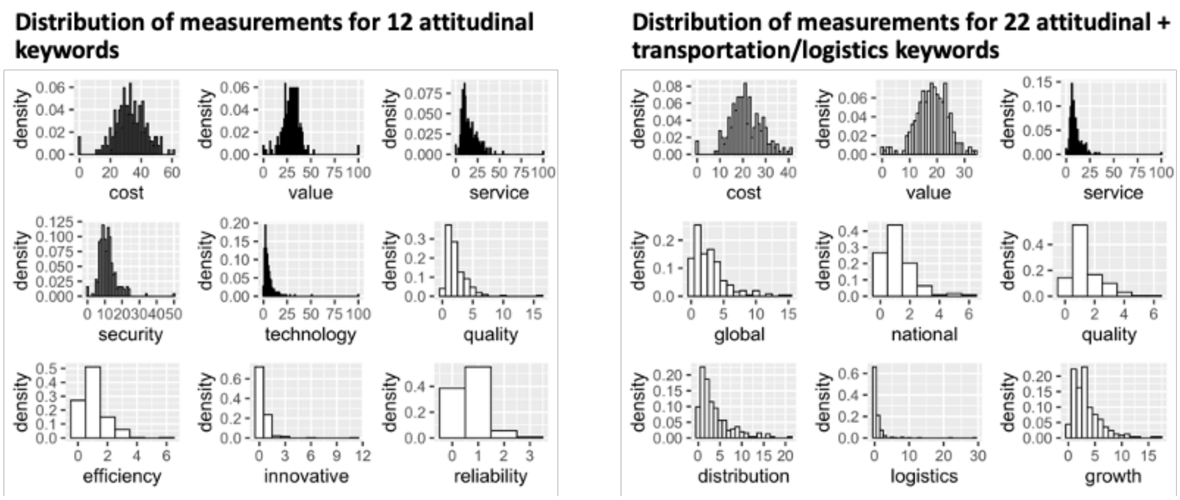
**Figure 29. Document statistics.**

## 4.4 Results: Attitudinal measurements

This section illustrates and discusses the attitudinal measurements for selected keywords.

### *4.4.1 SS-BOW measurements*

Attitudinal measurements for selected keywords using the SS-BOW algorithm range from zero to 100, which is expected due to the normalization process. The distribution of results for two keyword sets are illustrated, with each figure showing nine sub-plots. Figure 30 shows that cost, value and service are very commonly used terms in general, however with considerable variation among firms. For example, the keyword cost has a normalized value ranging from zero to about 60. Other keywords, such as efficiency and reliability, tend to be used much less frequently. Although sparsity can be problematic in quantitative analysis, rare observations can still be important determinants, so they are kept in the analysis for now.




*Note: Nine keywords were selected for each plot.*

**Figure 30. Histograms: SS-BOW measurement results.**

The right pane of Figure 30 provides a similar illustration, but is based on applying SS-BOW to the mixed keyword set with 22 keywords. Words related to shipping decisions tend to be used less often than the most common attitudinal keywords but more often than the irregular attitudinal keywords. For example, the distribution measurement ranges from zero to about 20, which is less than the range of cost (zero to 40) but greater than the range of quality (zero to about six). Because this set of words is a superset of the words from Figure 30, the new ranges are compressed – e.g., the maximum cost measurement has decreased from 60 to 40.

Figure 31 shows the attitudinal keywords listed in increasing order of their SS-BOW measurements. Cost, value, service, security and technology are the most common attitude words (based on the set that was pre-selected for this study). Cost and value are the most common, suggesting that companies are very cost-conscious and hoping to provide value to their customers. As suggested by Figure 30, several words (affordability, diversify, innovative, …) are less common. However, as discussed earlier, these may be important for distinguishing different attitudes or strategic focus areas of companies.

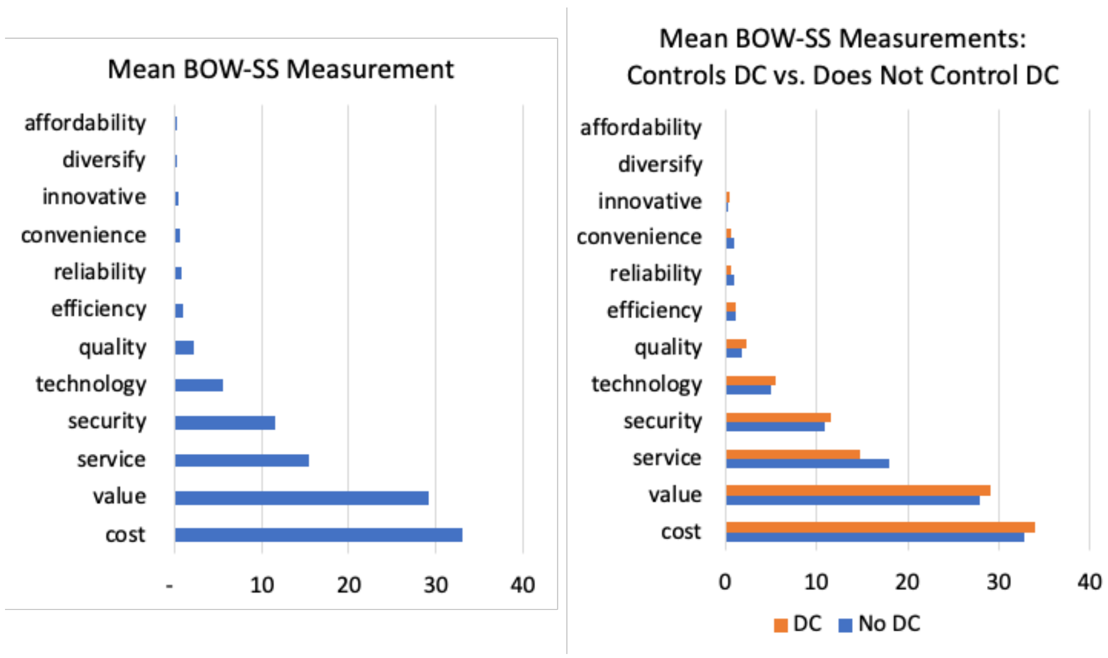



**Figure 31. Mean SS-BOW measurements of attitudinal keywords.**

For purposes of practical applications, determining whether the SS-BOW method uncovers differences between companies with different strategies is critical. This topic is explored in three ways. First, as the right pane of Figure 31 shows, the mean measurements for companies with different strategies can be visually inspected. Second, these means can be evaluated with statistical testing to infer whether these means are really different. This testing is conducted in Section 4.4.3. Third, differences in the shape of the SS-BOW measurement distributions can be examined.

In regards to the last point, Figure 32 illustrates some interesting differences in the SS-BOW measurement distributions with respect to private fleet ownership and DC control. First, companies with fleets mention *technology* less and *service* more. This is consistent with the intuition that trucking, as a relatively longstanding industry, has less potential for technological advancement compared to newer areas such as high-precision manufacturing. This also resonates with the author's experience that companies with private fleets differentiate themselves in part by controlling the quality of service throughout all aspects of delivery. Owning a private fleet allows these companies to control the delivery quality, for example, by ensuring that trucks are always available for high-priority, express deliveries.

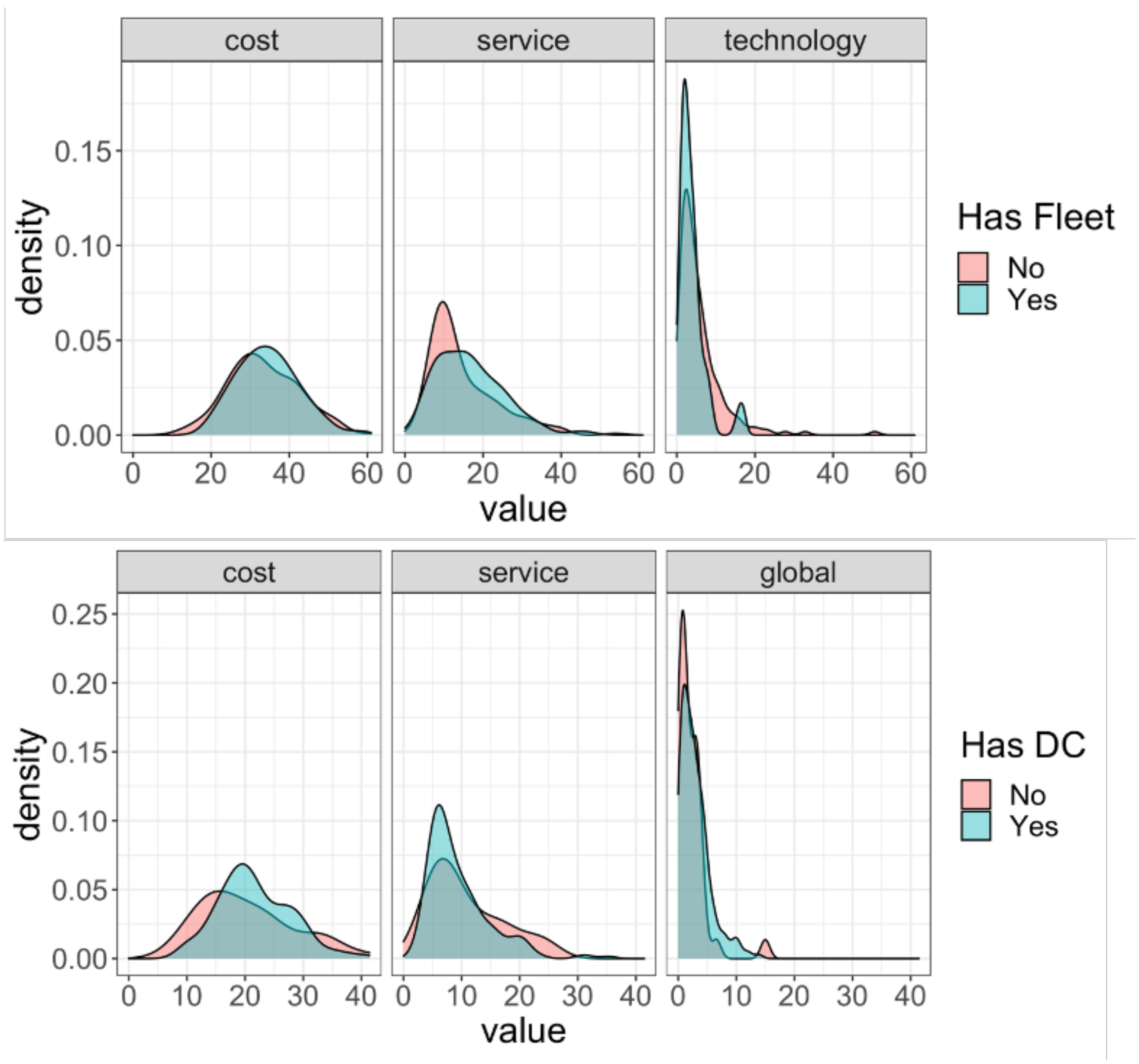



**Figure 32. Comparing SS-BOW measurements for select keywords; differences by type of logistics control.**

*Cost* has a similar distribution between companies with and without private fleets, suggesting that controlling costs is imperative to businesses regardless of the industry due the impact of costs on profitability. It is also possible that the similarity in distribution reflects outcomes that are at odds with each other. Specifically, private fleet ownership allows business to maintain consistent delivery costs; however, while delivery costs are relatively consistent, they are also relatively inflexible, making it harder to take advantage of lower costs that may arise in the for-hire market.

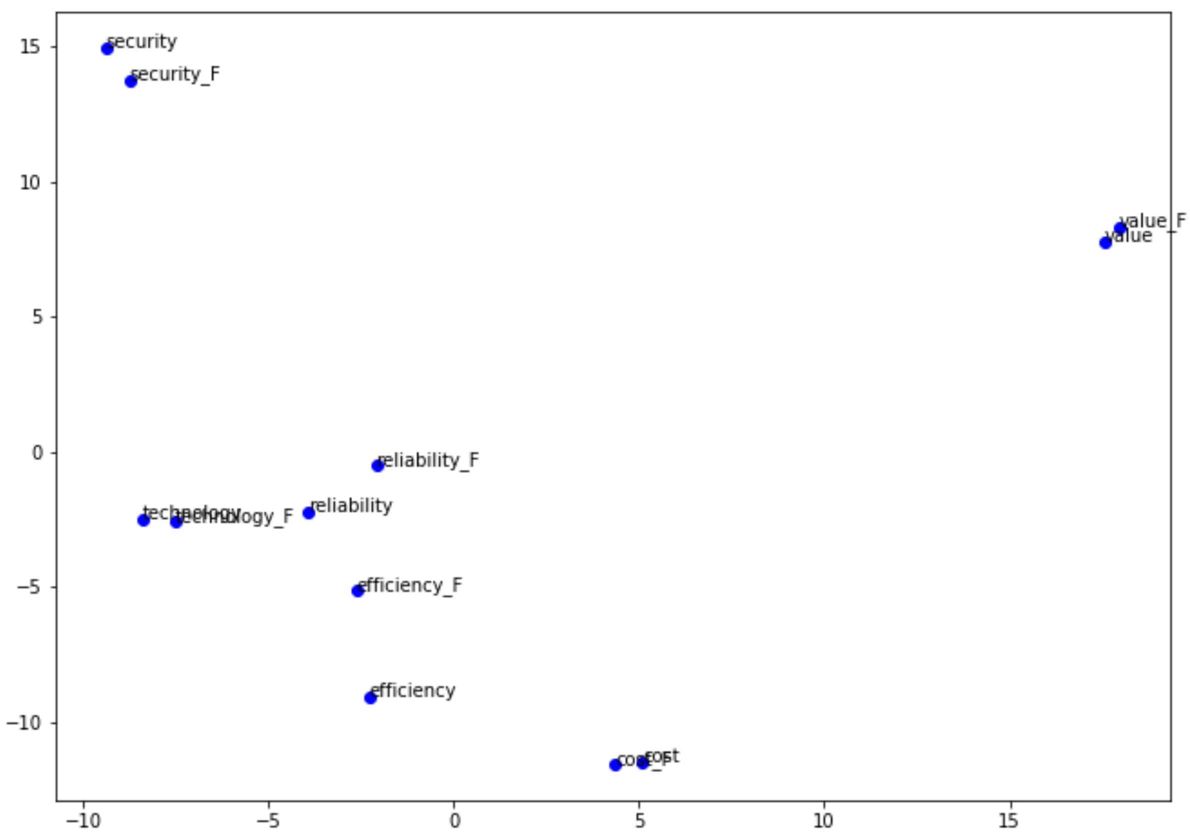
On the other hand, companies with DCs mention *cost* more and *service* less. This suggests that controlling distribution operations may allow companies greater control over costs, with somewhat less attention to controlling customer service. Additionally, companies with DCs generally mention the word *global* more regularly than companies without DCs, suggesting that control of distribution typically is advantageous to companies with global operations. The plot reveals one company without DCs that uses this word quite often (with an SS-BOW measurement of 15). However, further investigation reveals that this company is global and has many subsidiaries, therefore it is likely that distribution is actually handled by one of its subsidiaries rather than a third party.

### *4.4.2 W2v-PCA measurements*

This section demonstrates the remarkable empirical result that companies appear to use words differently, depending on the types of logistics controls that they exercise. This finding is presented and discussed from many viewpoints.

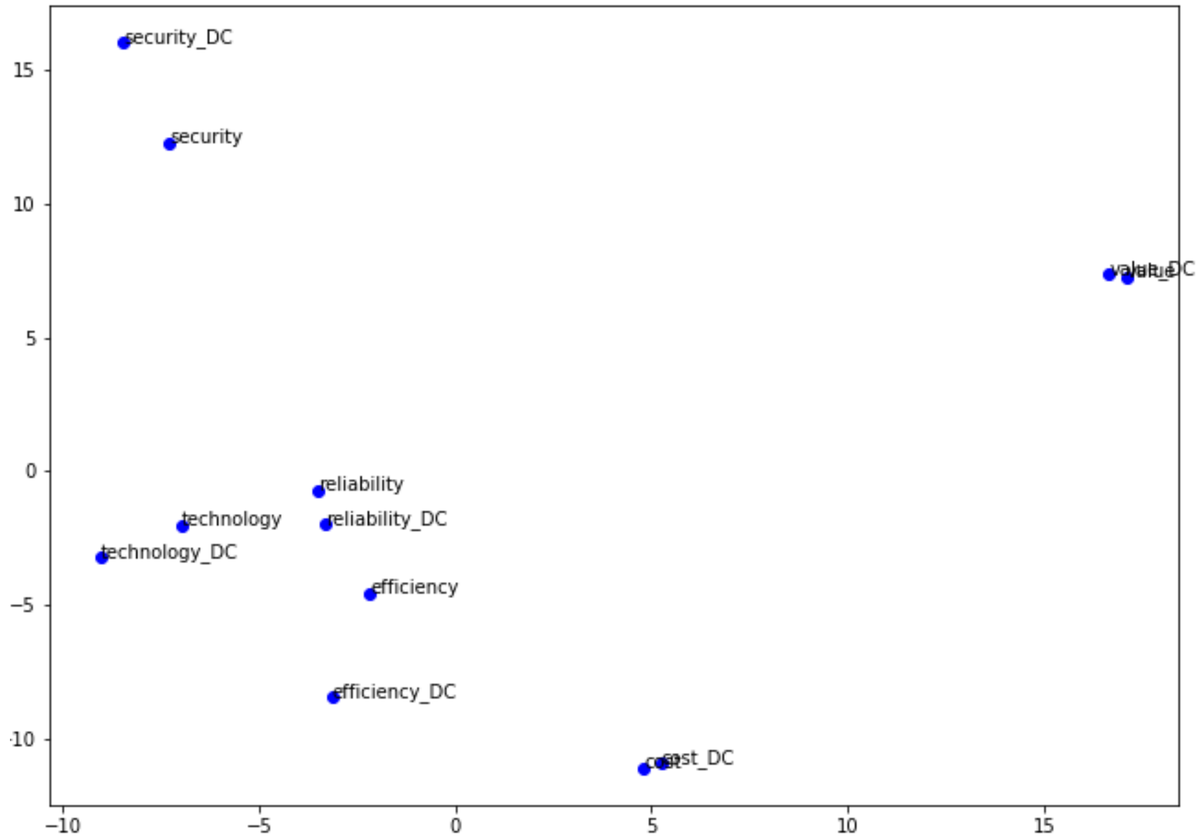
The first comparison analyzes the average keyword embedding for groups of companies that have the same logistics control status. To achieve this, Figure 33 and Figure 34 are generated using the same process that is outlined in Figure 25, but with one modification: the keyword tagging is not company-specific. Instead, keywords are appended with “_F” for all companies that have a private fleet (to generate Figure 33) and with “_DC” for all companies that control their distribution centers (Figure 34). Keywords for other companies are not labeled. The measurements are generated using a 100-dimensional word2vec analysis followed by a two-dimensional PCA analysis. These figures demonstrate that:

- The embeddings for each keyword pair (e.g., "security" and "security_F") are located in similar areas of each graph;
- For each keyword, there is a small difference in embedding location depending on the fleet (or DC control) status; and
- In some cases, the shift appears to be quite small, while in other cases it is larger.




"_F" denotes the average 2D word measurement across companies that have a private fleet; words not annotated are associated with companies that lack a private fleet.

**Figure 33. w2vPCA(2-dimensional) results: fleet vs. no fleet.**




"_DC" denotes the average 2D word measurement across companies that have DC control; words not annotated are associated with companies that do not.

**Figure 34. w2vPCA(2-dimensional) results: DC control vs. no DC control.**

The remaining figures and tables in this section are developed using company-specific word embeddings. These remaining graphics illustrate the results in various ways to highlight different aspects of their interpretation and practical implications.

Figure 35 illustrates the keyword measurement for all keywords in the attitudinal set. Symbology is based on a combination of the two logistics factors, with "T0" ("T1") denoting companies without (with) a private fleet and "D0" ("D1") denoting companies without (with) DC control. A "jitter" function is used to shift each point slightly away from its true location in a random direction, which improves the visual distinction between points.

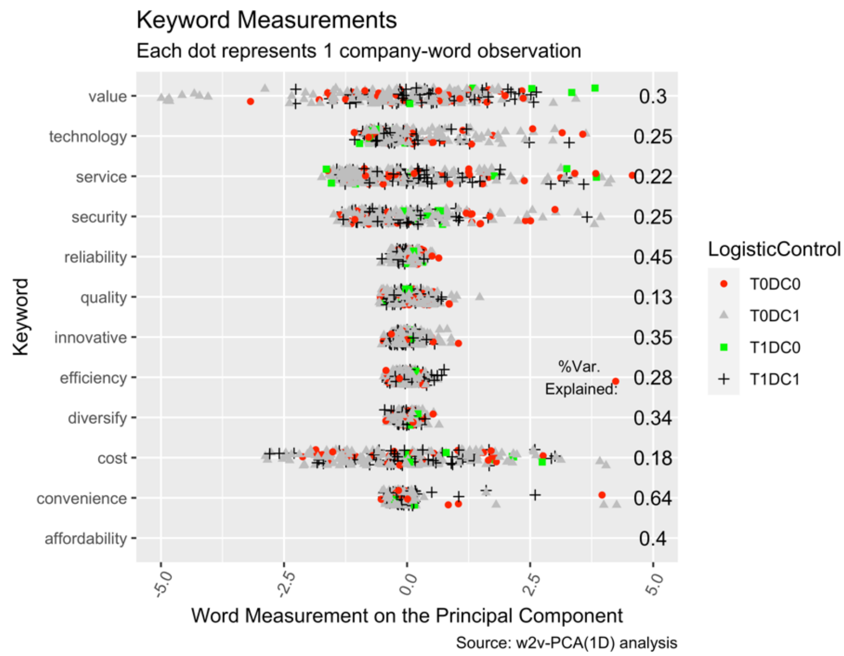



**Figure 35. "Jitter" plot of w2vPCA keyword measurements.**

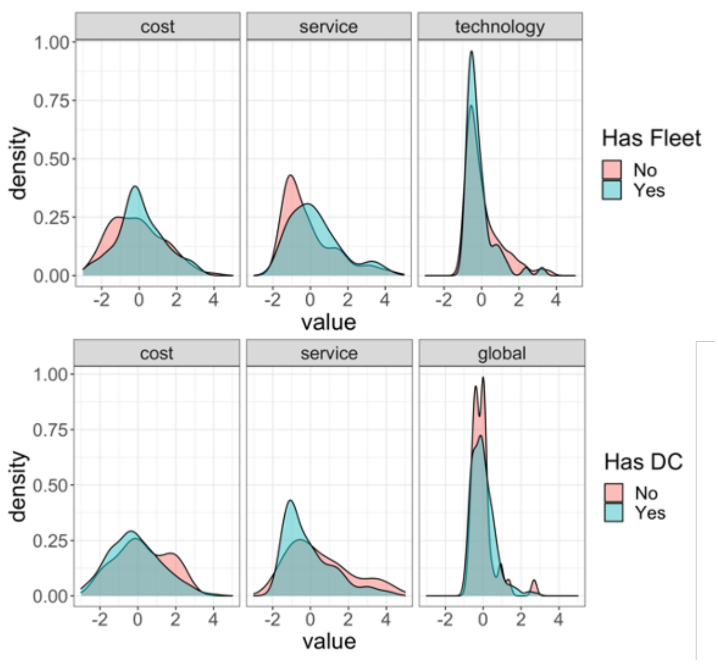
This figure illustrates several things. First, it illustrates the main output of w2vPCA: company-specific, quantitative measurements for each keyword. This demonstrates that the key objective of this study has been met: w2vPCA creates quantitative measurements of attitudes.

Second, the figure shows the spread in attitudinal measurements for each keyword. For instance, the measurement for "cost" ranges from about -3 to 4, while the measurement for "innovative" ranges from about -1 to 1. This implies that companies use the word "innovative" in a more limited range of contexts (compared to the word "cost"). Remarkably, the spread or differences in usage come from natural language that is written by the companies. This contrasts with the current, widespread practice of using a pre-imposed scale (e.g., 1 to 7) to rate survey responses.

Third, Figure 35 gives some insights into how word measurements differ by company strategy regarding the degree of logistics control. For instance, a handful of companies (with DC control but no private fleet, as indicated by gray triangles) on average use the word "value" quite differently from other companies, with measurements of about -4 to -5. This is similar to the point that is illustrated in Figure 33 and Figure 34, but Figure 35 shows differences in measurements across individual companies rather than groups of companies.

Fourth, the right-hand side of the plot displays a list of numbers. These numbers represent the amount of variance in that is captured by (in this case) the first principal component. For example, the measurement for "value" represents 30 percent of the total variance across embeddings of the word "value" among the companies. Although the implications of this result need further explanation, this study preliminarily suggests that this is a metric for the amount of "noise" in our w2vPCA measurements. It seems that ideally, the w2vPCA algorithm would capture 100 percent of the differences in word usage between companies, which would generate a PC metric of 1.0. Improving the process to pre-treat, or refine, the word and context selection process may improve this fit. This is suggested as a future direction.

Figure 36 displays smoothed histograms for select keywords, with differences illustrated for companies with different types of logistics control. For example, the top left image shows the observed frequency distributions of the word measurement for "cost" among companies with and without a private fleet. This illustrates provides visual indication that companies are using various keywords differently depending on their private fleet or DC control status.




*The top (bottom) panes use the 12 (22) keyword sets.*

**Figure 36. Comparing w2vPCA measurements for select keywords; differences by type of logistics control.**

### *4.4.3 Statistical analysis and comparison of results by method*

The discussion so far has focused on results and their interpretation. A t-test is now conducted for each keyword to test whether companies with different logistics strategies truly use the keyword in different ways. A two-tailed test is used since the difference in means, if such a difference exists, can be either greater than or less than zero.

The hypothesis test follows the format shown in Equation 8. The null hypothesis is that the mean measurement of keyword *k* is the same for Group 1 and Group 2. The alternative hypothesis is that the mean measurements are not equal. In this case, the group identification is based on logistics control status. In other words, each test measures whether there is a statistically significant difference between measurements depending on whether or not the company has control of the logistics element in question (for example, do companies with a private fleet on average have a different measurement of keyword *k* than companies without a private fleet?).

$$H_0: \mu_{k,1} - \mu_{k,2} = 0$$

$$H_{ALT}: \mu_{k,1} - \mu_{k,2} \neq 0$$

**Equation 8. Hypotheses for two-tailed t-test of means.**

The t-test results support direct comparisons of, for example, group means such as those that are illustrated in Figure 33 and Figure 34. More generally, the test results also provide insight into which keywords are best able to distinguish underlying company strategies regarding logistics control.

Tests are conducted for mean keyword measurement differences for each keyword set and each method. **Table 2** summarizes the p-value that results from each of the tests. A total of 136 tests are conducted, with four tests (based on two methods (SS-BOW; w2vPCA) with two types of logistics control (DC Control; Private Fleet)) conducted for each of the 34 keywords (having 12 (22) keywords from the attitudinal (mixed) keyword set). Note that while some words, such as “quality”, are present in both keyword sets, the process of measuring them alongside different sets of keywords means that the resulting measurements will differ in general. Therefore, a separate t-test must be used.

Each result in the table provides a conclusion to a hypothesis test. For example, within the attitudinal keyword set, the table shows a p-value, 0.018 for the word “value”. This is based on a mean w2vPCA measurement of -0.113 (0.639) for companies with (without) DC control, generating a t-statistic of 2.449 and its associated p-value (0.018) with about 48 degrees of freedom. This is highly statistically significant, exceeding the 95 percent level, providing sufficient evidence to conclude that the word “value” on average really is used differently by companies depending on whether they have DC control or not.

These results are used to guide inference into which of the selected keywords may perform well as attitudinal measures in predictive models. Specifically, any result with one or more stars is expected to perform well in representing or detecting differences in underlying one or both logistics control strategies. Several words are expected to perform less than optimally in this respect:

- Attitudinal: convenience, diversify, efficiency (although the statistical significance for efficiency is close to meeting the 70 percent threshold in three of the four cases)
- Mixed: efficiency (although its statistical significance again nears the 70 percent threshold in both w2vPCA cases)

The results can be summarized to compare SS-BOW to w2vPCA for the two keyword sets. A count of keywords that are have statistically significant differences in means is presented in Table 3 for each of the keywords sets and measurement methods. Using the Attitudes-only keywords set, the maximum possible score is 24 (12 words x 2 types of logistics control). The maximum possible score for the Mixed set is 44. So, in addition to showing the count of statistically significant differences, the table also shows (in parentheses) the count divided by the maximum possible value.

**Table 2. Statistical testing of the differences in means.**

| Keyword Set | Keyword | p-value | | | |
|---|---|---|---|---|---|
| | | SS-BOW | | w2vPCA | |
| | | DC Control | Private Fleet | DC Control | Private Fleet |
| Attitudinal | affordability | 0.091** | 0.068** | *0.31* | *0.44* |
| | convenience | *0.484* | *0.767* | *0.762* | *0.827* |
| | cost | *0.548* | *0.412* | 0.054** | *0.636* |
| | diversify | 0.18* | *0.828* | *0.63* | *0.59* |
| | efficiency | *0.995* | *0.308* | *0.336* | *0.353* |
| | innovative | *0.599* | 0.006*** | *0.723* | 0.019*** |
| | quality | 0.045*** | 0.084** | *0.972* | *0.734* |
| | reliability | 0.014*** | 0.289* | 0.025*** | *0.618* |
| | security | *0.347* | 0.258* | 0.153* | *0.667* |
| | service | 0.089** | 0.169* | 0.035*** | 0.079** |
| | technology | *0.734* | 0.003*** | *0.972* | 0.075** |
| | value | *0.359* | *0.735* | 0.018*** | 0.026*** |
| Mixed | cost | *0.377* | *0.894* | 0.043*** | *0.691* |
| | customer | 0.142** | *0.724* | *0.674* | 0.007*** |
| | delivery | *0.395* | 0.088** | 0.102** | 0.008*** |
| | distribution | 0.264* | 0.013*** | 0*** | *0.414* |
| | efficiency | *0.934* | *0.487* | *0.306* | *0.404* |
| | environmental | 0.103** | *0.804* | 0.018*** | *0.41* |
| | fleet | 0.256* | *0.481* | 0.165* | *0.516* |
| | global | 0.117** | 0.079** | *0.751* | *0.617* |
| | growth | 0.013*** | *0.969* | 0.15** | 0.198* |
| | logistics | *0.318* | *0.833* | 0.098** | 0.595 |
| | national | *0.563* | 0.033*** | 0.366 | 0*** |
| | provide | 0.011*** | *0.39* | 0.004*** | 0.006*** |
| | quality | 0.02*** | 0.008*** | *0.716* | *0.979* |
| | reliability | 0.057** | *0.353* | 0.164* | *0.802* |
| | safety | *0.591* | *0.579* | 0.031*** | 0.051** |
| | security | 0.151* | 0.016*** | 0.181* | *0.98* |
| | service | 0.233* | *0.508* | 0.036*** | 0.095** |
| | ship | *0.459* | 0.287* | 0.222* | *0.596* |
| | standard | *0.804* | 0.273* | 0.053** | 0.215* |
| | storage | 0.001*** | *0.581* | 0.002*** | 0.218* |
| | transportation | 0.021*** | 0.077** | 0.01*** | 0.075** |
| | value | 0.079** | *0.606* | 0.029*** | 0.025*** |

*Stars (***, **, *) denote significance at the 95, 85 and 70 percent levels.*

**Table 3. Summary of statistically significant differences in means.**

| Keyword Set | SS-BOW | w2vPCA |
|---|---|---|
| Attitudes only (% of max) | 12 (50%) | 8 (33%) |
| Mixed attitudes & transportation keywords (% of max) | 23 (52%) | 28 (64%) |

The SS-BOW method produces more statistically significant differences in means (12 keywords vs. eight) when the Attitudinal keyword set is used to distinguish differences between fleet owners and non-fleet owners, and companies with and without DC control. W2vPCA produces more statistically significant differences in means when the Mixed keyword set is used. These results, while informative for this case study, do not provide enough evidence to conclude that one method is superior in terms of performance. This is recommended as an area to explore in future studies.

The next section illustrates proof of concept applications of the various measurements and keyword sets. In doing so, the section provides further insights into the performance of various keywords and methods and their adequacy for serving as attitudinal measures.

## 4.5 Proof of Concept

A proof of concept application demonstrates that the resulting attitudinal measurements can be used in traditional applications that focus on latent variables. Exploratory and confirmatory factor analyses are conducted for both methods with different sets of keywords.

Four sets of keywords are tested initially (Table 4). To help inform the discussion, each keyword set is named based on (a) whether it contains attitudinal only keywords or a mix of attitudinal and logistics-related keywords and (b) the number of words in the set. For example, the keyword set named “Attitudinal (7)” contains seven attitudinal keywords.

**Table 4. Keyword sets for EFA.**

| | Keyword | convenience | cost | customer | delivery | distribution | diversify | efficiency | environmental | global | growth | innovative | logistics | national | provide | quality | reliability | safety | security | service | ship | standard | technology | value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Name | Attitudinal (11) | X | X | | | | X | X | | | | X | | | | X | X | | X | X | | | X | X |
| of | Attitudinal (7) | | | | | | | X | | | | X | | | | X | X | | X | X | | | X | |
| Keyword | Mixed (19) | | X | X | X | X | | X | X | X | X | | X | X | X | X | X | X | X | X | X | X | | X |
| Set | Mixed (9) | | X | | | X | | | | X | X | | X | X | | X | | | | X | | | | X |

### *4.5.1 Exploratory Factor Analysis (EFA) results*

Each of the four keyword sets is input to an EFA, with EFA performed for one, two and three latent factors. The resulting statistical estimates are shown in Table **5**. A lower Chi-Squared statistic (and a corresponding higher p-value) provide evidence that the number of factors that is used in the EFA is sufficient for adequately capturing the covariance among keywords (see Rosseel, 2012). In keeping with convention, a more parsimonious specification is also preferred to one with more variables. Based on a minimum p-value of 0.05 to identify statistical support for the number of factors tested, the w2vPCA results out-perform the SS-BOW results in all but the Mixed (19) EFA (where neither performs well). For instance, based on this criterion, only one factor is needed using the Attitudinal (7) set and w2vPCA measurements. In this case, two factors is much better statistically, therefore using two is more advisable.

Table 6 shows the factor loadings with an absolute value greater than 0.2 based on the Attitudes (7) keyword set. The table reveals that the SS-BOW measurements for efficiency, reliability and technology suggest the existence of one unique factor, while the existence of a second, unique factor is demonstrated by the loadings of innovative, quality, security and service on the factor. Greater magnitude indicates greater strength in the relationship between the word and the factor. The table also shows the factor loadings for the w2vPCA measurements. Statistics for these results are in Table **5**.

**Table 5. EFA results for three or fewer factors.**

| Keyword Set | EFA Parameters | | SS-BOW | | w2vPCA | |
|---|---|---|---|---|---|---|
| | Factors | DOF | Chi-Sq. Statistic | p-value | Chi-Sq. Statistic | p-value |
| Attitudinal (11) | 1 | 44 | 2219.4 | n/a (0.0...) | 123.2 | 2.00E-09 |
| | 2 | 34 | 2101.5 | n/a (0.0...) | 50.5 | 0.03 |
| | 3 | 25 | 1904.8 | n/a (0.0...) | 18.7 | 0.81 |
| Attitudinal (7) | 1 | 14 | 29.1 | 0.01 | 23.8 | 0.05 |
| | 2 | 8 | 8.5 | 0.38 | 6.9 | 0.55 |
| | 3 | 3 | 2.0 | 0.57 | 0.2 | 0.98 |
| Mixed (19) | 1 | 152 | 951.6 | 7.00E-116 | 1007.3 | 3.94E-126 |
| | 2 | 134 | 827.5 | 2.22E-100 | 715.2 | 3.75E-80 |
| | 3 | 117 | 709.3 | 1.53E-85 | 396.7 | 4.22E-32 |
| Mixed (9) | 1 | 27 | 161.9 | 3.39E-21 | 96.7 | 8.89E-10 |
| | 2 | 19 | 94.9 | 4.41E-12 | 46.5 | 0.0004 |
| | 3 | 12 | 55.6 | 1.41E-07 | 28.8 | 0.0043 |

**Table 6. Factor loadings for words in Attitudes (7) set.**

| | SS-BOW | | w2vPCA | |
|---|---|---|---|---|
| Keyword | Factor1 | Factor2 | Factor1 | Factor2 |
| efficiency | 0.39 | | 0.40 | |
| innovative | | 0.21 | | 0.38 |
| quality | | 0.35 | 0.22 | |
| reliability | 0.66 | | 0.51 | |
| security | | 0.51 | 0.38 | 0.46 |
| service | | -0.23 | 0.70 | -0.23 |
| technology | 0.33 | | 0.54 | |

### *4.5.2 Confirmatory Factor Analysis (CFA) results*

Two factors are identified for the remainder of this proof of concept analysis. The findings in Table 6 indicate that "innovative" and "security" will provide a good foundation for the former, with "efficiency", "reliability" and "technology" as the foundation for the latter. Based on additional experimentation with groupings in combination with judgment, "quality" is also assigned to the former while "service" is assigned to the latter. The factors are named Product Focus and Delivery Focus, respectively, in order to convey that they are believed to represent the companies' underlying strategies and whether they attempt to differentiate themselves from the competition by either product development and innovation or by providing excellent shipping and delivery to the customer.

CFA results for these factors and keyword assignments are shown in Table 7. The SS-BOW and w2vPCA based models have a Comparative Fit Index (CFI) of 0.849 and 0.917, respectively, which are close to the desired minimum threshold of 0.9 (Rosseel, 2012). All w2vPCA keyword variables are statistically significant to the 90 percent or higher level, with six of the seven variables being significant to the 99 percent level. The SS-BOW keywords are generally not as statistically significant, with most of them being significant at approximately the 85 percent level or better.

**Table 7. CFA results.**

| Latent Factor | Keyword | SS-BOW | | | | w2vPCA | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Estimate | Std.Err | z-value | P(>\|z\|) | Estimate | Std.Err | z-value | P(>\|z\|) |
| Product Focus | innovative | 0.20 | 0.14 | 1.39 | 0.16 | 0.15 | 0.09 | 1.63 | 0.10 |
| | quality | 0.74 | 0.47 | 1.56 | 0.12 | 0.23 | 0.09 | 2.50 | 0.01 |
| | security | 0.21 | 0.15 | 1.41 | 0.16 | 0.61 | 0.17 | 3.65 | 0.0 |
| Delivery Focus | efficiency | 0.22 | 0.13 | 1.77 | 0.08 | 0.44 | 0.08 | 5.74 | 0.0 |
| | reliability | 1.06 | 0.53 | 2.02 | 0.04 | 0.54 | 0.08 | 7.06 | 0.0 |
| | technology | 0.24 | 0.13 | 1.79 | 0.07 | 0.55 | 0.08 | 7.12 | 0.0 |
| | service | -0.06 | 0.07 | -0.91 | 0.36 | 0.59 | 0.08 | 7.65 | 0.0 |

Overall, the w2vPCA performs better than the SS-BOW model in the CFA from a statistical perspective. Testing of other specifications shows that this is due to using "service" in the Delivery Focus construct, where it is numerically supported based the w2vPCA EFA but not the SS-BOW EFA (Table 6). A better specification for SS-BOW involves shifting "service" to the Product Focus factor, resulting in a CFI of 0.926 for SS-BOW, but this worsens some aspects of the w2vPCA specification.

## 4.6 Conclusions

This study develops two innovative new methods to develop attitudinal measurement data. Each method is applied to large-scale text data that is passively collected, that is, the data exists and is readily available without the need to conduct surveys. The data for this particular study exist due to federal filing requirements for large US companies.

The bag-of-words with simple scaling (SS-BOW) method is based on relative frequency counts of a pre-selected set of keywords. The word2vec-Principal Components Analysis (w2vPCA) is based on differences in usage of words. The methods are designed to study differences in attitudes across individual agents, such as persons or companies.

This study applies the methods to a population of large, publicly owned US companies. Supplemental information on their logistics controls, particularly whether they own private fleets or operate their own distribution centers, inform the analyses. Attitudinal measurements across the companies are compared from a variety of perspectives, using various keyword sets and methods, focusing especially on whether the methods detect different uses of keywords as manifested in their strategic decisions around fleet and distribution controls.

The findings provide both empirical and statistical evidence that each method successfully identifies differences in the unobservable attitudes, or strategies, that companies adopt. The measurement results are tested in two different ways: a host of t-tests for differences in means as well as a proof of concept factor analysis application.

# 5 Behavioral Models of Strategic Decision-Making

## 5.1 Introduction

Generally speaking, multiple strategies and strategic decisions can be operative simultaneously to a company. In other words, companies are typically guided by multiple strategies and jointly make multiple strategic decisions. Acknowledging these factors, this chapter proposes and develops a methodology to incorporate strategies into the agent-based transportation paradigm, allowing multiple strategies to inform a set of decisions that are assessed jointly.

An additional, unique feature of the proposed methodology is its explicit treatment of choice set generation parameters as another type of strategic decision that is made by companies. For example, companies may simultaneously decide whether or not to own a warehouse, and if so, what is the minimum size that makes sense for its sales volumes. To complicate matters, warehouse ownership may only make sense if the company owns at least two trucks for delivery staging. A model that jointly includes both strategic decisions as well as these thresholds offers significant logical appeal and advances the state of the art in agent-based modeling for both passenger and freight transportation. These aspects are developed as part of the methodology for this study.

To be sure, previous studies have addressed some of these aspects as discussed in Chapter 2. However, these earlier works mostly focused on modeling only the decisions themselves. While these earlier efforts collected provided a trove of valuable insights and extremely powerful modeling methods, the current study introduces several novelties that adopt a different perspective and focus.

The proposed refinement to agent-based transportation models are expected to have far-reaching implications. First, various agent behaviors throughout the model stream currently can be disjointed or inconsistent. By unifying major decisions and choice set generation parameter estimation, this framework adds consistency to decisions and actions that are undertaken by the individual agent. Second, strategies (and their passenger counterpart, attitudes) are a key input to the strategic decisions, which include both actionable decisions (e.g., whether to own a fleet) and choice set parameters (e.g., minimum warehouse size). As such, the consistency of agent preferences and decisions throughout the model stream is further strengthened. Ultimately, by constructing better models, improvements in forecasting origin-destination flows, vehicle ownership, and other high-impact outcomes will be possible.

Finally, a proof of concept develops a set of strategic decisions for Fortune 500 companies in freight-intensive sectors, providing a demonstration of the methodology as well as a valuable input to a real-world model. The proof of concept applies the methods for the purpose of simulating fleets and distribution centers for the set of companies. This itself is a major contribution to freight transportation modeling, which currently has a gap in modeling these critical features of the freight landscape.

## 5.2 Approach

The methodology of this study comprises the following elements. The first aspect is the development of a mathematical framework for modeling strategies and strategic decisions. The second aspect is the solution approach for this mathematical framework. The final component is the proof of concept approach that is deployed to demonstrate the methodology.

### *5.2.1 Theoretical construct*

Figure 37 outlines the theoretical framework of the strategic decisions model. In keeping with convention, elements within ovals are latent or unobservable while rectangles denote manifest or observable elements. The construct includes:

- Attitudinal measurements, also known as indicators, which are treated as observable measurements of the latent variables;
- Latent variables or strategies;
- Exogenous input data, which are used to predict both the strategy scores and the strategic decisions (the latent variables are also input to the strategic decision model); and
- Strategic decisions, including truncated variables, continuous amount variables, and choice set thresholds.

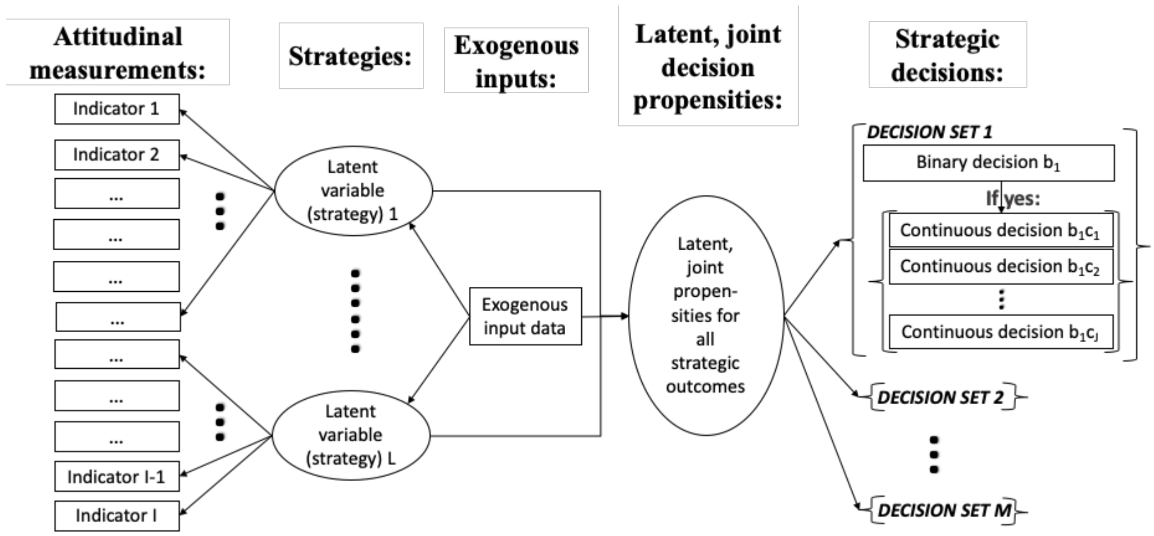



**Figure 37. Theoretical modeling construct for strategies and strategic decisions.**

### *5.2.2 Methodology to model agent selection of strategy bundles*

This subsection describes the methodology that is developed to model the agent's selection of a strategy "bundle." Let N be the number of observations, L (index l) denote the set of strategies, H (h) attitudinal measurements, D (d) strategic decisions using truncated variables, and C (c) strategic decisions regarding purely continuous decisions. Observable, explanatory variables are denoted below as the vectors *w* and *x*. Vectors of parameters to be estimated are denoted by $\gamma, \delta,$ and $\beta$. Error terms are denoted as $\eta, \varepsilon,$ and $\epsilon$.

Let $z_l^*$ be a latent continuous variable that represents strategy *l*. The structural equations for the latent variables or strategies are:

$$z_l^* = \gamma_l' w + \eta_l, \forall l \epsilon L \tag{9}$$

The model has the following measurement equations. First, it uses continuous attitudinal measurements that are considered the manifest outcome $y_h$ of unseen strategy *h*, with $\varepsilon_h \sim iid\ N(0,1)$ throughout the simulation:

$$y_h = \delta_h' z_l^* + \varepsilon_h, \forall l \epsilon L, h \epsilon H \tag{10}$$

The parameter subscripts of $\delta$ and *y* imply that each measurement is associated with only one strategy (therefore the potential subscript *l* is omitted). An intercept is not used currently.

Second, truncated continuous variables are used to represent three types of strategic decisions. Each can represent a decision that has two dimensions, one binary and the other continuous. This is an ideal setup for a binary decision that, if activated, is associated with a continuous amount (such as an input or output volume). Similarly, this is also an appropriate device to model a choice set parameter that may or not be relevant to a given individual – for example, minimum property size choice applies only for individuals that are choosing a property. Lastly, truncation limits can represent a simple binary or ordinal decision(s) for circumstances where an associated continuous amount is not available or not necessary. Each observed truncated variable, $y_d$, is associated with a latent continuous index, $y_d^*$ and threshold $\psi_d$ as follows:

$$y_d^* = \gamma_d' x + \delta_d' z^* + \varepsilon_d \text{ }, \forall d \epsilon D, y_d = \begin{cases} y_d^*, \ y_d^* > \psi_d \\ 0, otherwise \end{cases} \tag{11}$$

Alternative forms of this are easily integrated into this framework as follows. For any strategic decision is purely continuous, $y_d$ in the above formulation is rewritten as $y_d = y_d^*$. Additionally, the discreteness aspect of the outcome can be extended readily from a binary indicator (yes/no) to multiple ordered categories (e.g., $y_d^* < \psi_1, \psi_1 < \ y_d^* < \psi_2, \psi_2 < y_d^*$) using standard ordinal response methods.

Consider the vector $y^* = (y_1^*, y_2^*, .., y_D^*)'$. For individual *i*, create the Dx1 vector $y_\mathrm{i}^*$ to represent the strategic decisions model as a Seemingly Unrelated Regression (SUR), stacking the $y_d^*, x, z^*, \gamma_d, \delta_d$ and $\varepsilon_d$ terms into vectors accordingly:

$$y_i^* = x_i \gamma + z_i^* \delta + \varepsilon_i \tag{12}$$

Let $\mathrm{H} + \mathrm{D} = \mathrm{G}$. The reduced form of each $y_i^*$ is obtained by plugging in $z_i^* = \gamma_l' w_i + \eta_i$. Placing all observed explanatory variables into one vector, $x^M = \binom{x}{w}$, and placing all $\gamma$ and $\delta$ parameters into a single vector $\beta^M$, the entire measurement equation system is:

$$y_i^M = x_i^M \beta^M + \epsilon_i^M, \tag{13}$$

where $\epsilon_i^M \sim iid \ \mathrm{MVN}(0, \Sigma^\mathrm{M}),\ 0$ and $\mathrm{y_i^M}$ are Gx1 vectors and $\Sigma^\mathrm{M}$ is a GxG covariance matrix of the form:

$$\Sigma^{\mathrm{M}} = \begin{bmatrix} \Sigma_H & 0 \\ 0 & \Sigma \end{bmatrix}. \tag{14}$$

In this equation, $\Sigma_H \sim MVN(0, I)$ with an Hx1-dimensioned mean 0 vector and a variance-covariance matrix equivalent to the HxH identity matrix. The terms in $\Sigma_H$ are orthogonal to the terms in $\Sigma$, which is an unrestricted variance-covariance matrix for the strategic decision outcomes.

In this study, $\Sigma$ is estimable and is the mechanism that permits the interrelationship of strategic decisions for individual or company $i$. This study assumes that $\Sigma$ is multivariate normal and independent and identical across all $i$.

Let $y^D = (y_1, y_2, \dots, y_D)'$ and $y^H = (y_1, y_2, \dots, y_H)'$ be vectors of strategic decisions and attitudinal measurements (indicators), respectively, that are jointly observed for individual $i$. In the absence of indicator variables, the joint probability of $y^D$ is (subscripts for the individual are omitted):

$$\mathrm{P}(y^D | z^*, x, w, \delta_d, \gamma_d, \Sigma) \tag{15}$$

This probability expression is the basis of inferring the values of the model parameters. The notation P is used to imply that the outcomes contain both discrete and continuous aspects. However, as latent variables are used in the conditioning statement, the solution method must account for the fact that they are not observed. The most common approach utilized to mitigate this issue is to “integrate out” the latent variable, thus rewriting the probability expression as:

$$\int_{z*} \mathrm{P}(y^D | z^*, x, \delta_d, \gamma_d, \Sigma)\, g(z^* | w, \gamma_l) dz^* \tag{16}$$

Note that it is not necessary to include $y^*$ in this expression, since this quantity follows immediately from the conditioning variables.

It is also desirable to use information from the indicator measurements to express the joint probability of observing strategic decisions and indicators:

$$\mathrm{P}(y^D, y^H | x, w, \gamma_d, \gamma_l, \delta_d, \delta_h, \Sigma) = \int_{z*} \mathrm{P}(y^D | z^*, x, \delta_d, \gamma_d) f(y^H | z^*, \delta_h) g(z^* | w, \gamma_l) dz^* \tag{17}$$

At this point, the log-likelihood can be formed. The integral has one dimension for each latent variable. An application with multiple latent variables is often solved using MSL or MACML, but this work uses Bayesian estimation instead due to its advantages (Chapter 2). The solution method is presented in the next section.

### *5.2.3 Hierarchical Bayes: Gibbs sampling estimation method*

#### 5.2.3.1 Background and prior assumptions

Gibbs sampling is a specific case of the more general Markov Chain, Monte Carlo (MCMC) simulation methods. The Markov Chain aspect is due to using a one-step iteration process, with draws in each iteration having dependence on values in the previous iteration. The Monte Carlo aspect stems from the sampling nature, i.e., drawing values from a distribution in simulation.

Unlike classical inference, which treats an unknown parameter as single value, Bayesian inference treats the unknown parameter as a distribution from which any number of values may be realized. Bayesian statistics form the foundation of Bayesian inference, which has the following components. A prior distribution, which represents a belief regarding the distribution of the unknown parameter, is imposed. Observed sample data are used to update the belief. The output, which is referred to as the posterior distribution of the parameter conditional on the sample data, is used to infer certain qualities of the parameter (e.g., its mean).

This study uses the following prior distributions. Using assumptions or advance estimates, the prior distributions for $\gamma_l$ and $\delta_h$ are assumed to be $\gamma_l \sim N\left(\gamma^0, V_{\gamma_l}^0\right)$ and $\delta_h \sim N\left(\delta_h^0, V_{\delta_h}^0\right)$. The assumed prior distribution for $\Sigma$ is $\Sigma \sim IW(\nu, Q^0)$ where IW is the Inverse Wishart distribution, $\nu$ is the degrees of freedom and Q is the scale matrix. This simulation uses $\nu = \mathrm{N} + \mathrm{D}$ and $\mathrm{Q}^0 = \mathrm{I}$, the DxD identity matrix. Let $\beta_\mathrm{d}$ comprise all $\gamma_d$ and $\delta_\mathrm{d}$ parameters. The assumed prior distribution for $\beta_\mathrm{d}$ is normal: $\beta_d \sim MVN(\beta_d^0, V^0)$. To ensure that the prior for $\beta_\mathrm{d}$ is uninformative, this simulation uses $\beta_d^0 = (1,1, \dots ,1)'$ and a diagonal matrix with very large values ($V^0 = 100 \cdot I$), where the size of I is determined by the number of values in $\beta_\mathrm{d}$.

#### 5.2.3.2 Implementation

A Gibbs sampling simulation program runs through a number of iterations, drawing new parameter values in each iteration based on values of the other parameter values and observed (sample) observed outcomes. Solving the model with latent variables requires the formation of an augmented posterior distribution: P$(z^*, y^*, \gamma_l, \delta_h, \gamma_d, \delta_d, \Sigma|y^D, y^H)$. That is, the posterior distribution is augmented with the simulated latent variables, treating their distributions as additional distributions to be estimated.

In each iteration k, draws for each individual are made from the following set of full conditional distributions *h.* Subscripts for the individual are omitted for clarity of presentation. writing $y_d, y_h$ as $y$ and $\delta_d, \gamma_d$ as $\beta_d$:

$$y_d^{*(k)} \sim h\left(y_d^* \middle| z^{*(k-1)}, \gamma_l^{(k-1)}, \beta_d^{(k-1)}, \delta_h^{(k-1)}, \Sigma, y\right) \tag{18}$$

$$z^{*(k)} \sim h\left(z^* \middle| y_d^{*(k-1)}, \gamma_l^{(k-1)}, \beta_d^{(k-1)}, \delta_h^{(k-1)}, \Sigma, y\right) \tag{19}$$

$$\gamma_l^{(k)} \sim h\left(\gamma_l \middle| y_d^{*(k-1)}, z^{*(k-1)}, \beta_d^{(k-1)}, \delta_h^{(k-1)}, \Sigma, y\right) \tag{20}$$

$$\delta_h^{(k)} \sim h\left(\delta_h \middle| y_d^{*(k-1)}, z^{*(k-1)}, \gamma_l^{(k-1)}, \beta_d^{(k-1)}, \Sigma, y\right) \tag{21}$$

$$\beta_d^{(k)} \sim h\left(\beta_d \middle| y_d^{*(k-1)}, z^{*(k-1)}, \gamma_l^{(k-1)}, \delta_h^{(k-1)}, \Sigma, y\right) \tag{22}$$

$$\Sigma^{(k)} \sim h\left(\Sigma \middle| y_d^{*(k-1)}, z^{*(k-1)}, \gamma_l^{(k-1)}, \beta_d^{(k-1)}, \delta_h^{(k-1)}, y\right) \tag{23}$$

The sequence of draws converges to the joint posterior distribution of $(y^*, z^*, \gamma_l, \delta_h, \beta_d, \Sigma|y)$.

Since $z^*$ is not observed, information from its indicator measurements $(y_h)$ must be incorporated to make inference regarding $z^*$ and $\gamma_l$. This information is contained in the distribution $h(z^*|\, y_h)$ as illustrated in the following note.

Using linear algebra principles, matrix partitioning can generate useful statistics for conditional distributions. For $Z = (Z_1, Z_2)' \sim N(\mu, \Sigma)$ let $\mu = \begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}$ and $\Sigma = \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12}' & \Sigma_{22}\end{pmatrix}$. Then $Z_1|Z_2 \sim N(\mu_{1|2}, \Sigma_{1|2})$ with $\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(Z_2 - \mu_2), \Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}'$.

Assuming the following distribution, with I as the identity matrix of size HxH:

$$\begin{pmatrix} z^* \\ y_h \end{pmatrix} \sim MVN\left(\begin{bmatrix} \gamma_l' w \\ \delta_h \gamma_l' w \end{bmatrix}, \begin{bmatrix} 1 & \delta_h' \\ \delta_h & \delta_h \delta_h' + \mathrm{I} \end{bmatrix}\right), \tag{24}$$

a draw from the conditional probability $h(z^*|\gamma_l, \delta_h, y_h) \sim N\left(\mu_{z^*|y_h}, \sigma^2_{z^*|y_h}\right)$ is made for each observation. Matrix partitioning permits a mean and variance, conditional on the information in the measurement data, to be drawn:

$$\mu_{z^*|y_h} = \mu_{z^*|y_h} + \delta_h'[\delta_h \delta_h' + \mathrm{I}]^{-1}[y_h - \delta_h \gamma_l' w] \tag{25}$$

$$\sigma^2_{z^*|y_h} = \mathrm{I} - \delta_\mathrm{h}'[\delta_h \delta_h' + \mathrm{I}]^{-1}\delta_h \tag{26}$$

Since the data augmentation method generates values for the latent variables, they are treated as observable in their conditional distributions (Equations 17 and 18). As such, the conditional distributions for $\gamma_l$ and $\delta_h$ fit into a standard Bayesian regression construction:

$$h(\gamma_l | y_d^*, z^*, \beta_d, \delta_h, \Sigma, y) \sim N\left(\tilde{\gamma}_l, \tilde{V}_{\gamma_l}\right) \tag{27}$$

$$h(\delta_h | y_d^*, z^*, \gamma_l, \beta_d, \Sigma, y) \sim N\left(\tilde{\delta}_h, \tilde{V}_{\delta_h}\right) \tag{28}$$

It follows that draws are made using the closed form equations:

$$\tilde{V}_{\gamma_l} = \left(V^0_{\gamma_l} + w'w\right)^{-1}, \tilde{\gamma}_l = \tilde{V}_{\gamma_l}\left(V^{0^{-1}}_{\gamma_l} + w'z^*\right) \tag{29}$$

$$\tilde{V}_{\delta_h} = \left(V^0_{\delta_h} + z^{*\prime}z^*\right)^{-1}, \tilde{\delta}_h = \tilde{V}_{\delta_h}\left(V^{0\ -1}_{\delta_h} + z^{*\prime}y_h\right)$$

$\beta$ and $\Sigma$ are updated using the closed forms of their full conditionals: $\beta | y^*, \Sigma \sim N(\bar{\beta}, \bar{V})$ and $\Sigma | \mathrm{y}^*, \beta \sim IW\left(\nu + N, \Sigma_0 + (Y^* - X'\beta)'(Y^* - X'\beta)\right)$. The matrix X is in block diagonal form for each individual. A draw from these distributions is performed at each iteration.

$$\bar{V} = \left(V_0^{-1} + \sum\nolimits_N X'\Sigma^{-1}X\right)^{-1} \text{ and } \bar{\beta} = \bar{V}(V_0^{-1}\beta_0 + \sum\nolimits_N X'\Sigma^{-1}Y^*) \tag{30}$$

Draws from the full conditionals for $\mathrm{y}^*$ are more complex since they involve truncated distributions. Again, following previous researchers (Chapter 2), matrix partitioning is employed as a computational device to isolate one part of a joint distribution. For example, let us consider the objective of drawing a value of $y_{1i}^*$ conditional on $y_{2i}^*, \beta$ and $\Sigma$, knowing that the observed value $y_{1i} = 0$. Then

$$y_{1i}^* | y_{2i}^*, \beta, \Sigma \sim N\left(\mu_{1|\neg 1}, \sigma_{1|\neg 1}\right) \mathcal{I}_{y_{1i}^* < 0} \tag{31}$$

The last term is an indicator that equals one if the condition in the subscript is met and zero otherwise. In this case, the value is only drawn from the range (-Infinity, 0). The mean and variance are computed using the matrix portioning method described earlier, with the conditioning taken on all $y_{j \neq i}^*$, which are shaped into appropriately sized vectors and matrices to fit the problem.

The simulation draws $y_{\mathrm{i}}^*$ from a truncated multivariate normal distribution, using truncation thresholds that are based on the observed data. For example, for companies that own a private fleet, a draw from the joint distribution of medium-duty and heavy-duty truck fleet sizes is performed.

### *5.2.4 Proof of concept: Approach*

The methodology for modeling agent strategies is deployed in a proof-of-concept demonstration. In this demonstration, the parameters of the modeling system are first estimated using Gibbs sampling with 10,000 iterations. As many as 30,000 iterations were tested, but the additional iterations had negligible impact on the results. Finally, the models are applied to simulate fleets and DCs for the population of firms in this study.

#### 5.2.4.1 Model structure and specifications

This subsection presents the specifications of each model. Throughout the discussion, individual or company level subscripts are omitted for sake of clarity. Figure 38 shows the architecture of the entire model and the input, output and latent variables.

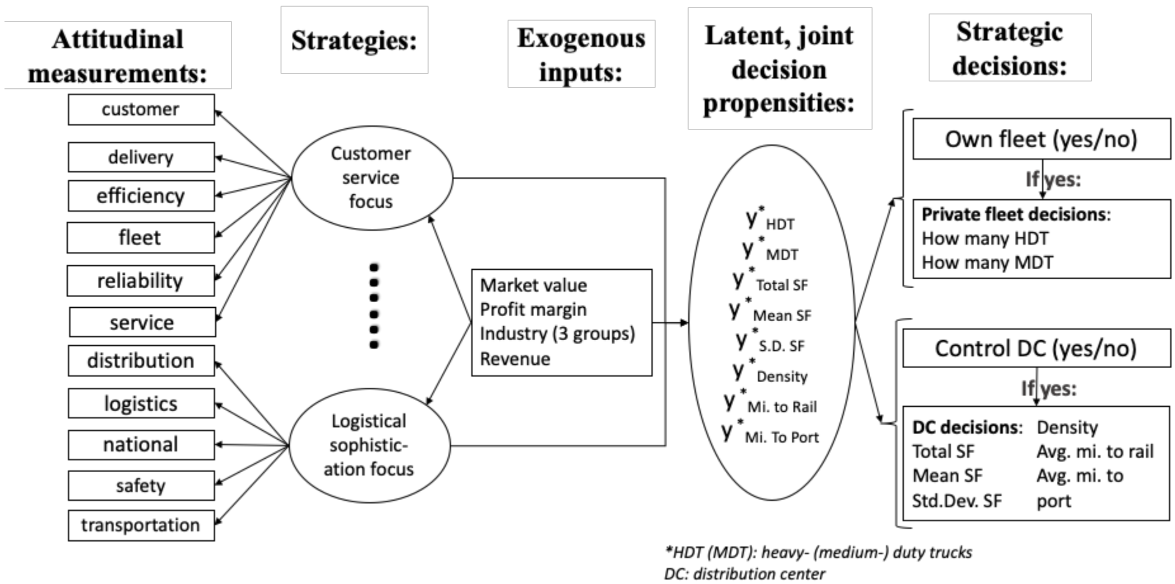



**Figure 38. Proof of concept model: Schematic**

Two latent strategies are hypothesized and tested:

- Customer service (CS) focus: Emphasizes reliability and efficiency in delivering products to the customer, focusing especially on last-mile delivery; and
- Logistical sophistication (LS) focus: Emphasizes large-scale distribution and transportation systems, which function together in logistical operations.

Scores for each strategy are predicted using market value (M) and percentage profit (P), which is computed as profit divided by revenue. These variables are used under the premise that larger companies may be more globally oriented and therefore oriented more towards logistical sophistication, while companies in less profitable industries may focus on excelling in customer service. These variables also perform better in the model, with higher statistical significance than alternatives that are tested (which include asset value and revenue). The latent strategy equations are thus specified as:

$$z_{CS}^{*} = \alpha_{CS} + \gamma_{CS,M}M + \gamma_{CS,P}P + \eta_{CS} \tag{32}$$

$$z_{LS}^{*} = \alpha_{LS} + \gamma_{LS,M}M + \gamma_{LS,P}P + \eta_{LS} \tag{33}$$

A total of 11 equations are used to measure strategy scores. The first six use the following keywords to measure CS: customer, delivery, efficiency, fleet, reliability and service. The remaining five equations use other keywords to measure LS: distribution, logistics, national, safety and transportation. Each equation has the form: $y_h = \delta_h z_l^* + \varepsilon_h$ (see Section 5.2.2). For example, one equation that relates strategy to an attitudinal measurement is:

$$y_{customer} = \delta_{customer} z_{CS}^* + \varepsilon_{customer} \tag{34}$$

The model predicts the following strategic decisions for each company: Number of heavy duty trucks owned (HDT), number of medium duty trucks owned (MDT), total DC square footage (SF), mean DC SF, standard deviation (SD) in SF, density (SF per person), average distance in miles to the closest major intermodal truck-rail facility (miles to rail), and average distance in miles to the closest major water port (miles to water). The same specification is tested to measure the latent outcome $y_d^*$ for each strategic decision *d*:

$$y_d^* = \alpha_d + \gamma_{ind1,d} \mathcal{I}_{ind1} + \gamma_{ind2,d} \mathcal{I}_{ind2} + \gamma_R R + \delta_{CS,d} z_{CS}^* + \delta_{LS,d} z_{LS}^* + \varepsilon_d, \tag{35}$$

where $R$ is annual revenue and $\mathcal{I}_{ind\,x}$ is a dummy variable that equals one if the company belongs to industry sector x. In the proof of concept, *ind1* includes (a) the food and beverage sector and (b) the petroleum and gases industries. *Ind2* includes the retail and wholesale sectors. The base or reference industry category includes manufacturing and processing sectors.

### 5.2.4.2 Variable transformations

For the estimation of the models, the number of heavy-duty trucks *H* owned by the company is transformed as $\log\left(1 + \frac{H}{1{,}000}\right)$. Adding one to the term in parentheses is necessary to avoid taking the logarithm of zero. The number of medium-duty trucks is transformed in the same way. Total DC square footage *T* is transformed as $\log\left(1 + \frac{T}{1{,}000{,}000}\right)$. The transformation of mean and standard deviation in square footage is similar but uses a factor of 100,000 instead of one million. The density quantity, $dens = \frac{T}{\sum_{\{z:T_Z>0\}} POP_Z}$, is transformed as $\log\left(1 + 10 \cdot \left(\frac{T}{dens}\right)\right)$, where $POP_Z$ is the total population of zone *z* and $T_Z$ is the company's total square footage in zone *z*.

The intermodal truck-rail terminal accessibility variable has a more complex transformation, which accounts for the continuous nature of the underlying latent variable for this strategic decision. Let $MI_{w,c}$

represent the distance in miles between distribution center $w \in W$ belonging to a given company $c \in C$ and the closest intermodal truck-rail (IMX) terminal. Denote the mean shortest distance for company $c$ as $\overline{MI_c}$. Figure 39 illustrates the underlying reasoning, which proceeds as follows. In Regime 1, companies that choose to not operate DCs have $y^*_{\overline{MI_c}} < 0$, since the accessibility of DCs to transportation is not part of their thinking. Regime 3 consists of similar companies, which have DCs but do not weigh accessibility to IMX very seriously. Regime 2 comprises companies for whom accessibility to IMX appears to be a very important when choosing DC locations. Ideally, Regime 1 should neighbor Regime 3; currently, these two regimes are disjointed.

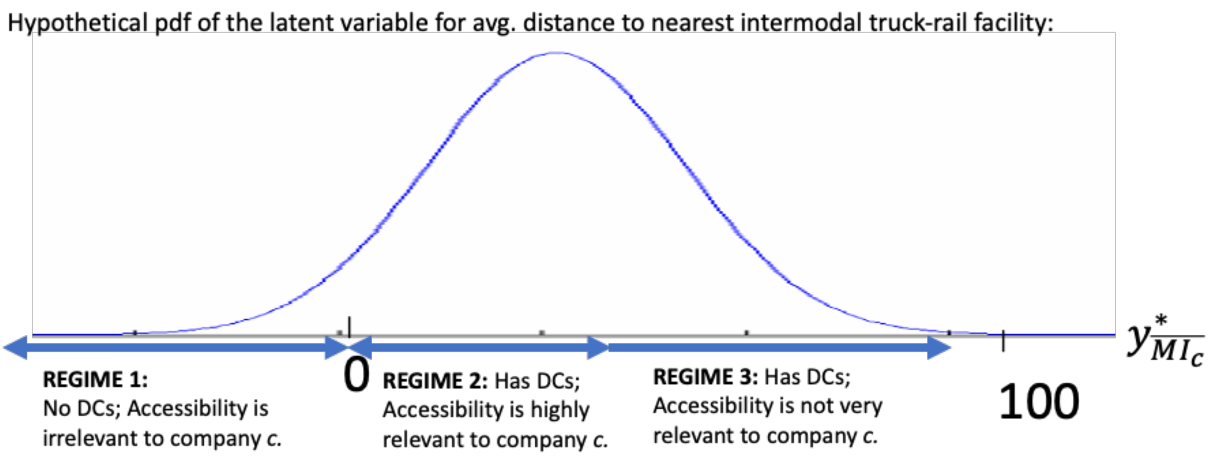



**Figure 39. Intuitive reasoning behind the transformation of accessibility variables.**

The following process is adopted to remedy this issue. First, compute the transformation $\overline{MI_c}^t = \log(1 + \overline{MI_c})$ and let $\overline{MI}^t_{MAX} = \max(\overline{MI_c}^t, c \in C)$. Then, for company c:

$$\overline{MI_c} = \begin{cases} \overline{MI}^t_{MAX} - \overline{MI}^t_c, & T_c > 0 \\ 0, & otherwise. \end{cases}$$

This transformation effectively creates the mirror image of positions $y^*_{\overline{MI_c}}$ values in Regimes 2 and 3. Subsequently, Regime 3 is adjacent to $y^*_{\overline{MI_c}} = 0$ and Regime 2 is positioned to the right. The transformation is exactly the same for the variable that represents proximity to the nearest water port, with the exception that its log transformation is $\log\left(1 + \frac{\overline{MI_c}}{10}\right)$.

Revenue *R* is transformed as 0.1*log(*R*) and Market Value *M* as 0.05*log(*M*). Percentage profit *P* is transformed as log(2+*P*), which ensures that the quantity is always positive.

#### 5.2.4.3 Simulation of fleets and nationwide distribution centers

A simulation is performed, applying the model to each company from the original dataset and simulating its strategic decisions regarding private fleets and DCs. The simulation predicts private fleet ownership and the number of HDT and MDT owned, as well as DC control and the attributes of its DC system (total square footage, mean square footage, etc.).

- Next, an algorithm is developed to assign the simulated DCs to a FAF zone within the US. This algorithm is not described in detail here, but more information is available from the author upon request. The main features include:
- Selecting a FAF zone *f* at random;
- Comparing the company's strategic density (+/- a buffer) to the hypothetical density for that company if it were to place either one or two DCs in *f*;
- Placing a DC in that zone if the hypothetical density falls within the buffered strategic density range.

This algorithm is later referred to as the Nationwide Zone Assignment Algorithm in Section 5.4.

### *5.2.5 Interpretation and characterization of model outcomes*

As discussed earlier, $y_d^*$ as a truncated variable provides information to simultaneously predict both binary and continuous strategic decisions. For example, if its predicted value is less than the threshold (zero in this case), then the decision is "No". Otherwise, the decision is "Yes" and the modeled outcome becomes the predicted amount (of total DC square footage, for instance).

Figure 40 characterizes the outputs of the model in this particular application. Estimates of fleet ownership can be used directly in the agent-based framework to represent the fleet size and mix for each company in the model population that chooses to own a fleet. As noted in Fang (2008), this outcome is characterized as a multiple discrete-continuous (MDC) outcome since the total consumption volume is determined for each category (vehicle type, in this case).

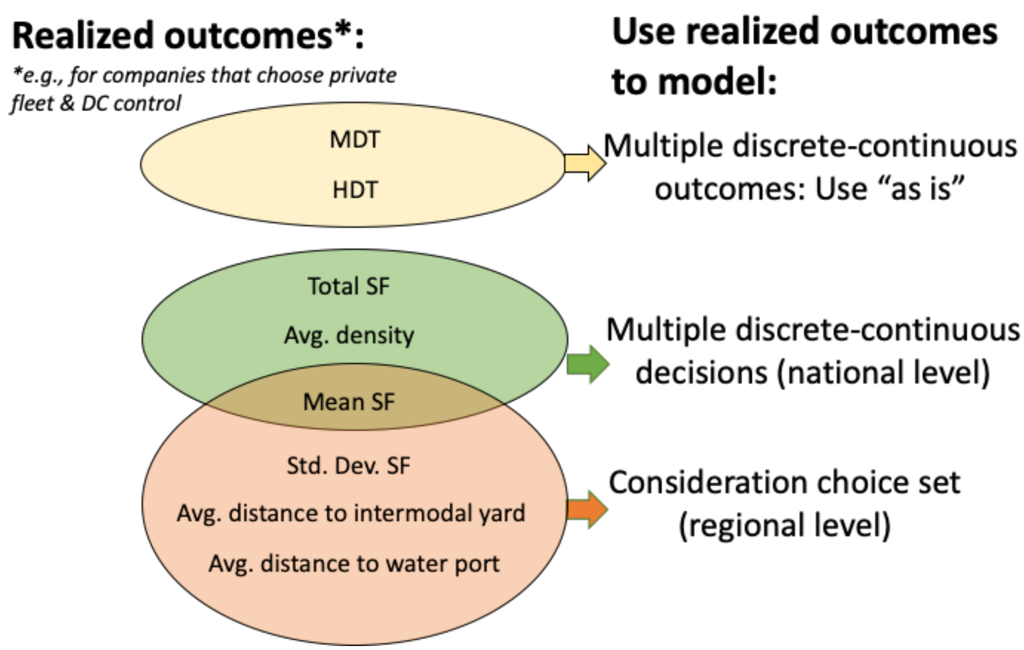



**Figure 40. Characterizing the model outputs.**

The estimated quantities of DC attributes inform a number of dimensions regarding the distribution structure of each company. The estimated total square footage is considered jointly with the company's average density per zone in order to estimate the amount of distribution space that the company assigns to each zone. This outcome is likewise characterized as an MDC outcome. A discretized version of this outcome is obtained by dividing the simulated total square footage by the mean square footage. The result of this calculation is another useful quantity: the total number of DCs that are operated by a company.

Finally, the mean and standard deviation in square footage, as well as the accessibility outcomes, are used as parameters that inform the generation of a consideration choice set. This is performed, for example, by winnowing down the list of all regional DC properties to a set that only includes properties with (a) square footage within one standard deviation of the mean, and (b) are within +/-10% of the average distance to major intermodal yards and water ports. Subsequently, the consideration choice set will be used as an input to a second-stage model that is developed specifically for the application of interest.

## 5.3 Data

This chapter uses the same company and attitudinal data sources as discussed in Chapter 4, which also contains a full description of the attitudinal data development process. The data in the current chapter contains more specificity on fleet and distribution center characteristics. The rest of this section describes and summarizes the various data inputs.

### *5.3.1 Company attributes and strategic decisions on logistics controls*

The 2017 Fortune 500 dataset contains the following attributes, which are used in this study: number of employees, revenue, profit, and market value. Other characteristics, including the value of assets and the share equity, are also tested but not used since the other financial measures (revenue, profit, market value) provide a better fit in statistical models that are developed later using the data. The analysis includes 260 companies from this list that belong to freight-intensive sectors.

This study evaluates two areas of strategic, logistics-related decisions. Private fleet ownership, including fleet characteristics, is one area of interest. The other area includes strategic decisions related to distribution centers. The rest of this section discusses data sources and input variables for these decisions.

The Top 500 listing compiled by FleetOwner magazine is used to identify which companies in the Fortune 500 list have a private fleet, and to identify the characteristics of these fleets. The following data items are available for the top 500 private fleets in the US:

- Company name
- Location of headquarters
- Total Vehicles
- Total Trucks
- Total Tractors
- Total Trailers

- Industry: Retail/Wholesale, Petroleum/Gases, Food Products, Manufacturing/Processing

This analysis assumes that tractors are heavy-duty trucks (HDT), and that the difference between Total Trucks and Total Tractors consists of medium-duty (MDT) trucks.

The CoStar real estate dataset is comprehensive, with coverage of over four million commercial properties, and contains information that is regularly updated and verified. The database identifies both the owner and, for properties that are leased, the tenant(s). This information is used to identify all distribution centers that are owned and/or leased by the companies in this study. In addition to the occupant's identity, the database provides information of the amount of space in the property (in square feet, or SF) that is occupied by the company. Finally, the use of the property (e.g., for manufacturing, retail, distribution, etc.) is available.

Although the database has comprehensive coverage of distribution centers and includes information on how the property is used, identifying distribution centers is not always straightforward. For example, examining the data reveals that sometimes a warehouse appears to be mislabeled as a distribution center in the original database. Therefore, a filtering process is used to distinguish properties that qualify as distribution centers for the current analysis. A manual inspection of roughly 100 properties, which included visual inspection of properties using satellite imagery, informed the criteria that are used in the filtering process.

The following process is used to identify distribution centers in the real estate database. First, records that have a property use of "Distribution", "Warehouse", "Light Distribution" and "Refrigeration/Cold Storage" are selected. All properties that are labeled as "Distribution" in the raw data are labeled as distribution centers for this study. Second, the size of the occupied space is used. When the occupied space meets or exceeds a certain threshold, the property is labeled as a distribution center; otherwise, the property is labeled as a warehouse. The thresholds are listed in Chapter 1.6.

Geospatial characteristics of the distribution centers—namely, accessibility to major transportation terminals and density normalized by population—are developed using data from the BTS, Census and FAF as mentioned earlier. Geospatial analysis generates the following fields for each company:

- Density (SF per population unit): total SF of distribution center space that is occupied by a company in a given FAF zone divided by the population of the zone
- Average distance (miles) to the nearest major rail-truck intermodal facility

- Average distance (miles) to the nearest major water port.

The last two items are computed as follows. The distance to the closest transportation facility is first computed for each distribution center. The set of distances is then averaged for each company.

Figure 41 shows the locations of distribution centers, intermodal rail-truck facilities, and major water ports that are used in this study. FAF zones are shown as well. Distribution centers are heavily concentrated in areas with high population, while farmland and desert/arid regions have fewer distribution centers. Water ports are naturally concentrated on the shores of oceans, rivers, and the Great Lakes. Intermodal truck-rail yards are relatively dispersed throughout the country.

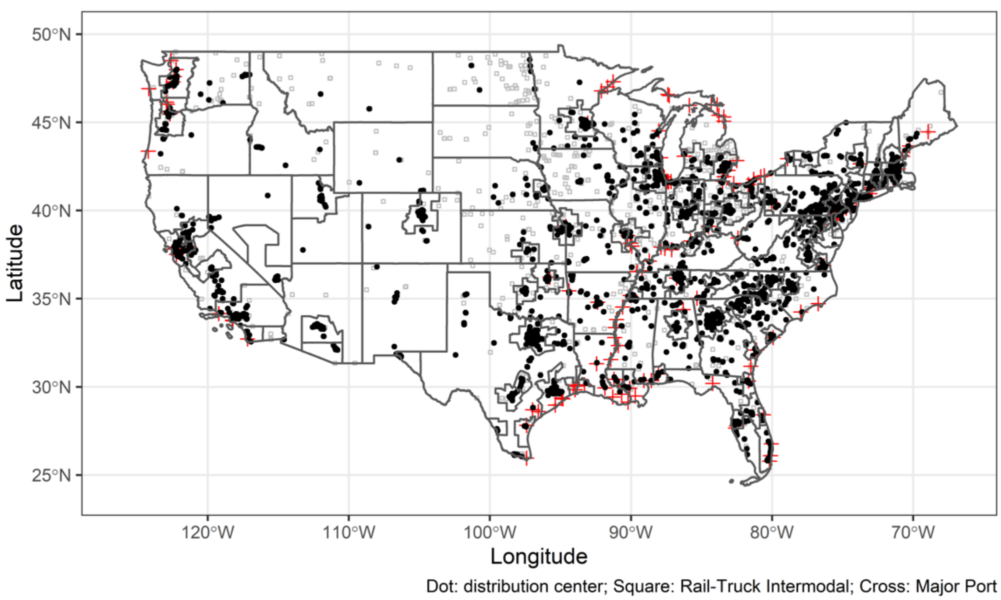



**Figure 41. Distribution centers and major intermodal transportation facilities with FAF zone boundaries.**

Table 8 summarizes the basic firm attributes, private fleet data and distribution center attributes across firms in the dataset. Two outliers are excluded to avoid skewing the results: one company with over 60,000 trucks (the next-highest number of trucks is about 15,000) and one company with fewer than 100 employees (the next-lowest is about 2,000). Companies with limited keyword usage, which is identified as using fewer than 50 percent of the keywords, are eliminated from the analysis. Seven privately-owned companies have no

10-K report and are likewise removed. Of the original 260 companies, a total of 247 companies remain for analysis.

**Table 8. Summary statistics: firm attributes.**

| Variable | Number of firms* | Mean | Min | Max | Std. Dev. |
|---|---|---|---|---|---|
| Basic firm attributes | | | | | |
| Fortune 500 rank (2017) | 247 | 239 | 1 | 496 | 141 |
| Revenue** | 247 | 26,199 | 5,197 | 485,873 | 44,473 |
| Profit | 247 | 1,421 | -5,763 | 16,540 | 2,790 |
| Percent profit | 247 | -56% | 6% | 81% | 13% |
| Assets | 247 | 29,892 | 1,497 | 365,183 | 48,317 |
| Market value | 247 | 37,846 | 241 | 423,031 | 58,905 |
| Employees (2017) | 247 | 67,345 | 1,770 | 2,300,000 | 161,174 |
| Industry | | | | | |
| Manufacturing/Processing | 147 | | | | |
| Retail/Wholesale | 65 | | | | |
| Petroleum/Gases | 22 | | | | |
| Food Products | 13 | | | | |
| Private fleet attributes | | | | | |
| Total number of trucks | 61 | 3,377 | 512 | 15,449 | 3,617 |
| Heavy duty trucks (HDT) | 53 | 1,175 | 1 | 8,271 | 1,997 |
| Medium duty trucks (MDT) | 61 | 2,354 | 4 | 13,497 | 3,012 |
| Distribution center attributes | | | | | |
| Total square feet (SF) | 209 | 6,879,151 | 9,367 | 159,668,069 | 16,318,431 |
| Number of DCs | 209 | 14 | 1 | 219 | 23 |
| Mean SF | 209 | 456,578 | 9,367 | 1,847,485 | 319,703 |
| Std.Dev.(SF) | 184 | 371,991 | 9,030 | 2,084,059 | 343,494 |
| Mean SF per person*** | 209 | 0.26 | 0.001 | 2 | 0 |
| Avg. miles to rail-truck intermodal | 209 | 9 | 0.27 | 81 | 8 |
| Avg. miles to major water port | 209 | 111 | 3 | 611 | 81 |

*Statistics shown are based on firms which have non-zero values for the variable.
**All financial figures in millions of 2017 USD
***Density in terms of mean square feet per person in the FAF zones in which the firm has 1 or more DCs

As the table indicates, considerable variation exists among the companies in terms of their size, financial characteristics, and logistics controls. Revenue, for instance, ranges from five billion annually to 500 billion. Number of employees ranges from about 2,000 to 2.3 million. About 25 percent of the companies are the retail or wholesale trade sectors, 15 percent in food or petroleum and gases industries, and the remainder are in manufacturing or processing industries. Food and the petroleum/gases sectors are combined for analysis since their individual sample sizes are somewhat low. About 25 percent of firms own a private fleet while 87 percent own and/or lease DCs. Fleet size ranges from about 500 to 15,000 and the total square feet that is owned or leased ranges from about 10,000 to 160 million. The mean and standard deviation in square footage of DCs for a company both range from about 9,000 to 2 million.

The geospatial analysis likewise reveals considerable variety among companies. On average, based on population in FAF zones where the company has DCs, the company has 0.26 square feet of DC space per person. This figure ranges as high as 2.21. On average, DCs are located nine miles from major rail-truck intermodal yards, but for some companies choose to locate much closer to such facilities on average. Similarly, some companies choose to locate their DCs extremely close to a major water port (10 miles or less), although overall proximity to a major water port appears to be less important with an average distance of over 100 miles.

### *5.3.2 Attitudinal measurement data*

Attitudinal measurements are developed for each company using publicly available, large-scale text data in the form of 10-K reports that were filed in year 2017 by publicly owned companies with US headquarters with the United States Securities and Exchange Commission (2020). These measurements are developed using the natural language-based, attitudinal measurement data generation process described in Chapter 4. Briefly, this methodology transforms text data to quantitative measurements of keyword usage that is unique to each company.

Figure 42 illustrates the attitudinal measurements of keywords that are used in this analysis. The 22 keywords represent a mix of attitudes and logistics functions. Four unique symbols are used in the plot, with one for each possible combination of two binary variables: owns/does not own a fleet (T1 and T0, respectively) and controls/does not control DC operations (DC1 and DC0, respectively. While the measurement values belong to $\mathbb{R}^1$ space (i.e., the measurements are real, one-dimensional numbers), the points in the plot are randomly shifted in a second dimension so that the reader may distinguish more clearly between points. For each keyword, the right side of the plot lists the measurement variance that is captured by the first principal component.

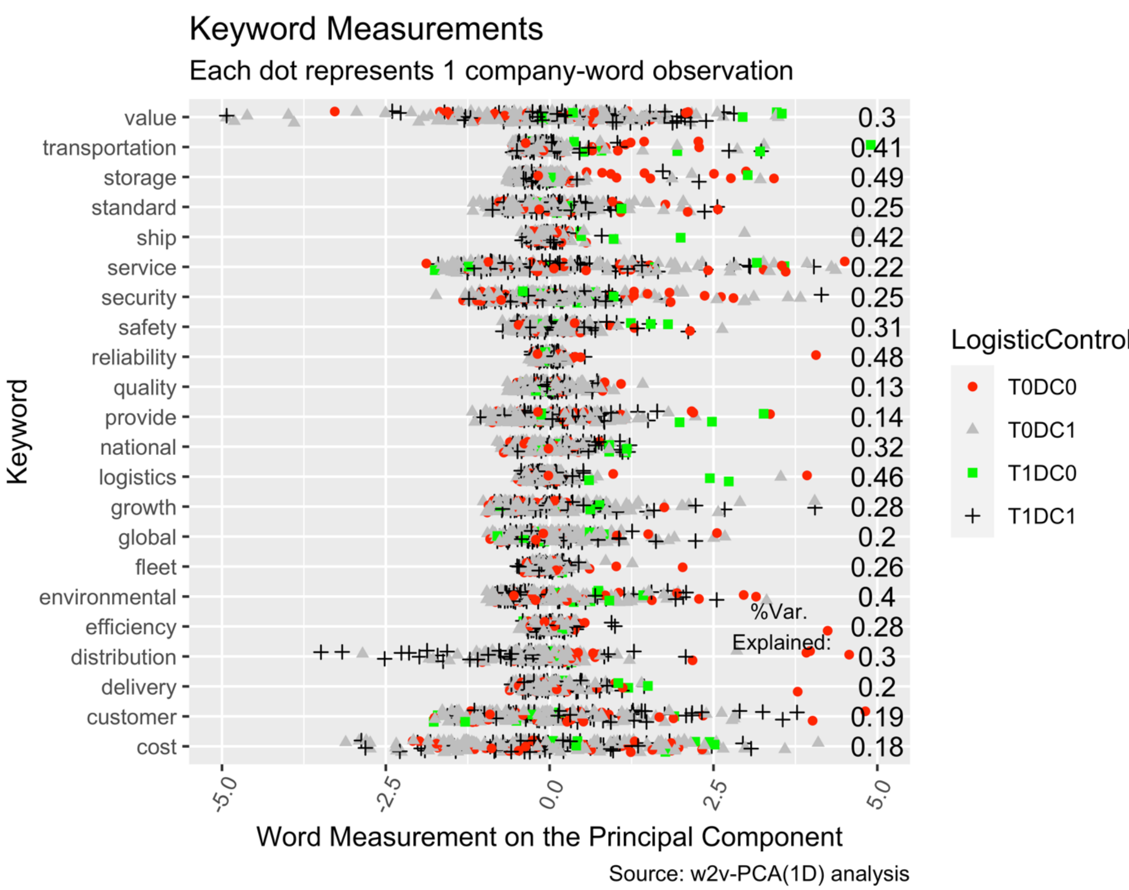



**Figure 42. "Jitter" plot of w2vPCA keyword measurements.**

There are two main points to make with respect to the interpretation of the numerical values. First, and most importantly, the w2vPCA attitudinal measurement algorithm identifies differences in word usage and results in measurements of those difference. However, the output measurements are not necessarily equivalent to “more” or “less” of some amount. In other words, the meaning of the directionality (e.g., positive, negative) in measurements comes from subsequent analysis, as will be shown in the discussion for Table 11 and Table 12 (Section 5.4.1).

In general, the spread of attitudinal measurements differs for various keywords. For example, the range of measurements for "national" is approximately -1 to 1, while the range for "cost" is about -3 to 4. As discussed in Chapter 4, the different ranges represent the implicit spread of differences in word uses by companies based on a sample of their natural language. The methodology for generating these measurements, in other words, does not constrain the measurement scale using pre-imposed values (such as 1=disagree, 2=neutral, 3=agree) as is common in current practice.

For the proof of concept, the keyword measurements in this figure are used to measure the two hypothesized, underlying firm strategies. For each company, the model estimates a quantitative value for each of these strategies and their impacts on strategic decisions.

## 5.4 Proof of Concept

This section presents the proof of concept results, including the estimated model parameters.

### *5.4.1 Estimation results*

The Gibbs Sampler estimation generates a different point estimate in each iteration. All model parameters, including the covariance terms, are point estimates based on the distribution of values that are generated throughout the simulation. The mean is used as the parameter estimate. As an example, Figure 43 shows the distribution of parameter estimates for two variables in the Total SF model: the retail/wholesale industry dummy variable and the LS strategy variable. The distribution is based on the 1,001-10,000$^{th}$ iterations of the process, disregarding the first 1,000 as "burn-in" estimates. The information in these distributions generates the standard error in addition to the point estimates of each parameter.

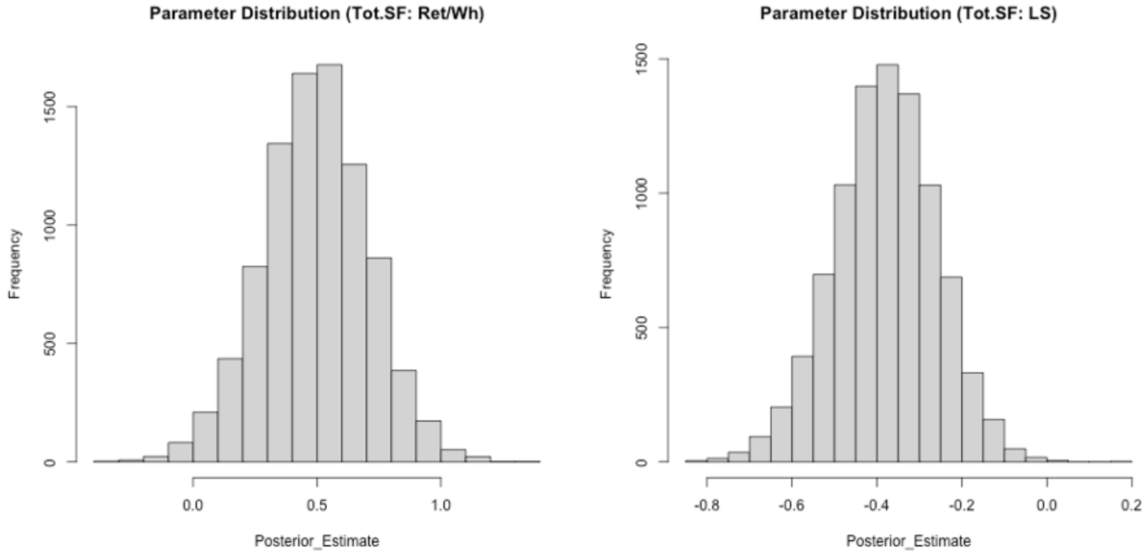



**Figure 43. Distribution of parameter estimates – examples.**

Table 9 presents the estimation results for the strategy model. The positive intercepts indicate both CS and LS scores for a firm with no market value and zero profit will be positive. An increase in market value will make each score decrease, with CS decreasing faster as market value grows. Similarly, both scores decrease as profit increases. Conversely, both scores increase as unprofitability (negative profit) grows. The standard errors on market value and percent profit are relatively wide, which implies a fair amount of noise or variability in the underlying data. However, the estimates are similar across various estimations of the model system, therefore they are considered to be reliable for purposes of this model. Future versions of the model will try to include more explanatory variables, which may have lower standard errors.

**Table 9. Estimation results: strategy model.**

| | *Customer service (CS) focus* | | | *Logistical sophistication (LS) focus* | | |
|---|---|---|---|---|---|---|
| *Variable (w)* | *Est.* | *(s.e.)* | *t-stat* | *Est.* | *(s.e.)* | *t-stat* |
| Intercept | 0.49 | (0.54) | *0.90* | 0.40 | (0.57) | *0.69* |
| Market Value | -0.20 | (0.74) | *-0.28* | -0.07 | (0.79) | -0.09 |
| % Profit | -0.55 | (0.72) | *-0.77* | -0.51 | (0.75) | *-0.68* |

*Dependent variable: z**

Table 10 shows the parameter estimates for the strategy measurement model. These results are truly remarkable. Indeed, the estimates for the "customer" word measurement and "service" word measurement stand out as the most prominent indicators of the CS strategy. Not only are their parameters relatively high, but both words have a relatively large spread (Figure 42)—as such, the parameter estimate multiplied by the measurement can be quite high, creating a relatively large impact in detecting CS. Other keywords that are used to measure this strategy are "delivery", "efficiency", "fleet" and "reliability". Their parameter estimates are lower than the other two, but they certainly are intuitively consistent with the notion of providing excellent customer service.

**Table 10. Estimation results: strategy measurement model.**

| *Measurement variable h:* | *z*: Customer service (CS) focus* | | | *Measurement variable h:* | *z*: Logistical sophistication (LS) focus* | | |
|---|---|---|---|---|---|---|---|
| | *Est.* | *(s.d.)* | *t-stat* | | *Est.* | *(s.d.)* | *t-stat* |
| customer | 0.77 | (0.08) | *9.50* | distribution | 0.74 | (0.09) | *8.29* |
| delivery | 0.18 | (0.07) | *2.64* | logistics | 0.19 | (0.07) | *2.74* |
| efficiency | 0.09 | (0.07) | *1.31* | national | 0.05 | (0.07) | *0.75* |
| fleet | 0.04 | (0.07) | *0.65* | safety | 0.16 | (0.07) | *2.17* |
| reliability | 0.07 | (0.07) | *1.02* | transportation | 0.35 | (0.07) | *4.66* |
| service | 1.06 | (0.09) | *11.99* | | | | |

*Dependent variable:* $y_h$

The findings for the LS strategy are similarly remarkable, with the highest parameters for "distribution", "transportation", and "logistics". These findings are highly consistent with the notion of a strong focus on logistics services as part of a company's strategy. "National" and "safety" also provide measures of the LS strategy, although are less impactful.

The discussion now focuses on the estimation results for the eight strategic decisions models. The two models related to private fleet are discussed first in Table 11. The reader is reminded that this model predicts the latent variable, $y_d^*$, which is manifest in binary (yes/no) and quantity decisions. As $y_d^*$ increases, the company is likely to both (a) have a private fleet and (b) have a larger private fleet. In other words, increases or decreases in $y_d^*$ impact both the binary yes/no and fleet size decisions in a similar way.

**Table 11. Estimation results: strategic decisions model (private fleet).**

| *Strategic decision d:* | *HDT* | | | *MDT* | | |
|---|---|---|---|---|---|---|
| *Variable (x, z*)* | *Est.* | *(s.e.)* | *t-stat* | *Est.* | *(s.e.)* | *t-stat* |
| Intercept | -3.04 | (0.85) | *-3.55* | -3.04 | (1.47) | *-2.07* |
| Food/Petro./Gases | 1.66 | (0.24) | *6.95* | 2.53 | (0.42) | *6.06* |
| Retail/Wholesale | 0.29 | (0.2) | *1.46* | 1.01 | (0.35) | *2.91* |
| Revenue | 1.97 | (0.82) | *2.40* | 1.04 | (1.47) | *0.71* |
| Customer service (CS) focus | 0.05 | (0.08) | *0.63* | 0.12 | (0.15) | *0.78* |
| Logistical sophistication (LS) focus | -0.11 | (0.08) | *-1.33* | -0.12 | (0.14) | *-0.86* |

*Dependent variable:* $y_d^*$

The baseline for this model is a manufacturer with zero revenue and zero CS and LS scores. All coefficients can be interpreted with respect to this baseline. The negative intercept indicates that the baseline company will have $y_d^* < 0$, and therefore will not own a private fleet. This is consistent with the data, in which 75 percent of companies do not own a private fleet. As revenue increases, the disposition towards fleet ownership and larger fleet size grows. This could be associated with geographic factors, as the highest revenue firms generally have widespread geographic coverage that could be well suited for handling a deep bench of drivers, trucks and trailers.

Membership in the food and beverage or petroleum/gases sectors is associated with much higher disposition to fleet ownership and larger fleets. Interestingly, a greater preference for MDT over HDT emerges (judging by the difference in parameter estimates). This suggests that large portion of the fleet for these companies is involved with urban delivery, which uses MDT, rather than long haul, which uses HDT. Similar effects occur but with less impact for companies in retail or wholesale sectors.

Based on the model output, underlying strategies do appear to influence strategic fleet-related decisions, but with less pronounced impacts than the above factors. Higher CS or lower LS scores are both associated with slightly higher disposition toward fleet ownership and larger fleets. The estimates have wide standard errors, which again indicates considerable variety in company behavior. Overall, the model results indicate that with respect to private fleet ownership and size, industry and company size are the most important factors, but that some companies do appear to own private fleets in part to differentiate themselves per their CS and LS strategies.

Table 12 shows the estimation results for the remaining six strategic decisions models, which are all related to DC control. The interpretation of $y_d^*$ and the baseline company are the same here as they are in the private fleet decisions models. In comparison to manufacturing firms, companies in food and beverage or petroleum/gases sectors are less likely to operate their own DCs, and when they do operate their own DCs, the total square footage and mean size are likely lower. In contrast, retail and wholesale firms are more likely to control their DCs, and they tend to control more total area than manufacturers. Firms with more revenue predominantly control their DCs, have more total area, and have larger DCs than lower-revenue firms. In regards to standard deviation in the size of individual Ds, manufacturers have about the same variability as retail/wholesale companies, and both have more variability than DCs belonging to companies in the food/beverage and petroleum/gases sectors. Higher values for CS and LS strategies are associated with lower disposition towards DC control, with lower total DC area, smaller DCs and less variability in DC size.

## Table 12. Estimation results: strategic decisions model (DC control).

| *Strategic decision d:* | *Tot. SF* | | | *Mean SF* | | | *SD, SF* | | |
|---|---|---|---|---|---|---|---|---|---|
| *Variable (x, z*)* | *Est.* | *(s.e.)* | *t-stat* | *Est.* | *(s.e.)* | *t-stat* | *Est.* | *(s.e.)* | *t-stat* |
| Intercept | -4.03 | (0.97) | *-4.14* | -0.54 | (0.72) | *-0.75* | -1.51 | (0.83) | *-1.82* |
| Food/Petro./Gases | -0.15 | (0.28) | *-0.55* | -0.64 | (0.21) | *-3.07* | -0.43 | (0.24) | *-1.80* |
| Retail/Wholesale | 0.49 | (0.21) | *2.32* | 0.07 | (0.16) | *0.46* | -0.03 | (0.18) | *-0.18* |
| Revenue | 5.08 | (1.01) | *5.04* | 1.85 | (0.75) | *2.48* | 2.46 | (0.86) | *2.86* |
| Customer service (CS) focus | -0.12 | (0.11) | *-1.14* | -0.17 | (0.08) | *-2.10* | -0.17 | (0.09) | *-1.84* |
| Logistical sophistication (LS) focus | -0.38 | (0.12) | *-3.14* | -0.26 | (0.09) | *-2.96* | -0.27 | (0.10) | *-2.63* |
| *Strategic decision d:* | *Density* | | | *Mi. to Rail (Reverse)* | | | *Mi. to Port (Reverse)* | | |
| *Variable (x, z*)* | *Est.* | *(s.e.)* | *t-stat* | *Est.* | *(s.e.)* | *t-stat* | *Est.* | *(s.e.)* | *t-stat* |
| Intercept | -1.61 | (0.68) | *-2.36* | 0.07 | (0.92) | *0.08* | -0.25 | (0.83) | *-0.30* |
| Food/Petro./Gases | -0.54 | (0.20) | *-2.75* | -0.44 | (0.26) | *-1.68* | -0.71 | (0.24) | *-2.99* |
| Retail/Wholesale | 0.03 | (0.15) | *0.22* | -0.09 | (0.2) | *-0.47* | -0.30 | (0.18) | *-1.67* |
| Revenue | 2.51 | (0.71) | *3.56* | 1.87 | (0.96) | *1.95* | 1.88 | (0.86) | *2.18* |
| Customer service (CS) focus | -0.15 | (0.08) | *-1.92* | -0.08 | (0.1) | *-0.78* | -0.10 | (0.09) | *-1.13* |
| Logistical sophistication (LS) focus | -0.24 | (0.08) | *-2.78* | -0.30 | (0.11) | *-2.70* | -0.25 | (0.10) | *-2.55* |

*Dependent variable:* $y_d^*$

The strategic model for density decisions suggests that companies in manufacturing, retail and wholesale sectors tend to have more DC space concentrated in more populated areas in comparison to food/beverage and petroleum/gases companies. This result is intuitive for petroleum and gases companies, but somewhat less intuitive for food and beverage companies. (Variations of the model with separate dummy indicators for petroleum/gases and food/beverage show that the result of this model is driven by the former, and that food and beverage companies actually tend to have DCs in more populated areas.) The transportation accessibility models suggest that relative to companies in other industries, manufacturers are most likely to locate their DCs near either intermodal truck-rail or water port facilities. This may be related to the global nature of supply chains among manufacturers in particular, relative to, for example, food and beverage companies, which are more likely to produce and sell products on the same continent.

Higher-revenue companies tend to have more DC space per unit population than lower-revenue companies. One possible explanation for this is that companies with higher revenue may generally be close to the final supply chain stage, consumer sales, and may therefore concentrate their distribution space near consumers. Companies with higher revenue are also located closer to rail and water ports than companies with lower revenue, suggesting that high-revenue companies in general have global supply chains that benefit from good access to rail and water, which provide good value for transporting goods over long distances. Finally, lower values of CS and LS translate into increased preference for dense DC operations and close proximity to rail and port facilities.

The results from Table 11 and Table 12 also give insight into how to identify strategy based on the CS and LS scores. As mentioned earlier, since the CS and LS scores are based attitudinal measurements for which the meaning of directionality is not known a priori, additional analyses (such as structural equation modeling or a model like the current one) must be used to decipher the meaning of positive and negative directions for each:

- The private fleet results suggest that a positive CS score means that a company has adopted a CS focus.
- The DC control results suggest that a negative LS score means that a company has adopted a LS focus.

Table 13 shows the estimated variance-covariance matrix for the strategic decisions model. Because of the variable transformations (Chapter 5.2.4.2), the variance of each falls in between 0.71 and 2.41. In general, the DC control decisions are highly correlated with each other but less so with private fleet decisions.

**Table 13. Estimated variance-covariance matrix ($\widehat{\Sigma}$): strategic decisions model.**

| | HDT | MDT | Total SF | Mean SF | SD SF | Density | Mi. to Rail (R.) | Mi. to Port (R.) |
|---|---|---|---|---|---|---|---|---|
| HDT | 0.71 | | | | | | | |
| MDT | 0.87 | 2.41 | | | | | | |
| Total SF | 0.44 | 0.45 | 1.80 | | | | | |
| Mean SF | 0.24 | 0.17 | 1.13 | 1.02 | | | | |
| SD SF | 0.31 | 0.31 | 1.34 | 1.01 | 1.33 | | | |
| Density | 0.22 | 0.16 | 1.03 | 0.84 | 0.89 | 0.89 | | |
| Mi. to Rail (R.) | 0.31 | 0.35 | 1.05 | 0.86 | 0.89 | 0.73 | 1.65 | |
| Mi. to Port (R.) | 0.25 | 0.22 | 0.98 | 0.85 | 0.88 | 0.69 | 1.08 | 1.33 |

*(R.) denotes 'Reverse'*

### *5.4.2 Application results*

The model results from Table 9 are applied to generate estimated values for the CS and LS strategy variables (Figure 44). The main purpose of this figure is to illustrate how strategy is quantified. The resulting values subsequently are input to the strategic decision models. For each strategy, the values mostly range from about -2 to 2, with slightly more skew in the LS model.

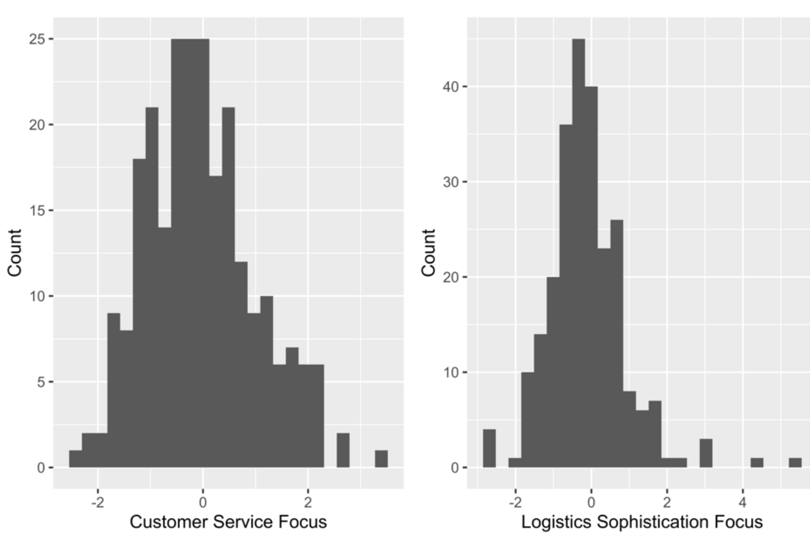



**Figure 44. Frequency distribution of company strategy scores.**

Figure 45 provides additional insight into the relationship between strategy and strategic decisions using pair plots with a best fit line. The plot on the left, the lower left pane in particular, shows the small but positive relationship between the CS strategy score and the latent decision variable for HDT ownership. The lower left pane of the plot on the right shows the greater relationship between LS and the Total SF decision.

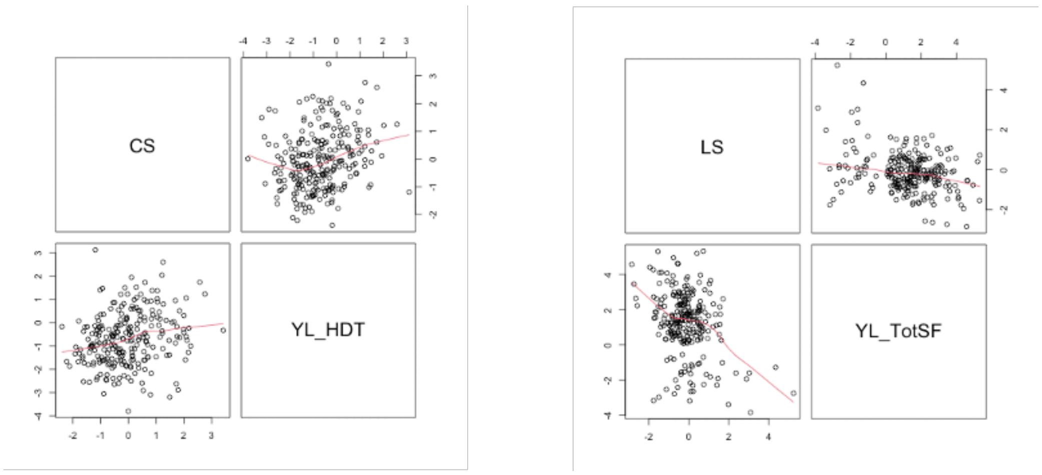



**Figure 45. Pair plots: strategies and latent strategic decisions.**

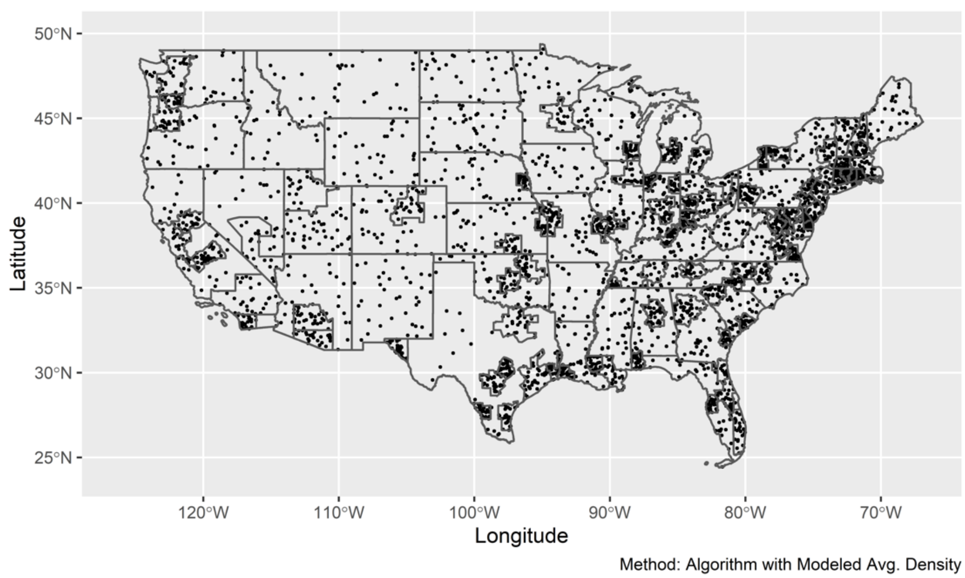
The last several plots demonstrate the practical significance of this model by simulating the national DC system, in addition to predicting fleet ownership. The strategic model outcomes, along with the Nationwide Zone Assignment Algorithm, are used in this demonstration. The models are applied to the input dataset. In follow-up work, the models should be applied to a set-aside sample for validation, but this exercise uses the entire sample for both model development and application so that a large enough sample is available for both.

Table 14 summarizes the characteristics of the total private fleet and DC populations that are simulated for the Fortune 500, freight-intensive companies in this modeling system. In the simulation, these companies collectively have a total of about 75,000 HDT, 192,000 MDT, and 1.4 billion SF of DC space. The mean DC size among these companies is predicted to be about 470,000 SF. These results are compared to their observed counterparts in the section discussing validation (Chapter 5.4.3) below.

**Table 14. Simulated and observed values of total fleet and DC population.**

| Outcome | Observed | Simulated | Difference | % Difference |
|---|---|---|---|---|
| HDT | 63,099 | 74,812 | 11,713 | 19% |
| MDT | 143,829 | 192,239 | 48,410 | 34% |
| Sum, Total SF | 1,437,657,569 | 1,407,933,827 | -29,723,742 | -2% |
| Mean SF | 385,991 | 470,724 | 84,733 | 22% |

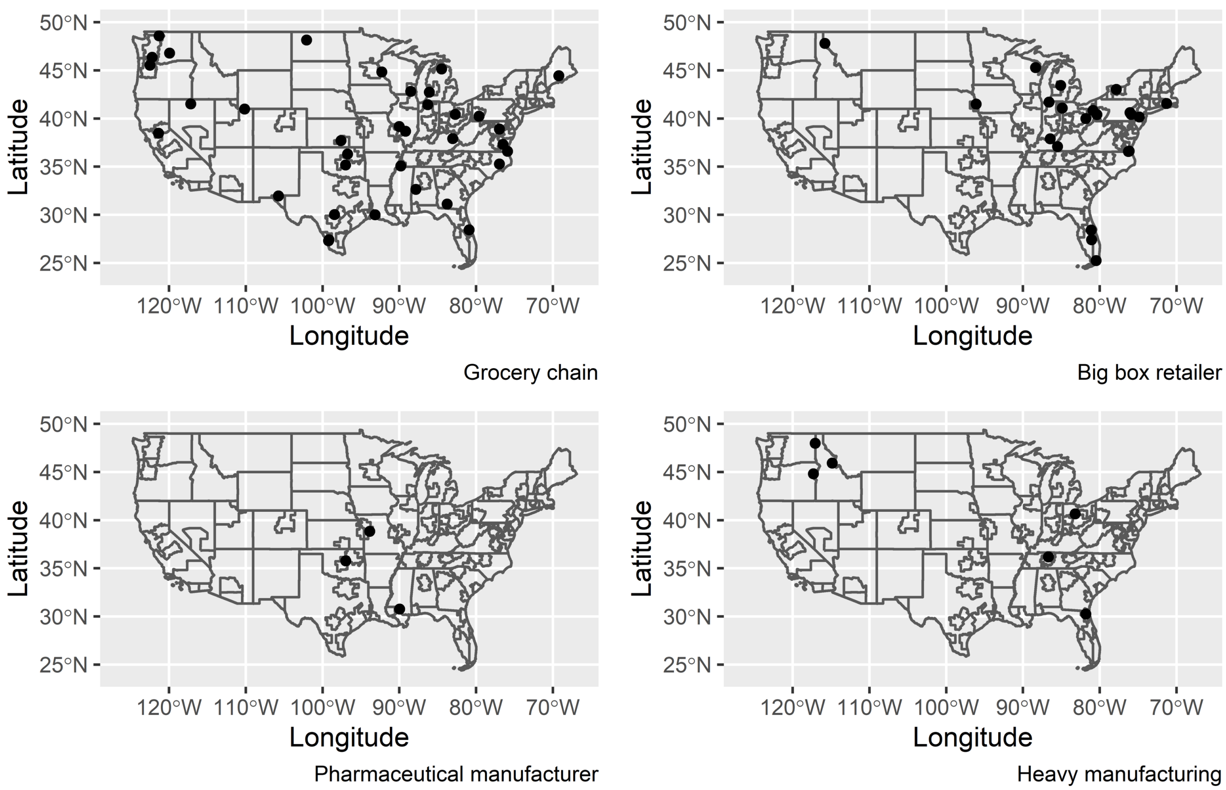
Figure 46 shows the geographic distribution of DCs across zones throughout the US. Comparing this map to Figure 41 shows that the model, and subsequently the algorithm used to assign DCs to locations, performs reasonably well by concentrating DCs in major metropolitan regions. The densest concentrations of DCs in both figures are along the east coast. However, the model and zone assignment algorithm predict too many DCs in rural areas (e.g., in the Upper Midwest). This can be improved by introducing new input variables to the model, such as sales data or new geospatial data, or by improving the algorithm.




*Locations within zones are randomized.*

**Figure 46. Simulation of nationwide DC locations.**

Figure 47 provides an example of simulated DCs that are generated for companies in four different industries. The grocer and big box retailer have more DCs and they are located almost entirely in metropolitan areas. The pharmaceutical manufacturer and the heavy manufacturer have fewer DCs, and their DCs are located in more rural areas. In general, the algorithm gives reasonably good geographic distribution throughout the US. While improving the geographic aspects of the algorithm is expected to be a major undertaking, the initial algorithm for now appears to deliver reasonable results that are useable in the agent-based model context.




*Locations within zones are randomized.*

**Figure 47. Simulated nationwide DC systems: companies in four sectors.**

Finally, Figure 48 illustrates how the model results are used to inform choice set generation. Example output for one company in the simulation is generated in Step 1. In Step 2, the strategic decisions for the company inform a series of choice set parameters: range of acceptable DC sizes, range of acceptable miles to rail facilities, and so on. In Step 3, the choice set is generated by identifying properties in the region that meet the criteria from Step 2. Following choice set generation, the candidate choice set is fed into a choice model by which a company selects a single property.

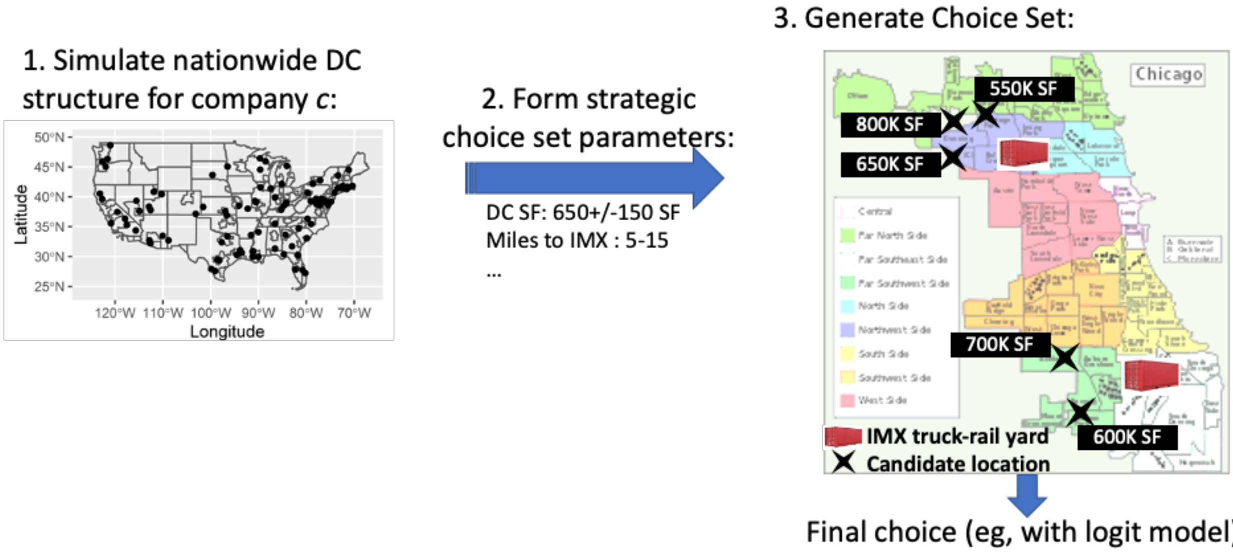



**Figure 48. Forming a consideration set for choice of specific DC location in the Chicago region (illustration).**

### *5.4.3 Validation*

This section establishes the validity of the model. Models results are summarized and compared to the observed data. The models are proven to produce reasonable and useful results.

Figure 49 summarizes the binary (yes/no) strategic decision outcomes for the observed and predicted number of companies in each of four categories: has neither fleet nor DC control, has fleet only, has DC control only, and has both. The number of companies in each category is printed on the figure. The results show that the model predicts the overall number in each category well, although with slight overprediction in the DC only category and fleet only categories and slight underprediction in the other categories.

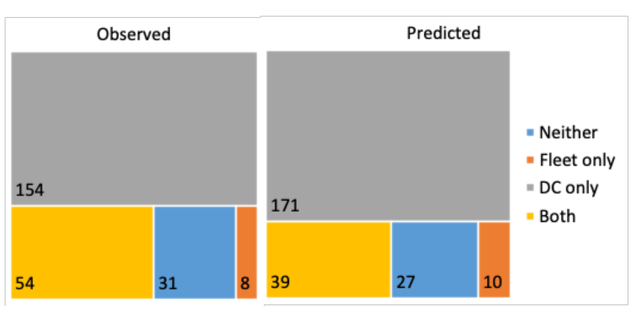



**Figure 49. Validation: logistics controls.**

Table 14 summarizes the observed and predicted outcomes for the main continuous strategic decision variables across all companies, including total numbers of HDT and MDT trucks, total DC area, and mean DC area. The model overpredicts both the HDT and MDT populations. The overprediction is slightly greater for MDT (34 vs. 19 percent) and should be examined further in future work to better calibrate the model before application. The total square footage, which is summed across all companies, matches the observed value (1.4 billion square feet) almost exactly. The mean DC size is overpredicted by 22 percent, which is not considered severe but warrants further calibration prior to application in a real-world setting. The total area divided by the mean implies that the resulting number of DCs is approximately 3,000, which is similar to the input total.

Figure 50 shows smoothed, normalized frequency distributions for the observed and simulated values of MDT and HDT. The HDT distribution matches very well. The MDT distribution does not fit as well as the HDT, but still fits the overall pattern reasonably well.

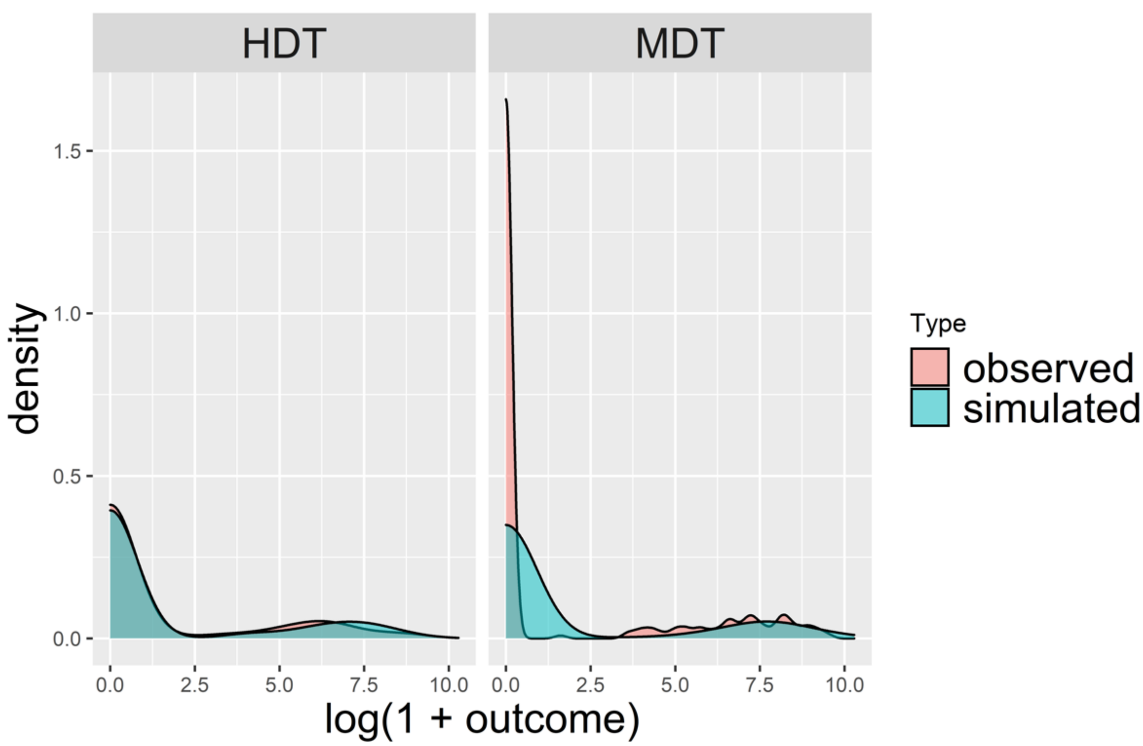



**Figure 50. Observed vs. Predicted outcome densities with smoothing (private fleet).**

Figure 51 shows similar validation curves for the DC control strategic decisions. The total square footage has excellent fit, meaning that the model produces a very realistic distribution of total DC square footage across companies. Distributions for the mean SF, SD SF and density are also matched well. Distributions for the two accessibility outcomes were not matched quite as well by the model, but the overall shape is visible. This may suggest some underlying variations that are not being captured, such as location selection that is based less on proximity to intermodal transportation and more on industrial land use availability, which often happens to be close to intermodal facilities (due to historical land use development).

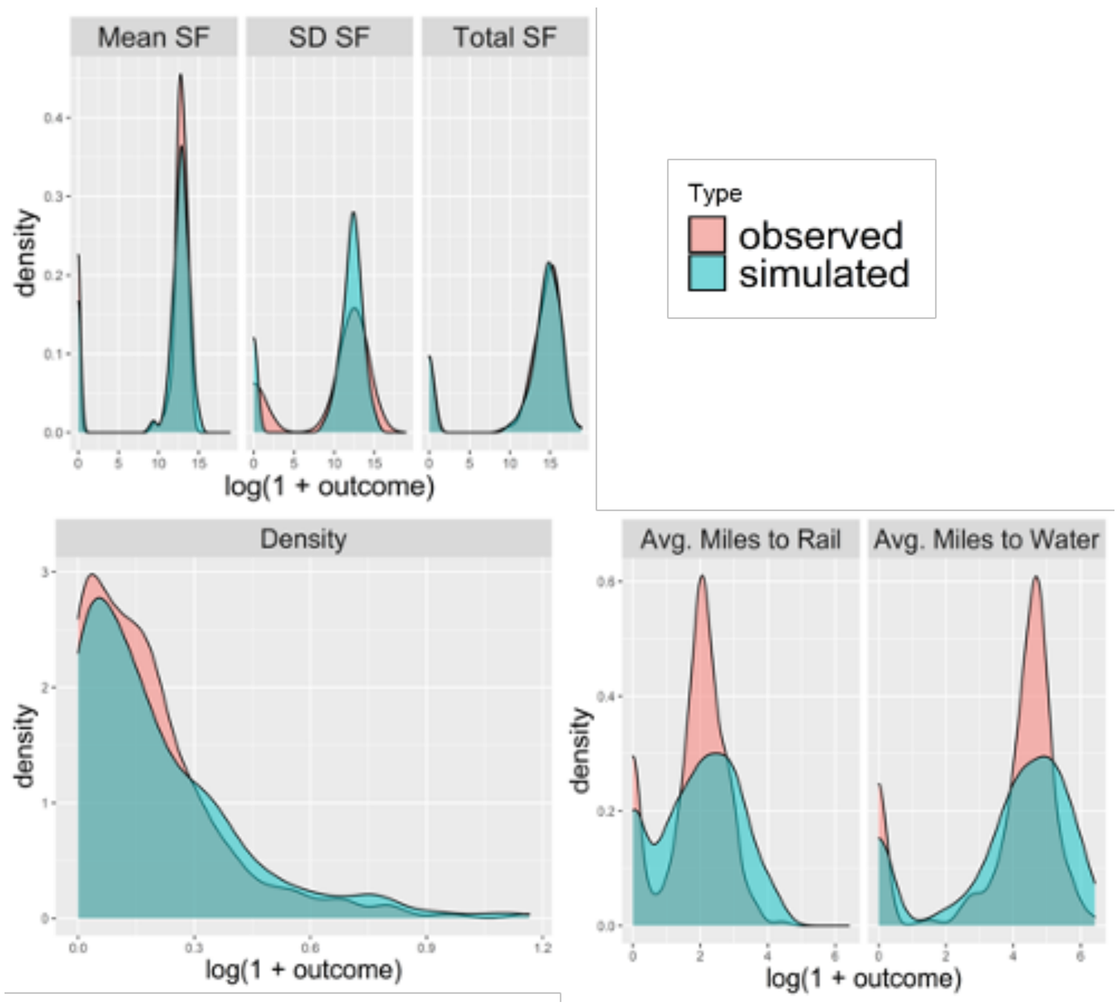



**Figure 51. Observed vs. Predicted outcome densities with smoothing (DC control).**

Figure 52 and Figure 53 show the performance of the model at the 10,000th iteration, focusing on the distribution of the strategic decision variables. Figure 52 portrays only the latent variables while Figure 53 portrays the latent variables for $y_d^* < 0$ and the observed values for $y_d^* > 0$. Ideally, the curves will be identical between the two plots, meaning that the latent predictions match the observed distributions. Overall, the curves match very well.

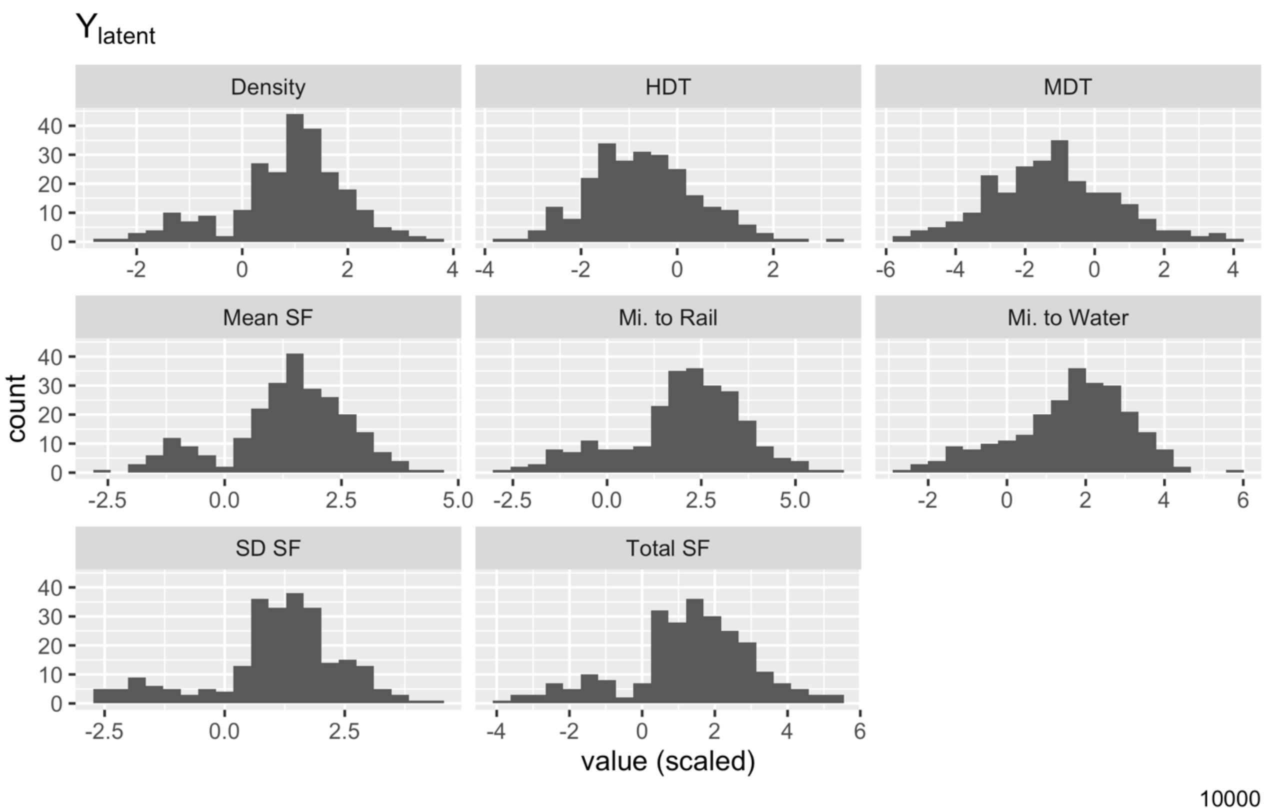



**Figure 52. Distribution of joint latent outcomes during estimation.**

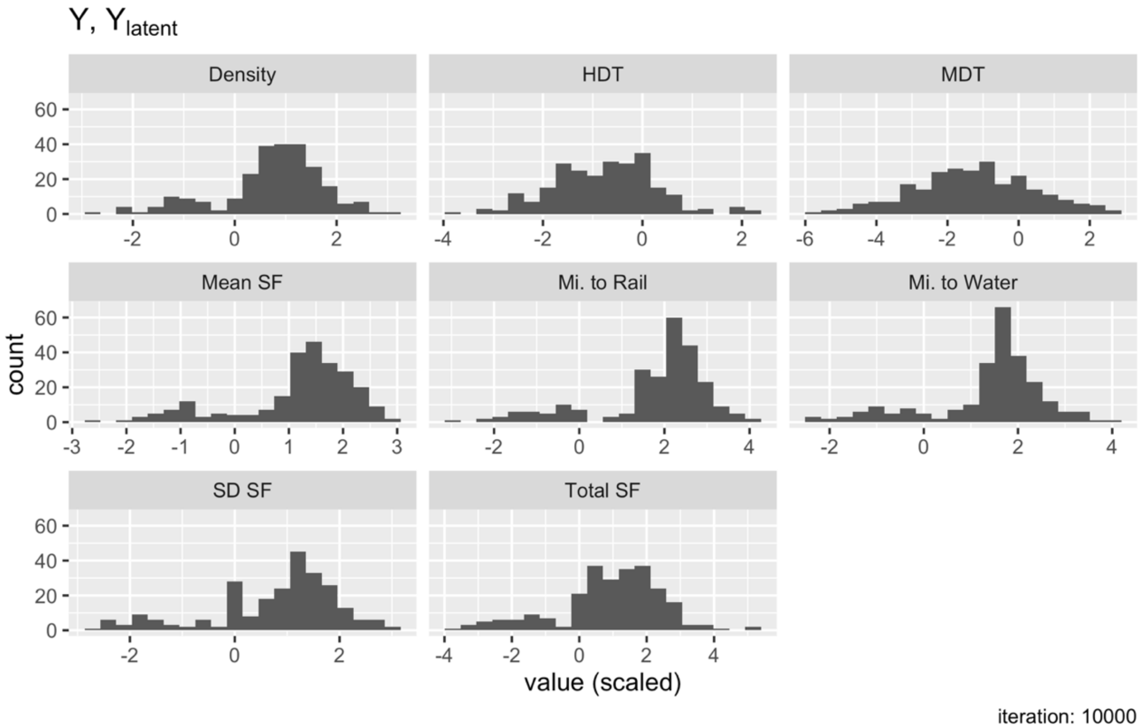



**Figure 53. Distribution of latent ($y_d = 0$) and observed ($y_d > 0$) joint outcomes during estimation.**

## 5.5 Conclusions

The contributions of this chapter are numerous and span multiple areas. First, the study develops a behavioral modeling framework to model strategies and strategic decisions of individual agents in a population. The study proposes a critical modification to the agent-based modeling paradigm, outlining and demonstrating how the strategies framework unifies downstream model decisions with more fundamental, major decisions and underlying attitudes of agents upstream. The study proposes and discusses how the modification applies to both passenger and freight agent-based modeling systems.

The methodology enables the joint simulation of three major types of strategic decisions. First, binary (yes/no) decisions, such as whether to own vehicles or not, can be modeled. Second, continuous quantities (such as vehicle miles traveled) can be modeled. Finally, this work proposes to model choice set generation parameters as a high-level, strategic decision jointly with these other types of strategic decisions, contributing a very unique and powerful approach to this challenging research area.

In order to operationalize the conceptual framework, the chapter develops a mathematical methodology to jointly model latent strategies and manifest strategic decisions. The methods utilize latent variable constructs for both strategies and strategic decisions, bundling the strategic decisions into a Seemingly Unrelated Regression (SUR). The entire system is solved using Bayesian estimation, in particular the Markov Chain Monte Carlo (MCMC) Gibbs sampling method, which utilizes data augmentation and neatly obviates the need for either multidimensional integration or approximations thereof.

Finally, the entire modeling system is applied to a real-world problem in agent-based freight modeling application to study the joint fleet ownership and distribution center decisions of large, freight-intensive companies in the US. For each firm, the number of heavy-duty and medium-duty trucks is modeled along with the number and characteristics, including geography, of its distribution centers. The interrelationship of strategic decisions is modeled explicitly in this application. Two strategic constructs, one emphasizing Customer Service (CS) and the other Logistical Sophistication (LS), are hypothesized and successfully modeled. The study finds that these two strategies have measurable impacts on fleet and distribution center decisions, although other factors (e.g., industry) have a greater impact overall. This application also successfully demonstrates the first real-world, behavioral modeling application of attitudinal measurements that are estimated using a novel natural language processing-based procedure that I develop earlier.

# 6 Linking Business Strategy to Transportation Energy Consumption and Emissions (TECE) Estimates: Effects of Automobile Assembly Location Strategy

## 6.1 Introduction

The premise of this thesis is that strategic behavior is important to capture in freight models, and that agent-based analysis presents a rich platform for modeling agent strategies and decisions. This chapter develops an agent-based sustainability analysis framework. The framework is applied in a proof of concept that clearly demonstrates the effects of company strategy, and changes thereof, on TECE. The supply chain decision involving automobile assembly location is studied for the Original Equipment Manufacturer (OEM) sector of the automotive industry. This study focuses on passenger autos sold in the US, which has one of the world's largest auto markets. The sustainability impacts of two strategies, near-sourcing assembly versus overseas assembly, of the Ford Focus passenger car are analyzed. Consequences are expected to be notable, since global shipping is known to be energy-intensive – e.g., the emissions generated by ocean shipping as reported in International Energy Agency (2017) reports that is about as much as what is generated by a industrialized country (Union of Concerned Scientists, 2017).

The remainder of this chapter is organized as follows. An agent-based framework for modeling freight TECE is presented first. The next section describes the study design, including data, methods and assumptions. Findings of the baseline analysis are presented, then the method is illustrated using a case study. Finally, key points of the study are summarized followed by a discussion of limitations and potential extensions.

## 6.2 Agent-based Modeling Framework for Freight TECE

An agent-based model is designed specifically for this case study, not unlike many models that government agencies use (Cambridge Systematics, Inc., 2011; Maricopa Association of Governments, 2018; Samimi, 2013). An agent-based modeling framework for freight TECE comprises the following steps:

Step 1. Define geographic regions for analysis

Step 2. Develop agent population

Step 3. Define agent strategies

Step 4. Form agent partnerships

Step 5. Evaluate mode and logistics choice

Step 6. Determine the flow of goods and vehicles

Step 7. Estimate TECE for inter-regional flows by mode

Step 1 is necessary for implementation of transportation demand models. Traffic analysis zones or parcels constitute the geographic unit of analysis in many models. States, regions, countries or continents are often used for national and international models.

Step 2 involves developing the population of decision-making agents. Agent attributes can size, industry, location and other factors.

Step 3 is integral to demonstrating the link between strategy and TECE outcomes. As an alternative to the comprehensive, agent-based freight framework is developed in Chapter 3, the flowchart below (Figure 54) illustrates a generic model stream that can be implemented quickly. The figure gives a sense of how the effects of strategy cascade downstream to TECE assessment.

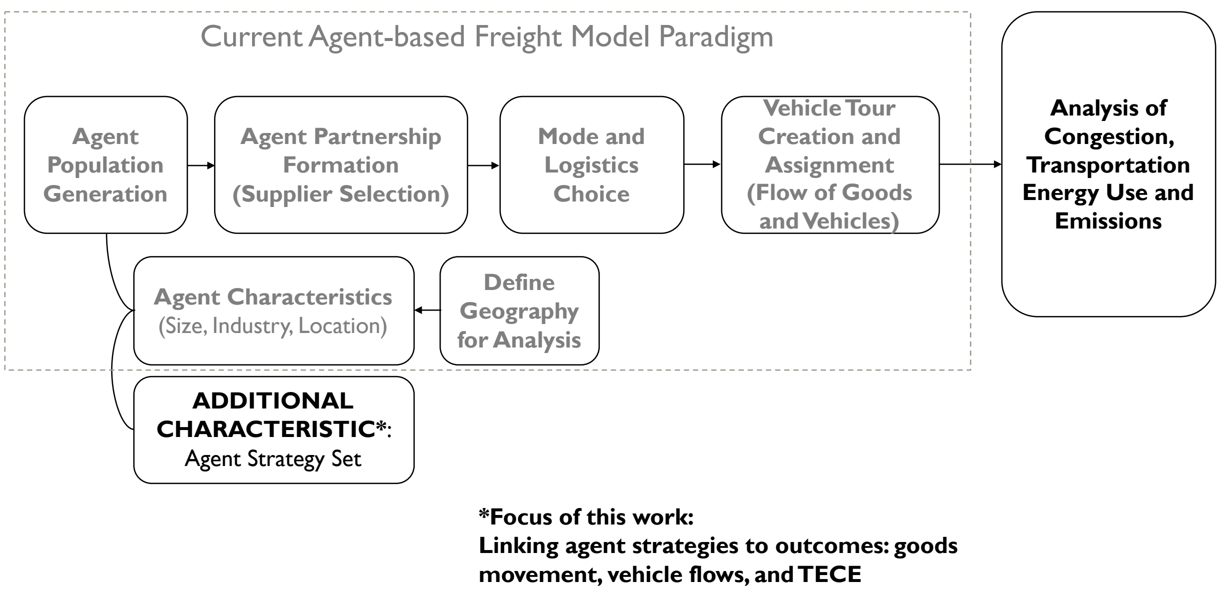



**Figure 54. Linking agent strategies to outcomes.**

Steps 4-6 are widely implemented in the agent-based freight demand modeling literature. Resulting flows of goods are outputs of the model.

This work also makes significant contributions to Step 7, the evaluation of energy and emissions impacts of freight transportation. A set of TECE rates by mode is assembled and a method to calculate TECE impacts is developed. A real-world application illustrates how TECE is estimated based on inputs from the previous steps.

As shown in the figure, this study contributes to this modeling framework by developing a method to link agent strategies to important model results such as traffic flows and TECE. Constrained by data availability, Steps 4-6 involve necessary assumptions to simplify the problem compared to the others such as Agent Population and Analysis of Congestion.

## 6.3 Study Design and Data

Using the automobile manufacturing sector as an example, this section describes step-by-step how the agent-based model is set up and applied, as well as assumptions made in the analysis and data. The analysis timeframe is the year 2017.

*Step 1. Define analysis regions*

The US was divided into seven regions for analysis (Figure 55). The rest of the world was divided into seven additional regions: Canada (Ontario), Mexico, South America, Europe, Africa, East Asia, Other Asia and Australia. Two US regions are large states (California and Texas). Others were determined based on population concentration and locations of existing automobile assembly plants, which are summarized next.

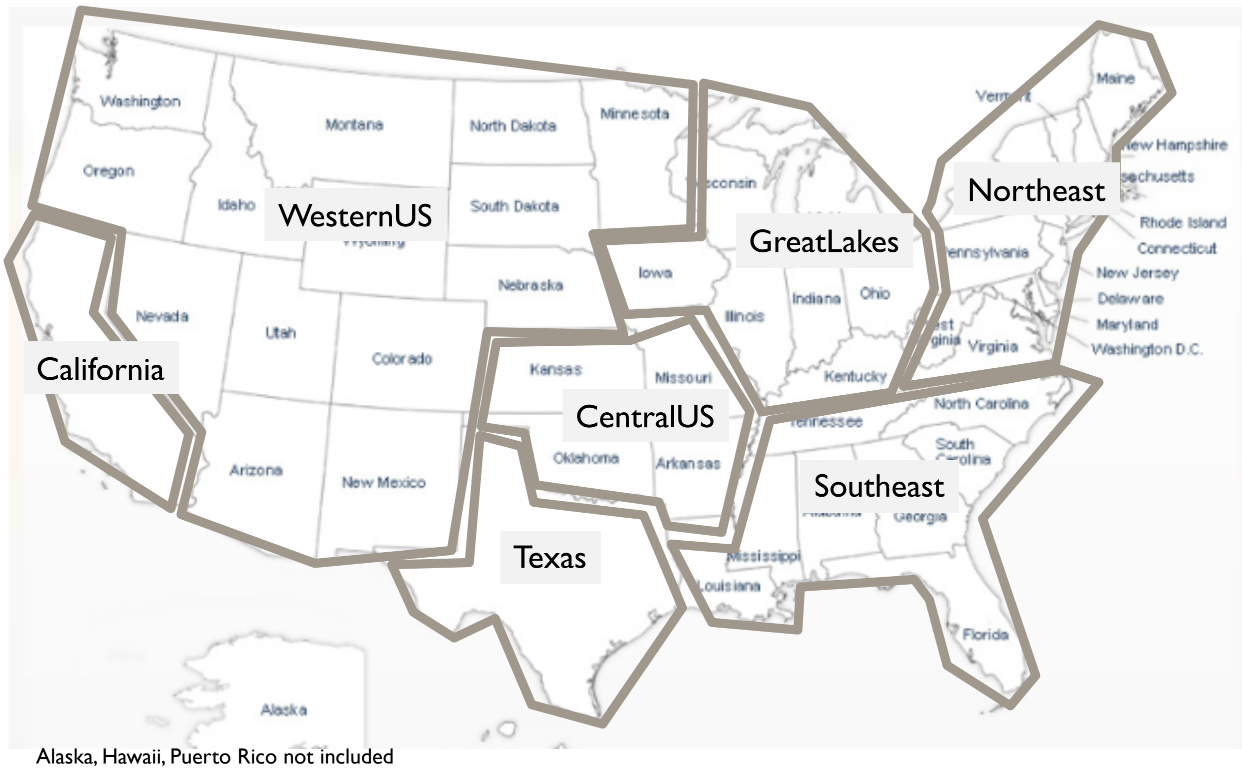



**Figure 55. US analysis regions.**

*OEM data: US sales and assembly locations by region*

Data on US sales and the number of assembly plants by OEM in each region were obtained from online sources. The findings are shown in Table 15. Several OEMs have plants in Australia, Africa and Other Asia, but these are not shown in the table due to space limitations and their negligible contributions to US auto sales. Abbreviations are: Cen US (Central US), SE US (Southeast US), EU (Europe), MX (Mexico) and E. Asia (East Asia).

## Table 15. Assembly plants and sales by OEM and region

| Company | Number of Assembly Plants | | | | | | | | Assembly Plant Sources | US Sales* | % US Sales |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Cen. US | E. Asia | EU | Great Lakes | MX | On-tario | SE US | Total | | | |
| Aston Martin | | | 1 | | | | | 1 | *(Aston Martin, 2018;* Tsui, 2018) | 1,869 | 0.01% |
| BMW | | 5 | 6 | | | | 1 | 16 | (Mitchell, 2017; BMW Group, 2018) | 354,110 | 2.1% |
| Daimler AG | | 4 | 10 | | 2 | | 5 | 27 | *(Daimler AG, 2018)* | 375,311 | 2.2% |
| FCA | | | | 5 | 3 | 2 | | 11 | (Fiat Chrysler Automobiles, 2018; *List of Chrysler factories, 2018)* | 2,073,073 | 12.0% |
| Ferrari | | | 1 | | | | | 1 | *(Ferrari, 2018)* | 2,518 | 0.01% |
| Ford | 1 | 6 | 6 | 10 | 3 | 1 | | 37 | (Ford Motor Company, 2018, BlueOvalNews.Com and AtomicFrog.Com, 1998) | 2,575,200 | 14.9% |
| GM | 2 | 12 | 7 | 8 | 2 | 2 | 1 | 40 | (GM Corporate Newsroom, 2018; GMAuthority.co, 2018; *List of General Motors factories, 2018)* | 3,002,241 | 17.4% |
| Honda | | 11 | 2 | 3 | 2 | 1 | 1 | 27 | (Honda Government Relations Office, 2018; *List of Honda Assembly Plants, 2018)* | 1,641,429 | 9.5% |
| Hyundai | | 6 | 1 | | | | 1 | 12 | (Hyundai Motor America, 2018) | 685,555 | 4.0% |
| Kia | | 7 | 1 | | 1 | | 1 | 11 | *(List of Kia Design and Manufacturing Facilities, 2018)* | 589,668 | 3.4% |
| Mazda | | 7 | | | 1 | | | 11 | *(Mazda, 2018)* | 289,470 | 1.7% |
| Mitsubishi | | 6 | | | | | | 7 | (Mitsubishi Motors Corporation Public Relations Department, 2018) | 103,686 | 0.6% |
| Nissan | | 16 | 4 | | 3 | | 2 | 32 | (NissanNews.com, 2018; Nissan Corporate Communications, 2018) | 1,593,464 | 9.2% |
| Subaru | | 4 | | 1 | | | | 5 | (Subaru, 2018; Turner, 2015) | 647,956 | 3.8% |
| Tata | | 1 | 3 | | | | | 17 | (Tata Motors, 2017) | 114,333 | 0.7% |
| Tesla | | | | | | | | 1 | *(Tesla Factory, 2018)* | 55,120 | 0.3% |
| Toyota | | 13 | 7 | 2 | 1 | 2 | 1 | 37 | *(Toyota-Global Newsroom, 2018; List of Toyota manufacturing facilities, 2018)* | 2,434,515 | 14.1% |
| Volkswagen | | 4 | 28 | | 1 | | 1 | 45 | *(List of Volkswagen Group factories, 2018)* | 625,068 | 3.6% |
| Volvo | | 3 | 2 | | | | | 5 | *(List of Volvo Car production plants, 2018)* | 81,507 | 0.5% |
| Total | | | | | | | | 343 | | 17,246,093 | 100% |

**US Car Sales Data (2018) and Wall Street Journal Market Data Center (2018)*

*Population data*

Population totals were computed for each region using (US Census Bureau, 2018) with year 2017 data. These data were used to apportion total US sales volumes. Population totals for the various regions are (percentages are with respect to the contiguous US total):

- Northeast: 75 million (23%);
- Southeast: 66 million (20%);
- Great Lakes: 54 million (17%);
- Western US: 45 million (14%);
- California: 40 million (12%);
- Texas: 28 million (9%); and
- Central US: 16 million (5%).

*Step 2. Develop agent population*

In order to estimate freight activity at the agent level, it is first necessary to establish the population of agents that are making decisions. For this study, the population of agents is the set of OEMs that sell vehicles in the US. Since this set is small with fewer than two dozen agents, it was possible to use observed agents and their characteristics for the study rather than a synthetic population as is routinely done with travel demand models. A web search was used to collect data (Table 1), which then were used to develop the population of agents and their characteristics. The output of this stage is the set of OEMs by number and geographic location of assembly plants, size as measured by US sales, and industry. All agents in this work are in the same industry class.

Assembly location strategies, which affects import/export volumes, are another important OEM characteristic. This is discussed next.

*Step 3. Define agent strategies*

Several OEM strategies were considered for analysis in this study: assembly location decision, mode of transport used, and port used for imports. Assembly plant location was selected as the strategy to focus on in developing the methodology. While this decision is based partly on quantitative optimization, it also is driven by strategies weighing the anticipated direction of market demand and other factors (Coia and Ludwig, 2016).

This strategy has major ramifications for the flow of goods and subsequently TECE. This strategy is initially treated as a fixed input but is varied later to demonstrate the impact of a change in strategy on TECE. Mode and port strategies can be examined in subsequent work.

The rest of this subsection discusses the data and assumptions supporting this aspect of the study. Only automobiles being sold in the US are included in this study.

Detailed data regarding the number of assembly plants used by each OEM in each region was gathered. However, complete information on plant-specific production volumes for US-sold vehicles, which are partly based on strategy, was difficult to obtain. To address this constraint, for this study assumptions were made regarding the volumes produced by location for sales in the US. Three assumptions were made. First, for vehicles assembled in the US, it is assumed that the percentage assembled by origin region is proportional to the number of plants in origin region. Second, OEMs with assembly plants both in the US and elsewhere were assumed to conduct the assembly mostly in the US. Third, only Canada, Mexico, East Asia and Europe were assumed to be assembly locations for imported vehicles. Reports on the auto industry support these assumptions (for example, Isidore, 2016). The percentages of US-sold vehicles produced by each OEM by region are assumed to be as followed (percentages in parentheses):

- Aston Martin: Europe (100);
- BMW: Southeast (100);
- Daimler AG: Southeast (100);
- FCA: GreatLakes (60), Mexico (10), Ontario (30);
- Ferrari: Europe (100);
- Ford: GreatLakes (60), CentralUS (5), Mexico (20), Ontario (25);

- GM: GreatLakes (50), Southeast (30), CentralUS (10), Mexico (7.5), Ontario (2.5);
- Honda: GreatLakes (50), Mexico (10), Southeast (10), E_Asia (10), Ontario (20);
- Hyundai: Southeast (40), E_Asia (60);
- Kia: Southeast (20), E_Asia (40), Mexico (40);
- Mazda: E_Asia (80), Mexico (20);
- Mitsubishi: E_Asia (100);
- Nissan: Southeast (80), Mexico (10), E_Asia (10);
- Subaru: GreatLakes (60), E_Asia (40);
- Tata: Europe (80), S_Am (20);
- Tesla: California (100);
- Toyota: Southeast (40), GreatLakes (40), Mexico (10), Ontario (10);
- Volkswagen: Southeast (40), Mexico (10), Europe (50); and
- Volvo: Europe (60), E_Asia (40).

*Step 4. Define agent partnerships*

This aspect of the analysis can be done at the agent level using auto dealerships or wholesalers, who receive shipments of assembled vehicles, but for this study was treated with less detail to focus on other developments. Total US sales by each OEM is apportioned to the various regions based on regional population. This is a reasonable assumption considering vehicle shipping costs paid by consumers: while actual shipping charges may vary by region, the consumer pays the same destination charge regardless following Palermo (2013). Average vehicle weight is assumed to be two tons (Lowrey, 2011).

*Step 5. Evaluate mode and logistics choice*

Like the previous step, this step can be treated with high resolution but here is modeled using a higher-level analysis to keep the focus on other aspects of the study. Modal information is based on Freight Analysis Framework (FAF) ton-mile estimates of rail, truck and multiple mode use by domestic and import SCTG 36 flows using distance bands (Freight Analysis Framework Data Tabulation Tool, 2018). Over-the-ground distances are based on regional centroids. According to the FAF estimates, 99% of SCTG 36 US flows by ton-mile are delivered by three modes: truck, rail and multiple modes (e.g., truck-trail intermodal). This data source was used to develop distance-based mode assumptions for this study as follows. Shipments are assumed to always use truck for journeys of 500 miles or less. Longer-distance flows are assumed to use a mix of truck and rail. Shipments traveling between 500 and 1,500 miles are assumed to use 80% truck and 20% rail, those between 1,500 and 2,000 miles are assumed to use 60% truck and 40% rail, and those traveling longer distances 55% rail and 45% truck. This approach also is consistent with other studies (for example, Sharma & Associates, Inc., 2018) demonstrating that rail starts to become cost-effective relative to truck for ground-based transport distances of 400 miles or more.

Rail is only available for North American-based assembly. All other foreign assembly locations are assumed to require ocean shipping for transport of vehicles to the US. Upon reaching the US, these foreign-assembled vehicles are assumed to use either all-truck or rail-with-truck per the assumptions described above.

Port-of-entry assumptions for imports used in this study are based on reports from various automotive industry sources (including Coia, 2015). Scanning these sources shows that a thorough accounting of ports used by each OEM can be developed. However, based on the resource constraints of this study, it was decided to focus the most detailed data collection on OEM-specific automobile assembly location and to use assumptions for port-of-entry.

Two major automobile import ports were selected for use in this study to represent the port-of-entry of imported vehicles. The primary objective was to represent whether import flows enter on an East Coast port or a West Coast port. Imported vehicles destined to the Northeast or Southeast are assumed to always enter through the Port of Baltimore, which is located in the Northeast. Vehicles destined to California or the Western US region are assumed to always enter through the Port of Los Angeles. Vehicles destined to other areas of the US are assumed to use whichever port (Baltimore or Los Angeles) generates lower distance. Vehicles from South America destined to the California or Western US, for example, are assumed to enter through the Port of Los Angeles while those going elsewhere in the US are assumed to enter through the Port of Baltimore. Of course, numerous other ports are used for vehicle imports (for example, the Port of Jacksonville or JAXPORT

(Jacksonville Port Authority, 2018). A more detailed representation of ports would have some impact on the results, but the assumptions used here are for demonstrating the methodology.

Distance between each Origin-Destination pair was obtained from two sources: Mapping-Tools.Com (2018) for over-ground routes and SeaDistances.org (2018) for sea routes.

*Step 6. Determine the flow of goods and vehicles*

The inputs from preceding steps were used to compute interzonal and intrazonal volumes by mode and logistics path for each OEM. Volumes were measured in number of vehicles, tons and ton-miles by mode for each OEM.

*Step 7. Estimate TECE for inter-regional flows by mode*

A detailed data assembly process was undertaken to establish the necessary emission and energy use rates for freight modes.

Ship energy use and emissions data were obtained from an international study in Olmer et al. (2017).

Rail TECE rates were obtained from two sources.  Locomotives information from the US Environmental Protection Agency (2019) was the source of rail-related emissions in this study. A large RRD, line-haul only, and year 2017 were assumed when retrieving the information. A fuel efficiency of 471 ton-miles per gallon is used based figures from CSX Corporation (2016), which illustrated its calculation of fuel consumption per revenue ton-mile in a regulatory reporting process.

Truck TECE rates were gathered from various sources. First, truck loaded weight is assumed to be about 72,000 pounds or 36 tons. After accounting for the weight of the semi-tractor and a typical trailer, which is about 15 tons total according to Truckers Report Forum (2006), this suggests that about 10 sedans or around seven light-duty trucks would be transported on each carrier, which is in line with a typical commercial carrier (‘Car carrier trailer’, 2018).

CO2 and NOX emissions rates for a 36-ton truck were obtained from federal testing of two heavy-duty trucks in Boriboonsomsin et al. (2017). Fuel efficiency was assumed to be 50 loaded US-ton miles per gallon of fuel based on the upper end of a 40-50 ton-mi/gal range cited in Souten (2004). PM rates for heavy trucks were obtained from ICF Consulting (2017) using MOBILE6.2, using 52 mph as the average urban highway speed.

The truck SO2 rate was obtained from a MOVES2014 (US Environmental Protection Agency, 2014) run with 1,000 heavy-duty intercity trucks traveling a one-mile section of an unrestricted highway in Cook County, IL in Oct. 2017 during the 10-11 AM period. The MOVES rate is 0.614293 g/mi of SO2 emitted, or 0.000614293 g/mi per truck. Assuming about 36 tons per truck leads to the rate in the table.

The full set of TECE rates used in analysis for each mode are shown in Table 16.

**Table 16. TECE rates used in analysis.**

| Measurement | Ocean | Rail | Truck |
|---|---|---|---|
| Fuel efficiency (loaded US ton-miles per gallon of fuel) | 1,179 | 471 | 50 |
| Energy consumption: gallons of fuel per US ton-mile | 0.00085 | 0.00212 | 0.02 |
| CO2 emissions (g/ton-mi) | 8.5 | 21.7 | 61.1 |
| NOX emissions (g/ton-mi) | 0.2085 | 0.2484 | 0.2328 |
| SO2 emissions (g/ton-mi) | | 0.004 | 0.00017 |
| SOX emissions (g/ton-mi) | 0.1218 | | |
| PM10 emissions (g/ton-mi) | 0.0172 | 0.0062 | 0.0047 |

## 6.4 Results

Using the TECE rates shown here and the flows computed in Step 6, TECE was evaluated. The TECE analysis focuses primarily on ton-miles traveled (or simply ton-miles), which can be considered the freight equivalent of vehicle-miles traveled since it is a weighted sum of miles traveled. TECE for ton-miles on each leg of the journeys was computed by multiplying TECE rates by the ton-miles. Values were summed across legs to obtain Total TECE.

### *6.4.1 Predicted annul vehicle production and sales*

This section presents the results and discusses the main findings. Figure 56 shows the predicted annual number of new vehicles produced in each region according to the model. Roughly 17,000,000 new passenger vehicles in total were sold in the US in 2017 (Table 15). Based on the data and assumptions, most of the vehicles come from the US with around 1.5-2 million vehicles each from Ontario (Canada), East Asia, and

Mexico, and a smaller number coming from Europe. For example, Volvo sells about 81,000 vehicles in the US each year (Table 15) with approximately 60% coming from Europe and the rest from East Asia. Multiplying 81,000 by 60% yields 48,600 Volvo vehicles originating from Europe to be sold in the US. This calculation was performed for all OEMs.

Similarly, the modeled number of vehicles sold in the US by OEM is pictured (Figure 57). In this work, the modeled number of vehicles by OEM was constrained to match the observed US sales volumes by OEM. Ferrari and Aston Martin sales are too low to observe on the chart, but their sales information is noted above (Table 15).

The modeled number of vehicles sold in the US by destination region is pictured (Figure 58). In this work, the total vehicles sold is constrained to match observed US sales data.

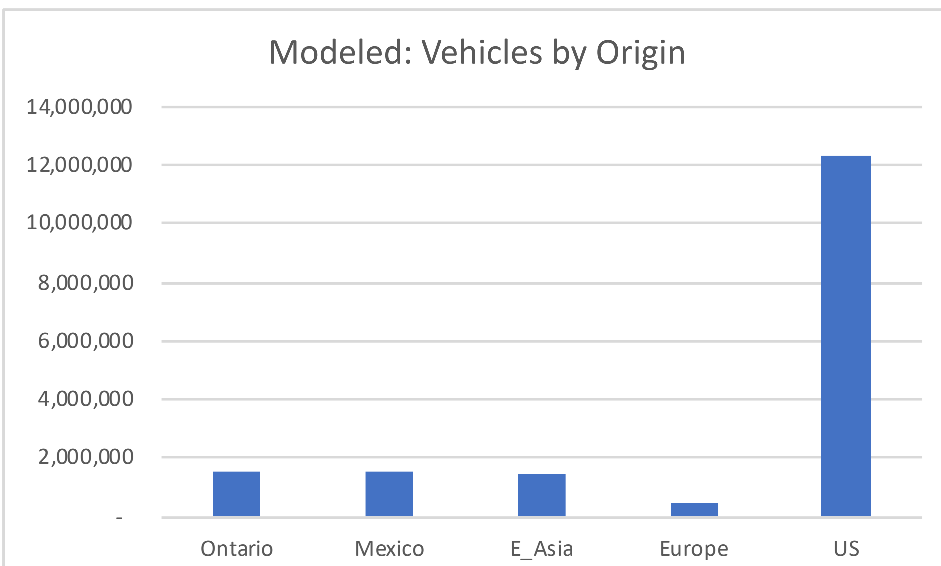



**Figure 56. Modeled vehicle origins by country/region.**

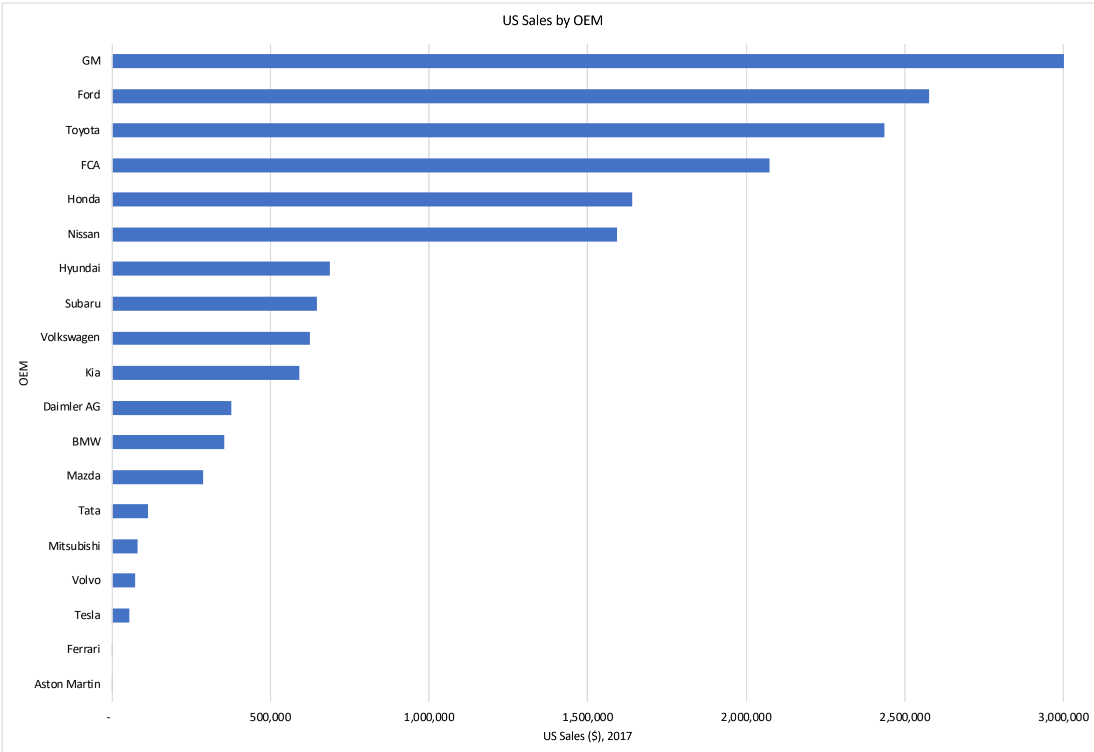



**Figure 57. Modeled vehicles sold in the US by OEM.**

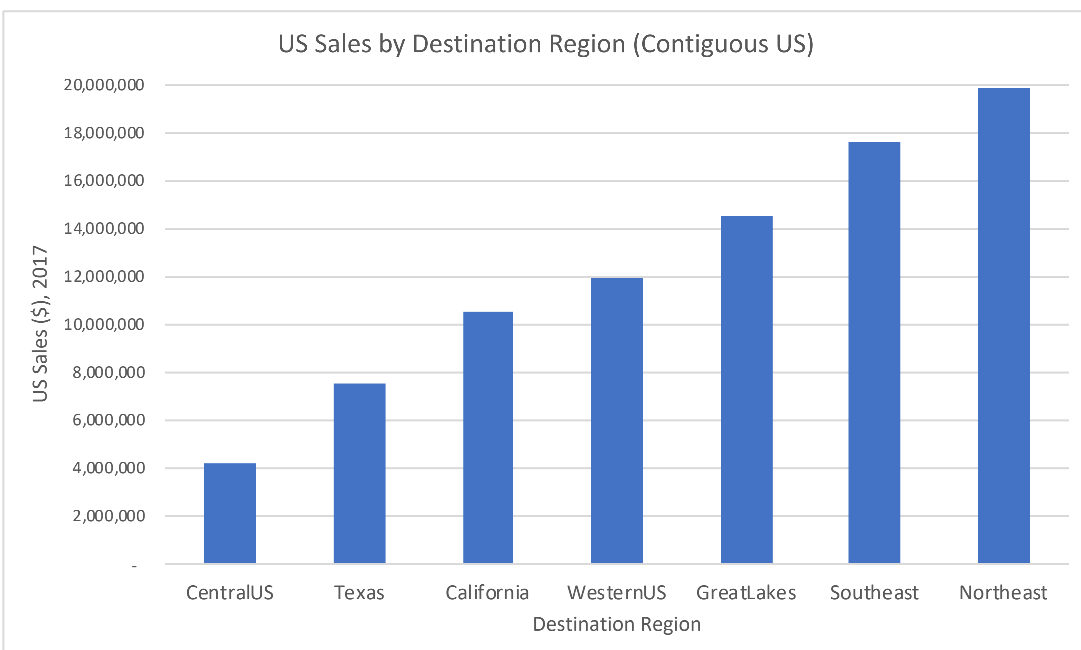



**Figure 58. Modeled vehicles sold in the US by destination region.**

Modeled volumes by foreign origin were compared with the 2012 FAF data (Oak Ridge National Laboratory, 2017), which is based on the most recent US commodity flow survey. However, the FAF data is quite aggregate. For example, in the FAF data, SCTG 36 (which is the finest level of commodity detail; see USDOT and US Department of Commerce, 2017) covers shipments of:

- Passenger vehicles;
- Large commercial vehicles such as tractors and semi-trailers; and
- Automobile parts.

To develop a meaningful comparison with the modeled passenger vehicle, the tonnage of passenger vehicles within SCTG 36 of FAF needed to be better identified. This was accomplished as follows. First, the 2012 Commodity Flow Survey (CFS) Microdata (US Census, 2015) sample was used to estimate the total tonnage of shipments sent by companies belonging to North American Industry Classification System (NAICS) category 336, which includes automobile manufacturers. Only shipments over 2,500 pounds (about the size of a compact car) that were valued at over $8,000 were included. Based on this, about 40% of SCTG 36 shipments are composed of new vehicles. This figure includes not only passenger vehicle shipments but also shipments of motor homes, semi-trailers, etc. For now, it is expected that the FAF-based estimate should be higher than the modeled estimate in order to account for this. Further, US automobile sales were in 2017 were about 20% higher than in 2012 (Statistica, 2018), which means the model estimates should be higher than the FAF estimates accordingly.

While this comparison is not perfect, it underscores the value of agent-based approaches for modeling transportation flows. When aggregate data such as regional commodity flows are used instead, development and application of crude factors is the only way to utilize the data for more detailed analysis.

Despite these issues, the adjusted FAF data provide a reasonable benchmark estimate for comparison with the modeled values. The resulting flows by origin region corresponding to Modeled and FAF, respectively, are:

- Ontario: 1,509,173 and 1,766,797;
- Mexico: 1,511,467 and 1,658,213;

- East Asia: 1,467,111 and 1,671,894;
- Europe: 457,292 and 536,979;
- Total: 4,945,044 and 5,633,883.

Despite the caveats noted above, this comparison shows that the modeled and FAF-based estimates are indeed in the same general range.

### *6.4.2 Predicted distribution of modes for automobile shipments*

Figure 59 shows the ton-miles of shipments that are estimated to use each mode. Ocean-based FAF ton-miles are not shown since the FAF data account for only the US-based portion of travel, meaning that the FAF data do not include the vast majority of ocean-based ton-miles accrued by European and East Asian imports.

Despite the shortcoming in comparison ocean-based travel, comparisons can be made for rail and truck travel. The comparisons indicate that proportion of truck and rail ton-mile estimates based on the model are similar in the model outputs and FAF, with a ratio of about two ton-miles shipped truck for every ton-mile on rail in each. Overall, modeled truck ton-miles are about 22% higher than FAF, which is expected due to growth in US sales between 2012 and 2017. Modeled rail ton-miles exceed the expected 20% (see earlier discussion regarding 2012 vs. 2017 sales), suggesting that a next step for the model approach would be to improve the modal estimation process.

Another point worth noting is that only 10% of shipments are estimated to use ocean modes. However, as shown above (Figure 59), this 10% translates into significant percentage of total ton-miles due to the long distances traveled across oceans. This highlights the importance of assembly plant location for its effects on the flow of goods.

### *6.4.3 Predicted TECE by mode*

TECE results from this analysis are shown below in Table 17. Most of the fuel consumption is incurred by truck, which exceeds rail fuel consumption by about a 17:1 ratio. This is much greater than the 2:1 ratio in ton-miles, highlighting the relative fuel efficiency of rail. Similarly, ocean and rail fuel consumption are similar despite the greater ton-miles shipped by ocean, which is due to the relative fuel efficiency of ocean modes compared to rail. by which the energy use is about evenly split. Truck shipping generates the greatest

CO2 emissions in this analysis, which is largely attributable to the reliance on truck for a majority of over-the-ground ton-miles. Ocean shipping produces a minority of the total CO2 emissions. Ocean and truck shipping produce most of the NOX. However, for PM10 and sulfur-related emissions, ocean shipping is the largest producer. Recent legislation should significantly reduce these emissions (DieselNet with M. Pedersen, 2018).

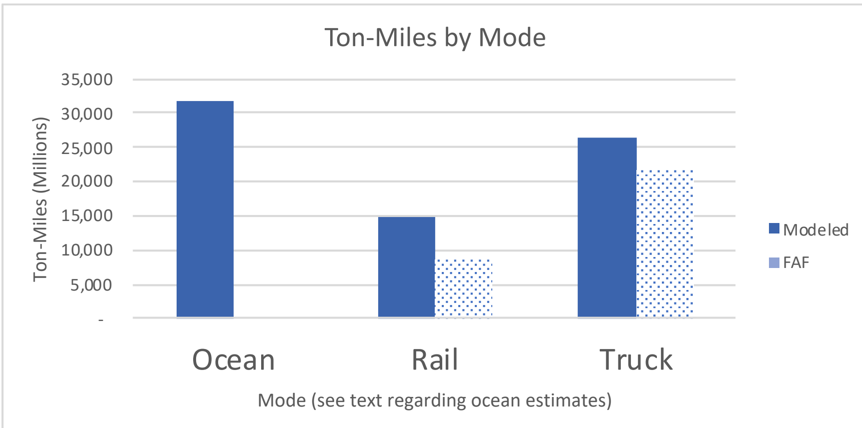



**Figure 59. Millions of ton-miles by mode accrued by automobile shipments.**

**Table 17. TECE results.**

| **Measurement** | **Ocean** | **Rail** | **Truck** | **Total, All US Sales** |
|---|---|---|---|---|
| Energy consumption: gallons of fuel (1,000,000s) | 27 | 31 | 527 | 585 |
| Emissions (1,000,000s of grams) | | | | |
| CO2 emissions | 270,701 | 320,625 | 1,610,561 | 2,201,887 |
| NOX emissions | 6,624 | 3,672 | 6,135 | 16,430 |
| SO2 emissions | - | 59.0 | 0.4 | 59.4 |
| SOX emissions | 3,869 | - | - | 3,869 |
| PM10 emissions | 547 | 91 | 124 | 762 |

## 6.5 Case Study: Change in Agent Strategy

A case study is presented to demonstrate the potential TECE impacts of a strategic move in assembly location. As the US market has been consuming more sports-utility vehicles and trucks and fewer small cars, in 2017, Ford made the strategic decision to shift production of the Ford Focus from Michigan to China in order to free up capacity at its Great Lakes plants for larger vehicles (Durbin, 2017) (Figure 60). For this analysis, it is assumed that US Ford Focus sales in 2019 will be the same as in 2017 (about 158,000 cars).

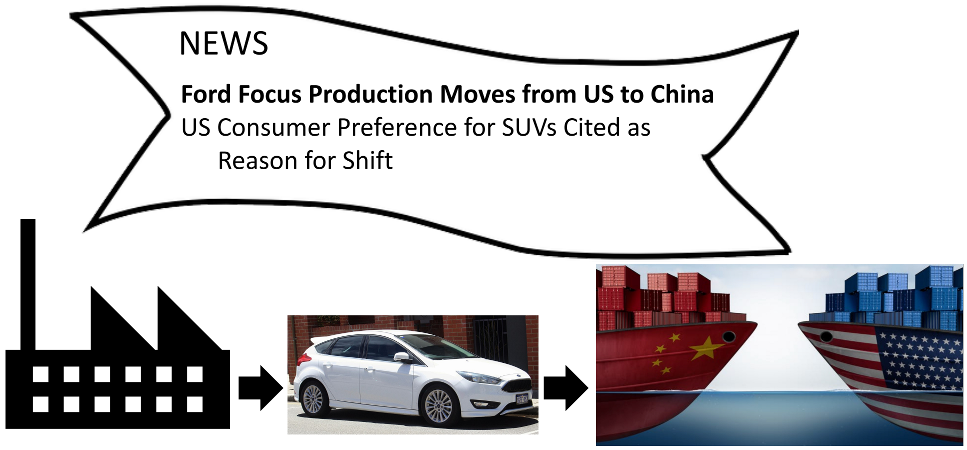



**Figure 60. Ford Focus assembly to move from Michigan to China.**

This analysis considers only Ford Focus vehicles that are sold in the US to demonstrate the effectiveness of the methodology in this study. A broader accounting would also include the effects on TECE associated with shipments of Ford Focus sales in other countries and shifts in SUV and pickup truck production.

The TECE results of this analysis are shown below in

Table **18**. This uses the same assumptions stated above—e.g., that destinations in the Northeast and Southeast US use the Port of Baltimore while other US destinations use the Port of Los Angeles. Total ton-miles by mode would increase as follows:

- Ocean: an increase from zero to 2.9 billion; and
- Rail: about 40% (from 107 million to 148 million).

Ton-miles by truck is predicted to undergo little change, with approximately 239 million ton-miles in each scenario. Compared to baseline TECE associated with transporting the Ford Focus from Michigan, energy consumption would increase by 51%; $CO_2$ emissions would more than double; and large increases would occur in NOX, PM10 and Sulfur-related emissions.

**Table 18. TECE analysis of Ford Focus case study.**

| Measurement | Scenario | Ocean | Rail | Truck | Total | Change | % Change |
|---|---|---|---|---|---|---|---|
| Energy consumption: gallons of fuel (1,000,000s) | Baseline | - | 0.23 | 4.77 | 5.00 | 2.54 | 51% |
| | New | 2.45 | 0.32 | 4.78 | 7.54 | | |
| Emissions (1,000,000s of grams) | | | | | | | |
| CO2 emissions | Baseline | - | 2,333 | 14,586 | 16,919 | 25,503 | 151% |
| | New | 24,609 | 3,222 | 14,591 | 42,423 | | |
| NOX emissions | Baseline | - | 27 | 56 | 82 | 612 | 744% |
| | New | 602 | 37 | 56 | 695 | | |
| SO2 emissions | Baseline | - | 0.4 | 0.0 | 0.4 | 0.16 | 38% |
| | New | - | 0.6 | 0.0 | 0.6 | | |
| SOX emissions | Baseline | - | - | - | - | 351.73 | N/A |
| | New | 352 | - | - | 352 | | |
| PM10 emissions | Baseline | - | 0.7 | 1.1 | 1.8 | 49.94 | 2791% |
| | New | 49.7 | 0.9 | 1.1 | 51.7 | | |

These results demonstrate that strategic decisions regarding supply chain strategies—in this case, near-sourcing vs. foreign or low-cost country sourcing—have important implications for TECE. A single strategy and one OEM was examined here. Next steps could include finding the new TECE associated with shipping sports utility vehicles and trucks from the Great Lakes plants, but this is beyond the scope of the current work.

## 6.6 Conclusions

This chapter demonstrates the importance of assessing impacts of business strategy on goods movement and related outcomes. The study bridges a gap in the existing literature. This work has major implications for agent-based freight modeling since strategies are the guiding force behind many business decisions that shape the flow of goods and vehicles. An agent-based, sketch planning approach is developed in addition. The approach is applied to evaluate the energy and emissions impacts of a strategic shift in auto assembly location.

# 7 Conclusions

## 7.1 Summary

This thesis provides an innovative new paradigm for agent-based modeling of freight transportation. The work is motivated by numerous, fundamental gaps in extant, agent-based freight transportation behavioral models. Existing working systems use the establishment as the agent and do not consider collections of establishments, also known as firms. By and large, existing frameworks also do not model critical transportation decisions such as private fleet ownership, fleet size and fleet composition; and distribution center decisions such as location. However, decades of passenger modeling experience have shown that agent collections (e.g., the household) are important to model along with a more fundamental agent (the person), and that household location decisions and fleet ownership are critical inputs to everyday transportation decisions such as mode choice and tour patterns. By analogy, just as both home and workplace locations inform the passenger tour, factory and distribution center (for example) locations inform the freight tour. Based on this experience, and on insights from the business domain, remedying these gaps while simultaneously developing innovative features is a major thrust of this research.

A theoretical, agent-based modeling architecture is developed to fulfill this vision (Figure 61). Each layer plays a different functional role in agent behavior. The highest, strategic level covers decisions that have an enduring impact such as investment in fixed assets. The tactical layer covers decisions that can be altered more easily, such as shipment decisions. The operational layer covers everyday decisions including scheduling of vehicle tours and adjustment of routes in real-time. The design of the architecture supports the operationalization the famous push-pull boundary that is witnessed in supply chains. This boundary is related to information sharing among agents, or visibility of the end consumer, and it has important effects on production, consumption, inventory, and shipment size decisions. Since information availability continues to grow and change, it is more important than ever to model this phenomenon. An “Effect of Information” is integrated into the architecture for this reason. Interactions between freight and passenger travel, such as with retail goods movement and crowd-shipping, are also considered. Parts of the theoretical framework are operationalized in a proof of concept where several of the framework features, including distribution path choice and interactions of freight and passenger retail activity, are also implemented.

The architecture is centered around firm strategy, which is an unseen policy that the agent adopts to guide its decisions and activities. Strategy plays a key role in unifying upstream and downstream decision-making by the agent. The architecture operationalizes firm strategy by deploying a novel mathematical system that jointly considers firm strategy and strategic decisions for a number of dimensions. The framework uses a Seemingly Unrelated Regression (SUR) system, which is well established in the econometrics literature, and extents the SUR model by introducing latent variables, which are firm strategies. A Gibbs Sampling solution method is implemented to estimate the model parameters of this system. The outcomes, strategic decisions, are used either "as is" or are fed into downstream model decisions such as facility location decisions.

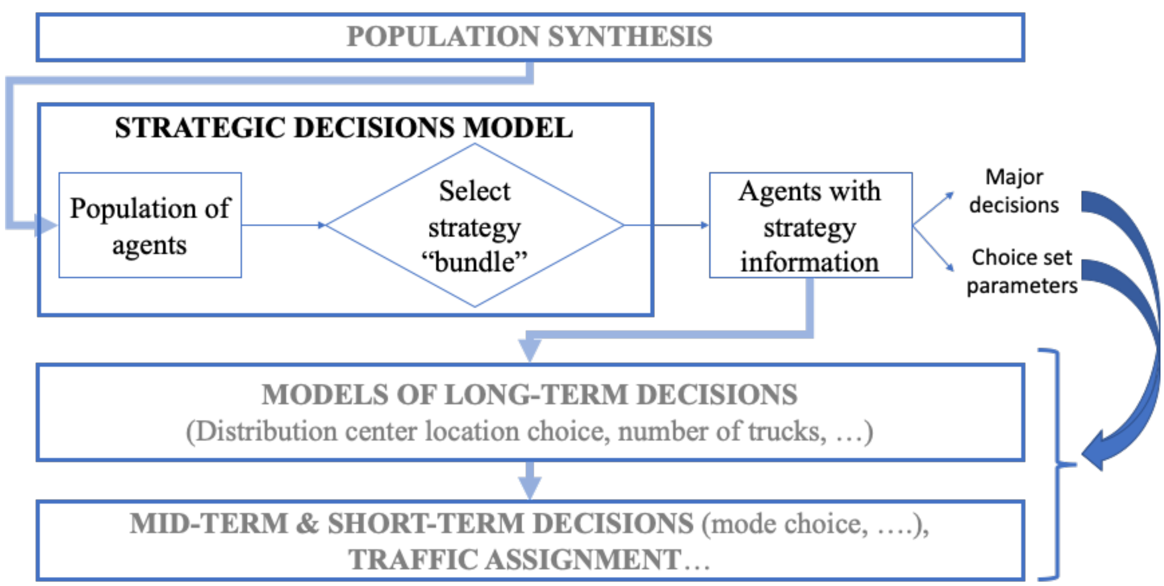



**Figure 61. Summary figure: strategy in the agent-based modeling context.**

A proof of concept demonstrates the value of this approach by modeling two strategies, Logistical Sophistication and Customer Service focus, jointly with eight strategic decisions that involve various facets of private fleet ownership and distribution center (DC) control. The strategies are shown to impact the strategic decisions along with exogenous firm characteristics including industry sector, providing convincing evidence with respect to the jointness of these decisions. To demonstrate the connection between strategy and downstream decisions, the DC-related strategic decisions, which in this case operationalize firm strategies as they relate to fixed asset investments, are applied to analyze a multiple discrete-continuous decision (total DC area in each FAF zone throughout the US). A new algorithm is developed and applied to use the strategic decision outcomes for purposes of estimating both the national distribution center structure of a firm. A mock-up showing how to further use the outcomes as choice set generation parameters, for distribution center location choice in this case, is also illustrated. The proof of concept itself provides valuable model results in an under-researched area by providing a way to estimate firm-level private fleet and distribution center structure simultaneously. It provides statistical evidence that strategy does, in fact, impact strategic logistics control decisions, and it gives important insights more generally into how companies make decisions regarding internalization versus outsourcing of logistics controls.

Two innovative methods are developed to generate measurements of strategy using natural text. This development is motivated by the complete gap of quantitative strategy data for firms. The Simple Scaled Bag-of-Words (SS-BOW) algorithm creates measurements of firm strategy based on relative usage of select words, while the word2vec-Principal Components Analysis (w2vPCA) algorithm creates measurements based on quantifying relative differences in word use among firms. The concept is proven by applying each method to a sample of large-scale text data, which are generated by the Attitudinal Data Development Engine (ADDE) that is designed and implemented expressly for this purpose. A battery of evidence, including visualization, statistical testing of means, and factor analysis, is compiled to prove the value and reasonableness of the new methods. The measurements are also input to the SUR model that is described above, which produces intuitive results regarding underlying strategy.

Finally, a global sustainability analysis demonstrates that modeling agent strategies is extremely relevant to predicting the impacts of business activity on transportation flows, energy use and emissions. The main purpose of the sustainability analysis is to demonstrate how strategy can have major, far-reaching impacts for transportation-related outcomes. In the process, a method to evaluate the transportation energy consumption and emissions (TECE) specific to global, multimodal transportation—and additionally to the automotive manufacturing industry in particular—based on various supply chain configurations and strategies is proposed and demonstrated. A host of data is collected to conduct the analysis. A case study shows that near-sourcing in the automotive assembly context—that is, assembling automobiles near their sales locations—as a firm strategy supports sustainability. In contrast, sourcing from foreign assembly locations can have significant, detrimental impacts on energy use and emissions.

## 7.2 Extensions

The first priority is implementing the remaining aspects of the new architecture. Chapter 3 discusses the implementation to date. The next steps are to model the firm structure among establishments and update their trade collaborations at the firm level; to implement the strategic logistics decisions model results directly into the computational system (POLARIS); and, for firms with no private fleet or distribution control, to model their relationships with carriers and 3PLs. The carriers model, with decisions resulting in transport and logistics service offerings, needs to be developed. E-commerce is already implemented but can be updated with recent data. The Information Effect and push-pull boundary need to be coded. The mathematical details for the above elements need to be worked out. Truck touring models are currently external to the main computational program, so these need to be integrated directly.

Future efforts can extend and improve the strategy data algorithms in several ways. Meta-analysis methods can be developed to improve on the process of keyword selection. Procedures to compare SS-BOW and w2vPCA methods can be further developed, then applied to determine which method is superior. Researchers can apply the methods in different contexts to strengthen empirical evidence of their validity. A rigorous comparison of the methods and their results to established methods in survey-based attitudinal measurements should be undertaken. Mathematical proofs for these methods should be developed. Processes to improve the preparation of input data should be developed.

The strategy model can be extended in several directions. The methodology can be extended to include ordinal decisions in addition to binary and continuous decisions. Additional research can shed light into what decisions are best characterized as strategic. This will have implications for the modeling structure, as strategic decisions are proposed to be modeled early in the model stream and other decisions downstream. Differences in this application for passenger and freight contexts can be explored. The application shown here can be extended to utilize a larger sample size and more independent variables, with related improvements being the use of separate samples for estimation and validation. The various models that are estimated can be improved with additional calibration and validation.

The sustainability estimation assesses the impacts of shipping in the final stages of the automotive supply chain. A more complete accounting of energy consumption and emissions would involve applying the methods to the supply chains of other industries and across all stages of the supply chain. More immediate potential improvements relate to using a finer level of geographic and network detail and improving the mode and port choice processes.

Finally, adaptations to passenger models are logical extension of this work. The concepts of strategy and strategic decisions as unifying elements for passenger modeling decisions can be explored. The strategy model can be deployed for passenger-specific applications. The attitudinal data development algorithms can be applied to develop attitudinal data for individual persons. Unbiased sources of strategy data that are analogous to annual, mandatory company reports will need to be located.

## Cited Literature

Abdelwahab, W. & M. Sargious. (1992). Modelling the demand for freight transport: A New Approach. *Journal of Transport Economics and Policy*, *26*(1), 49.

Adnan, M., Pereira, F. C., Azevedo, C. L., Basak, K., Lovric, M., Raveau, S., Zhu, Y., Ferreira, J., Zegras, C., & Ben-Akiva, M. (2016). *SimMobility: A Multi-scale Integrated Agent-Based Simulation Platform*. Transportation Research Board 95th Annual Meeting.

Albert, J. H., & Chib, S. (1993). Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association*, *88*(422), 669–679. JSTOR.

Alho, A., Bhavathrathan, B. K., Stinson, M., Gopalakrishnan, R., Le, D.-T., & Ben-Akiva, M. (2017). A multi-scale agent-based modelling framework for urban freight distribution. *Transportation Research Procedia*, *27*, 188–196.

Amore, M. D., & Murtinu, S. (2019). Tobit models in strategy research: Critical issues and applications. *Global Strategy Journal*, 1–25.

Anowar, S., Eluru, N., & Miranda-Moreno, L. F. (2014). Alternative Modeling Approaches Used for Examining Automobile Ownership: A Comprehensive Review. *Transport Reviews*, *34*(4), 441–473.

Aramark. (2017). Form 10-K: For the Fiscal Year Ended September 29, 2017. US Securities and Exchange Commission Archives. Retrieved August 4, 2020 from https://www.sec.gov/edgar.shtml.

Aston Martin (2018). Aston Martin Reports Q4 and Full Year 2016 Results. https://media.astonmartin.com/aston-martin-reports-q4-and-full-year-2016-results/. Accessed April 5, 2018.

Auld, J., Hope, M., Ley, H., Sokolov, V., Xu, B., & Zhang, K. (2016). POLARIS: Agent-based modeling framework development and implementation for integrated travel demand and network and operations simulations. *Transportation Research Part C: Emerging Technologies*.

Auld, J., & Mohammadian, A. (2012). Activity planning processes in the Agent-based Dynamic Activity Planning and Travel Scheduling (ADAPTS) model. *Transportation Research Part A: Policy and Practice*, *46*(8), 1386–1403.

Autonation, Inc. (2017). Form 10-K: For the Fiscal Year Ended December 31, 2016. US Securities and Exchange Commission Archives. Retrieved August 4, 2020 from https://www.sec.gov/edgar.shtml.

Barbosa, L., & Feng, J. (2010). Robust Sentiment Detection on Twitter from Biased and Noisy Data. *Coling 2010: Posters*, 36–44.

Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, *1*(4), 371–394.

Beagan, D., Fischer, M., & A. Kuppam. (2007). *Quick Response Freight Manual II* (Final Report FHWA-HOP-08-010). US Department of Transportation Federal Highway Administration.

Ben-Akiva, M. (2010). Planning and Action in a Model of Choice. In S. Hess & A. Daly (Eds.), *Choice Modelling: The State-of-the-art and The State-of-practice* (pp. 19–34). Emerald Group Publishing Limited.

Ben-Akiva, M, & Boccara, B. (1995). Discrete choice models with latent choice sets. *International Journal of Research in Marketing*, *12*, 9–24.

Ben-Akiva, M, Bolduc, D., & Park, J. Q. (2013). Discrete Choice Analysis of Shippers' Preferences. In Moshe Ben-Akiva, H. Meersman, & E. Van de Voorde (Eds.), *Freight Transport Modelling* (pp. 119–141). Emerald Group Publishing Limited.

Ben-Akiva, M., & de Jong, G. (2008). The Aggregate–Disaggregate–Aggregate (ADA) Freight Model System. In M. Ben-Akiva, H. Meersman, & E. van de Voorde (Eds.), *Recent Developments in Transport Modelling* (pp. 117–134). Emerald Group Publishing Limited.

Ben-Akiva, M., & Lerman, S. (1985). *Discrete Choice Analysis*. MIT Press.

Ben-Akiva, M., Mcfadden, D., Train, K., Walker, J., Bhat, C., Bierlaire, M., Bolduc, D., Boersch-Supan, A., Brownstone, D., Bunch, D. S., Daly, A., De Palma, A., Gopinath, D., Karlstrom, A., & Munizaga, M. A. (2002). Hybrid Choice Models: Progress and Challenges. *Marketing Letters*, *13*(3), 163–175.

Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A Neural Probabilistic Language Model. *Journal of Machine Learning Research*, *3*, 1137–1155.

Bhat, C. R. (2008). The Multiple Discrete-Continuous Extreme Value (MDCEV) Model: Role of Utility Function Parameters, Identification Considerations, and Model Extensions. *Transportation Research Part B*, *42*(3), 274–303.

Bhat, C. R. (2011). The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. *Transportation Research Part B: Methodological*, *45*(7), 923–939.

Bhat, C. R. (2015a). A new generalized heterogeneous data model (GHDM) to jointly model mixed types of dependent variables. *Transportation Research Part B: Methodological*, *79*(C), 50–77.

Bhat, C. R. (2015b). A comprehensive dwelling unit choice model accommodating psychological constructs within a search strategy for consideration set formation. *Transportation Research Part B: Methodological*, *79*, 161–188.

Bhat, C. R., Astroza, S., Bhat, A. C., & Nagel, K. (2016). Incorporating a multiple discrete-continuous outcome in the generalized heterogeneous data model: Application to residential self-selection effects analysis in an activity time-use behavior model. *Transportation Research Part B: Methodological*, *91*(C), 52–76.

Bhat, C. R., Pinjari, A. R., Dubey, S. K., & Hamdi, A. S. (2016). On accommodating spatial interactions in a Generalized Heterogeneous Data Model (GHDM) of mixed types of dependent variables. *Transportation Research Part B: Methodological*, *94*(C), 240–263.

BigMackTrucks.com (2017). *2017 Fleet Owner 500—America's Top Private Fleets—Trucking News*. Retrieved January 8, 2020, from https://www.bigmacktrucks.com/topic/48661-2017-fleet-owner-500-americas-top-private-fleets/

BlueOvalNews.Com and AtomicFrog.Com (2018). Who Makes It and Where? A Guide to Ford Manufacturing Sites All Over the World., 1998. Accessed April 5, 2018 at http://www.atomicfrog.com/mirrors/www.blueovalnews.com/who_makes_it_and_where.htm.

BMW Group (2018). BMW Worldwide Production https://www.bmwusfactory.com/manufacturing/bmw-worldwide/. Accessed April 5, 2018.

Boerkamps, J. & A. van Binsbergen. (1999). *GoodTrip—A new approach for modelling and evaluation of urban goods distribution*. Urban Transport Conference, 2nd KFB Research Conference, Lund.

Bollen, K. A. (1989). *Structural equations with latent variables* (pp. xiv, 514). John Wiley & Sons.

Boriboonsomsin, K., G. Scora, G. Wu and M. Barth (2017). Improving Vehicle Fleet, Activity, and Emissions Data for On-Road Mobile Sources Emissions Inventories. Prepared for FHWA.

Cambridge Systematics, Inc. (2011, June 30). *A Working Demonstration of a Mesoscale Freight Model for the Chicago Region: Final Report and User's Guide* (Report No. CS 8413.001). Prepared for the

Chicago Metropolitan Agency for Planning with contributions from A. Mohammadian, K. Kawamura, and J. Auld.

Cambridge Systematics, Inc., MORPACE International, Perteet, Inc., & Transportation Management & Design, Inc. (2007). *South Cook County – Will County Service Restructuring Initiative Market Research: Final Report*. Prepared for Pace Suburban Bus Service.

Caplice, C. (2006, December). *Transportation Management—Course ESD.260/15.770/1.260, Logistics Systems* [Course lecture]. https://ocw.mit.edu/courses/engineering-systems-division/esd-260j-logistics-systems-fall-2006/lecture-notes/lect25.pdf

Caplice, C. (2016). *SCM.260: Logistics Systems, Course Documents*.

Car carrier trailer (2018). Wikipedia. https://en.wikipedia.org/wiki/Car_carrier_trailer. Accessed April 10, 2018.

Castro, M., Eluru, N., Bhat, C. R., & Pendyala, R. M. (2011). Joint Model of Participation in Nonwork Activities and Time-of-Day Choice Set Formation for Workers: *Transportation Research Record*, *2254*(1), 140–150. https://doi.org/10.3141/2254-15

Cavalcante, R. & Roorda, M. (2013). Freight Market Interactions Simulation (FREMIS): An Agent-Based Modeling Framework. *Procedia Computer Science(19)*, 867–873. doi: 10.1016/j.procs.2013.06.116

CC BY-SA (2020). *Creative Commons—Attribution-ShareAlike 3.0 Unported—CC BY-SA 3.0*. Retrieved August 4, 2020, from https://creativecommons.org/licenses/by-sa/3.0/

CC BY-NC-SA (2020). *Creative Commons—Attribution-NonCommercial-ShareAlike 3.0 Unported—CC BY-NC-SA 3.0*. Retrieved August 5, 2020, from https://creativecommons.org/licenses/by-nc-sa/3.0/

Chen, A. (2014, June 5). *Inventory Management I: Economic Order Quantity (EOQ)* [Course lecture]. 15.734 Intro to OM, MIT. Accessed October 2018 at http://www.mit.edu/~anniecia/OMrecitation1_AnnieChen_MIT.pdf

Chen, L., & Lee, H. L. (2017). Modeling and Measuring the Bullwhip Effect. In A. Y. Ha & C. S. Tang (Eds.), *Handbook of Information Exchange in Supply Chain Management* (Vol. 5, pp. 3–25). Springer International Publishing.

Chicago Metropolitan Agency for Planning (CMAP). (2012). *Metropolitan Chicago's Freight Cluster: A Drill-Down Report on Infrastructure, Innovation, and Workforce: Full Report.*

Choo, S., & Mokhtarian, P. L. (2008). How do people respond to congestion mitigation policies? A multivariate probit model of the individual consideration of three travel-related strategy bundles. *Transportation*, *35*(2), 145–163.

Choo, S., & Mokhtarian, P. L. (2012). Individual responses to congestion policies: Modeling the consideration of factor-based travel-related strategy bundles. *KSCE Journal of Civil Engineering*, *16*(5), 822–834.

Chow, J. Y., J, Yang, C. H., & Regan, A. C. (2010). State-of-the art of freight forecast modeling: Lessons learned and the road ahead. *Transportation; New York*, *37*(6), 1011–1030.

Coia, A. (2016, July 29). North American port survey: Ports get an extra lift from a rise in imports. Automotive Logistics.

Coia, A. and C. Ludwig (2016, June 10). Plant Locations: Right Place, Right Time. Automotive Logistics.

Collins, C., Hasan, S., & Ukkusuri, S. (2013). A Novel Transit Rider Satisfaction Metric: Rider Sentiments Measured from Online Social Media Data. *Journal of Public Transportation*, *16*(2), 21–45. https://doi.org/10.5038/2375-0901.16.2.2

CoStar. (2020). *CoStar: Commercial Real Estate Information Company*. Retrieved July 21, 2020, from https://www.costar.com/docs/default-source/wireframe-docs/costar-suite-brochure.pdf?sfvrsn=8cfc22a4_2

Council of Supply Chain Management Professionals (CSCMP). (2013, August). *CSCMP Supply Chain Management Definitions and Glossary*. Definitions Compiled by Kate Vitasek (Www.Scvisions.Com). https://cscmp.org/CSCMP/Educate/SCM_Definitions_and_Glossary_of_Terms.aspx

Cowles, M. K., Carlin, B. P., & Connett, J. E. (1996). Bayesian Tobit Modeling of Longitudinal Ordinal Clinical Trial Compliance Data With Nonignorable Missingness. *Journal of the American Statistical Association*, *91*(433), 86–98. JSTOR. https://doi.org/10.2307/2291385

Crastes dit Sourd, R., Zawojska, E., Mahieu, P.-A., & Louviere, J. (2018). Mitigating strategic misrepresentation of values in open-ended stated preference surveys by using negative reinforcement. *Journal of Choice Modelling*, *28*, 153–166. https://doi.org/10.1016/j.jocm.2018.06.001

CSX Corporation. Fuel Efficiency (2016). https://www.csx.com/index.cfm/about-us/the-csx-advantage/fuel-efficiency/. Accessed April 10, 2018.

Daimler AG. Wikipedia. https://en.wikipedia.org/wiki/Daimler_AG. Accessed April 5, 2018.

Daziano, R. A. (2015). Inference on mode preferences, vehicle purchases, and the energy paradox using a Bayesian structural choice model. *Transportation Research Part B: Methodological*, *76*, 1–26. https://doi.org/10.1016/j.trb.2015.02.012

Drop shipping. (2019). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Drop_shipping&oldid=888388103

de Bok, M., & Tavasszy, L. (2018). An empirical agent-based simulation system for urban goods transport (MASS-GT). *Procedia Computer Science*, *130*, 126–133. https://doi.org/10.1016/j.procs.2018.04.021

de Jong, G., & Ben-Akiva, M. (2007). A micro-simulation model of shipment size and transport chain choice. *Transportation Research Part B: Methodological*, *41*(9), 950–965. https://doi.org/10.1016/j.trb.2007.05.002

de Jong, G., Vierth, I., Tavasszy, L., & M. Ben-Akiva. (2013). Recent developments in national and international freight transport models within Europe. *Transportation*, *40*(2), 347–371.

Demetsky, M. J. (1974). Measurement of Urban Commodity Movements. *Transportation Research Record: Journal of the Transportation Research Board*, *496*, 57–67.

DieselNet with M. Pedersen (2018). International: IMO Marine Engine Regulations. https://www.dieselnet.com/standards/inter/imo.php. Accessed 4/15/2018.

Durbin, D. (2017, June 20). Ford Focus: Made in China. US News. https://www.usnews.com/news /business /articles/2017-06-20/ford-to-import-focus-small-car-from-china-in-2019. Accessed April 15, 2018.

Ehrler, V. C., Seidel, S., Lischke, A., Davydenko, I., Dobers, K., Lewis, A., & Val, S. (2018, April 16). *Standardization of Transport Chain Emission Calculation – Status Quo and What Is Needed Next*. 7th Transport Research Arena TRA 2018, Vienna, Austria.

Estall, R. (1985). *A modern geography of the United States. Quoted in D. Hall and A. Braitwaite, Development of Thinking in Supply Chain and Logistics Management in Handbook of Logistics and Supply Chain Management, Edited by A.M. Brewer et al., 2001*. New York: Times Co.

Everitt, B. S., & Dunn, G. (1991). *Applied Multivariate Data Analysis*. Halsted Press.

Fang, H. A. (2008). A discrete–continuous model of households' vehicle choice and usage, with an application to the effects of residential density. *Transportation Research Part B: Methodological*, *42*(9), 736–758.

Federal Highway Administration (2020). *Development of a Behavioral-Based National Freight Demand Model and an Innovative Freight Data Collection Method*. The Exploratory Advanced Research Program Fact Sheet FHWA-HRT-17-122 HRTM-30/06-18(WEB)E.

Ferrari (2018). Wikipedia. https://en.wikipedia.org/wiki/Ferrari. Accessed April 5, 2018.

Fiat Chrysler Automobiles (2018). FCA US LLC Manufacturing Operations. http://media.fcanorthamerica.com/newsrelease.do?id=9117&mid=. Accessed April 5, 2018.

Fischer, M. J., Outwater, M. L., Cheng, L. L., Ahanotu, D. N., & Calix, R. (2005). Innovative Framework for Modeling Freight Transportation in Los Angeles County, California. *Transportation Research Record*, *1906*(1), 105–112. https://doi.org/10.1177/0361198105190600113

FleetSeek. (2017). *FleetSeek*. Retrieved July 21, 2020, from https://fleetseek.com/

Ford Motor Company (2018). Ford Lio Ho Motor Co., Ltd. (Joint venture – 70% Ford/30% Lio Ho Group). https://corporate.ford.com/content/ford-corporate/en/company/plant-detail-pages/ford-lio-ho-motor-co-ltd.html. Accessed April 5, 2018.

Forrester, J. W. (1961). *Industrial dynamics.* M.I.T. Press.

Fortune. *Fortune 500*. Retrieved January 18, 2020, from https://fortune.com/fortune500/

Friedrich, H. (2010, July 11). Simulation of logistics in food retailing for freight transportation analysis. *World Conference in Transportation Research*.

Freight Analysis Framework Data Tabulation Tool (2018). Developed by Oak Ridge National Laboratory with funding from the Bureau of Transportation Statistics and the Federal Highway Administration. https://faf.ornl.gov/faf4/Extraction0.aspx. Accessed July 25, 2018.

Geweke, J. (2005). *Contemporary Bayesian Econometrics and Statistics*. John Wiley & Sons, Inc.

Ghiassi, M., Skinner, J., & Zimbra, D. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with Applications*, *40*(16), 6266–6282. https://doi.org/10.1016/j.eswa.2013.05.057

GM Corporate Newsroom (2018). List of Global Assembly Plants. http://media.gm.com/media/us/en/gm/plants-facilities.html. Accessed April 5, 2018.

GMAuthority.com (2018). General Motors Manufacturing Plants. http://gmauthority.com/blog/gm/gm-manufacturing/. Accessed April 5, 2018.

Gonzalez-Feliu, J. (2019). *Logistics and Transport Modeling in Urban Goods Movement*. IGI Global.

Google Code Archive. (2013). *Word2vec (archive)* [C]. Retrieved July 19, 2020, from https://code.google.com/archive/p/word2vec/

Grap - Own work (2010). CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=9755150. Retrieved May 20, 2019, from https://en.wikipedia.org/wiki/Bullwhip_effect with Creative Commons license CC BY-SA 3.0, https://creativecommons.org/licenses/by-sa/3.0/.

Hall, D., & Braitwaite, A. (2001). Development of Thinking in Supply Chain and Logistics Management. In *Handbook of Logistics and Supply Chain Management, Edited by A.M. Brewer et al.*

Harris, F. (1915). *Operations and cost*. A. Shaw and Company.

Heaver, T. D. (2001). Perspectives on Global Performance Issues,. In *Handbook of Logistics and Supply Chain Management, Edited by A.M. Brewer et al.*

Hensher, D. A. (2014). Attribute processing as a behavioural strategy in choice making. In *Handbook of Discrete Choice Modelling, Hess, S. and Daly, A.J. (eds.): Vol. Chapter 4.1*. Edward Elgar.

Hinton, G. E. (1986). Learning distributed representations of concepts. *Proceedings of the Eighth Annual Conference of the Cognitive Science Society*, 1–12.

Holmgren, J., Davidsson, P., Persson, J. A., & Ramstedt, L. (2012). TAPAS: A multi-agent-based model for simulation of transport chains. *Simulation Modelling Practice and Theory*, *23*, 1–18. https://doi.org/10.1016/j.simpat.2011.12.011

Hoekstra, S., & Romme, J. (1992). *Integral Logistic Structures: Developing Customer-oriented Goods Flow*. Industrial Press Inc.

Honda Government Relations Office (2018). Honda employs more than 31,000 associates in the U.S. https://hondainamerica.com/employment/. Accessed April 5, 2018.

Hong, S., Pourabdollahi, Z., Ehlen, M., Stinson, M., Jeon, K., You, D., & Zhu, H. (2017, May). *A behavioral agent-based supply chain and freight transportation model: Application for Arizona Sun Corridor Megaregion*. TRB Applications Conference, Raleigh, North Carolina.

Hunt, J. D., & Stefan, K. J. (2007). Tour-based microsimulation of urban commercial movements. *Transportation Research Part B: Methodological*, *41*(9), 981–1013. https://doi.org/10.1016/j.trb.2007.04.009

Hyundai Motor America (2017). Hyundai Motor America Reports 2017 December and Total Year Sales. Acessed https://www.prnewswire.com/news-releases/hyundai-motor-america-reports-2017-december-and-total-year-sales-300577146.html. Accessed April 5, 2018.

ICF Consulting (2017). Assessing the Effects of Freight Movement on Air Quality at the National and Regional Level. Prepared for FHWA. https://www.fhwa.dot.gov/Environment/air_quality/research/effects_of_freight_movement/chapter00.cfm. Accessed April 5, 2018.

International Energy Agency. (2017). *Plotting a more sustainable course for the international shipping sector*. Retrieved August 9, 2020, from https://www.iea.org/news/plotting-a-more-sustainable-course-for-the-international-shipping-sector

Isidore, C. (2016, Nov. 16). Ford will be the first automaker to import cars from India to the U.S. CNNMoney. https://money.cnn.com/2016/11/16/news/companies/ford-cars-import-india/index.html. Accessed April 12, 2018.

Jacksonville Port Authority. Cargo: Autos. http://www.jaxport.com/cargo/autos. Accessed July 20, 2018.

Jin, X., & Shams, K. (2016). *Examining the Value of Travel Time Reliability for Freight Transportation to Support Freight Planning and Decision-Making: Final report*. Florida Department of Transportation. Project Number BDV29-977-15.

Jin, Y., Williams, I., & Shahkarami, M. (2005). *Integrated regional economic and freight logistics modeling: Results from a model for the Trans-Pennine Corridor, UK*. European transport conference, Strasbourg.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, *374*(2065).

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263–291. JSTOR.

Kamargianni, M., & Polydoropoulou, A. (2013). Hybrid Choice Model to Investigate Effects of Teenagers' Attitudes toward Walking and Cycling on Mode Choice Behavior. *Transportation Research Record*, *2382*(1), 151–161.

Kaplan, D. (2009). *Structural Equation Modeling (2nd ed.): Foundations and Extensions*. SAGE Publications, Inc.

Kaplan, S., Shiftan, Y., & Bekhor, S. (2012). Development and estimation of a semi-compensatory model with a flexible error structure. *Transportation Research Part B: Methodological*, *46*(2), 291–304.

Kraus, W., de Souza, F. A., Carlson, R. C., Papageorgiou, M., Dantas, L. D., Camponogara, E., Kosmatopoulos, E., & Aboudolas, K. (2010). Cost Effective Real-Time Traffic Signal Control Using the TUC Strategy. *IEEE Intelligent Transportation Systems Magazine*, *2*(4), 6–17.

Lai, K., Ngai, E. W. T., & Cheng, T. C. E. (2002). Measures for evaluating supply chain performance in transport logistics. *Transportation Research Part E: Logistics and Transportation Review*, *38*(6), 439–456.

Lavieri, P. S., Garikapati, V. M., Bhat, C. R., Pendyala, R. M., Astroza, S., & Dias, F. F. (2017). Modeling Individual Preferences for Ownership and Sharing of Autonomous Vehicle Technologies. *Transportation Research Record: Journal of the Transportation Research Board*, *2665*, 1–10.

Lee, M. (2018, Fall). *Course Notes, IDS 566: Advanced Text Analytics for Business*.

Levy, O., Goldberg, Y., & Dagan, I. (2015). Improving Distributional Similarity with Lessons Learned from Word Embeddings. *Transactions of the Association for Computational Linguistics*, *3*, 211–225.

Liedtke, G., Matteis, T., & Wisetjindawat, W. (2015). Impacts of Urban Logistics Measures on Multiple Actors and Decision Layers: Case Study. *Transportation Research Record: Journal of the Transportation Research Board*, *2478*(1), 57–65.

List of Chrysler factories (2018). Wikipedia. https://en.wikipedia.org/w/index.php?title=List_of_Chrysler_factories. Accessed April 5, 2018.

List of General Motors factories (2018). Wikipedia. https://en.wikipedia.org/wiki/List_of_General_Motors_factories. Accessed April 5, 2018.

List of Honda Assembly Plants (2018). Wikipedia. https://en.wikipedia.org/wiki/List_of_Honda_assembly_plants. Accessed April 5, 2018.

List of Kia Design and Manufacturing Facilities (2018). Wikipedia. Accessed April 5, 2018. https://en.wikipedia.org/wiki/List_of_Kia_design_and_manufacturing_facilities

List of Toyota manufacturing facilities (2018). Wikipedia. https://en.wikipedia.org/wiki/List_of_Toyota_manufacturing_facilities. Accessed April 5, 2018.

List of Volkswagen Group factories (2018). Wikipedia. https://en.wikipedia.org/wiki/List_of_Volkswagen_Group_factories. Accessed April 5, 2018.

List of Volvo Car production plants (2018). Wikipedia. https://en.wikipedia.org/wiki/List_of_Volvo_Car_production_plants. Accessed April 5, 2018.

Lowrey, A. (2011, June 27). Your Big Car Is Killing Me. Slate. http://www.slate.com/business/2018/04/its-time-for-the-mortgage-interest-deduction-to-go.html. Accessed April 20, 2018.

Mapping-Tools.Com (2018). Pre-Computed Mileage Charts. https://www.mileage-charts.com. Accessed April 15, 2018.

Manski, C. (1977). The Structure of Random Utility Models. *Theory and Decision*, *8*(3), 229-254.

Maricopa Association of Governments. (2018). *Mega-Regional Multi-Modal Agent-Based Behavioral Freight Model MAG Next Generation Freight Demand Model: Final Report*.

Mazda (2018). Wikipedia. https://en.wikipedia.org/wiki/Mazda. Accessed April 5, 2018.

McCulloch, R., & Rossi, P. (1994). An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics*, *64*(1–2), 207–240.

Meersemen, H., & Van de Voorde, E. (2001). International logistics: A Continuous Search for Competitiveness. In *Handbook of Logistics and Supply Chain Management, Edited by A.M. Brewer et al.*

Mele, J. (2017, February 7). 2017 Fleet Owner 500. *FleetOwner*. Accessed August 23, 2020 at https://www.fleetowner.com/truck-stats/article/21695385/2017-fleet-owner-500.

Menon, A., Choi, J., & Tabakovic, H. (2018). What You Say Your Strategy Is and Why It Matters: Natural Language Processing of Unstructured Text. *Academy of Management Proceedings*, *2018*(1), 18319.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *ArXiv:1301.3781 [Cs]*. http://arxiv.org/abs/1301.3781

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. *ArXiv:1310.4546 [Cs, Stat]*, *2*, 3111–3119. http://arxiv.org/abs/1310.4546

Mintzberg, H. (1987). The Strategy Concept II: Another Look at Why Organizations Need Strategies. *California Management Review*, *30*(1), 25–32.

Mitchell, A. (2017, June 25). Pushing 9,000 employees, BMW workforce is on the rise. The Greenville News.

Mitsubishi Motors Corporation. *Manufacturing Centers and Related Facilities*. Public Relations Department. https://www.mitsubishi-motors.com/en/company/facilities/. Accessed April 5, 2018

Moscardi, C. (2019, April 9). *Using Natural Language Processing to Improve the Commodity Flow Survey*. Innovations in Freight Data Workshop, Irvine, California.

Nissan Corporate Communications (2017, Feb. 16). Fact Sheet: Nissan Canton Vehicle Assembly Plant. https://www.onenewsnow.com/media/8726626/nissan-canton-2-16-17.pdf. Accessed April 5, 2018.

NissanNews.com (2018, Jan. 3). Fact Sheet: Vehicle Assembly Plant and Battery Plant - Smyrna, Tennessee. http://nissannews.com/en-US/nissan/usa/channels/Plant-Fact-Sheets/releases/vehicle-assembly-plant-smyrna-tennessee. Accessed April 5, 2018.

NLTK Project. (2020, April 13). *Natural Language Toolkit—NLTK 3.5 documentation*. https://www.nltk.org/

Norwich University Online. (2017, February 1). *Understanding the Just-In-Time Inventory System*. https://online.norwich.edu/academic-programs/resources/understanding-just-time-inventory-system

Oak Ridge National Laboratory. (2017). *Freight Analysis Framework*. For FHWA Freight Management and Operations. Retrieved July 31, 2019, from https://ops.fhwa.dot.gov/freight/freight_analysis/faf/

Olmer, N., B. Comer, B. Roy, X. Mao and D. Rutherford (2017, October). Greenhouse gas emissions from global shipping, 2013-2015. The International Council on Clean Transportation.

Outwater, M., Smith, C., Wies, K., Yoder, S., Sana, B., & Chen, J. (2013). Tour based and supply chain modeling for freight: Integrated model demonstration in Chicago. *Transportation Letters*, *5*(2), 55–66. https://doi.org/10.1179/1942786713Z.0000000009

Palermo, N. (2013, August). New Car Delivery or Destination Charges Explained. Autotrader. https://www.autotrader.com/car-tips/new-car-delivery-or-destination-charges-explained-213280. Accessed April 20, 2018.

Paleti, R., Bhat, C., & Pendyala, R. (2013). Integrated model of residential location, work location, vehicle ownership, and commute tour characteristics. *Transportation Research Record*, *2382*, 162–172. https://doi.org/10.3141/2382-18

Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, *10*(1), 439–446. https://doi.org/10.32614/RJ-2018-009

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pennington, J., Socher, R., & Manning, C. D. (2014). *GloVe: Global Vectors for Word Representation*. https://nlp.stanford.edu/projects/glove/

Phillips 66. (2020). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Phillips_66&oldid=965337879. Accessed Augut 4, 2020.

Porter, M. (1980). *Competitive Strategy: Vol. (Citied in Wikipedia, accessed at https://en.wikipedia.org/wiki/Strategic_management on July 15, 2020)*. Free Press.

Punel, A., & Stathopoulos, A. (2018, January 7). *Exploratory Analysis of Crowdsourced Delivery Service Through a Stated Preference Experiment*. Transportation Research Board Annual Meeting, Washington, D.C.

Rashidi, T. H., Auld, J., & Mohammadian, A. (Kouros). (2012). A behavioral housing search model: Two-stage hazard-based and multinomial logit approach to choice-set formation and location selection. *Transportation Research Part A: Policy and Practice*, *46*(7), 1097–1107.

Rashidi, T. H., & Roorda, M. J. (2018). A business establishment fleet ownership and composition model. *Transportation*, *45*(3), 971–987.

Rehurek, R. (2020). *gensim: Topic modelling for humans* [Python3, C.]. https://radimrehurek.com/gensim/models/word2vec.html

Resource Systems Group, University of Illinois at Chicago, & Bowman, J. (2012). *Tour-based and Supply Chain Freight Forecasting Framework: Final Report*. FHWA BAA DTFH61-10-R-00013.

Rich, N. (1995). *Supply stream 'responsiveness' project. Quoted in Peter Hines, Nick Rich, (1997),"The seven value stream mapping tools", International Journal of Operations & Production Management, Vol. 17 Iss: 1 pp. 46–64.* (Supply Chain Development Programme I Workshop #6). Tesco Stores Limited.

Richardson, L. (2020). *beautifulsoup4: Screen-scraping library* (4.9.1) [Python]. Retrieved July 19, 2020, from http://www.crummy.com/software/BeautifulSoup/bs4/

Rodrigue, J.-P. (2020). *The Geography of Transport Systems*. Hofstra University, Department of Global Studies & Geography. https://transportgeography.org

Roorda, M. J., Cavalcante, R., McCabe, S., & Kwan, H. (2010). A conceptual framework for agent-based modelling of logistics services. *Transportation Research Part E: Logistics and Transportation Review*, *46*(1), 18–31. https://doi.org/10.1016/j.tre.2009.06.002

Rong, X. (2014). *Word2vec Parameter Learning* Explained. arXiv:1411.2738.

Rosseel, I. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1–36.

Rumelt, R. (2011). *Good strategy/bad strategy: The difference and why it matters (quoted in David J. Teece (2019) A capability theory of the firm: An economics and (Strategic) management perspective, New Zealand Economic Papers, 53:1, 1-43)*. Crown Business.

Sadeghi, A., Talan, D. M., & Clayton, R. (2016). *Establishment, firm, or enterprise: Does the unit of analysis matter?* U.S. Bureau of Labor Statistics: Monthly Labor Review. Retrieved August 4, 2020, from https://www.bls.gov/opub/mlr/2016/article/establishment-firm-or-enterprise.htm

Sakai, T., Romano Alho, A., Bhavathrathan, B. K., Chiara, G. D., Gopalakrishnan, R., Jing, P., Hyodo, T., Cheah, L., & Ben-Akiva, M. (2020). SimMobility Freight: An agent-based urban freight simulator for evaluating logistics solutions. *Transportation Research Part E: Logistics and Transportation Review*, *141*, 102017. https://doi.org/10.1016/j.tre.2020.102017

Samimi, A., Mohammadian, A. (Kouros), & Kawamura, K. (2013). A behavioral freight movement microsimulation model: Method and data. *Transportation Letters*, *2*(1), 53–62.

Sea-Distances.org (2018). https://sea-distances.org. Accessed 4/15/2018.

Shabani, K., Outwater, M., & Murray, D. (2018). *Behavioral/Agent-Based Supply Chain Modeling Research Synthesis and Guide* (FHWA-HOP-18-003). Resource Systems Group, American Transportation Research Institute, and Kittelson & Associates, Inc. for Federal Highway Administration.

Shapiro, C. (1989). The Theory of Business Strategy. *The RAND Journal of Economics*, *20*(1), 125–137.

Sharma & Associates, Inc. (2013). Market Analysis: Validation of a 70-ton Higher Speed Freight Truck Design for Operations up to 125 MPH. Publication DOT/FRA/ORD-13/32. FRA, US Department of Transportatio.

Shlens, J. (2005). *A Tutorial on Principal Component Analysis*. Accessed at http://www.cs.cmu.edu/~elaw/papers/pca.pdf in July, 2016.

Sillaparcharn, P. (2007). Vehicle Ownership and Trip Generation Modelling: A Case Study of Thailand. *IATSS Research*, *31*(2), 17–26. https://doi.org/10.1016/S0386-1112(14)60218-1

Socher, R. (2017, Winter). *Course Notes, CS224d: Deep Learning for Natural Language Processing*. Accessed August 4, 2020 at https://cs224d.stanford.edu/index.html.

Souten, D. (2004, Oct. 24-26). Trucks, Trains, Ships, and Planes: An Update on Goods Movement-Related Emissions. Presented at Linking Goods Movement to Prosperity and Environmental Quality, Lake Arrowhead, California.

Statistica (2018). Light vehicle retail sales in the United States from 1977 to 2017 (in 1,000 units). https://www.statista.com/statistics/199983/us-vehicle-sales-since-1951/. Accessed July 27, 2018.

Stinson, M. A. (2016). *A Proposed Agent-Based Freight Modeling Framework for SimMobility's Mid-Term and Long-Term*. Working document shared with the SMART SimMobility Freight Modeling Team and MIT EAR Innovative Freight Data Collection Team.

Stinson, M., Auld, J., Mohammadian, A., & Rousseau, A. (2018, October 17). *An Agent-based Model of Freight Transportation with Emerging Trends in POLARIS*. 3rd VREF Conference on Urban Freight, Gothenburg, Sweden.

Stinson, M., Auld, J., & Mohammadian, A. (Kouros). (2020, April 6). *A large-scale, agent-based simulation of metropolitan freight movements with passenger and freight market interactions*. 9th International Workshop on Agent-based Mobility, Traffic and Transportation Models, Methodologies and Applications (ABMTRANS), Warsaw, Poland.

Stinson, M., Enam, A., Moore, A., & Auld, J. (2019). Citywide Impacts of E-Commerce: Does Parcel Delivery Travel Outweigh Household Shopping Travel Reductions? *Proceedings of the Association for Computing Machinery - Smart Cities Conference in Portland, Oregon*.

Stinson, M., Lin, J., & Mohammadian, A. (Kouros). (2019, January 13). *Linking Business Strategy to Transportation Energy Consumption and Emissions Estimates: Effects of Automobile Assembly Location Strategies*. Transportation Research Board Annual Meeting, Washington D.C.

Stinson, M., & Mohammadian, A. (Kouros). (2020, November 3). Modeling Firm Transportation Strategy using Big Text Data. *Forum on Integrated and Sustainable Transportation Systems*. Forum on Integrated and Sustainable Transportation Systems, Delft, The Netherlands (Virtual).

Stinson, M., Pourabdollahi, Z., Livshits, V., Jeon, K., Nippani, S., & Zhu, H. (2017). A joint model of mode and shipment size choice using the first generation of Commodity Flow Survey Public Use Microdata. *International Journal of Transportation Science and Technology*, *6*(4), 330–343.

Strategic Management Journal. Wiley Online Library. https://doi.org/10.1002/(ISSN)1097-0266

Subaru (2018). Corporate Profile: Facilities. https://www.subaru.co.jp/en/outline/office.html. Accessed April 5, 2018.

Swait, J. D., & Ben-Akiva, M. (1986). Constraints on Individual Travel Behavior in a Brazilian City. *Transportation Research Record*, *1085*. https://trid.trb.org/view/283761

Tata Motors (2017, March 29). *Tata Motors*. http://www.tata.com/company/profileinside/Tata-Motors. Accessed April 5, 2018.

Tavasszy, L., van der Vlist, M. J. M., & Ruijgrok, C. J. (1998, July 12). *Scenario-wise analysis of transport and logistics systems with a SMILE*. 8th WCTR Conference, Antwerp, Belgium.

Teece, D. J. (2019). A capability theory of the firm: An economics and (Strategic) management perspective. *New Zealand Economic Papers*, *53*(1), 1–43.

Tesla Factory (2018). Wikipedia. Accessed https://en.wikipedia.org/wiki/Tesla_Factory. Accessed April 5, 2018.

Thaller, C., Dahmen, B., Liedtke, G., & Friedrich, H. (2016). Freight Transport Demand Modelling. In U. Clausen, H. Friedrich, C. Thaller, & C. Geiger (Eds.), *Commercial Transport* (pp. 39–54). Springer International Publishing.

Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables. *Econometrica*, *26*(1), 24.

Toyota-Global Newsroom (2018). Worldwide Operations: Overseas Manufacturing Companies. Toyotanews.pressroom.toyota.com, Accessed April 5, 2018.

Train, K. E. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press.

Transportation Research Board of the National Academies. (2003). *A Concept for a National Freight Data Program: Special Report 276*.

Truckers Report Forum (2006). How much does the tractor weight? https://www.thetruckersreport.com/truckingindustryforum/threads/how-much-does-the-tractor-weigh.13716/. Accessed April 10, 2018.

Tsui, C. (2018, Jan. 4). Selling 58 percent more cars year-over-year, Aston sales exceeded 5,000 cars for the first time since 2008. The Drive.

Turner, K. (2015, Sept. 21). Subaru to add 1,200 jobs in $140M expansion. Indystar.

Union of Concerned Scientists. (2020). Published Jul 16, 2008 and Updated May 11, 2020. *Each Country's Share of CO2 Emissions*.

Urban, M. S., Beagan, D., & Fischer, M. (2012, May). *A Working Demonstration of a Mesoscale Freight Model for the Chicago Region*. 4th Transportation Research Board Conference on Innovations in Travel Modeling, Tampa, Florida.

US Census Bureau. (2020a). *Commodity Flow Survey (CFS): Technical Documentation*. 2017 Commodity Flow Survey Overview and Methodology.

US Census Bureau. (2020b). *Data*. Census.Gov. Retrieved July 21, 2020, from https://www.census.gov/data

US Census Bureau. (2020c). *County Business Patterns*. Census.gov: Our Surveys & Programs. Accessed September 2019 at https://www.census.gov/programs-surveys/cbp.html.

US Census Bureau (2015). 2012 Commodity Flow Survey (CFS) Public Use Microdata (PUM) File Data Users Guide: Technical Documentation.

US Department of Transportation. (2017). *Beyond Traffic: 2045 Final Report*.

US Census Bureau (2018). US Census America Fact Finder Tool. https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml. Accessed April 20, 2018.

USDOT Bureau of Transportation Statistics (2020). *Geospatial Data: National Transportation Atlas Database (NTAD)*. Retrieved March 31, 2020, from https://data-usdot.opendata.arcgis.com

USDOT and US Department of Commerce (2017). 2017 Commodity Flow Survey Standard Classification of Transported Goods. Accessed July 25, 2018 at https://www.gov/econ/cfs/2017/CFS-1200_17.pdf.

US Environmental Protection Agency (2009). Emission Factors for Locomotives. Publication EPA-420-F-09-025.

US Environmental Protection Agency (2014). MOVES and Other Mobile Source Emissions Models. Accessed April 5, 2018 at https://www.epa.gov/moves.

US Securities and Exchange Commission (SEC). (2009). *Form 10-K*. Retrieved July 21, 2020, from https://www.sec.gov/fast-answers/answers-form10khtm.html

United States Securities and Exchange Commission (2020). *SEC.gov Company Search Page*. EDGAR: Company Filings. Retrieved January 18, 2020, from https://www.sec.gov/edgar/searchedgar/companysearch.html

Vij, A., & Walker, J. L. (2016). How, when and why integrated choice and latent variable models are latently useful. *Transportation Research Part B: Methodological*, *90*, 192–217.

Walker, J. L. (2001). *Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables* [PhD Dissertation]. Massachusetts Institute of Technology.

Waters, D. (2001). Inventory Management. In *Handbook of Logistics and Supply Chain Management, Edited by A.M. Brewer et al.*

Wisconsin Department of Transportation. (2017). *Comparing the Models: Insights from Freight Behavior Data Improve Wisconsin's Freight Models* (FHWA-HOP-17-035; Freight Demand Modeling and Data Improvement Implementation Support). https://ops.fhwa.dot.gov/freight/freight_analysis/fdmdi/index.htm

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data* (2nd ed.). The MIT Press.

Xu, J., Hancock, K., & Southworth, F. (2003). Simulation of regional freight movement with trade and transportation multi-networks. *Transportation Research Record: Journal of the Transportation Research Board*, *1854*, 152–161.

**Publisher's Policy On Use Of Published Materials (For Chapter 3.4)**

Link to license: https://creativecommons.org/licenses/by-nc-nd/4.0/

Creative Commons — Attribution-NonCommercial-NoDerivatives 4.0 International — CC BY-NC-ND 4.0 9/11/20, 12:32 PM

This page is available in the following languages:

# Creative Commons License Deed

**Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)**

This is a human-readable summary of (and not a substitute for) the license.

## You are free to:

**Share** — copy and redistribute the material in any medium or format

The licensor cannot revoke these freedoms as long as you follow the license terms.

## Under the following terms:

**Attribution** — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

**NonCommercial** — You may not use the material for commercial purposes.

**NoDerivatives** — If you remix, transform, or build upon the material, you may not distribute the modified material.

**No additional restrictions** — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

## Notices:

You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation.

No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as publicity, privacy, or moral rights may limit how you use the material.

# Publisher's Policy On Use Of Published Materials (For Chapter 4.2.1, 4.2.4)

Apt #: 161121

## IEEE COPYRIGHT AND CONSENT FORM

To ensure uniformity of treatment among all contributors, other forms may not be substituted for this form, nor may any wording of the form be changed. This form is intended for original material submitted to the IEEE and must accompany any such material in order to be published by the IEEE. Please read the form carefully and keep a copy for your files.

**TITLE OF PAPER/ARTICLE/REPORT, INCLUDING ALL CONTENT IN ANY FORM, FORMAT, OR MEDIA (hereinafter, "the Work"):** Modeling Firm Transportation Strategy using Big Text Data

**COMPLETE LIST OF AUTHORS:** Monique Stinson*, Abolfazl (Kouros) Mohammadian

**IEEE PUBLICATION TITLE (Journal, Magazine, Conference, Book):** 2020 Forum on Integrated and Sustainable Transportation System

### COPYRIGHT TRANSFER

1. The undersigned hereby assigns to The Institute of Electrical and Electronics Engineers, Incorporated (the "IEEE") all rights under copyright that may exist in and to: (a) the above Work, including any revised or expanded derivative works submitted to the IEEE by the undersigned based on the Work; and (b) any associated written or multimedia components or other enhancements accompanying the Work.

### CONSENT AND RELEASE

** 2. In the event the undersigned makes a presentation based upon the Work at a conference hosted or sponsored in whole or in part by the IEEE, the undersigned, in consideration for his/her participation in the conference, hereby grants the IEEE the unlimited, worldwide, irrevocable permission to use, distribute, publish, license, exhibit, record, digitize, broadcast, reproduce and archive, in any format or medium, whether now known or hereafter developed: (a) his/her presentation and comments at the conference; (b) any written materials or multimedia files used in connection with his/her presentation; and (c) any recorded interviews of him/her (collectively, the "Presentation"). The permission granted includes the transcription and reproduction of the Presentation for inclusion in products sold or distributed by IEEE and live or recorded broadcast of the Presentation during or after the conference.

** 3. In connection with the permission granted in Section 2, the undersigned hereby grants IEEE the unlimited, worldwide, irrevocable right to use his/her name, picture, likeness, voice and biographical information as part of the advertisement, distribution and sale of products incorporating the Work or Presentation, and releases IEEE from any claim based on right of privacy or publicity.

4. The undersigned hereby warrants that the Work and Presentation (collectively, the "Materials") are original and that he/she is the author of the Materials. To the extent the Materials incorporate text passages, figures, data or other material from the works of others, the undersigned has obtained any necessary permissions. Where necessary, the undersigned has obtained all third party permissions and consents to grant the license above and has provided copies of such permissions and consents to IEEE.

☐ Please check this box if you do not wish to have video/audio recordings made of your conference presentation.
See reverse side for Retained Rights/Terms and Conditions, and Author Responsibilities.

Monique A. Stinson — July 27, 2020

** Presenter signs for 2 & 3 — Date

### GENERAL TERMS

- The undersigned represents that he/she has the power and authority to make and execute this assignment.
- ~~The undersigned agrees to indemnify and hold harmless the IEEE from any damage or expense that may arise in the event of a breach of any of the warranties set forth above.~~
- In the event the above work is not accepted and published by the IEEE or is withdrawn by the author(s) before acceptance by the IEEE, the foregoing copyright transfer shall become null and void and all materials embodying the Work submitted to the IEEE will be destroyed.
- For jointly authored Works, all joint authors should sign, or one of the authors should sign as authorized agent for the others.

(1) * Yvette Woell by Nicholas Lundvick — July 27, 2020

**Author/Authorized Agent for Joint Authors** — **Date**

Nicholas Lundvick
Scholarly Communications Librarian
for UChicago Argonne, LLC, Operator of Argonne National Laboratory

### U.S. GOVERNMENT EMPLOYEE CERTIFICATION (WHERE APPLICABLE)

This will certify that all authors of the Work are U.S. government employees and prepared the Work on a subject within the scope of their official duties. As such, the Work is not subject to U.S. copyright protection.

(2)____________________

**Authorized Signature** — **Date**

(Authors who are U.S. government employees should also sign signature line (1) above to enable the IEEE to claim and protect its copyright in international jurisdictions.)

### CROWN COPYRIGHT CERTIFICATION (WHERE APPLICABLE)

This will certify that all authors of the Work are employees of the British or British Commonwealth Government and prepared the Work in connection with their official duties. As such, the Work is subject to Crown Copyright and is not assigned to the IEEE as set forth in the first sentence of the Copyright Transfer Section above. The undersigned acknowledges, however, that the IEEE has the right to publish, distribute and reprint the Work in all forms and media.

(3)____________________

**Authorized Signature** — **Date**

(Authors who are British or British Commonwealth Government employees should also sign line (1) above to indicate their acceptance of all terms other than the copyright transfer.)

*rev. 032515*

* Signed on behalf of Argonne employees only
(Stinson)

# IEEE COPYRIGHT FORM *(continued)*

## RETAINED RIGHTS/TERMS AND CONDITIONS

### General

**1.** Authors/employers retain all proprietary rights in any process, procedure, or article of manufacture described in the Work.

**2.** Authors/employers may reproduce or authorize others to reproduce the Work, material extracted verbatim from the Work, or derivative works for the author's personal use or for company use, provided that the source and the IEEE copyright notice are indicated, the copies are not used in any way that implies IEEE endorsement of a product or service of any employer, and the copies themselves are not offered for sale.

**3.** In the case of a Work performed under a U.S. Government contract or grant, the IEEE recognizes that the U.S. Government has royalty-free permission to reproduce all or portions of the Work, and to authorize others to do so, for official U.S. Government purposes only, if the contract/grant so requires.

**4.** Although authors are permitted to re-use all or portions of the Work in other works, this does not include granting third-party requests for reprinting, republishing, or other types of re-use. The IEEE Intellectual Property Rights office must handle all such third-party requests.

**5.** Authors whose work was performed under a grant from a government funding agency are free to fulfill any deposit mandates from that funding agency.

### Author Online Use

**6. Personal Servers.** Authors and/or their employers shall have the right to post the accepted version of IEEE-copyrighted articles on their own personal servers or the servers of their institutions or employers without permission from IEEE, provided that the posted version includes a prominently displayed IEEE copyright notice and, when published, a full citation to the original IEEE publication, including a link to the article abstract in IEEE Xplore. Authors shall not post the final, published versions of their papers.

**7. Classroom or Internal Training Use.** An author is expressly permitted to post any portion of the accepted version of his/her own IEEE-copyrighted articles on the author's personal web site or the servers of the author's institution or company in connection with the author's teaching, training, or work responsibilities, provided that the appropriate copyright, credit, and reuse notices appear prominently with the posted material. Examples of permitted uses are lecture materials, course packs, e-reserves, conference presentations, or in-house training courses.

**8. Electronic Preprints.** Before submitting an article to an IEEE publication, authors frequently post their manuscripts to their own web site, their employer's site, or to another server that invites constructive comment from colleagues. Upon submission of an article to IEEE, an author is required to transfer copyright in the article to IEEE, and the author must update any previously posted version of the article with a prominently displayed IEEE copyright notice. Upon publication of an article by the IEEE, the author must replace any previously posted electronic versions of the article with either (1) the full citation to the IEEE work with a Digital Object Identifier (DOI) or link to the article abstract in IEEE Xplore, or (2) the accepted version only (not the IEEE-published version), including the IEEE copyright notice and full citation, with a link to the final, published article in IEEE Xplore.

## INFORMATION FOR AUTHORS

### Author Responsibilities

The IEEE distributes its technical publications throughout the world and wants to ensure that the material submitted to its publications is properly available to the readership of those publications. Authors must ensure that their Work meets the requirements as stated in section 8.2.1 of the IEEE PSPB Operations Manual, including provisions covering originality, authorship, author responsibilities and author misconduct. More information on IEEE's publishing policies may be found at http://www.ieee.org/publications_standards/publications/rights/authorrightsresponsibilities.html. Authors are advised especially of IEEE PSPB Operations Manual section 8.2.1.B12: "It is the responsibility of the authors, not the IEEE, to determine whether disclosure of their material requires the prior consent of other parties and, if so, to obtain it." Authors are also advised of IEEE PSPB Operations Manual section 8.1.1B: "Statements and opinions given in work published by the IEEE are the expression of the authors."

### Author/Employer Rights

If you are employed and prepared the Work on a subject within the scope of your employment, the copyright in the Work belongs to your employer as a work-for-hire. In that case, the IEEE assumes that when you sign this Form, you are authorized to do so by your employer and that your employer has consented to the transfer of copyright, to the representation and warranty of publication rights, and to all other terms and conditions of this Form. If such authorization and consent has not been given to you, an authorized representative of your employer should sign this Form as the Author.

### IEEE Copyright Ownership

It is the formal policy of the IEEE to own the copyrights to all copyrightable material in its technical publications and to the individual contributions contained therein, in order to protect the interests of the IEEE, its authors and their employers, and, at the same time, to facilitate the appropriate re-use of this material by others. The IEEE distributes its technical publications throughout the world and does so by various means such as hard copy, microfiche, microfilm, and electronic media. It also abstracts and may translate its publications, and articles contained therein, for inclusion in various compendiums, collective works, databases and similar publications.

**THIS FORM MUST ACCOMPANY THE SUBMISSION OF THE AUTHOR'S MANUSCRIPT.**
**Questions about the submission of the form or manuscript must be sent to the publication's editor.**
**Please direct all questions about IEEE copyright policy to:**
**IEEE Intellectual Property Rights Office, copyrights@ieee.org, +1-732-562-3966 (telephone)**

# Curriculum Vitae

## MONIQUE STINSON, PHD

Department of Civil and Materials Engineering
UIC Translab
University of Illinois at Chicago
842 West Taylor Street, MC 246
Chicago, IL 60607
Phone: (312) 996-3428
mstins2@uic.edu

Vehicle and Mobility Systems Group
Argonne National Laboratory
9700 South Cass Avenue, Building 222
Lemont, Illinois 60439
Phone: (630) 252-6562
mstinson@anl.gov

Home:
Chicago, Illinois 60613
Phone: (773) 853-6707
quest16@gmail.com

## Education

**University of Illinois at Chicago** (2018-2020), Doctor of Philosophy in Civil and Materials Engineering
Thesis*: Strategic Decisions in Agent-Based Freight Transportation Models: Methods and Data*

**Massachusetts Institute of Technology** (2015-2017), graduate coursework

**University of Texas at Austin** (2000-2002), Master of Science in Civil Engineering
*Thesis: A Discrete Choice Exploration of the Route Choice Preferences of Commuter Bicyclists*

**University of Texas at Austin** (1996-2000), Bachelor of Science in Civil Engineering

**Pennsylvania State University** (1995-1996), undergraduate coursework

## Experience

**Computational Transportation Scientist, Argonne National Laboratory**, 2018-present

**Research Assistant, Massachusetts Institute of Technology**, 2015-2018

**Senior Planner, RS&H, Inc.,** 2014-2015

**Senior Associate, Cambridge Systematics, Inc.,** 2004-2014

**Transportation Management Cadre, Chicago Area Transportation Study,** 2002-2004

**Research Assistant, University of Texas at Austin**, 2000-2002

## Publications in Peer-Reviewed Journals and Conference Proceedings

1. **Stinson, M.**, and A. (Kouros) Mohammadian. *Modeling Firm Transportation Strategy using Big Text Data*. IEEE Forum on Integrated and Sustainable Transportation Systems (Forum ISTS 2020), Nov. 3-5, 2020, in Delft, The Netherlands (accepted for presentation and publication in proceedings).
2. **Stinson, M.,** J. Auld and A. (Kouros) Mohammadian**.** *A large-scale, agent-based simulation of metropolitan freight movements with passenger and freight market interactions*. Procedia Computer Science, pp. 771-778 Proceedings of the 9th International Workshop on Agent-based Mobility, Traffic and Transportation Models, Methodologies and Applications (ABMTRANS-20), (originally scheduled for April 6 - 9, 2020, Warsaw, Poland).
3. **Stinson, M.A.**, J. Auld and A. (Kouros) Mohammadian. *Light Duty Vehicle Choice Models Examining Alternative Fuel Technology Preferences among Commercial Fleet Owners*. Transportation Research Procedia, Volume 46, 2020, Pages 309-316.
4. **Stinson, M.,** A. Enam, A. Moore, and J. Auld. *Citywide Impacts of E-Commerce: Does Parcel Delivery Travel Outweigh Household Shopping Travel Reductions?* Proceedings of the Association for Computing Machinery - Smart Cities Conference in Portland, Oregon. September 2019.

5. Alho, A., Bhavathrathan, B. K., **Stinson, M.**, Gopalakrishnan, R., Le, D. T., & Ben-Akiva, M. (2017). *A multi-scale agent-based modelling framework for urban freight distribution*. Transportation Research Procedia, 27, 188-196.
6. **Stinson, M.A.,** Z. Pourabdollahi, V. Livshits, K. Jeon, S. Nippani and H. Zhu. *A Joint Model of Mode and Shipment Size Choice Using the First Generation of CFS Public Use Microdata*. International Journal of Transportation Science and Technology, Volume 6, Issue 4, pp. 330-343, 2017.
7. **Stinson, M.A.,** C. Porter, K. Proussaloglou, R. Calix and C. Chu. *Modeling the Impacts of Bicycle Facilities on Work and Recreational Bike Trips in Los Angeles County*, California, Transportation Research Record No. 2468, Washington D.C., 2014.
8. **Stinson, M.A.,** Bhat, C.R. *Frequency of Bicycle Commuting: Internet-Based Survey Analysis*, Transportation Research Record No. 1878, Washington D.C., 2004.
9. **Stinson, M.A.,** Bhat, C.R. *An Analysis of Commuter Bicyclist Route Choice Using a Stated-Preference Survey*, Transportation Research Record No. 1828, Washington D.C., 2003.

## Submitted Manuscripts

10. **Stinson, M.**, and A. (Kouros) Mohammadian. *Two Algorithms for Measuring Attitudes using Natural Language* (in submission to Transportation Research Board 2021 Annual Meeting), virtual.
11. **Stinson, M.**, and A. (Kouros) Mohammadian. *A Behavioral Model for Strategic Freight Transportation Decisions* (in submission to Transportation Research Board 2021 Annual Meeting), virtual.
12. Sahin, O. and **Stinson, M.**. *Analysis of Urban Freight Flows and Last-Mile, Retail Goods Movement Over Time Using Inference with GPS Trajectory Data* (in submission to Transportation Research Board 2021 Annual Meeting), virtual.
13. **Stinson, M.,** B. Zou, A. Mohammadian, D. Briones, and A. Manjarrez. *Vehicle Ownership Models for a Sharing Economy with Autonomous Vehicle Considerations*. Under review by journal.

## Book Chapters

14. Enam, A. with F. de Souza, O. Verbas, **M. Stinson** and J. Auld. Autonomous Vehicles and Transportation Modeling. *Encyclopedia of Transportation*, 12. August 2019
15. Cheah, L., F. Zhao, **M. Stinson**, F. Lu, J. Ding-Mastera, V. Marzano, M. Ben-Akiva (2018). Next-Generation Commodity Flow Survey: A Pilot in Singapore: Modeling and Planning Initiatives. In E. Taniguchi and R. Thompson (Eds.) *City Logistics 2: Modeling and Planning Initiatives*. Wiley-ISTE, London.
16. **Urban, M. [Stinson]** & Petersen, E. (2014). Transit fares. In M. Garrett (Ed.), *Encyclopedia of transportation: Social science and policy* (Vol. 1, pp. 1436-1439). Thousand Oaks, CA: SAGE Publications, Inc.

## Selected Consultant Technical Reports

17. RS&H, Inc. (Project Manager). *Mega-Regional Multi-Modal Agent-Based Behavioral Freight Model: MAG Next Generation Freight Demand Model*, Prepared for the Maricopa Association of Governments with assistance from Federal Highway Administration (FHWA) SHRP2 C20 IAP Grant, 2018.
18. RS&H, Inc. (Lead Author). *SWOT Analysis of TRANSEARCH And FAF Data*. Project No.: TRANSTAT TWO 7, CONTRACT NO.: C9B48, 2016.
19. Cambridge Systematics, Inc. (Project Manager). *An Interagency Fare Choice Model for the Chicago Region*, prepared for the Regional Transportation Authority, 2013.
20. Cambridge Systematics, Inc. (Project Manager). *CTA Ridership and Revenue Model: Draft Final Report*, prepared for the Chicago Transit Authority, 2012.
21. Cambridge Systematics, Inc. (Deputy Project Manager). *A Working Demonstration of a Mesoscale Freight Model for the Chicago Region: Final Report and User's Guide*. Chicago Metropolitan Agency for Planning, 2011.
22. Cambridge Systematics, Inc. (Contributor). *Bicycle Model Development–Phase I*, prepared for the Los Angeles County Metropolitan Transportation Authority, 2011.

23. Cambridge Systematics, Inc. (Deputy Project Manager), *Illiana Expressway Economic Opportunities Analysis: Final Report*, prepared for the Illinois Department of Transportation, 2010.
24. Cambridge Systematics, Inc. (Contributor). *Regional Freight System Planning Recommendations Study: Final Report*, prepared for the Chicago Metropolitan Agency for Planning (CMAP), 2010.
25. Cambridge Systematics, Inc. (Contributor). *I-95/I-395 Transit/TDM Study: Final Report*, prepared for the I-95/I-395 Transit/TDM Technical Advisory Committee led by Virginia Dept. of Rail and Public Transportation, 2008.
26. Cambridge Systematics, Inc. (Contributor) *San Joaquin Valley (SJV) Goods Movement Study: Phase II*, prepared for the Kern Council of Governments and presented to the Council of Fresno County Governments, 2007.
27. Cambridge Systematics, Inc. (Contributor). *Market Analysis and Service Planning Tool*, prepared for Pace Suburban Bus, 2006.
28. Cambridge Systematics, Inc. (Contributor). *Mid-City Freightway: Evaluation of Alternative Alignments and Tolls*, prepared for the Chicago Department of Transportation, 2006.
29. Cambridge Systematics, Inc. (contributor). *MnPass System Study: Final Report*, prepared for the Minnesota DOT, 2005.
30. Cambridge Systematics, Inc. (Contributor). *Cook-DuPage Corridor Travel Market Analysis Study: Existing and Future Conditions: Final Report*, 2005.

## Selected Presentations

31. **Stinson, M.** E. Rahimi and A. Mohammadian**.** *A Carrier Load Pricing Model with Seasonality and Forward vs. Back-Haul Pricing Differences***.** METRANS International Urban Freight Conference (I-NUF) in Long Beach, California**,** October 16-18, 2019.
32. **Stinson, M.A.** with A. Enam and A. Mohammadian. *Supply Impacts of Disruptive Technologies: Vehicle Ownership as an Investment Decision*. International Choice Modeling Conference, Kobe, Japan, August 19-21, 2019.
33. **Stinson, M.**, B. Pandey, A. Enam, A. Rousseau, J. Auld. *Spatiotemporal Analysis of the Freight Analysis Framework Data*. Innovations in Freight Data Workshop, April 9-10, 2019, Irvine, CA.
34. **Stinson, M.,** J. Lin, A. Mohammadian. *Linking Business Strategy to Transportation Energy Consumption and Emissions Estimates: Effects of Automobile Assembly Location Strategies*. Transportation Research Board Annual Meeting, Washington D.C., Jan. 13-17, 2019.
35. **Stinson, M.**, J. Auld and A. Mohammadian. *An Agent-based Model of Freight Transportation with Emerging Trends in POLARIS.* 3rd VREF Conference on Urban Freight, Gothenburg, Sweden, October 17-19, 2018.
36. Ding-Mastera, J., P. Jing, **M. Stinson**, E. Manzi, F. Zhao, V. Marzano, L. Cheah, M. Ben-Akiva. *Combining GPS Data Collection with Assisted Machine Learning to Enhance Freight Vehicle and Driver Surveys: Methodology and Demonstration*. Transportation Research Board Annual Meeting, Washington D.C., Jan. 13-17, 2018.
37. Cheah, L. **M. Stinson,** Z. Chen, V. Marzano, F. Zhao, M. Ben-Akiva. *Future Freight Surveys of Shipments*. 7$^{th}$ METRANS International Urban Freight Conference (I-NUF), Long Beach, California, October 17-20, 2017.
38. Lee, Y. and **M. Stinson**. *Modeling Urban Freight Vehicle Choice in Retail Supply Chains*. Transportation Research Board Annual Meeting, Washington D.C., Jan. 8-12, 2017.
39. **Stinson, M.,** et al. *An Integrated Approach to Freight Data Collection*, Future Urban Mobility Symposium 2016, Singapore, July 25-26, 2016.
40. **Stinson, M.A.** with Z. Pourabdollahi, R. Tillery and K. Zuehlke. *A Tour-Based Freight Model for the Tampa, Florida Metropolitan Region*. TRB Planning Applications Conference, Atlantic City, NJ, May 17-21, 2015.

## Invited Talks

1. **M. Stinson**. *E-Commerce and Parcel Delivery: VMT and Energy Consumption*. Chicago Metropolitan Agency for Planning Freight Advisory Committee Meeting, February 24, 2020.
2. **M. Stinson**. *Inter/Intracity Freight Movement Using Data-Driven Agent-Based System Simulation*. 21$^{st}$ Century Truck Program Freight Operations Efficiency Tech Team Meeting, Jan. 21, 2020.

3. **M. Stinson**. *Assessing the e-commerce effect: parcel delivery vs. household shopping*. Improving Last-Mile and 50-Feet Logistics with Smart Initiatives to Improve Freight Mobility Workshop, Transportation Research Board Annual Meeting, Washington D.C., Jan. 13-16, 2020.
4. **M. Stinson**. *E-Commerce Impacts on Regional Travel and Energy Use: Household Shopping and Parcel Delivery Tradeoffs*. FHWA Talking Freight Seminar, December 19, 2019.
5. **M. Stinson.** *Modeling Future Mobility Impacts Using the POLARIS Agent-based Transportation Simulator*. Society of Automotive Engineers Innovations in Mobility Conference, Novi, Michigan, Oct. 29-31, 2019.
6. **M. Stinson**. *CAV Modeling Structure and Planning Applications at Argonne National Laboratory*. CAV Expert Workshop, University of California-Davis, April 29-30, 2019.

## Professional Leadership

**Paper Review Coordinator**, Freight Transportation Planning and Logistics Committee, Transportation Research Board, 2/2020-present (**Member** since 3/10/2017)

**Member**, Freight Data Committee, Transportation Research Board, 02/2020 - present

**Co-Chair**, Transportation Students Group Center for Transportation and Logistics Distinguished Speaker Seminar Series, 2017

**Member**, Bicycle Transportation Committee, Transportation Research Board, 2003-2004

**Chairperson**, Transport Chicago Conference, 2003

**Poster Session Coordinator**, Transport Chicago Conference, 2004-2005

**President**, Institute of Transportation Engineers (ITE) Student Chapter (The University of Texas at Austin), 2000-2001

## Honors and Awards

***Best Paper Award*** for "A large-scale, agent-based simulation of metropolitan freight movements with passenger and freight market interactions", 9th International Workshop on Agent-based Mobility, Traffic and Transportation Models, Methodologies and Applications (ABMTRANS-20)

***Full Tuition Waiver***, University of Illinois at Chicago (2018-2020)

***David Boyce Graduate Scholarship*** (2019), University of Illinois at Chicago

***Women's Transportation Seminar-Boston Chapter Ann Hershfang Graduate Scholarship*** (2018-2019)

***"Best Data Fusion Application"*** (2017) by the *Transportation Research Board (TRB) Innovations in Freight Data Workshop* Planning Committee, Team Member with Maricopa Association of Governments

***Full Tuition Waiver*** and ***Graduate Stipend***, Massachusetts Institute of Technology (2015-2017)

***2nd Place, Student Paper Competition***, Transport Chicago Conference (2003)

***Full Tuition Waiver*** and ***Graduate Stipend***, The University of Texas at Austin (2000-2002)

***Huff Endowed Presidential Fellowship*** (2000-2002)

***College of Engineering Advanced Institute Graduate Fellowship*** (2000-2002)

***Engineering Honors Program*** (1997-2000)

***Edward and Rebecca Case Endowed Presidential Scholarship*** (1999-2000)

***H. H. Dalrymple Endowed Presidential Scholarship*** (1998-1999)