Title
Commodity Based Freight Demand Modeling Framework using Structural Regression Model

Author
SUN, YUE

Publication Date
2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Commodity Based Freight Demand Modeling Framework using Structural Regression Model

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Civil and Environmental Engineering

by

Yue Sun

Dissertation Committee:
Professor Stephen G. Ritchie, Chair
Professor Michael G. McNally
Professor David Brownstone

2018

DEDICATION

To

my beloved family

for their endless love, trust and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

It is an extremely memorable experience for me to study at UCI-ITS. At the moment when I finally completed my Ph.D. degree, I would like to express my deepest appreciation to the persons for providing me great encouragement and support.

First and foremost, I want to express my sincere gratitude to my advisor, Professor Stephen G. Ritchie, for choosing me as a trans-major student, for encouraging me to find out my research area, for providing me opportunities to learn different skills, for supporting me during my internship, and most significant, for guiding me to become a qualified doctor.

I am grateful for other two member of my committee, Professor Michael G. McNally and Professor David Brownstone, for their valuable suggestions and constructive comments on my dissertation. I would like to thank Professor Professor Will Recker, Professor R. Jayakrishnan, Professor Jean-Daniel M. Saphores and Professor Wenlong Jin for their teaching and instructions.

My appreciation is extended to Dr. James Marca for helping me a lot on data analysis and programming, Dr. Andre Tok for sharing the knowledge and hands-on experience on freight transportation planning and helping me revise the paper, and Dr. Fatemeh Ranaiefar for sharing useful information and data. I also want to thank other group members I worked with: Dr. Craig Ross Rindt, Dr. Suman Kumar Mitra, Dr. Kyung Hyun, Dr. Kyungsoo Jeong, Koti Reddy Allu, Youngeun, and Junhyeong Park. We are both colleagues and friends by sharing research thoughts and life experience.

I thank all my friends in ITS: Dr. Zhe Sun, Qinglong Yan, Xuting Wang and Chenying Qin for their kind friendship.

Many thanks go to my best friends Jielin Sun, Dr. Hao Yang, Dr. Qijian Gan, Dr. Shan Jiang, Nicole Mi, Rufei Huang, and two sweet couples, Dr. Xu Gao and Cong Zhang, Jin Yang and Lu Xu. I feel so lucky to meet and hold profound friendship with them. Their thoughtful concern on me and my family all help me reach where I am.

Last but not least, I would like to express my deep appreciation to my family. I greatly thank my parents, Lianyong Sun and Hongying Zhang, for raising me up and continuous encouragement during each stage of my life. My special thanks are offered to my wife, Biling Liu, for your determination being my life mate from the very beginning, for your companionship during the last 12 years, and for our best treasure you brought to me, our son Robin, who was just born during the final stage of my Ph.D study. Surrounded by all your love, I will never walk alone!

CURRICULUM VITAE

YUE SUN

EDUCATION

Ph.D. in Civil Engineering – Transportation System Engineering 2018

University of California, Irvine

M.S., Civil Engineering – Transportation System Engineering 2014

University of California, Irvine

B.S., Mathematics and Applied Mathematics 2012

Nankai University, Tianjin, China

AREA OF EXPERTISE

Transportation Planning, Travel Demand Modeling, Freight Forecasting Modeling, Statistical Data Analysis, Optimization, Machine Learning, Geographic Information System (GIS) Application

RESEARCH EXPERIENCE

- Research Assistant, *University of California, Irvine* 01/2017-03/2018

  California Natural Gas Vehicle Incentive Program (NGVIP)

  funded by California Energy Commission

- Research Assistant, *University of California, Irvine* 06/2016-10/2016

  Development of Truck Activity Monitoring System (TAMS) *funded by ARB and Caltrans*

- Research Assistant, *University of California, Irvine* 10/2015-06/2016

  State Rail Travel Forecasting and Economics Analysis *funded by Caltrans*

- Research Assistant, *University of California, Irvine* 08/2014-09/2015

  California Air Resources Board Vehicle Activity Database (CALVAD) *funded by ARB*

- Research Assistant, *University of California, Irvine* 09/2013-12/2013

  California Statewide Freight Forecasting Model (CSFFM) *funded by Caltrans*

PROFESSIONAL EXPERIENCE

- Engineering Intern, *FEHR and PEERS* 2017-2018
- Intern, Modeling and Forecasting Department, Southern California Association of Governments (SCAG) 2016-2017

PUBLICATIONS

- Tok, A., Hyun, K. K., Hernandez, S., Jeong, K., Sun, Y. E., Rindt, C., & Ritchie, S. G. (2017). Truck Activity Monitoring System (TAMS) for Freight Transportation Analysis (No. 17-06807).
- Research on New Commercial Pension Insurance Based on the Urban Working-class Assets Replacement. ISSN 1003-6040, Qi Lu Zhu Tan, 06/2011.

AWARDS

- Honorable Mention, Mathematical Contest in Modeling, 2011

TECHNICAL SKILLS

- Planning Software: TransCAD (GISDK), Cube (Voyager and Cargo), TransModeler, Synchro, ArcGIS, QGIS

- Database and Data Analysis: R, SQL, Matlab, SAS, SPSS, Stata, MS Access, Excel
- Programming Language: Python, C, C++, Java, JavaScript
- Knowledge of Data Source: PeMS, HPMS, AirSage, CHTS, Transearch, FAF, ATRI , VIUS, WIM, TAMS

ABSTRACT OF THE DISSERTATION

Commodity Based Freight Demand Modeling Framework using Structural Regression Model

By

Yue Sun

Doctor of Philosophy in Civil and Environmental Engineering

University of California, Irvine, 2018

Professor Stephen G. Ritchie, Chair

Among the main freight modeling approaches, commodity-based models stand out in their ability to incorporate all travel modes and capture the economic mechanisms driving freight movements. However, challenges still exist on the effective use of public freight data and the ability to accurately reflect the supply chain relationships between commodities. In this research, a commodity-based framework for freight demand forecasting using a Structural Regression Model (SRM) is explored, and applied to the original California Statewide Freight Forecasting Model (CSFFM) using the Freight Analysis Framework Version 4 (FAF4) data.

The framework developed in this study contains four innovative components: (1) mathematical approach for determining freight economic centroids; (2) the aggregation of commodities using the Fuzzy C-means clustering algorithm; (3) employing weighted travel distance by commodity group (CG) instead of highway skim to provide a more representative travel distance across multiple modes; and (4) the forecasting of freight demand using SRM method to comprehensively consider the direct effect, indirect effect and latent variables. The SRM is adopted in both the total generation model and domestic direct demand model. The

application results are further compared with the original CSFFM forecasts in 2012 to illustrate the advantages of the proposed framework.

1 INTRODUCTION

1.1 Research Motivation

With the increasing trend of goods movements, there is a growing need for effective freight demand forecasting tools to meet freight planning policy making needs due to the significant role of freight transport in the global and domestic economy, highway operation, travel safety, logistics management, pavement maintenance and emissions. In addition, freight movement is multi-modal: it is not only limited to truck-based movements, but includes rail, air, water, pipeline and intermodal transport as well.

Freight demand has been of research interest since 1980s (Chiang, et al., 1981; Jiang, et al. 1999; Regan and Garrido, 2002; Turnquist, 2006; Holguín-Veras and Jaller, 2014), and subsequently led to studies investigating data sources, modeling methods and application scope (Abdelwahab and Sargious, 1992). The National Cooperative Highway Research Program (NCHRP) and National Cooperative Freight Research Program (NCFRP) have also published several reports for guiding future freight studies such as freight model development, freight data collection and usage and freight infrastructure investment (NCHRP Report 586, 606, NCFRP Report 32, 33, 37, etc.).

In 2013, the Freight Demand Modeling and Data Improvement (C20) plan was published as part of the second Strategic Highway Research Program (SHRP 2) with the following objectives (Chase, et al., 2013):

- Identify the freight modeling and data collection needs.
- Summarize and encourage innovative freight research and data collection technologies.

- Strengthen the relations between freight transportation planning and land use, infrastructure investment, logistics operation and other related areas.
- Establish a regular procedure to motivate and support the inventive modeling approaches, data collection and analysis methods as a basis for sustainable freight research.

The approach and scale of freight models greatly depend on the availability of data. While proprietary companies can provide freight shipment data at disaggregate levels (zip code level, county level, business economic area (BEA) region level), the associated cost usually is a significant barrier in the development of a regional or state level model that may require regular future updates (U.S. Department of Transportation, 2015). In addition, the reliability of private data cannot be easily verified due to lower transparency in data collection and processing methodologies (Mani and Prozzi, 2004). Public freight data sources are able to provide general information including commodity types, destination flows and transportation modes at coarse geographic levels (state level, Freight Analysis Framework (FAF) zone level which always contains several counties) and adopted in the development of many statewide models (NCHRP Report 606, 2008; Chow, et al., 2010). Jones and Sharma developed a sub-model for agricultural grain shipments under Nebraska Statewide Truck Model using Census of Agriculture data and considered both truck and rail modes (Jones and Sharma, 2002). The Alabama Statewide Freight Flow Model adopted both published Commodity Flow Survey (CFS) data and a freight transportation survey with the collection of employment, O-D, mode and freight projections in future five and ten years. The model estimated freight flow by mode from the industry and size using regression analysis (Anderson, et al., 2007). The State of Florida started the development of a Florida Multi-Modal Statewide Freight Model in 2012 using FAF data, Transearch, establishment surveys and truck surveys. It combined a national supply chain model and regional tour-based truck models (RSG, 2012 and 2013). A continuous challenge for researchers is how to effectively capture the supply chain relations among commodities using public freight data.

Freight is not only transported via trucks, but can be hauled by a variety of modes depending on freight's properties (commodity type, weight, value, required travel distance, required arrival time, etc.). All freight activities are essentially generated as a sequence of decisions on commodities by producers, shippers and consumers (Wisetjindawat, et al., 2005). Commodity-based models focus on modeling commodity flows in weight or value for multiple modes and approaches to capture the connections between freight flows and economic activities (employment, Gross Domestic Product (GDP), etc.).

Emissions are a significant concern among the array of freight movement impacts. According to statistics from the U.S. Environmental Protection Agency (EPA) in 2015, the transportation sector produces 27 percent of the total Greenhouse Gas (GHG) emission in U.S. and acts as a major air pollutant contributor for NOx, VOCs and PM (EPA, 2017). Among all the freight transportation modes, medium and heavy duty trucks play the most significant role in emissions (Figure 1-1). Facing such emissions challenges, the federal government published several regulations (Greenhouse Gas Reporting Program, National Emission Standards for Hazardous Air Pollutants (NESHAP), Final Rule for Greenhouse Gas Emissions and Fuel Efficiency Standards for Medium- and Heavy- Duty Engines and Vehicles. etc.), to monitor and control total air pollution from transportation and protect public health and welfare (EPA, 2017). At the state level, California has passed Assembly Bill (AB) 32 (California's Global Warming Solutions Act of 2006) and Senate Bill (SB) 375 to direct the California Air Resources Broad (CARB) to set regional plans and achieve GHG reduction targets.

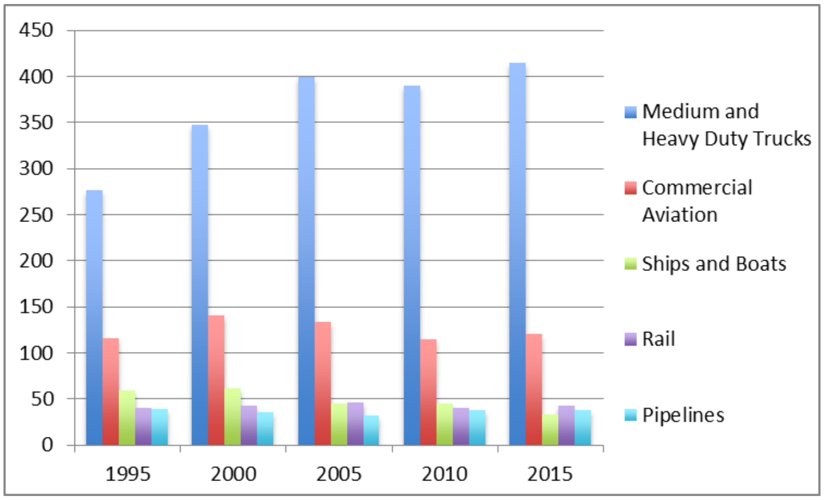


Figure 1-1Total GHG Emission (Tg CO2 Equivalent) of Freight Transportation Modes from 1995 to 2015 (EPA, 2017)

Commercial vehicles are the primary contributors of transportation-related emissions and represent one-third of fuel consumption across all transportation modes (EPA, 2017). The adoption of alternative fuel technologies in commercial trucks and trains is gaining increasing attention due to the depletion of petroleum sources and increases in fuel cost (Adolf, et al., 2016; Lischke, 2017). In general, there are five major types of alternative fuel power-train technologies in freight transport (Bachman, 2011):

- Electric power system: consumes non fossil fuels to generate electricity and store in battery packs, and is currently appropriate for short distance travel.
- Hybrid electric power system: combines an electric motor and battery packs with traditional engine which can help to save fuel by 35%.
- Hydraulic hybrid power system: In the acceleration process, a high-pressure accumulator releases the fluid to activate the motor. Approximate 70 percent of the kinetic energy during the braking is captured and stored to use in the regeneration. Eaton Company developed and applied this technology in trains (Eaton, 2011).

- Natural Gas: a clean fossil fuel that is compressed mainly from methane. Natural gas can reduce up to 25 percent carbon pollution compared with diesel fuel, but needs to be stored in heavy tanks in compressed natural gas (CNG) or liquefied natural gas (LNG) form.
- Biodiesel: is usually produced from food processing as an organic and renewable fuel instead of diesel fuel, and produces low carbon emissions.

The usage and effects of alternative fuels on freight transportation have been investigated over the last couple of decades. An early study in 1998 tested the emissions of trucks using Fischer-Tropsch (F-T) diesel, a synthesized diesel fuel extracted from natural gas, biomass and coal, and found that the application of F-T diesel reduced four regulated emissions including oxides of nitrogen, particulate matter, carbon monoxide and hydrocarbon (Norton, et al., 1998). The National Renewable Energy Laboratory (NREL) conducted an operational study of using B20 (20 percent biodiesel and 80 percent petroleum diesel) in United States Postal Service (USPS) cargo vans and trucks, and obtained useful operating cost and engine maintenance data (Fraer, 2005). Recently, the California Energy Commission (CEC) sponsored the Natural Gas Vehicle Incentive Project (NGVIP) to encourage the adoption of new natural gas commercial trucks and analyze their travel patterns. The project is administered by the Institute of Transportation Studies at the University of California, Irvine since 2015 (CEC, 2017). One of the project purposes is to facilitate the integration of the California Statewide Travel Demand Model (CSTDM) and the California Statewide Freight Forecasting Model (CSFFM) in order to estimate statewide impacts of alternative fuel freight modes including flows, trip length, vehicle miles traveled (VMT) and emissions. The California Department of Transportation (Caltrans) funded the development of the original CSFFM which was completed in 2013.

The State of California is a significant freight origin and destination in the U.S. for intrastate, interstate and international economic trade. Two of ten U.S. top metropolitan port complexes in total value are located in California, including the Los Angeles-Long Beach port complex in the south and the San Francisco-Oakland port complex in the north (Tomer and Kane, 2015).

CSFFM is a freight forecasting tool coded in Citilabs Cube software with the ability to estimate freight movements by commodity group (CG) and mode. The development of the original CSFFM model was based on the Freight Analysis Framework Version 3 (FAF3) data. FAF3 defined five zones in California. These five zones were further disaggregated into 97 Freight Analysis Zones (FAZ) at the county and sub-county level (Figure 1-2). In addition, the original CSFFM also contained 38 import/export gateways (including eleven seaports, eight airports and nineteen Mexican border land ports) and 31 Transport Logistic Nodes (including thirteen airports and eighteen rail terminals). The rest of the 118 FAF3 domestic zones and eight international regions were maintained in the CSFFM zone system as outside-California zones. The 43 SCTG two-digit commodity types in FAF3 were aggregated into fifteen commodity groups in forecasting freight demand. The original CSFFM was calibrated for a base year of 2007, and provides future freight forecasts in 2020 and 2040 (CSFFM, 2015).

Figure 1-2 Disaggregation of FAF3 zones to FAZs in California (CSFFM, 2015)

There are five core modules in the original CSFFM: Commodity Module, Mode Split Module, Transshipment Module, Seasonality and Payload Factor Module, and Network Module. The contents and functions of each module are summarized in Table 1-1. The model final outputs are truck and rail flows at a network level, and water, air and pipeline tonnage flows at an O-D level.

Table 1-1 CSFFM Modules (CSFFM, 2015)

Module	Content	Module Output
Commodity Module	Total generation	Total production and attraction
	Domestic flow distribution	Domestic OD flows by Commodity Group (CG)
	Import/export gateway distribution	Import/export OD flows by CG
Mode Split Module	Determine mode-share in each OD pair	OD flows by mode (Truck only, rail only, Rail-truck, Air-truck, Water only, Pipeline)
Transshipment Module	Split intermodal trips by mode	Truck, rail and air OD flows by CG
Seasonal and Payload Factor Module	Produce seasonal and payload factor	Seasonal and annual truck flows by truck class and CG
Network Module	Route choice and traffic assignment	Truck link flows by CG and rail link flow by railcars

A freight generation model is used to estimate total commodity production and consumption at an FAZ level in future years. In the original CSFFM, the path analysis in a Structural Equation Modeling (SEM) approach was integrated into a generation model for certain CGs with the purpose of considering both direct relations and causal effects among dependent variables (freight production and consumption) and independent variables (demographic, socioeconomic and industrial related variables).

Using a direct demand model approach, Structural Equations for Multi-Commodity OD Distribution (SEMCOD) was developed for the domestic freight distribution step by combining conventional generation and distribution procedures. Hence, SEMCOD is able to avoid the estimation error from a separate generation model compared with a traditional gravity distribution model (Ranaiefar, 2013). SEMCOD defined four O-D clusters based on distance, including one intra-zonal cluster and three inter-state clusters (Table 1-2).

Table 1-2 OD Clusters in SEMCOD (CSFFM, 2015)

OD Cluster	Description
1	Intra-state flows
2	Interstate flows between adjacent zones or Interstate flows between non adjacent zones ≤ 500 miles
3	500 miles ≤Interstate flows between non adjacent zones ≤ 1500 miles
4	Interstate flows≥ 1500 miles

Although the application of path analysis of SEM in freight generation and direct demand models is an innovation of the original CSFFM, the model still suffers from the following limitations:

- **Location of the centroid**

The zone centroid is a special node located in each traffic analysis zone (TAZ) as the origin and destination of each trip. In the original CSFFM, the geometric center of each zone was selected as the centroid. Further adjustments were manually processed based on needs (CSFFM, 2015). However, the geometric center cannot represent the freight economic center (Park, et al., 2011) and may lead to incorrect route choice in traffic assignment, especially for the original CSFFM which has a coarse zone system. Figure 1-3 presents an example in the State of Oregon to illustrate such an issue. A zoomed-in map is shown in Figure 1-4.

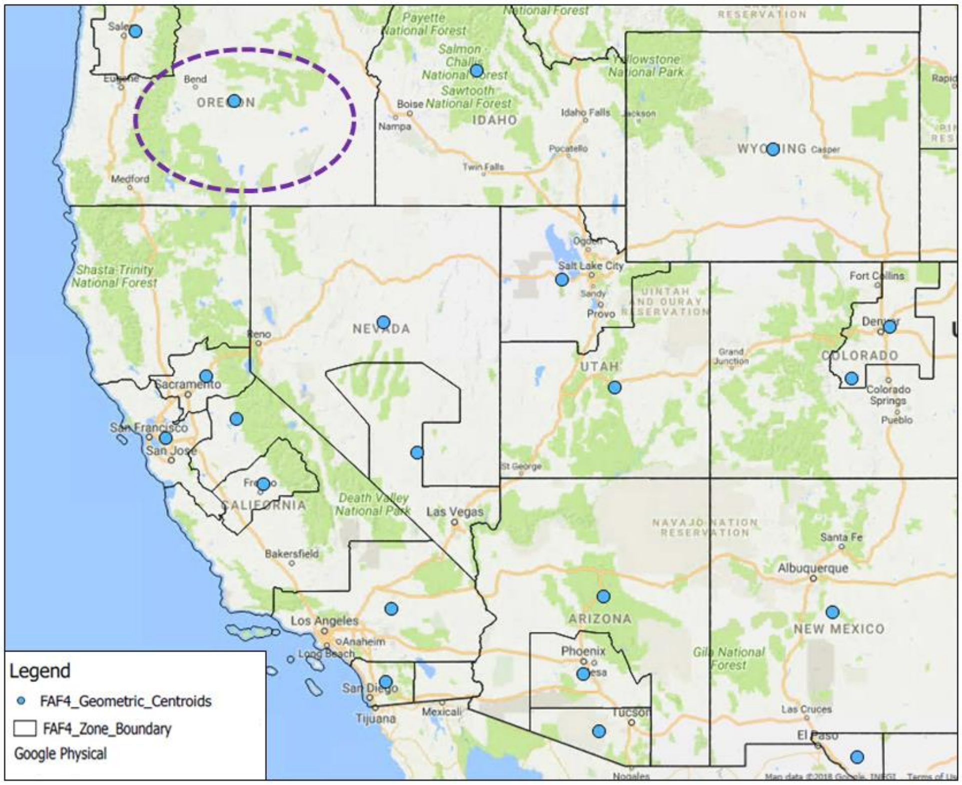


Figure 1-3 FAF4 Centroid Map (West U.S.)

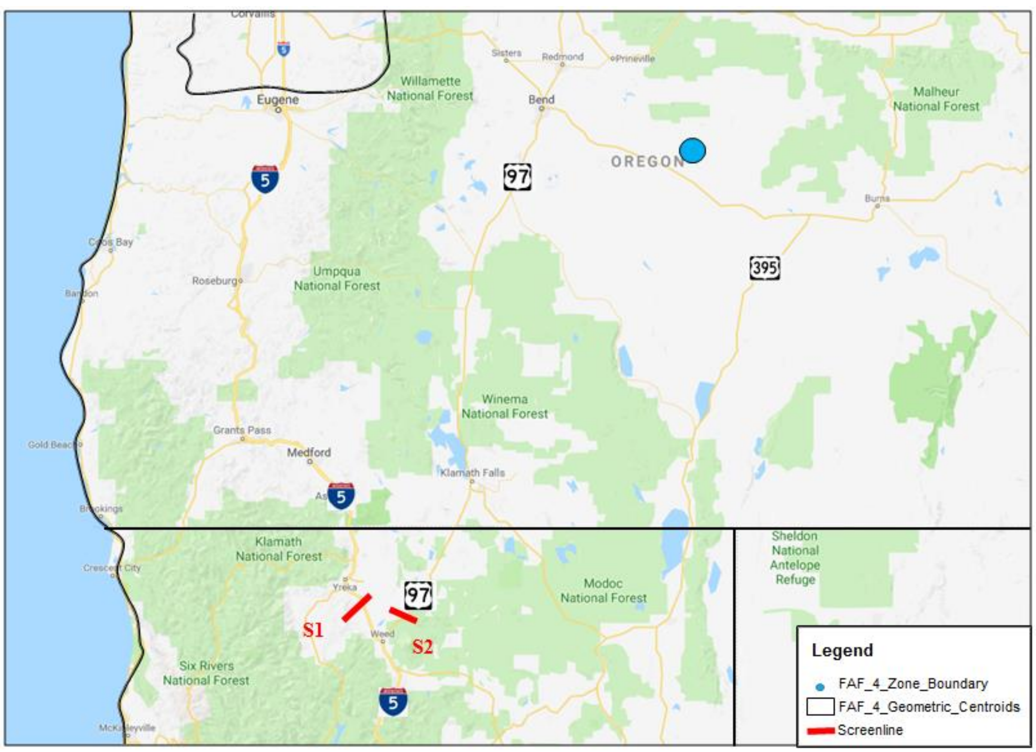


Figure 1-4 FAF4 Centroid Map (Zoomed in Remainder of Oregon)

The State of Oregon is a major external region of California. I-5 and U.S. Route 97 are two highways connecting California and Oregon, and S1 and S2 are the corresponding screenlines of each route. In 2010, the bidirectional annual average daily truck traffic (AADT) at S1 and S2 was 3918 and 1140, respectively (Caltrans, 2010). However, by using congestion speed in the California Statewide Travel Demand Model (CSTDM) and the geometric centroids, the 2010 scenario traffic assignment results of CSFFM show that the daily two-way flows at S1 and S2 are 1760 and 4199. The model results are significantly different from the condition reported by Caltrans.

- **Classification of commodity group**

In the original CSFFM, the two-digit SCTG 43 commodity types used in FAF3 were aggregated into fifteen CGs based on mode share and trip length distribution (TLD). However, this grouping method lacks a statistical analysis to test the hypothesis of similarity. Besides, the TLD of each commodity was developed based on the highway skim and cannot accurately represent the trip length of non-truck modes. In the most recent FAF4 data, the Census aggregated two-digit SCTG commodities into nine groups to avoid disclosure issues in the Foreign Trade Division (FTD) dataset (FHWA, 2015). The fifteen CGs used in the original CSFFM splits the sequence of nine SCTG groups (Table 1-3 and 1-4).

Table 1-3 CSFFM Commodity Groups (CSFFM, 2015)

CG ID	Two-digit SCTG	Description
G1	01-04	Agriculture products
G2	26-29	Wood, paper and printed products
G3	16	Crude petroleum
G4	17-19	Fuel and oil products
G5	10-13	Gravel/sand and other nonmetallic minerals
G6	14-15	Coal and metallic mineral ores
G7	05-09	Food , beverage, tobacco products
G8	24, 30, 39, 40, 42, 43	Manufactured products
G9	20-23	Chemical/pharmaceutical products
G10	31	Nonmetal mineral products
G11	32-34	Metal manufactured products
G12	41	Waste material
G13	35, 38	Electronics
G14	36-37	Transportation equipment
G15	25	Logs and lumber

Table 1-4 FAF4 SCTG Groups (BTS, 2017)

SCTG Group Code	Two-digit SCTG Code	Description
1G	01-05	Agriculture products and fish
2G	06-09	Grains, alcohol, and tobacco products
3G	10-14	Stones, non-metallic minerals, and metallic ores
4G	15-19	Coal and petroleum products
5G	20-24	Pharmaceutical and chemical products
6G	25-30	Logs, wood products, and textile and leather
7G	31-34	Base metal and machinery
8G	35-38	Electronic, motorized vehicles, and precision instruments
9G	39-43,99	Furniture, mixed freight, misc. manufactured products, and commodity unknown

- **Distance skim**

In travel demand modeling, a skim matrix represents the impedances (distance, travel time or travel cost) between zone pairs. The distance skim for estimating the domestic direct demand model used in the original CSFFM was based on the highway skim and does not adequately represent the travel distance of non-truck modes and commodities.

- **Model method**

The original CSFFM adopted the path analysis method in SEM to forecast freight demand. Path analysis is a special case of SEM by considering causal connections among dependent and independent variables (Xue, 2007). However, path analysis requires that all variables are observed variables and assumes that the observed variables are measured without error (Garson, 2013). Furthermore, the causal relationship between variables can only be unidirectional (Crossman, 2017).

In the SEMCOD, there is no statistical support for the method of determining OD clusters by distance. In addition, SEMCOD cannot estimate the freight flow distribution for CG3 (Crude petroleum) and CG6 (Coal and metallic mineral ores) since no models were successfully developed for them. Moreover, SEMCOD considers both origin and destination variables when estimating O-D flow distribution. As a consequence, the estimation of intra-zonal movements suffers from the double-counting of zonal variables, where the origin and destination are the same.

1.2 Research Statement

In this research, a commodity-based freight demand forecasting framework is proposed and applied in updating the original CSFFM using FAF4 data. The innovations of this framework are:

- Provide a mathematical approach to determine zonal freight economic centroids representing origins and destinations of freight movements.
- Aggregate commodities into CGs using a Fuzzy C-means (FCM) clustering algorithm.
- Employ weighted travel distance by CG as a more representative travel distance for

multi-modal freight transport.

- Adopt SRM method in total generation model and domestic direct demand model to capture the direct effect, indirect effect of exogenous variables on endogenous variables, and the underlying effect of latent variables (LV) on endogenous variables.

1.3 Organization of the Study

This dissertation is organized into five chapters. A brief introduction of each chapter is presented below：

Chapter one provides a background on the significance of freight movements and research objectives.

Chapter two first presents the literature review for conventional freight modeling approaches and freight data sources. FAF data – the primary data sources used in the model development – is introduced in detail followed by a brief review of freight generation and distribution models. Path analysis, Confirmatory Factor Analysis (CFA) and Structural Regression Model (SRM) are subsequently introduced in this chapter.

Chapter three introduces the fundamentals, mathematical formulas and explanation of each technique in the proposed framework including freight economic centroids selection, commodity-based average travel distance calculation, grouping of commodities and freight demand model development using SRM.

Chapter four contains the details, results and analysis of updating original CSFFM using the proposed framework. A comparison between the updated and original models is presented and analyzed.

Chapter five summarizes the research contributions and future research topics.

2 BACKGROUND AND LITERATURE REVIEW

With the growing realization of the importance of freight demand forecasting, many categories of freight demand models have been explored and developed for different states and metropolitan areas. Freight forecasting models can be summarized into five categories: Direct Facility Flow Factoring Method, Origin-Destination Factoring Method, Truck Model, Four-Step Commodity Model and Economic Activity Model (NCHRP 606, 2008). Similar to passenger travel demand models, the generation-distribution approach has been adopted in several freight studies (SCAG, 2008; Rwakarehe et al., 2014). While generation refers to the cargo or vehicle trips generated for freight movement, distribution describes the flow of goods between freight zones. In this chapter, section one first summarizes the state-of-the art freight modeling approaches. Section two presents a brief review of freight data sources in the United States and focuses on the Freight Analysis Framework (FAF) data since it is used in the framework application of this study. The freight generation and distribution models are reviewed along with their strengths and inadequacies in sections three and four independently. Structural Regression Model (SRM) is the full format of SEM and serves as the methodology of generation model and direct demand model in this framework. The model components and fitness measures of SEM will be introduced in in section five.

2.1 Approaches of Freight Modeling

NCHRP Report 606 summarizes freight forecasting approaches into five classes: Direct Facility Flow Factoring Method, Origin-Destination Factoring Method, Truck Model, Four-Step Commodity Model and Economic Activity Model. Table 2-1 presents the model components, characteristics and application cases of each model class (Kuzmyak, 2008; RSG, 2015; Doustmohammadi et al., 2016).

Table 2-1 Summary of Freight Forecasting Models

Model Class	Model Component	Characteristic	Application
Facility Factoring Method	Facility flows	Short-term forecast Least data needed Limited applicability	Minnesota's Highway 10 Truck Forecasting Model Florida Heavy Duty Truck Freight Model for Ports
O-D Factoring Method	O-D flows Mode split Traffic assignment	Applying growth rate in O-D table	Ohio DOT Interim Freight Model
Truck Model	Trip generation Trip distribution Traffic assignment	3 Step Model Single truck mode	SCAG Heavy Duty Truck Model
Four-Step Commodity Model	Trip generation Trip distribution Mode split Traffic assignment	Commodity flows Multiple modes Large region applicability	CSFFM
Economic Activity Model	Trip generation Trip distribution Mode split Traffic assignment Economic/Land use modeling	A combination of economic/land use model and freight demand model	Oregon Statewide Passenger and Freight Forecasting Model

The Direct Facility Flow Factoring Method is a straightforward method to apply growth factors to the existing freight link-by-link flow to obtain short-terms forecasts (Chow, et al., 2010). This method requires the least data but lacks of explanation power (RSG, 2015). NCHRP Report 606 also pointed out that the Direct Facility Flow Factoring Method is limited to link level analysis and cannot forecast freight movements of a whole transportation system.

The Origin-Destination Factoring Method forecasts future O-D table by applying growth factors to an existing O-D table (Doustmohammadi, et al., 2016). This approach also has additional mode choice and traffic assignment steps compared with Direct Facility Flow

Factoring Method. This method was applied to estimate O-D flows in the Ohio Department of Transportation Interim Freight Model (TRC, 2005).

The Truck Model approach represents single mode models that estimate truck O-D tables by trip generation and distribution steps and subsequently assigns truck trips onto the network. A well-designed truck model is predictive for long term forecasting but requires more detailed truck data (Chow, et al., 2010). An example of a truck model is the Southern California Association of Governments' (SCAG) Heavy Duty Truck (HDT) Model. SCAG's model categorizes trucks into 3 types: light heavy duty truck, medium heavy duty truck and heavy-heavy duty truck. The External Trip Generation and Distribution Model obtains the Traffic Analysis Zone (TAZ) level daily truck trips converted from Transearch annual tonnage flows. In the Internal Trip Generation and Distribution Model, trips were generated from the trip rates which refer to the number of trips per household or employment. The trip distribution was processed by using a gravity model developed from Global Positioning System (GPS) survey data. This HDT model also included the intermodal rail model and port model as special trip generators. The multiclass Trip assignment was processed by incorporating truck trip tables and passenger trip tables in the SCAG Travel Demand Model. The forecasting results can be used in both the analysis of freight movements and emissions (SCAG, 2008). The truck models are able to estimate empty trips, but are not applicable to modeling freight movements in multiple modes (Jones and Sharma, 2002).

Four-Step Commodity Models share a similar procedure with passenger demand models, including generation, distribution, mode choice and traffic assignment. Compared with Truck Models, Commodity Models are able to consider other freight modes besides truck and can capture the economic mechanisms that generate freight movements (Doustmohammadi, et al., 2016). The four-step commodity model structure was adopted in several statewide freight forecasting practices. The Wisconsin Statewide Freight Model considered twenty-five commodities and four freight modes including truck, rail, air and water and modeled the freight movements at a county level. The generation model used generation rate and attraction rate for each commodity developed from corresponding two-digit Standard Industrial Classification employment data. The annual tonnage O-D tables by commodity were obtained from a gravity distribution model and were calibrated using county-level Transearch data. The commodity flows were further converted to truck flows using the payload factors generated from Wisconsin's Vehicle Inventory and Use Survey (Proussaloglou et al., 2007). A commodity-based California Statewide Freight Forecasting Model (CSFFM) was developed with SEM structure in its total generation model and domestic Structural Equation Multi-Commodity OD Distribution (SEMCOD) model as introduced in Chapter one (CSFFM, 2015). Other statewide practices also include the Commodity Flow Processing part of Iowa Statewide Freight Commodity Model and the Pennsylvania Statewide Commodity-based Freight Model (Iowa DOT, 2008; White, Anater & Wall, 2007).

Economic Activity Models utilize a land-use sub-model to forecast the economic inputs prior to a four step model. The estimated transport cost also has feedback on the economic land use model. Oregon State developed their statewide model by integrating passenger and freight models using the economic activity model approach (Hunt et al., 2001).

2.2 Freight Data Source

The availability of comprehensive freight data is highly expected in developing freight forecasting models (NCFRP Report 19, 2012). In fact, it is always prohibitive or expensive to acquire the equivalent level of data as passenger models due to the privacy of commodity transaction firms, especially for statewide areas (Chow, et al., 2010). Facing such limitations, a critical challenge is how to extract more useful information from public data sources in forecasting freight movement patterns. FAF is a freight movement data source created through the integration of the Commodity Flow Survey (CFS) and other out-of-scope (OOS) data sources (U.S. Department of Agriculture (USDA), 2002 Vehicle Inventory and Use Survey (VIUS), Fisheries of the United States, Solid Waste Management Annual Report by State, etc.) , and was used in the development of the original CSFFM (CSFFM, 2015). In this study, the recently published FAF Version 4 (FAF4) data is selected to update the original CSFFM in framework application and is reviewed in detail in this section.

2.2.1 Freight Data Review

Currently, there are several public and private freight-related data sources for freight modeling use, with different collection method, data type, spatial coverage, modes and commodity types (Tok, et al., 2011; CDM Smith, 2015). Table 2-2 provides a characteristics summary of core and new freight data sources.

Table 2-2 Freight Data Source Review

Data Source	Data Type	Coverage	Temporal	Mode	Commodity	Model Usage
WIM	Detector-based;Truck characteristics	National	Daily	Truck	None	Validation
TAMS	Detector-based; Vehicle characteristic	California	Daily	Truck	Agriculture, Beverage, Livestock, Logging	Validation
VIUS	Survey-based; Truck characteristics	National	Every five years (discontinued)	Truck	None	Calibration, Validation
FAF	Model-based; Cargo movement	National	Every five years	Truck, Rail, Air, Water, Pipeline, Multiple modes and mail, Other	SCTG commodities	Estimation, Calibration
Transearch	Model-based; Cargo movement	Flexible	Annual	Truck, Rail, Air, Water	STCC commodities	Estimation, Calibration, Validation
ATRI	GPS-based; Truck touring	Flexible	Flexible	Truck	Unknown	Estimation, Calibration, Validation

Weigh-in-Motion (WIM) system is a sensor-based technology to collect detailed Commercial Motor Vehicle (CMV) data without requesting a vehicle to stop (Hernandez, 2014). The data collected at WIM sites includes vehicle gross weight, axle spacing, speed and vehicle classification using Federal Highway Administration (FHWA) 13 vehicle classification scheme (Lu, Q., et al., 2002; Caltrans, 2016). In the Heavy Duty Truck Model of Southern California Association of Governments (SCAG), WIM data was used to update the time-of-day factors in trip assignment step (SCAG, 2012)

The Truck Activity Monitoring System (TAMS) is a novel temporally continuous and spatially comprehensive truck data source hosted on an interactive GIS enabled web-based user interface for freight related planning and decision making. The initial truck classification model was developed in a pilot study from 2012 funded by CARB, and then continually sponsored by Caltrans from 2015 to enhance the truck classification models and expand data collection sites to over 90 locations along major truck corridors across California, encompassing state borders, regional cordons and metropolitan areas. The core technology of TAMS is inductive signature technology, an evolution of conventional inductive detector technology. TAMS has two detector solutions: standalone inductive signature technology at existing Inductive Loop Detector (ILD) sites which can provide four categories incorporating over 40 truck body classifications, and integrated WIM and Inductive Signature Technology at existing WIM sites with the capacity to distinguish 63 vehicle classifications which is a great improvement compared to the original FHWA 13 vehicle category scheme (TAMS, 2016). With the detailed classification scheme, TAMS also has the ability to distinguish trucks from some specific industries, such as logging, livestock and agriculture. Table 2-3 shows the Truck Classification Scheme used in TAMS (Tok, et al., 2017).

Table 2-3 Truck Classification Scheme in TAMS

Units Type	Body Class	Units Type	Body Class
Single-Units without Trailer	Conventional Van/Platform	Semi-Tractor Trailers	Enc. Van
	Cab Over Van/Platform		53ft Container
	30ft Bus		40ft Container
	20ft Bus		40ft Container Reefer
	Multi Stop Van/RV		20ft Container
	Utility/Service		Platform
	Concrete		Tank
	Dumpster Transport		Open Top Van
	Garbage		Auto
	Bobtail		Low Boy Platform
	Dump Triple Rear		Drop Frame Van
	Street Sweeper		Dump
	Dump/Tank		Logging
Single-Units with Trailer	SU small trailer		Livestock
	Dump-Dump		Agriculture
	RV with Towed Vehicle		Beverage
	Concrete w/Lift Axle		Enclosed Van Reefer
	Tank-Tank		Platform/Tank
	Platform-Platform		Dump
	Tow Truck with vehicle	Multiple Semi	Pneumatic Tank

	Dump with Lift Axle	Tractor Trailer Combination	Hopper
Multi-Units with Semi Tractor Trailers	Enclosed Van		Agricultural Van
	Enc. Van Reefer		Low Chassis Van

The Vehicle Inventory and Use Survey (VIUS) is a representative survey-based truck data source produced by United States Census Bureau and quinquennially updated from 1963 to 2002 (Jeong, et al., 2016). It collected physical and operational truck characteristics including vehicle age, body type, average weights, commodity type and travel distance (Bureau of the Census, 2004). VIUS can provide payload factors to convert tonnages to truck trips for commodity-based models (Maks Inc., 2016). Although with detailed truck information, the small survey sample size cannot fully capture all trucks' activities. Only about 0.16 percent of total registered trucks in the United States were surveyed in 2002 VIUS (UCI-ITS, 2013). Moreover, the most recent VIUS was conducted in 2002 which can now only be used as a backcasting validation source (Hernandez, 2014).

Conducted by the Federal Highway Administration (FHWA), FAF is a freight movement database primarily developed based on the Bureau of Transportation Statistics Commodity Flow Survey (CFS), and integrated various data including VIUS, United States Department of Agriculture (USDA), Annual Retail Trade Survey, Service Annual Survey (SAS), etc. (FHWA, 2015). FAF covers state, FAF zone, foreign regional and major gateways' freight transportation activities, and provides estimation of commodity movement by origin-destination, commodity types and modes in both weight and values (FHWA, 2011; Schmitt, 2008). As the data source selected for this research's application, FAF will be introduced and discussed in more detail in the following section.

Transearch and American Transportation Research Institute (ATRI) are two private freight data sources. Transearch can provide freight information including O-D, commodity, mode, tonnage and value (Liao, 2014; IHS Markit, 2017). Transearch data have been employed in the development of statewide freight models in Florida, Iowa and Vermont (NCHRP Report 606, 2008). ATRI provides a truck Global Positioning System (GPS) database providing truck trajectory information and can be used in the development of truck touring models (ATRI, 2017; CDM Smith, 2016). Pinjari et al. explored the methodologies for extract the average speed, trip duration, trip length, truck flow and O-D tables from ATRI raw data for freight modeling and planning uses (Pinjari, et al., 2014). However, private databases are costly, and do not have well documented data collection and estimation assumptions and methodologies (Mani and Prozzi, 2004).

2.2.2 Freight Analysis Framework (FAF)

FAF is a publicly available freight database produced by FHWA and the Bureau of Transportation Statistics (BTS) providing tonnage, value and ton-mile estimates by mode and commodity type. Until 2017, there are four versions of FAF products (FHWA, 2002, 2009, 2012, 2015). The introductions of each FAF generation are presented as follows.

2.2.2.1 FAF Generations

FHWA started the FAF project in 1999 in order to develop a national database of freight transportation flows for main freight transportation modes (Wurel, Erin, et al., 2009). The first version, FAF1, was carried out with freight flow estimation of 1998 and forecasts for 2010 and 2020, for commodity types defined by the Standard Transportation Commodity Code (STCC). FAF1 contained four modes including highway, railroad, water and air (FHWA, 2005). The forecasting results showed that the domestic freight movement would increase by 65 percent from 1998 to 2020, and the international freight movement would have a faster growth trend than domestic freight, with a 2.6 percent annual growth rate (FHWA, 2002). However, FAF1 was developed relying on proprietary data (1998 Transearch) and made little use of CFS (Cambridge Systematics, 1997), which led to inconsistencies between FAF1 and CFS (Donnelly, 2010). In addition, FAF1 can only estimate long distance freight movement and has limitations in both geographical details and estimation method transparency (FHWA, 2011).

FAF2 is the second version of FAF with base year in 2002 and future year in 2035. The coverage of FAF2 includes 114 regions of 2002 CFS, 17 international gateways and 7 international regions (FHWA, 2002). The domestic zone map of FAF2 is presented in Figure 2.1 (Donnelly, 2010). FAF2 was developed primarily based on the 2002 CFS and considered more complex transportation modes, including truck, water, rail, truck plus rail, pipeline and unknown, air plus truck and other (Lupa, 2008). In addition, FAF2 used the two-digit format commodity types defined in the Standard Classification of Transported Goods (SCTG) which contained 43 commodity types (FHWA, 2007) instead STCC. FAF 2 was applied to the development of a Statewide Freight Study and Action Plan in Alabama, and the Kansas Freight Analysis Framework (KFAF) (Alabama DOT, 2010, Wurfel, Erin, et al., 2009).

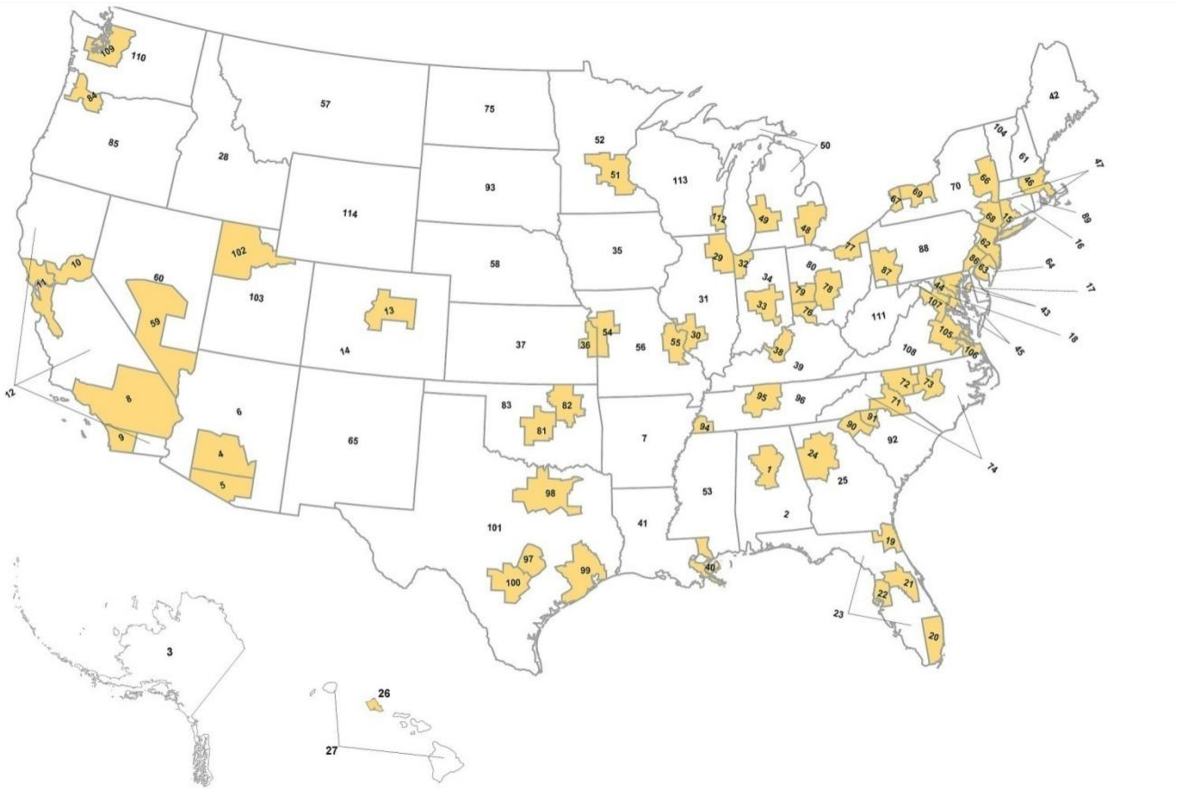


Figure 2-1 Domestic Zones Map of FAF2 (Donnelly, 2010)

As the third generation, FAF3 was developed with a 2007 base year and future year 2015 through 2040 with five-year intervals. Similar to FAF2, FAF3 is generated using 2007 CFS as a primary data source. FAF3 contains 123 domestic zones (Figure 2-2) and 8 international zones used for imports and exports including Canada, Mexico, Rest of Americas, Europe, Africa, Southwestern, Central Asia, Southeastern Asia and Oceania. Thus there are totally 131 origins and destinations existing in the FAF3 dataset (FHWA, 2011).

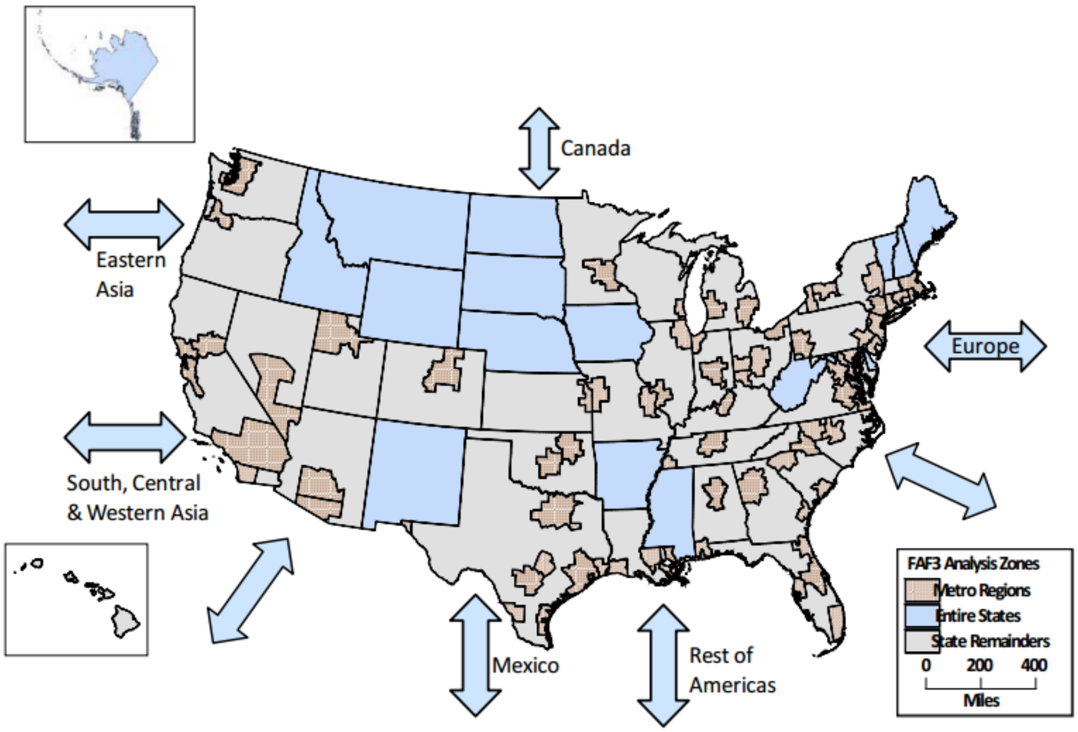


Figure 2-2 Domestic Zones Map of FAF3 (FHWA, 2011)

FAF3 still follows the two-digit SCTG 43 commodity types (Table 2-4) with modified definitions. In addition, FAF3 contains 7 mode classifications that are defined slightly differently from FAF2 (Table 2-5). Since FAF1 and FAF2 were removed from the FAF official website and are no longer available, FAF3 also predicts back year freight movement in 1997 and 2002 in current state-level data format for public use (FHWA, 2012). A FAF3 tabulation tool can provide origin-destination datasets in 4 trade types: total flows, domestic flows, import flows and export flows. Compared with the previous FAF1 and FAF2, FAF3 first estimated domestic ton-miles besides tonnage and value of freight movement at the state level.

Table 2-4 SCTG two-digit Commodity Types (FHWA, 2012)

Code	Commodity Description	Code	Commodity Description
1	Live animals and live fish	23	Chemical products and preparations, n.e.c.*
2	Cereal grains	24	Plastics and rubber
3	Other agriculture products	25	Logs and other wood in the rough
4	Animal feed and products of animal origin, n.e.c.*	26	Wood products
5	Meat, fish, seafood, and their preparations	27	Pulp, newsprint, paper, and paperboard
6	Milled grain products and preparations, bakery products	28	Paper or paperboard articles
7	Other prepared foodstuffs and fats and oils	29	Printed products
8	Alcoholic beverages	30	Textiles, leather, and articles of textiles or leather
9	Tobacco products	31	Nonmetallic mineral products
10	Monumental or building stone	32	Base metal in primary or semi-fished forms and in finished basic shapes
11	Natural sands	33	Articles of base metal
12	Gravel and crushed stone	34	Machinery
13	Nonmetallic minerals n.e.c.*	35	Electronic and other electrical equipment and components and office equipment
14	Metallic ores and concentrates	36	Motorized and other vehicles(including parts)
15	Coal	37	Transportation equipment, n.e.c.*
16	Crude petroleum	38	Precision instruments and apparatus
17	Gasoline and aviation turbine fuel	39	Furniture, mattresses and mattress supports, lamps, lighting fittings, and illuminated signs
18	Fuel oils	40	Miscellaneous manufactured products
19	Coal and petroleum products,n.e.c.*	41	Waste and scrap
20	Basic chemicals	43	Mixed freight
21	Pharmaceutical products	99	Commodity unknown
22	Fertilizers		

FAF3 has been applied in several researches such as Oklahoma Freight Flows Analysis and Coastal Region Metropolitan Planning Organization (CORE MPO) Freight Goods Movement Assessment and CSFFM (Parsons Brinkerhoff, 2012; CDM Smith, 2014; CSFFM, 2015).

Table 2-5 FAF3 Modes (FHWA, 2012)

Code	Mode	Description
1	Truck	Includes private and for-hire truck. Does not include truck that is part of ***Multiple Modes and Mail*** or truck moves in conjunction with domestic air cargo.
2	Rail	Includes any common carrier or private railroad. Does not include rail that is part of ***Multiple Modes and Mail***.
3	Water	Includes shallow draft, deep draft, Great Lakes and intra-port shipments. Does not include water that is part of ***Multiple Modes and Mail***.
4	Air(includes truck-air)	Includes shipments typically weighing more than 100 pounds that move by air or a combination of truck and air in commercial or private aircraft. Includes air freight and air express. Does not include shipments weighing 100 pounds or less which are typically classified with ***Multiple Modes and Mail***, in the case of imports and exports by air, domestic moves by ground to and from port of entry or exit are categorized with ***Truck***.
5	Multiple Modes and Mail	Includes shipments by multiple modes and by parcel delivery services, U.S. Postal Service, or countries. This category is not limited to containerized or trailer-on-flatcar shipments.
6	Pipeline	Includes crude petroleum, natural gas, and product pipelines. Notes: Does include flows from offshore wells to land which are counted as ***Water*** moves by the U.S. Army Corps of Engineers. Does not include pipeline that is part of ***Multiple Modes and Mail***.
7	Other and Unknown	Includes movements not elsewhere classified such as flyaway aircraft, and shipment for which the mode cannot be determined.
	No Domestic Mode	Includes shipments that have an international mode but no domestic mode and is limited to import shipments of crude petroleum transferred directly from inbound ships to a U.S. refinery at the zone of entry. This is done to ensure a proper accounting of import flows, while avoiding assigning flows to the domestic transportation network that do not use it.

In October 2015, the fourth version of FAF, FAF4 was officially released and published by FHWA. FAF4 estimates the freight movement in 2012 mainly based on the 2012 CFS and provides forecasts in 2013, 2014, 2015, and 2020 through 2045 in five-year intervals (FHWA, 2015). FAF4 has 132 domestic regions and 7 freight modes plus an additional non-domestic mode referring imports. In this study, since FAF4 is the data source for updating the original CSFFM which was developed from FAF3, the differences between FAF4 and FAF3 will be reviewed in the next section.

2.2.2.2 Comparison between FAF3 and FAF4

Geography

For domestic regions, FAF4 has 132 domestic zones, which is consistent with the 2012 CFS (Table 2-6), while FAF3 has 123 zones. The updates include 10 new FAF zones and 1 removed Remainder of New Jersey zone (Table 2-7). Besides, due to population and economic pattern changes from 2007 to 2012, the boundaries of some FAF4 zones are also different from those in FAF3 (Figure 2-3, Hwang, H. L. et al., 2016). For the foreign areas, there are still 8 foreign regions in FAF4 the same as in FAF3.

Table 2-6 Summary of CFS Area Changes: 2007 to 2012 (CFS, 2015)

Type of CFS Area	**Example**		**2007 Total**	**2012 Total**	**Number with Definition Changes**
Metropolitan Area	In 2007 and 2012	Portland(OR part)	74	74	46
	New for 2012	Portland(WA part)		10	N/A
Remainder of State	Full State	Idaho(Remainder of ID)	15	13	31
	Less than full state	Remainder of OR	34	35	
Total			123	132	

Table 2-7 New Zones in FAF4 (FHWA, 2015)

Code	**Regions**	**State**	**Type**
65	Fresno-Madera, CA CFS Area	CA	C
101	Philadelphia-Reading-Camden, PA-NJ-DE-MD CFS Area	DE	C
183	Fort Wayne-Huntington-Auburn, IN CFS Area	IN	C
202	Wichita-Arkansas City-Winfield, KS CFS Area	KS	C
211	Cincinnati-Wilmington-Maysville, OH-KY-IN CFS Area	KY	C
311	Omaha-Council Bluffs-Fremont, NE-IA CFS Area (NE Part)	NE	C
331	Boston-Worcester-Providence, MA-RI-NH-CT CFS Area (NH Part)	NH	C
423	New York-Newark, NY-NJ-CT-PA CFS Area (PA Part)	PA	C
473	Knoxville-Morristown-Sevierville, TN CFS Area	TN	C
532	Portland-Vancouver-Salem, OR-WA CFS Area (WA Part)	WA	C

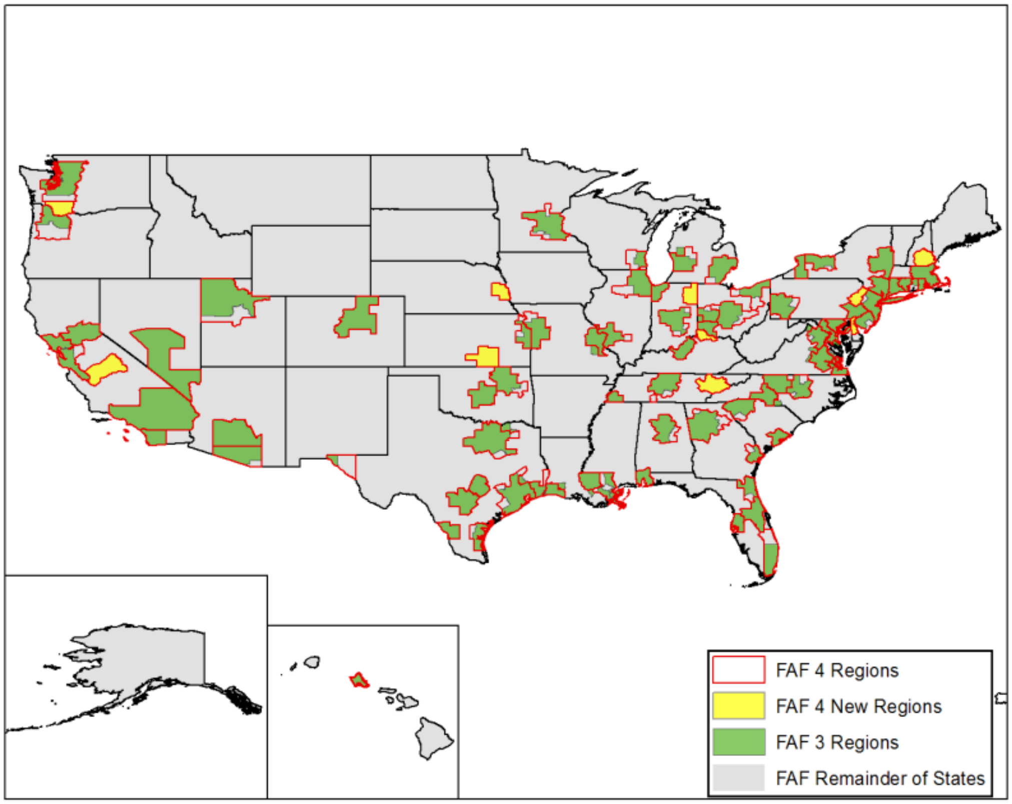


Figure 2-3 FAF3 and FAF4 Domestic Regions Comparison (Hwang, H. L. et al., 2016)

The FAF4 zones in California are magnified in Figure 2-4 for detailed discussion. First, a new zone Fresno-Madera (Code 065) is separated from Remainder of California. Second, the boundary of San Jose-San Francisco-Oakland zone is extended by including Stockton County.

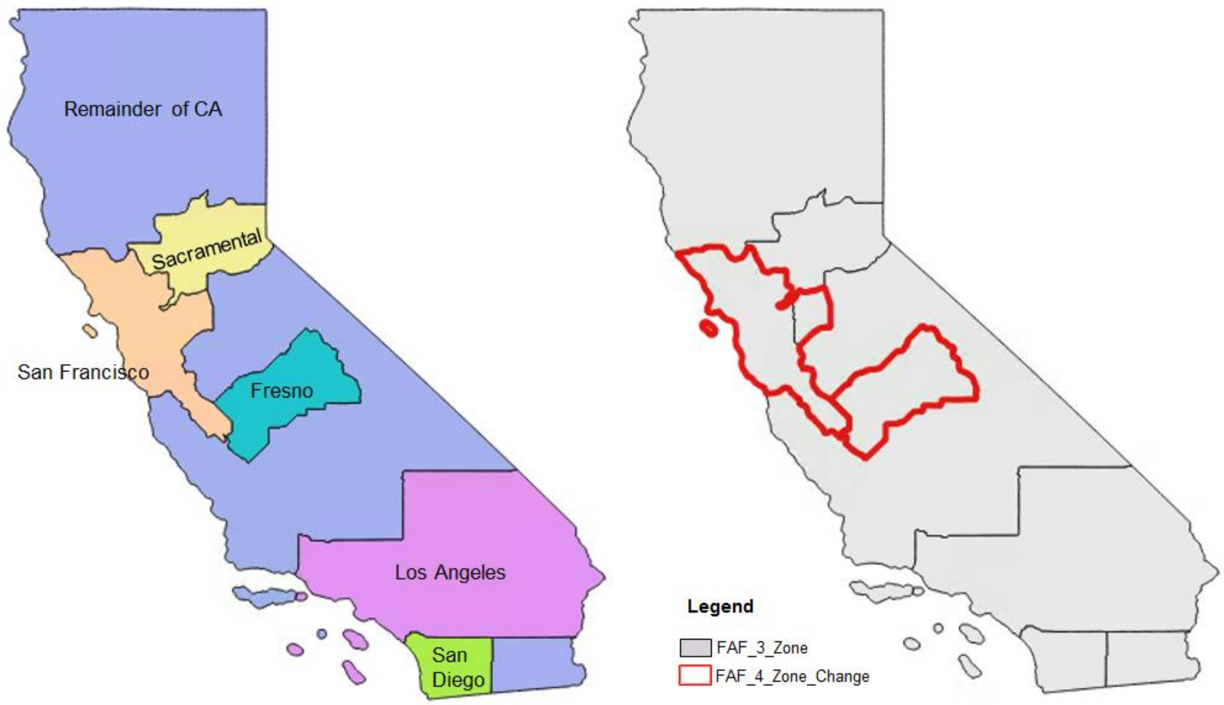


Figure 2-4 FAF4 Geographic Change in California

Commodity Classification

The commodity types used in FAF4 still follows SCTG two-digit level commodity classifications, but with several updates in the following commodities:

1) Fuel ethanol is separated from SCTG 08 (Alcoholic beverages and denatured alcohol) and reclassified under SCTG 17 (Gasoline, aviation turbine fuel and ethanol).

2) In FAF3, Kerosene was included in SCTG 19(Other coal and petroleum products). But in FAF4, it is also reclassified in SCTG 17.

3) The SCTG 18 (Fuel oils) in FAF3 was re-defined as Fuel oils (includes Diesel, Bunker C, and Biodiesel) in FAF4 in order to reflect the recent development of Biofuels. The subcategories of SCTG 18 are listed in Table 2-8 (Hwang, H. L. et al., 2016).

Table 2-8 2012 SCTG 18 Subcategories (Hwang, H. L. et al., 2016)

SCTG Code	Definition
18100	Fuel oil (includes diesel, distillate heating oil, Bunker C excludes biodiesel)
18200	Blends of fuel oils (includes 5% or less biodiesel by volume, B5, or less)
18210	Blends of fuel oils with more than 5% biodiesel by volume, (excludes B100)
18220	Biodiesel (derived from vegetable oils or animal fats), B100 (excludes mixtures of biodiesel and diesel fuel)

Modes

In the 2012 CFS along with FAF4, the "Unknown" mode is reassigned to "Truck" or "Rail" (Table 2-9), but its mode name is still maintained for consistency with FAF3 mode classes. In addition, all shipments of 150 thousand pounds or more in "Multiple Modes and Mail" are reassigned to "For-hire Truck" under "Truck" mode (Hwang, H. L. et al., 2016).

Table 2-9 Reassignment of "Unknown" Mode in 2012 (Hwang, H. L. et al., 2016)

New Mode in 2012	Shipment Weight
Truck	less than 2 truckloads or 80 thousand pounds
Rail	more than 2 truckloads or 80 thousand pounds

Estimation of Non-CFS Domestic Flows (Out-of-Scope OOS industries)

Although CFS represents the majority of freight movements in the United States for all freight modes, there are still some out-of-scope industry activities that are not covered by CFS. Therefore other datasets are required to fill in the data gaps. The OOS industries include (Hwang, H. L. et al., 2016):

- Farm-based agricultural shipments
- Fisheries and logging
- Municipal solid waste and construction and demolition debris
- Retail, services and household/business moves
- Crude petroleum

- Natural gas

The methodologies for integrating other data sources in FAF4 are similar to FAF3 for most OOS industries except for "Farm-based agricultural shipments", which is a significant OOS field in CFS. The United States Department of Agriculture (USDA) is an available data source that provides agriculture statistics for 2012. Due to differences in commodity types, USDA's results were regrouped into SCTG commodity categories. In FAF3, farm-based agricultural shipments were regrouped into SCTG 01(Animal and fish), 02 (Cereal grains) and 03(Agricultural products). FAF4 has more comprehensive considerations by including SCTG 04 (animal feed, eggs, honey and other products of animal origin) and 07 (Other prepared foodstuff, fats and oils) as well (BTS, 2015). Meanwhile, FAF4 considers the relationships between related products and their farm-based commodities when determining the farm-based shipment destinations. For example, in FAF3, only the CFS areas shipped out SCTG 03 (tobacco harvested) were used as the farm-based destination of the same commodity SCTG 03. While in FAF4, the CFS areas shipped out SCTG 09 (tobacco products) were also considered to be the potential destination of SCTG 03.

Ton-miles

FAF4 provides ton-miles in FAF zone level (132×132), while FAF3 only has state level ton-miles (51×51) estimation.

2.3 Freight Generation Model

The purpose of generation models is to predict the quantity of freight activities produced or attracted by each freight analysis zone (FAZ). The NCFRP Report 19 summarizes the generation models into two categories: freight generation (FG) and freight trip generation (FTG). There is a significant distinction between these two definitions. FG refers to cargo generation, and is usually measured in cargo weight or values, while FTG is the generation of trips, which reflects the decisions to transport FG according to shipment size, mode and delivery frequency (Holguín-Veras, José, et al., 2012). In this section, the approaches for both FG and FTG are reviewed, including linear regression, cross classification, spatial regression and SEM.

A linear regression approach was applied in multiple freight generation modeling studies (Southworth 2003; QRFM, 2007; Bastida and Holguin-Veras, 2009; SCAG, 2012). The freight production and consumption were assumed to have linear relationships with zonal level land use data. However, Novak pointed out that there was a high correlation between productions and consumptions for different commodities from the analysis of CFS data which cannot be directly captured using a linear regression method (Novak, et al., 2011).

To address the limitations of linear regression, alternative approaches have been proposed and applied in the generation step. An urban truck generation model was developed for the Brooklyn and Manhattan area using the cross-classification method (Bastida and Holguin-Veras, 2009). Cross-classification can help in improving the model fitness, but may result in losing explanatory power (Ranaiefar, 2013). Another approach is spatial regression in order to correct the spatial correlation among variables. The main problem of the spatial regression method is that the spatial structure is fixed and cannot be applied to other geographic zones (Novak et al., 2011; Chun et al. 2012).

Ranaiefar designed a freight generation model using path analysis in SEM and applied it in the original CSFFM. The path analysis generation model (PAGM) can capture the dependencies between one endogenous variable and another endogenous variable, such as the causal effects of production and consumption of one CG on the production of another CG. It can also reflect the supply chain elasticity compared with linear regression model (Ranaiefar, 2013). A review of SEM will be presented in the following section.

2.4 Freight Distribution Model

As the second step of a commodity-based four-step model, freight distribution is the process of distributing the total known freight from each origin to all destination zones and finally produce O-D tables with predicted flows for each OD-pair (Masoud and Hyun, 2012).

The Fratar model, which is also known as the Growth Model, was the first generation of distribution models and requires a base year O-D matrix to forecast the future year conditions. The main limitation of this approach is the disregard of travel pattern changes over years caused by the land use change and increasing freight demand (Levinson and Ajay, 1995).

Economic Input-Output models connect commodity flows with economic activities and can reflect the economic mechanism in freight movements, but are still unable to capture empty truck trips and local truck movements. Moreover, Input-Output models require extensive data which may not be available at the traffic analysis zone (TAZ) level (NCFRP, 2010).

The gravity model was originally influenced by Isaac Newton's law of universal gravitation and used in both passenger and freight distribution process. The first rigorous use of a gravity model was proposed by Casey in 1955 to estimate the shopping trips between towns in a region with consideration of population and distance (Ortúzar and Willumsen, 2011). Wilson (1967) redefined the gravity model by using the Entropy Maximization derivation and developed the standard formulation for doubly constrained gravity models, which was subsequently applied in freight distribution practice (NCHRP Report 606, 2008). Gravity models use an impedance term representing the travel cost (distance, travel time, etc.) between each zone pair when distributing flows. The built-in packages of gravity distribution models in transportation planning software (TransCAD, Cube, etc.) spread the application of this method. Nevertheless, Wisetjindawat, et al. indicated that simulating commodity movements as spatial distributions of particles lacked theoretical fundamentals in a gravity model (Wisetjindawat, et al., 2006).

Direct demand models aim to estimate flow directly from demographic, land use and economic parameters (Talvitie, 1973). In freight modeling, a few studies have applied the direct demand modeling approach to estimate freight flows. Jonnavithula (2004) designed a SEM framework for O-D flow estimation by determining the relationships among dependent variables. In the original CSFFM, SEMCOD was developed for domestic freight distribution. SEMCOD integrated the generation and distribution steps and considers the structural relations between dependent and independent variables and the cross-relations among CGs (Ranaiefar, 2014). This model can reflect the supply chain relations for certain CGs, but still has limitations in model methodology as introduced in Chapter one.

2.5 Structural Equation Model

SEM is a statistical modeling method to estimate and test the cross-relationship between both measured variables and latent variables. Generalized from linear regression, SEM has a more flexible linear-in-parameters format that allows endogenous variables to serve as causal variables for other endogenous variables, which provides the opportunities to understand the comprehensive correlation and covariance among multiple variables (Kline, 2011). The basic definitions in SEM are listed below (Suhr, 2006; Kenny, 2011):

Measured Variable (MV): directly measured variable or observation.

Latent Variable (LV): construct that cannot be directly measured and is defined based on MVs. For example, “Education Level” is a latent variable that may be constructed based on highest degree obtained, school ranking and Grade Point Average (GPA).

Exogenous Variable: A variable in the model that is not caused by other variables, and usually causes variables.

Endogenous Variable: A variable in the model that is caused by other variables (exogenous variable or other endogenous variable).

Direct Effect: a directional relationship from independent variable to dependent variable.

Indirect Effect: a relation from independent variable to dependent variable through one or more intervening variables.

Total Effect: total of direct effects and indirectly effects of one independent variable on the dependent variable.

Path Diagram: a graphical representation of SEM. The common symbols used in Path Diagram are presented in Figure 2-5.

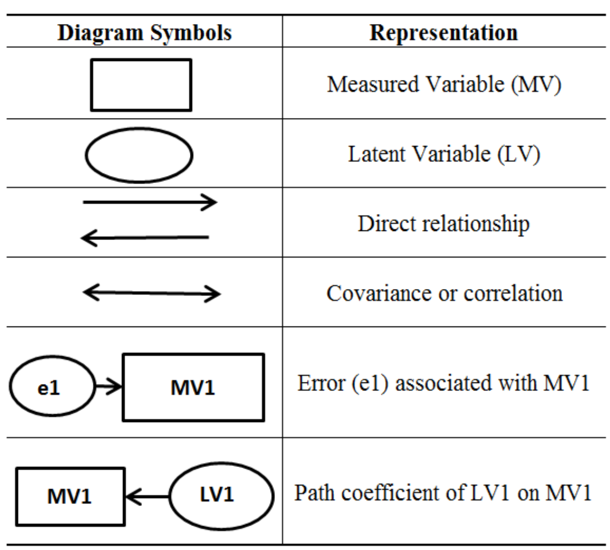


Figure 2-5 Path Diagram Symbols

Free Parameter: a parameter that is estimated from the data.

Fixed Parameter: a parameter that is not estimated from the data. The value of a fixed parameter is usually fixed to 0 or 1.

Specification: formulating a hypothetical statement of a SEM with a set of parameters. Specification is the first and most importance step of SEM since the following results are all based on the assumption that the model is basically correct.

Identification: if a unique solution for all parameters exists, then the model is identified, otherwise it is not identified.

There are various explorations on the procedures of SEM. In this research, a five steps procedure of SEM is summarized from Kline’s and Crockett’s studies (Kline, 2011; Crockett, 2012) and listed below:

Step 1: Model specification.

Step 2: Model identification. If the model is not identified, then back to step 1.

Step 3: Model estimation. This is an iterative step to evaluate the model fitness. If the model fitness is poor, then the model needs to be re-specified from the beginning.

Step 4: Model testing. This step refers to the analysis of fitness of entire model and individual model parameters.

Step 5: Model modification. This is the final step to further improve the model fitness by using new parameters in the model if meaningful.

The framework of SEM procedure is shown in Figure 2-6.

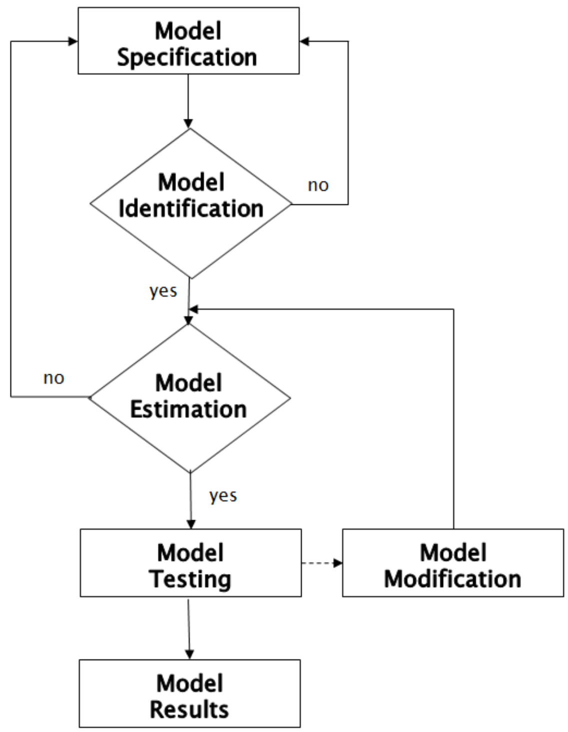


Figure 2-6 Framework of SEM Procedure

Compared with the linear-regression approach, SEM is flexible and able to analyze the relationships among both MVs (observed variables) and LVs (unobserved variables), including the relationships between dependent and independent variables, the effects of independent variables on another independent variable, the correlation between two independent variables, and the relationships among error terms. In addition, SEM considers both measurement and structural errors, while linear regression always ignores the measurement error. A drawback is that the formulation of SEM is based on researchers' hypotheses of variables and pathways, which may not be accurate.

As a multivariate statistical modeling method, SEM has been used in many fields of research (e.g. economic trend analysis, investment decisions, health issues, et al.) and also been gradually accepted in travel behavior studies for analyzing the household travel demand, the influence of neighborhood type land use system and car ownership on travel behaviors (Golob, 2003; Bagley and Mokhtarian, 2002; Acker et al. 2007; Acker and Witlox, 2010). However, the SEM application in forecasting freight movement is still limited.

In this section, the SEM technologies include Path Analysis, Confirmatory Factor Analysis (CFA) and Structural Regression Model (SRM) will be introduced respectively followed by the fitness measures.

2.5.1 Path Analysis

Path Analysis, which is also known as Causal Modeling, is a special case of SEM which only deals with measured variables and assumes no measurement errors. This method was first developed by a geneticist Swell Wright who attempted to study genetic influences in the 1920s, and then gained great popularity in social science research (Xue, 2007). Compared to linear regression models, Path Analysis considers the correlation between dependent variables and is able to distinguish direct effects and indirect effects directly from a path diagram (Garson, 2013).

2.5.2 Confirmatory Factor Analysis

Confirmatory Factor Analysis (CFA) is another special case of SEM to specify both the number of factors (LVs) and their corresponding indicators (MVs). CFA is the statistical procedure to form the measurement model in SEM (Brown, 2014). Brown and Moore summarized the advantages of CFA as follows:

- Enables researchers to evaluate their hypothesis of the relationships between MVs and the underlying constructs of them.
- LVs can help reduce variable dimension, represent the impact of unobserved factors, and improve the model's explanatory power and sensitivity.
- The measurement error can be considered and corrected in CFA.

2.5.3 Structural Regression Model

SRM, which is called full SEM model, is a core technology of SEM (Kline, 2011). As a synthesis of a measurement model and structural model, a SRM can test the hypotheses of direct effects and indirect effects similarly to Path Analysis, and can also involve latent variables by using CFA. Figure 2-7 provides a SRM path as an example to illustrate the relationships.

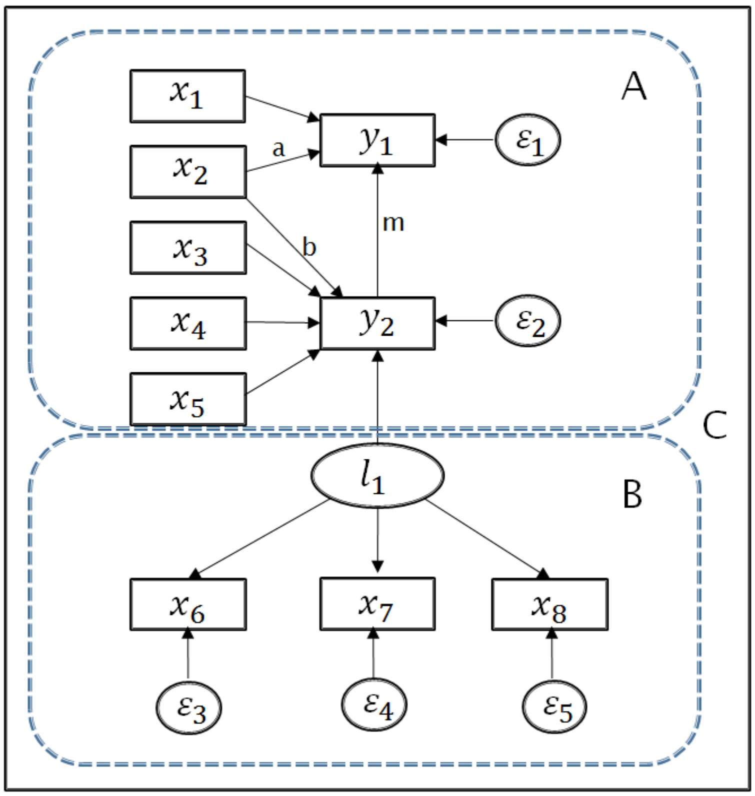


Figure 2-7 An Example of SRM

In this SRM, y_1 and y_2 are two dependent variables to be predicted. x_1 through x_8 are MVs representing observations. ε_1 to ε_5 are error terms for corresponding variables. y_1 is determined by three indicators: x_1, x_2 and y_2, while y_2 is predicted by x_3, x_4, x_5 and a latent variable l_1 which is measured by x_6, x_7 and x_8. a, b and m are coefficients between variables linked by the arrows. The model in the dashed box A is an example of path analysis. In this model, a is the direct effect from x_2 to y_1. The correlation existing between y_1 and y_2 also brings indirect effect b×m from x_2 to y_1. Thus the total effect of x_2 to y_1 is a+bm. Dashed box B presents a CFA model of l_1. The combination of structural model A and measurement model B is the final SRM.

2.5.4 Model Fitness Measures

The purpose of fitness measures is to evaluate the consistency of a SEM with the sample data and test hypothesis. Currently there are dozens of fit statistics for SEM, and new statistic are still coming out (Hooper, 2008). It is never necessary to report every fit statistic for evaluating a SEM, and also, there is no definitive fitness measure. The selection of fit statistics is in terms of modeling objective, model structure, and sample data size and distribution. It should be noted that the value of fit indices only represents the overall fit of a SEM (Kline, 2005). As a result, a SEM still can be inadequate even though the values of fitness measures are good. In addition, a SEM may not be theoretically significant with fit statistics since the direction of paths may be in the unexpected way (Chen, 2007). The final rationality of a model should be determined by both model fitness and case background. In this section, various fitness measures that commonly used in SEM will be introduced.

2.5.4.1 Absolute Fit Index

Chi-square(χ^2), which is also called likelihood ratio chi-square, is an original and typical absolute fit index representing the overall fit "badness" of a SEM which means the more significant the chi-square is, the more unacceptable the model is (Kline, 2005; Moss, 2017). The formula of chi-square is shown below (Equation 2-1).

$$\chi^2(df) = (N-1)F[S, \sum(\hat{\theta})] \qquad \text{Equation 2-1}$$

Where

χ^2: chi-square that ranges from 0 to +∞ ($\chi^2 = 0$ refers to perfect fit while $\chi^2 = +\infty$ represents poor fit)

df: degrees of freedom

N: sample size

S: unrestricted sample variance or covariance matrix

$\sum(\hat{\theta})$: variance or covariance matrix restricted by the SEM

It is obvious that the value of χ^2 is mainly affected by the sample size. For a sample size from 75 to 200, χ^2 can perform as a reasonable fit index. However, if the same size is too large (e.g. greater than 400), the χ^2 will always be significant which will lead to the rejection of the SEM (Kenny, et al., 2015). Whereas, when dealing with a small sample size, χ^2 may have a lack of power since the good and poor fitting models cannot be recognized. A model with complex structure and a lot of parameters will be more likely to have an acceptable χ^2 fit. In addition, the use of χ^2 test also requires multivariate normality assumption of the variables (Hooper, 2008).

Root mean square error of approximation (RMSEA) is another absolute measure of fit developed in the 1980s (Steiger, 1990). RMSEA measures the difference between observed and hypothesized covariance matrices per degree of freedom and has become popular in recent SEM studies (Chen, 2007). The RMSEA is estimated using Equation 2-2:

$$RMSEA = \frac{\sqrt{x^2 - df}}{\sqrt{df\,(N-1)}} \qquad \text{Equation 2-2}$$

The value of RMSEA depends on χ^2. RMSEA will be set to 0 when χ^2 is less than the degree of freedom. Similar to χ^2 , RMSEA is also an index of poorness of a model. It is suggested not to report RMSEA for models with low degrees of freedom since there will have greater simple error in RMSEA (Kenny, et al., 2015). Generally the value of RMSEA from 0 to 0.08 is acceptable for a well-fitting model (Hooper, 2008).

Other absolute fit indices also include Goodness-of-fit (GFI), Adjusted Goodness-of-fit (AGFI), Relative chi-square which equals to $\frac{\chi^2}{df}$, Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC) and Standardized Root Mean Square Residual (SRMR) etc. Most of them have the same issues as χ^2 since they are all transformations of χ^2, except SRMR.

SRMR is defined as the average standardized residuals between the observed covariance matrix and hypothesized covariance matrix (Chen, 2007). The calculation of this absolute fit index can be indicated as:

$$SRMR = \sqrt{\frac{\sum_{i=1}^{p}\sum_{j=1}^{i}[(s_{ij} - \hat{\sigma}_{ij})/(s_{ii}s_{jj})]^2}{p(p+1)/2}} \qquad \text{Equation 2-3}$$

Where

p : number of observed variables;

s_{ii} and s_{jj} : observed standard deviations;

s_{ij} : observed covariance;

$\hat{\sigma}_{ij}$: reproduced covariance.

SRMR requires large data sample size and high parameter quantity (Hooper, 2008). The range of SRMR is from 0 to 1, where 0 represents perfect fit. A value of 0.08 is accepted as the highest value for a SEM (Hu and Bentler, 1999).

2.5.4.2 Incremental Fit Index

Normed-fit Index (NFI) and Comparative-fit Index (CFI) are two main types of incremental fit measures that represents the fit of a target model (model of interest) and a null model (model with uncorrelated variables) (Moss, 2017). The mathematic expressions of these two indices are:

$$NFI = \frac{\chi_n^2 - \chi_t^2}{\chi_n^2} \qquad \text{Equation 2-4}$$

$$CFI = 1 - \frac{\max[(\chi_t^2 - df_t), 0]}{\max[(\chi_t^2 - df_t), (\chi_n^2 - df_n), 0]} \qquad \text{Equation 2-5}$$

NFI is a ratio that is determined by the chi-square of the null model and a target model. Compared with NFI, CFI also has the concern of degree of freedom. Since incremental fit measures have the assumption that all the MVs are uncorrelated in the null model, the chi-square of the null model with a large number refers to poor fit of the model. The range for both NFI and CFI is from 0 (poor fit) to 1(good fit). A NFI or CFI that greater than 0.95 is suggested to be a criteria for determining a good fit (Hu and Bentler, 1999).

3 FRAMEWORK DEVELOPMENT METHODOLOGIES

In this chapter, the methodologies for developing the proposed freight demand forecasting framework are presented in detail with the fundamentals, formulas and explanations. Since this framework will be applied in the California Statewide Freight Forecasting Model (CSFFM) based on the Freight Analysis Framework Version 4 (FAF4) data, the formulas are given in FAF zone format. However, the applications of these methodologies are not limited to FAF zone levels and can be applied to other scales of zones as well.

3.1 Freight Economic Centroid

In the transportation modeling network, a centroid is usually defined as the geometric center of each TAZ representing the origin or destination of each trip (LSA, 2008). In the recent decade, a population weighted centroid was created with the purpose to accurately reflecting the personal activities concentration area, and has been applied in many GIS-based researches including migration, health condition and other personal activity studies (Sarkar and Vick, 2011; Rushton, et al., 2007). This concept is also relevant to freight movement forecasting which usually has a coarser zone system than passenger demand models. The locations of freight centroids have a non-negligible impact on route choice and total vehicle miles traveled (VMT). Park et.al made initial effort to estimate centroids weighted from city employment data for FAF2 zones (Park, et al., 2011). However, this method didn't provide constraint to limit the location of weighted centroid within the same zone boundary. Therefore, based on the location of geometric centroids, this research proposed freight economic centroids to represent the center of freight activities by considering employment effects using an optimization method. The formula used for computing the freight economic centroid is shown as follows:

Objective: $Min\ f\big(Z(X,Y)\big) = \left(\sum_i e_i x_i - EX\right)^2 + \left(\sum_i e_i y_i - EY\right)^2$

Subject to: $Z(X, Y) \in \text{zone j, for any county i in FAF zone j}$ Equation 3-1

Where

e_i: the total employment in county i;

E: the total employment in FAF zone j;

(x_i, y_i): the geometric centroid in county i.

In a FAF zone system, some reminder of state zones consist of several nonadjacent zones. The constraint $\mathrm{Z(X, Y)} \in \text{zone j}$ aims to ensure that the centroid falls within the same analysis zone. For instance, the Remainder of California zone in FAF4 contains three separate areas (Figure 2-4). Without the constraint, the centroid of Remainder of California will be located inside the Fresno-Madera zone (zone 065). While by applying the constraint, the freight economic centroid is adjusted to the south of the Fresno-Madera zone and still within the Reminder of California area (Figure 3-1).

Figure 3-1 Centroid Comparison in Remainder of California (FAF4)

3.2 Aggregated Commodity Groups using Fuzzy C-means Clustering Method

The two-digit Standard Classification of Transported Goods (SCTG) code is a standard commodity classification used in the United States which includes 43 commodities. However, the explanatory variables for estimating shipment behaviors are limited for some commodities (Ranaiefar, 2013). In this research, an aggregated commodity group (CG) clustering method is developed using Fuzzy C-means based on the similarity of mode share of each commodity with the following principles:

1) Aggregate the two-digit SCTG commodities based on the similarity of mode share distribution of each commodity.

2) Keep the sequence of 9 SCTG groups defined for FAF4.

3) Match the available explanatory variables in model estimation for each CG.

Fuzzy C-means is a clustering algorithm used to find the similarity patterns of data points and classify them into more than one cluster (Bezdek, Et al., 1984). Fuzzy C-means has been applied in many fields such as bioinformatics, marketing management and image processing (Lecca, 2011; Chou, 2008; Shihab, 2001; Chuang, et al. 2006). Compared with non-fuzzy clustering methods where each data point is only assigned to one group, Fuzzy C-means provides potential cluster choices for each data point by measuring their connectivity and intensity (Wolfram, 2017).

The Fuzzy C-means algorithm aims to minimize the objective function in an iterative optimization process shown below (Chuang, et al., 2006):

$$\text{Min } J_m = \sum_{i}^{N}\sum_{j}^{C}\mu_{ij}^{m} \| x_i - c_j \|^2 ,\ 1 < m < \infty \qquad \text{Equation 3-2}$$

Where

m is any real number greater than 1. It defines the degree of fuzzification;

x_i is the ith data in d-dimensional measured data set X;

c_j is the center of the cluster j;

μ_{ij} is the probability of sample x_i in the cluster j;

$\| x_i - c_j \|$ is measured the distance (or similarity) between x_i and c_j.

3.3 Weighted Travel Distance

Highway skims measure truck travel distance but are not representative of the average travel distance for all good movement modes. Additionally, freight analysis zones (FAZ) are always much larger than TAZs due to the lower resolution of available freight data. For example, the California Statewide Travel Demand Model (CSTDM) has approximately 5,400 TAZs, while the corresponding study area only resulted in only 97 FAZs for the California Statewide Freight Forecasting Model (CSFFM). Thus, for each OD pair, the travel distances of different commodities can vary significantly. Since FAF4 provides FAF4 zone level Tons and Ton-miles information of each O-D pair by mode and commodity, the weighted travel distance skim can be calculated by the following formulations:

$$\overline{D_{ij}^{mn}} = \frac{TM_{ij}^{mn}}{T_{ij}^{mn}}$$

$$\overline{D_{ij}^{n}} = \frac{\sum_{m=1}^{M} T_{ij}^{mn} * \overline{D_{ij}^{mn}}}{\sum_{m=1}^{M} T_{ij}^{mn}} = \frac{TM_{ij}^{n}}{T_{ij}^{n}} \qquad \text{Equation 3-3}$$

Where

$\overline{D_{ij}^{mn}}$: The average travel distance between zone i and zone j of mode m and commodity n;

$\overline{D_{ij}^{n}}$: The average travel distance between zone i and zone j of commodity n;

TM_{ij}^{mn}: Ton-miles between zone i and zone j of mode m and commodity n;

TM_{ij}^{n}: Ton-miles between zone i and zone j of commodity n;

T_{ij}^{mn}: Tons between zone i and zone j of mode m and commodity n;

T_{ij}^{n}: Tons between zone i and zone j of commodity n;

3.4 Freight Demand Forecasting Model using Structural Regression Model

Corresponding to the framework application in CSFFM, the freight demand forecasting model contains two parts: total freight generation model and domestic direct demand model that integrated freight generation and distribution steps. The models are developed using the Structural Equation Model (SEM) function in the STATA program. STATA provides four estimation methods of SEM (StataCorp, 2013):

- Maximum likelihood (ML): default method and can report SEM fitness measures including chi-square, RMSEA, CFI and SRMR. ML has the assumption that all variables (both observed variables and latent variables) are joint normalized.
- Quasimaximum likelihood (QML): QML has more relaxed normality assumptions, and can report SRMR as main SEM fitness index.
- Asymptotic distribution free (ADF): doesn't have normality assumption but requires large sample size.
- Maximum likelihood with missing values (MLMV): aims to mine information from the observations that contain missing data.

The estimation method chosen for each model will be discussed below.

3.4.1 Total Freight Generation Model

The purpose of the freight generation step is to estimate the total cargo production and consumption at each zone by identifying the effects of explanatory variables (demographic, socioeconomic, industrial characteristics, etc.). In this research, a Structural Regression Model (SRM) structure was developed in order to capture direct effects, indirect effects and underlying relations among variables. The general formula of total freight generation model is:

$$P_i^m = \underbrace{\sum_q \alpha L_{in}}_{\text{\{Latent Variables\}}} + \underbrace{\sum_k \beta X_{ik}}_{\text{\{Zonal Variables\}}} + \underbrace{\sum_l \gamma P_i^n}_{\text{\{Production of Other CGs\}}} + \underbrace{\sum_p \delta C_i^n}_{\text{\{Consumption of Other CGs\}}} + \underbrace{e_{mi}}_{\text{\{Error term\}}}$$

Equation 3-4

$$C_i^m = \underbrace{\sum_q \alpha' L_{in}}_{\text{\{Latent Variables\}}} + \underbrace{\sum_k \beta' X_{ik}}_{\text{\{Zonal Variables\}}} + \underbrace{\sum_l \gamma' P_i^n}_{\text{\{Production of Other CGs\}}} + \underbrace{\sum_p \delta' C_i^n}_{\text{\{Consumption of Other CGs\}}} + \underbrace{e'_{mi}}_{\text{\{Error term\}}}$$

Equation 3-5

The estimation method of SRM is in terms of the normality and sample size of input data. Bentler and Chow, and Barnes all discussed the non-normal distribution of socioeconomic data in their studies (Bentler and Chow, 1987; Malthouse, 2001). According to Ranaiefar, the measured variables used to develop the total generation model in the original CSFFM had non-normal distributions (Ranaiefar, 2013). Under this situation, Kline suggests that ADF and QML can be the estimation methods in dealing with non-normalized data, but ADF requires a large sample size (Kline, 2011). In the application in CSFFM, the total generation model is built based on FAF4 data which only has a small sample size of 132. As a result, QML was finally selected as the estimation method.

3.4.2 Domestic Flow Direct Demand Model

To forecast the domestic freight flow between each zone, a direct demand modeling approach with SRM structure was developed in this research. This modeling method can avoid accumulated error in separate generation and distribution approaches, and also accords with the determination process of freight flow which is a synthetic action of producers, shippers and consumers. The origin zone variables and destination zone variables are both used in the estimation of zonal flow. For intra-zonal movements, the origins and destinations are the same zone. To avoid duplicate consideration of effects of zonal attributes, it is necessary to divide the domestic direct demand model into two scale levels: intra-zonal model and inter-zonal model. The mathematical formulations of domestic direct demand model are:

Intra-zonal Model:

$$f_{ij}^{m} = \underbrace{\sum_{n \neq m} \alpha^{n} f_{ij}^{n}}_{\text{\{Relationship with other CGs}} + \underbrace{\sum_{l} \beta_{l}^{m} X_{il}}_{\text{\{Zonal Variables\}}} + \underbrace{\sum_{r} \delta_{r}^{m} L_{r}}_{\text{\{Latent Variables\}}} + \underbrace{\lambda^{m} dist_{ij}^{m}}_{\text{\{CG Based Distance\}}} + \underbrace{\varepsilon_{ij}^{m}}_{\text{\{Error term\}}}$$

Equation 3-6

Inter-zonal Model:

$$Ln(f_{ij}^{m}) = \sum_{n \neq m} \alpha^{n} Ln(f_{ij}^{n}) + \sum_{l} \beta_{l}^{m} Ln(X_{il}) + \sum_{k} \gamma_{k}^{m} Ln(X_{jk}) + \sum_{r} \delta_{r}^{m} L_{r} + \lambda^{m} Ln(dist_{ij}^{m}) + \varepsilon_{ij}^{m}$$

{Relationship with other CGs + {Origin Variables} + {Destination Variables} + {Latent Variables} + {CG Based Distance} + {Error term}

Equation 3-7

Facing the same non-normalized data issue as the total generation model, the intra-zonal model is also estimated using the QML method. In the inter-zonal model, a log-linear format is used for explanatory variables to improve their normalization which makes it suitable for ML estimation.

4 FRAMEWORK APPLICATION: AN UPDATE OF CALIFORNIA STATEWIDE FREIGHT FORECASTING MODEL

The California Statewide Freight Forecasting Model (CSFFM) is a commodity-based freight forecasting tool developed by the Institute of Transportation Studies (ITS) at the University of California, Irvine and funded by the California Department of Transportation (Caltrans). In this chapter, the proposed exploratory framework is applied to update the original CSFFM using Freight Analysis Framework Version 4 (FAF4) data. Results of the original CSFFM and the new framework are also compared to illustrate the advantages of the proposed framework.

4.1 Data Preparation

4.1.1 FAF4 Data

FAF4 provides tonnage, ton-miles and values information on FAF zone level (132 FAF zones) and state level (50 states in the United States and District of Columbia) by commodity and mode. FAF4 has the base year of 2012 and complete database of 2013 to 2015. In addition, FAF4 includes 5-year forecasts from 2020 to 2045. Table 4-1 summarized the FAF4 data and usage in this research.

Table 4-1 FAF4 Data Used in this Research

Year	Geographic level	Data Description	Usage
2012	132 FAF4 Zones	Shapefile	Centroid identification
2012	132 FAF4 Zones	Total flow	Calibration of total generation model
2012	132 FAF4 Zones	Domestic flow	Calibration of domestic direct demand model
2012	132 FAF4 Zones	Ton-miles	Calculation of commodity-based average distance

4.1.2 Demographic and Socioeconomic Data

The official 2012 demographic and socioeconomic data are collected and imputed by California Department of Transportation (Caltrans) with the following updates (Table 4-2):

Table 4-2 2012 Observation Data

List of Variables	2012 Status
Population	updated in county level
Employment	updated in county level
Establishment	updated in county level
Harvested Land	updated in county level
GDP	updated in county level
Sold Livestock	updated in county level
Capacity of Refineries	use 2007 data
Fuel Price	updated in county level
Payload by FHWA	use 2007 data
Empty Factor	use 2007 data
Rail Price	use 2007 data
Toll	updated slightly. Not significant

The county level data was then aggregated to FAF4 zone levels. It should be noted that in the above table, the capacity of refineries, fuel Price, payload factor, empty factor and rail price are not updated by Caltrans. Only the updated data were used in model development in this study.

4.2 Identification of Freight Economic Centroids

Using the methodology introduced in Chapter Three, the freight economic centroids are identified for each FAF zone in the United States. An overall comparison map of geometric centroids and freight economic centroids is shown in Figure 4-1.

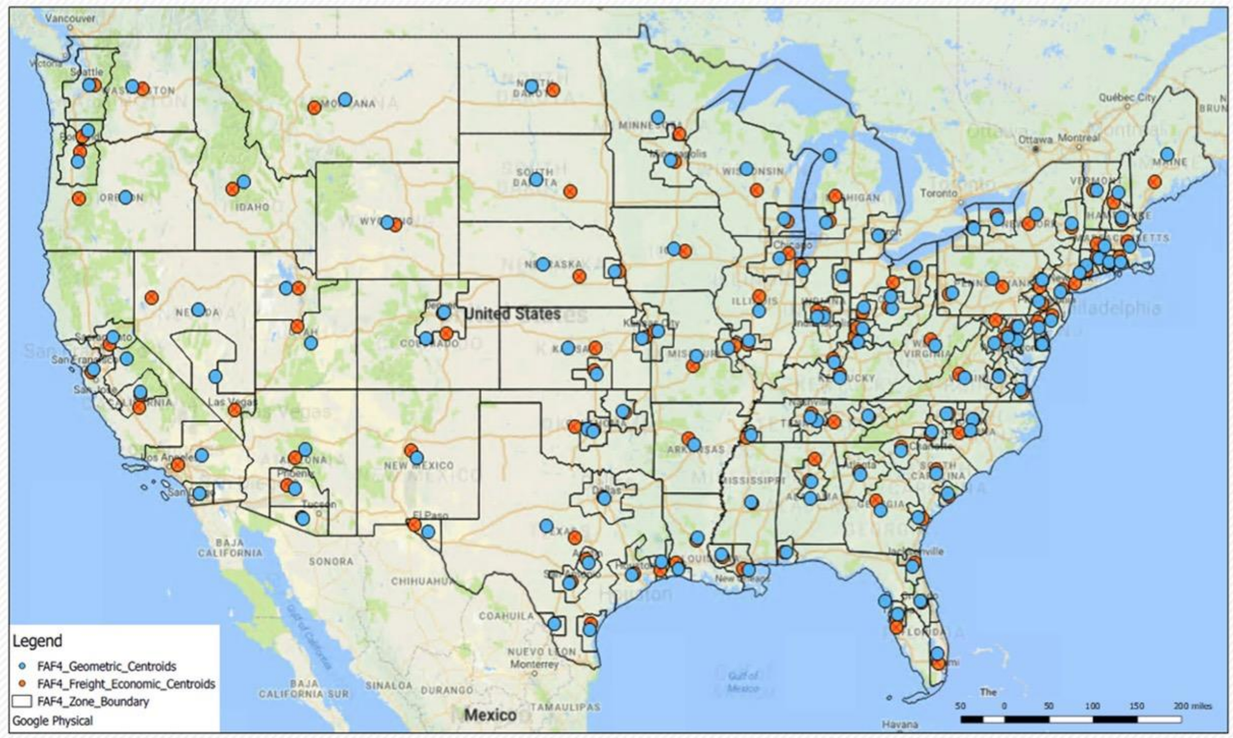


Figure 4-1 Freight Economic Centroids Location Map

In above map, blue points represent the locations of geometric centroids for each FAF zone, and the red points are the corresponding freight economic centroids. In general, the new centroids are closer to the highway system compared to geometric centroids and provide potential convenience for transporting goods using the highway system.

An amplifying map of California and surrounding states is presented in Figure 4-2 for detailed analysis. In FAF zone 061 (Los Angeles-Long Beach), the original geometric centroid was located in San Bernardino County. The freight economic centroid is now seated close to Los Angles Port and Long Beach Port where has frequent freight movement activities. Another obvious centroid change is in FAF zone 069 (Remainder of California) which includes three separate areas located in northern, central and southern California. The new centroid moves to the south of FAF zone 065 (Fresno-Madera) and near I-5.

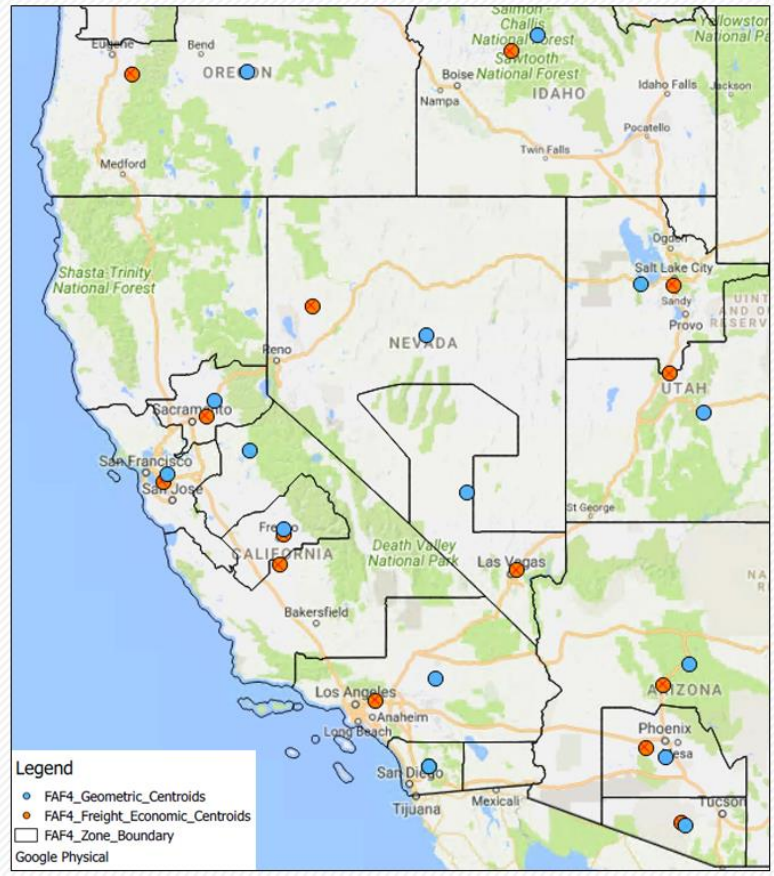


Figure 4-2 Freight Economic Centroids Map of California and Surrounding States

Significant changes also occur in the states surrounding California. In Nevada state, the geometric center of FAF zone 321 (Las Vegas-Henderson) is seated in the deserted area which is not reasonable to represent the origin and destination of freight flows. The new centroid moves south to Las Vegas downtown and is located near I-15. In FAF zone 329 (Remainder of Nevada), the freight economic centroid is situated near I-80 and the boundary of the State of California. Regarding to the Remainder of Oregon zone (Figure 4-3), the freight economic centroid moves west and is close to I-5.

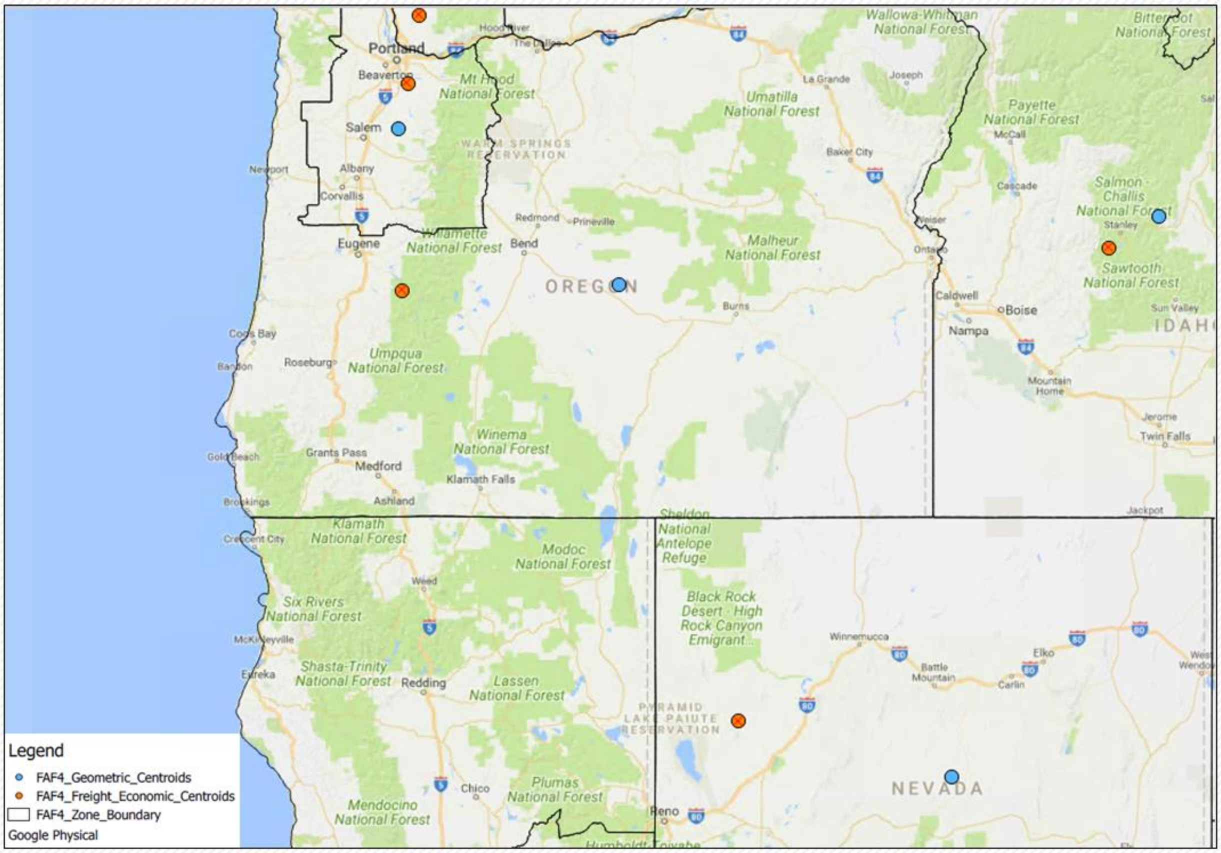


Figure 4-3 Freight Economic Centroids Location Map of Remainder of Oregon

4.3 Aggregation of Freight Commodity Groups

To aggregate freight commodities using the Fuzzy C-means (FCM) algorithm, a significant step is to set a criterion to recognize similarity of different commodities. In the original CSFFM, the Mode Split Module is the second module to decide the mode share of each origin-destination (O-D) by commodity group (CG). Considering similar mode shares in commodity aggregation can help reduce errors in the mode split step and overcome the shortage of explanatory variables for certain commodities (Ranaiefar, 2013).

In this research, the dominant mode share is used to classify CGs. The mode share distribution of each commodity was calculated as input for FCM. Three clusters are pre-defined before the clustering process: cluster 1 for truck dominant commodity, cluster 2 for rail dominant commodity and cluster 3 for pipeline and other dominant commodity. The FCM results and mode share distribution of each commodity group are presented from Table 4-3 to Table 4-14, and Figure 4-4 to Figure 4-15. In the result tables, column two to four show the commodity's probability to be classified to each cluster. The cluster ID in column five is the recommended cluster provided by FCM. Besides, the FCM algorithm also provides a potential Neighbor Cluster ID as reference for users when the Cluster ID is not acceptably limited by other constraints.

Table 4-3 FCM Results of CG1 (Agriculture products and fish)

2-Dig SCTG	Cluster 1	Cluster 2	Cluster 3	Cluster ID	Neighbor Cluster ID
1	97%	2%	1%	1	2
2	86%	11%	3%	1	2
3	100%	0%	0%	1	-
4	100%	0%	0%	1	-
5	98%	2%	0%	1	2

SCTG: Standard Classification of Transported Goods

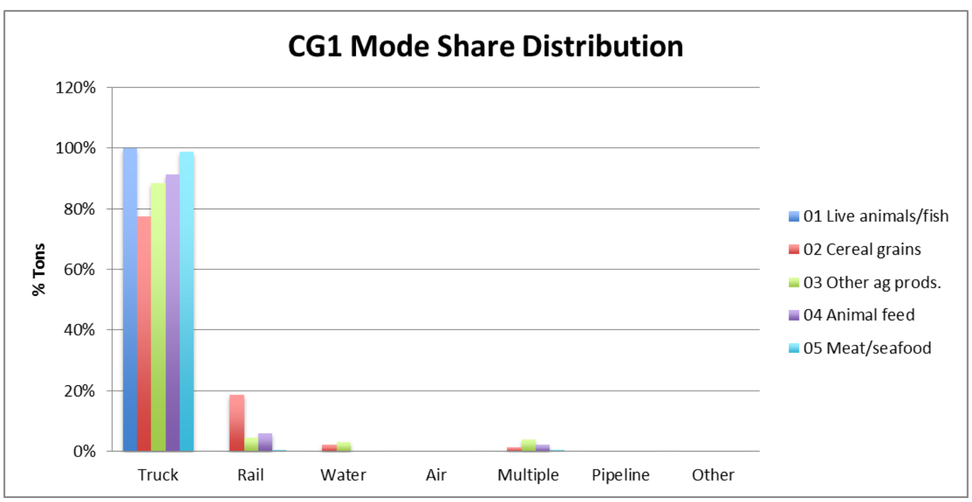


Figure 4-4 Mode Share Distribution of CG1 (Agriculture products and fish)

The aggregation of CG1 (Agriculture products and fish) is a good example to illustrate the clustering function of FCM. In Table 4-3, all of the five SCTG commodities have the highest probability (all above 85 percent) falling in cluster 1 (truck dominant). The obvious pattern shown in CG1 mode share distribution presents the rationality to group them into the same CG. The final grouping result of CG1 is also consistent with SCTG Group 1.

Table 4-4 FCM Results of CG2 (Food, alcohol, and tobacco products)

2-Dig SCTG	Cluster 1	Cluster 2	Cluster 3	Cluster ID	Neighbor Cluster ID
6	96%	3%	1%	1	2
7	100%	0%	0%	1	-
8	100%	0%	0%	1	-
9	98%	1%	1%	1	2

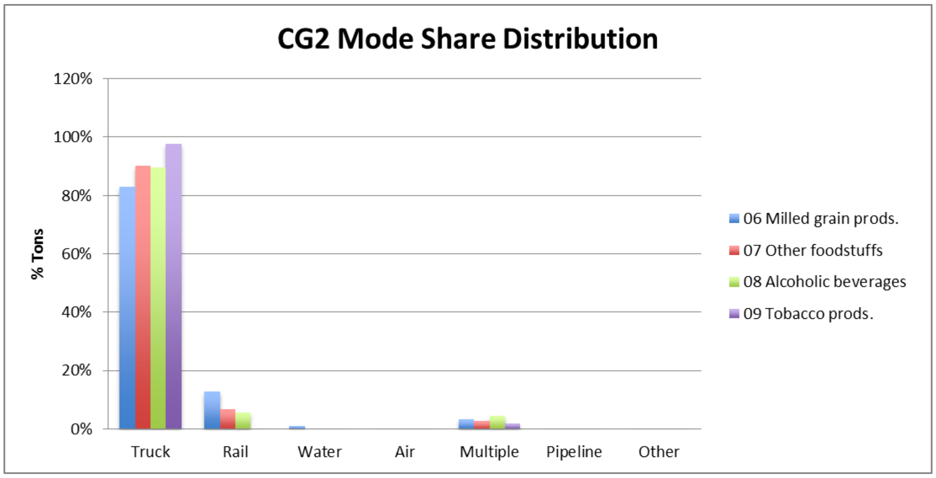


Figure 4-5 Mode Share Distribution of CG2 (Food, alcohol, and tobacco products)

CG2 represents Food, alcohol, and tobacco products which are dominantly transported using truck. The results of FCM shows significant similarity pattern in cluster 1 (the probabilities of all four commodities in CG2 are above 95 percent). The commodities in CG2 are the same as SCTG Group 2.

Table 4-5 FCM Results of CG3 (Stones, non-metallic minerals)

2-Dig SCTG	Cluster 1	Cluster 2	Cluster 3	Cluster ID	Neighbor Cluster ID
10	99%	1%	0%	1	2
11	99%	1%	0%	1	2
12	100%	0%	0%	1	2
13	91%	7%	2%	1	2

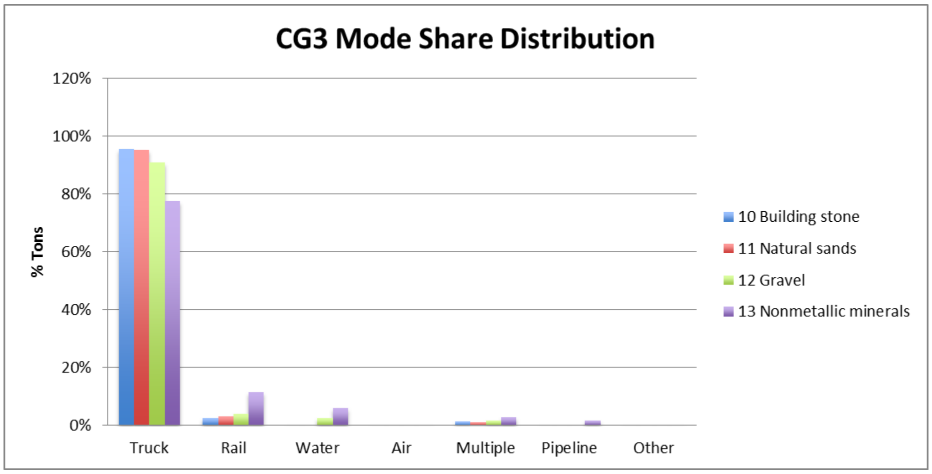


Figure 4-6 Mode Share Distribution of CG3 (Stones, non-metallic minerals)

Table 4-6 FCM Results of CG4 (Coal and metallic mineral ores)

2-Dig SCTG	Cluster 1	Cluster 2	Cluster 3	Cluster ID	Neighbor Cluster ID
14	8%	84%	8%	2	1
15	5%	90%	5%	2	1

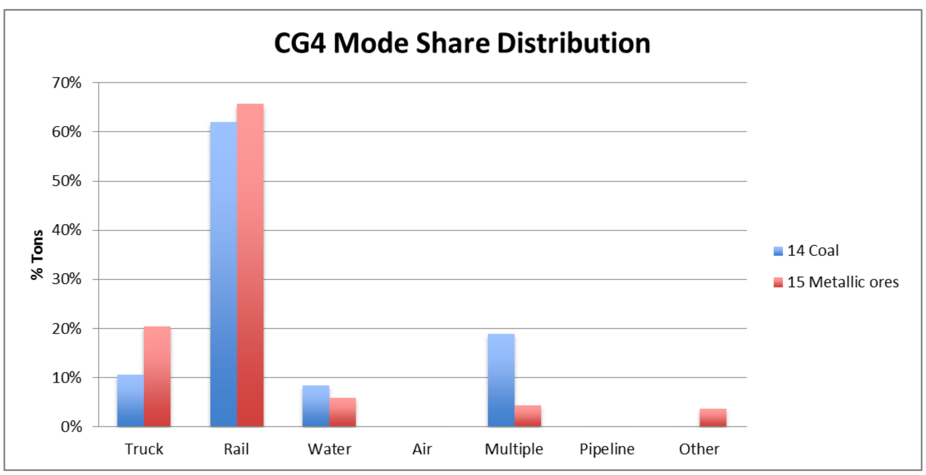


Figure 4-7 Mode Share Distribution of CG4 (Coal and metallic mineral ores)

SCTG Group 3 covers SCTG 10 to 14 for Stones, non-metallic minerals, and metallic ores. Informed by the FCM results, SCTG 10 to 13 are mainly carried on trucks and grouped to

CG3. SCTG 14 (Coal) has a different dominant transport mode in rail and is classified together with SCTG 15 (Metallic ores) to form CG4 in this study.

Table 4-7 FCM Results of CG5 (Crude petroleum and other coal and petroleum that not elsewhere classified)

2-Dig SCTG	Cluster 1	Cluster 2	Cluster 3	Cluster ID	Neighbor Cluster ID
16	1%	2%	97%	3	2
19	1%	1%	98%	3	2

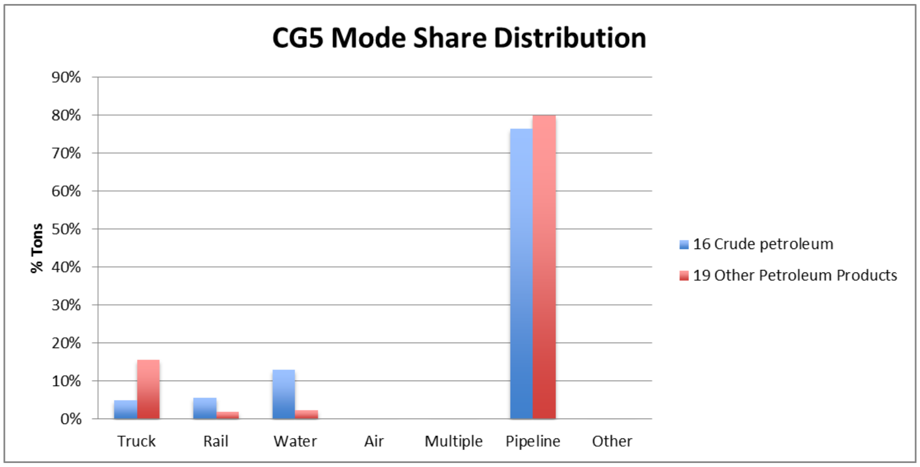


Figure 4-8 Mode Share Distribution of CG5 (Crude petroleum and other coal and petroleum that not elsewhere classified)

Although SCTG 16 (Crude petroleum) and 19 (Other coal and petroleum that not elsewhere classified) are not sequential commodities, they are grouped to CG5 due to the similar primary mode share using Pipeline and other modes.

Table 4-8 FCM Results of CG6 (Fuel and oil products)

2-Dig SCTG	Cluster 1	Cluster 2	Cluster 3	Cluster ID	Neighbor Cluster ID
17	43%	27%	30%	1	3
18	35%	28%	29%	1	3

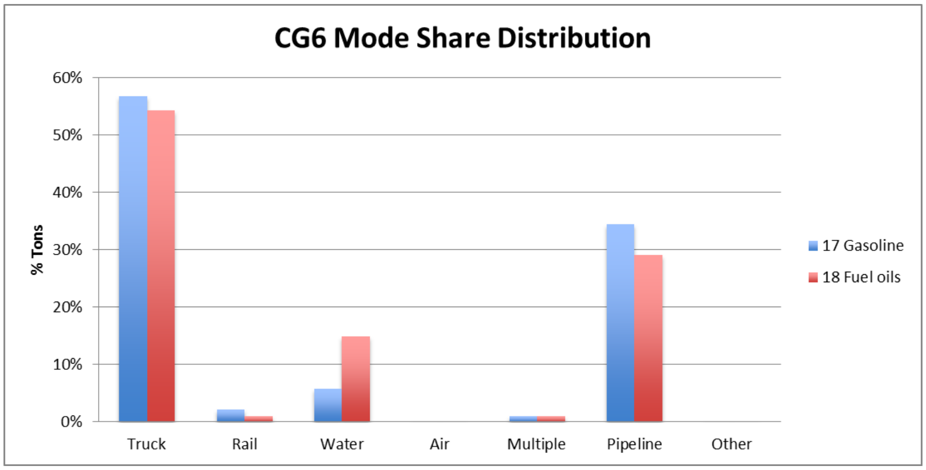


Figure 4-9 Mode Share Distribution of CG6 (Fuel and oil products)

Unlike previous CGs that have an obvious and dominant travel mode, SCTG 17 (Gasoline) and 18 (Fuel oils) in CG6 have even mode share. The FCM results of recommended Cluster ID and potential Neighbor Cluster ID show that assigning them in cluster 1 or cluster 3 are both acceptable. Considering the similar liquid properties and usage of them, they are grouped as CG6 finally.

Table 4-9 FCM Results of CG7 (Pharmaceutical and chemical products)

2-Dig SCTG	Cluster 1	Cluster 2	Cluster 3	Cluster ID	Neighbor Cluster ID
20	30%	60%	10%	2	1
21	97%	2%	1%	1	2
22	45%	48%	7%	2	1
23	100%	0%	0%	1	-
24	64%	29%	6%	1	2

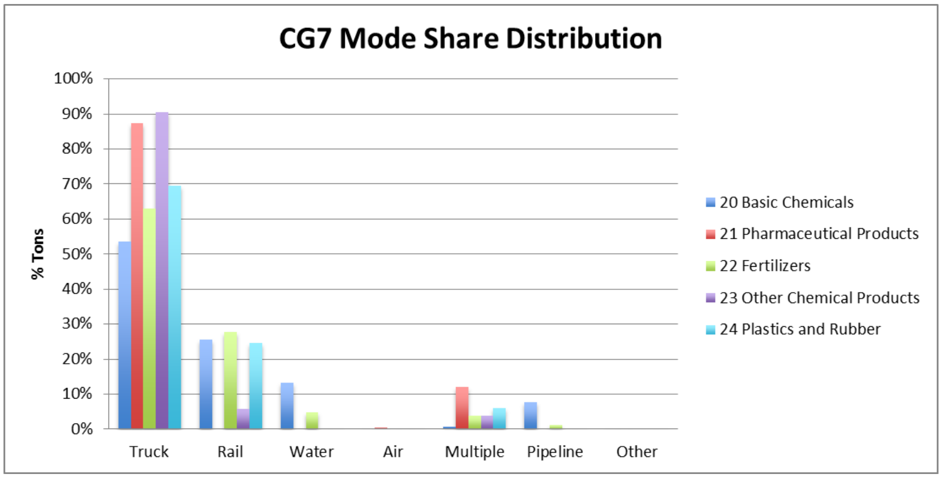


Figure 4-10 Mode Share Distribution of CG7 (Pharmaceutical and chemical products)

SCTG Group 5 includes SCTG 20 to 24 as Pharmaceutical and chemical products. However, observed from the FCM results in Table 4-9, there are two patterns mode share of these five commodities. SCTG 21 (Pharmaceutical products), 23 (Other chemical products) and SCTG 24 (Plastics and rubber) have dominant truck transport mode. SCTG 20 (Basic chemicals) and 22 (Fertilizers) are mainly shipped using rail. In this research, Caltrans only collected employment data for 3-digit North American Industry Classification System (NAICS) 325 (Chemical Manufacturing) from United States Census County Business Patterns (CBP) to keep consistent with original CSFFM. To match the same level employment data, the classification of SCTG Group 5 is kept and renamed to CG7 in this study.

Table 4-10 FCM Results of CG8 (Logs, wood products, and textile and leather)

2-Dig SCTG	Cluster 1	Cluster 2	Cluster 3	Cluster ID	Neighbor Cluster ID
25	97%	2%	1%	1	2
26	100%	0%	0%	1	-
27	86%	11%	3%	1	2
28	99%	1%	0%	1	2
29	99%	1%	0%	1	2
30	99%	1%	1%	1	2

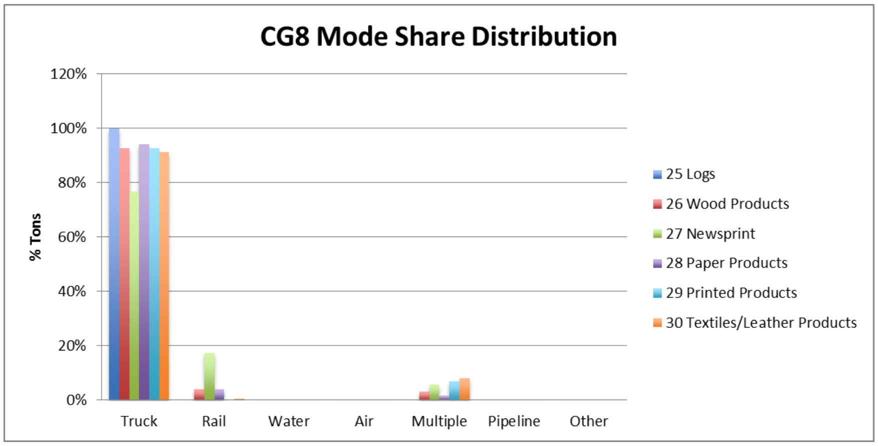


Figure 4-11 Mode Share Distribution of CG8 (Logs, wood products, and textile and leather)

An obvious truck dominant mode share pattern was found from SCTG 25 to 29 and led to the formation of CG8 (Logs, wood products, and textile and leather). This grouping is consistent with SCTG Group 6.

Table 4-11 FCM Results of CG9 (Base metal and machinery)

2-Dig SCTG	Cluster 1	Cluster 2	Cluster 3	Cluster ID	Neighbor Cluster ID
31	99%	1%	0%	1	2
32	96%	3%	1%	1	2
33	100%	0%	0%	1	-
34	99%	1%	0%	1	2

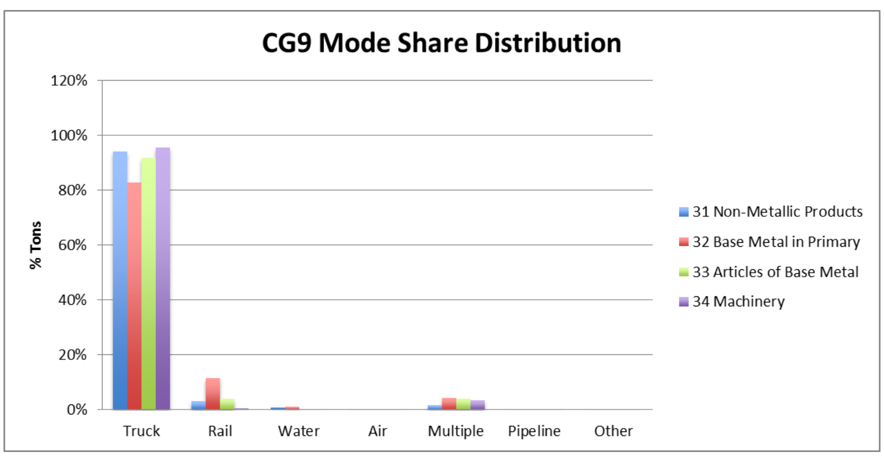


Figure 4-12 Mode Share Distribution of CG9 (Base metal and machinery)

CG9 represents Base metal and machinery including SCTG 31 (Non-metallic products), 32 (Base metal in primary), 33(Article of base metal) and 34 (Machinery). The primary mode of CG9 is truck which is suitable for carrying commodities with heavy weight.

Table 4-12 FCM Results of CG10 (Electronic, motorized vehicles, and precision instruments)

2-Dig SCTG	Cluster 1	Cluster 2	Cluster 3	Cluster ID	Neighbor Cluster ID
35	99%	1%	0%	1	2
36	99%	1%	0%	1	2
37	36%	56%	8%	2	1
38	62%	26%	12%	1	2

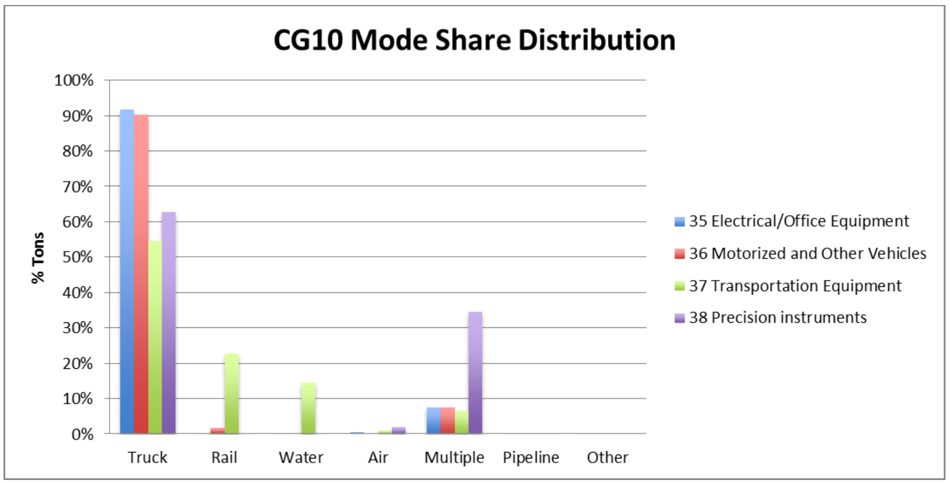


Figure 4-13 Mode Share Distribution of CG10 (Electronic, motorized vehicles, and precision instruments)

The FCM results in Table 4-12 support the aggregation of SCTG 35 (Electrical/office equipment), 36 (Motorized and other vehicles) and 38 (Precision instruments) whose dominant mode is truck. However, when considering the available employment data, motor vehicle manufacturing (NAICS 3361) is a sub-classification of Transportation equipment manufacturing (NAICS 336) in NAICS. But the 4-digit NAICS data was not collected by Caltrans. The Neighbor Cluster ID suggests that SCTG 37 (Transportation equipment) can be classified in the same group of the other three. The aggregation of CG10 is consistent with SCTG Group 8.

Table 4-13 FCM Results of CG11 (Furniture, mixed freight, misc. manufactured products, and commodity unknown)

2-Dig SCTG	Cluster 1	Cluster 2	Cluster 3	Cluster ID	Neighbor Cluster ID
39	99%	1%	0%	1	2
40	99%	1%	0%	1	2
41	100%	0%	0%	1	-
43	98%	1%	1%	1	2
99	11%	82%	7%	2	1

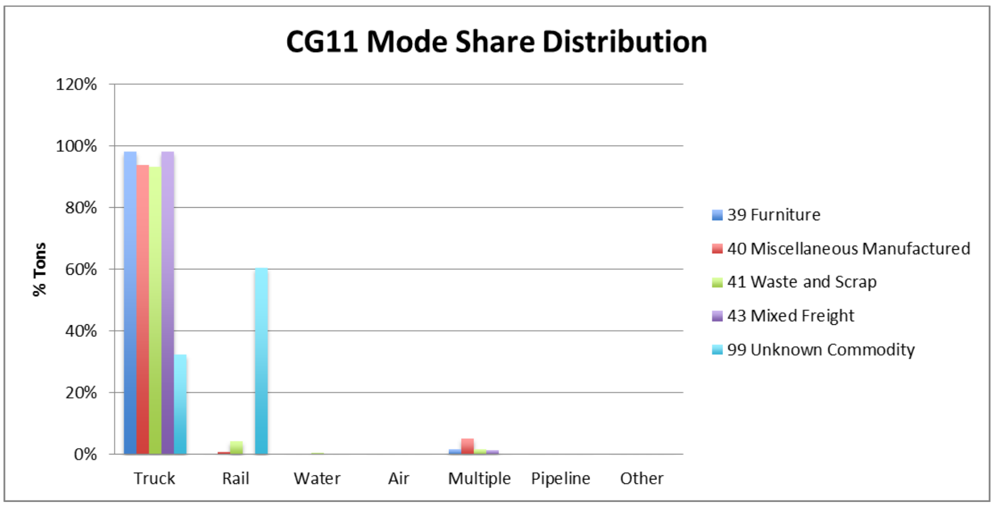


Figure 4-14 Mode Share Distribution of CG11 (Furniture, mixed freight, misc. manufactured products, and commodity unknown)

CG11 contains mixed commodities types covering from SCTG 39 to 43. Although the mode share distribution of SCTG 99 (Unknown commodity) is different from others and recommended to be grouped in cluster 2, it is still kept in CG11 in accord with SCTG Group 9.

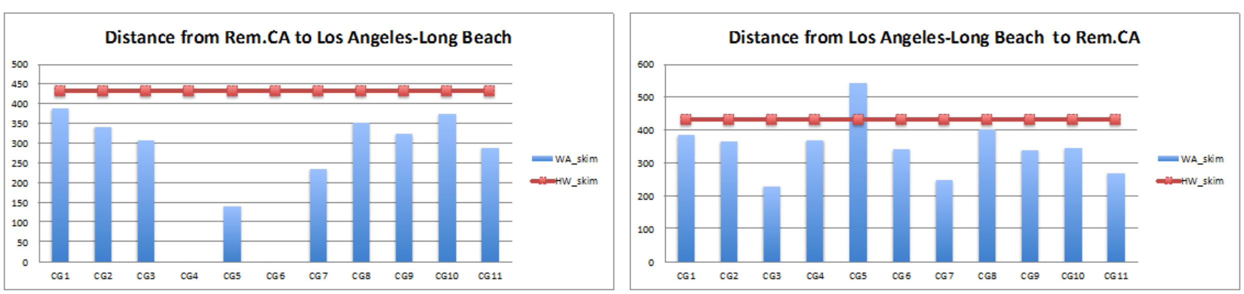
A summary of aggregation results of CGs are presented in Table 4-14. Seven of the eleven CGs are consistent with SCTG groups. CG3 to CG6 are the re-combinations of SCTG Group 3 and 4 based on the FCM results.

Table 4-14 Summary of Commodity Groups

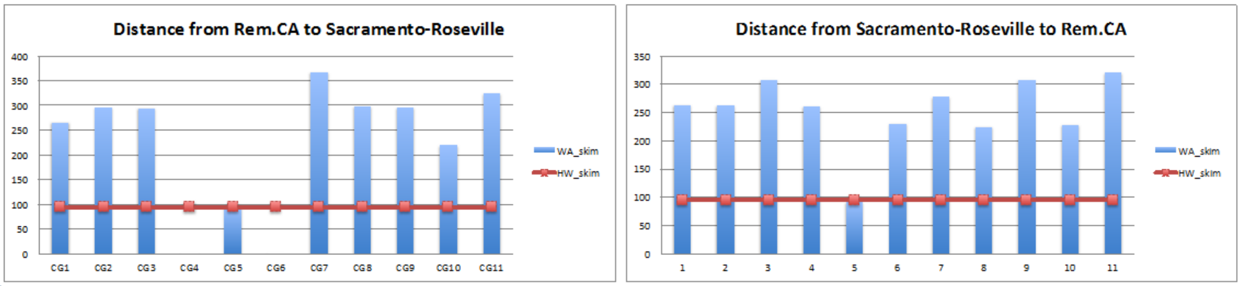
CG	SCTG two-digit covered	Description	% Share of Total
1	1-5	Agriculture products and fish	12.01%
2	6-9	Food, alcohol, and tobacco products	5.43%
3	10-13	Stones, non-metallic minerals	15.26%
4	14-15	Coal and metallic mineral ores	7.81%
5	16,19	Crude petroleum and Other not elsewhere classified	17.19%
6	17-18	Fuel and oil products	12.26%
7	20-24	Pharmaceutical and chemical products	5.96%
8	25-30	Logs, wood products, and textile and leather	6.03%
9	31-34	Base metal and machinery	9.42%
10	35-38	Electronic, motorized vehicles, and precision instruments	1.54%
11	39-43	Furniture, mixed freight, misc. manufactured products, and commodity unknown	7.09%

4.4 Calculation of Weighted Travel Distance

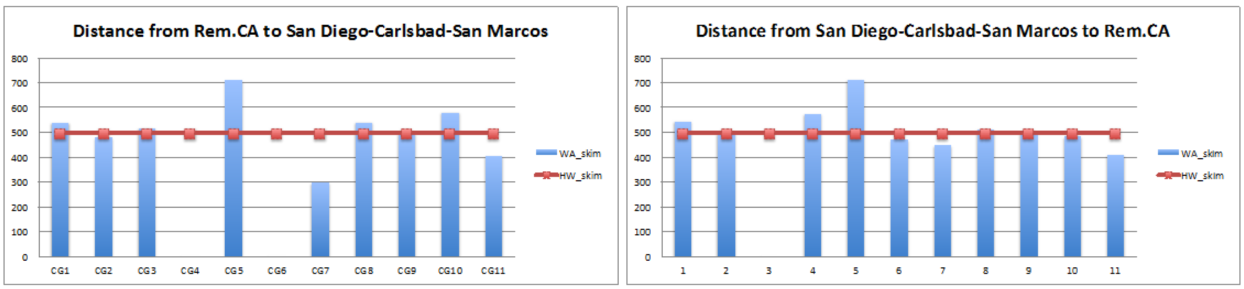
Using the methodology introduced in Chapter three, the weighted travel distance is calculated using Tons and Ton-miles for each FAF O-D pair by CG. Figure 4-15 shows the weighted travel distance between the Remainder of California and each California FAF zone. The highway distances are also shown in red for comparison.



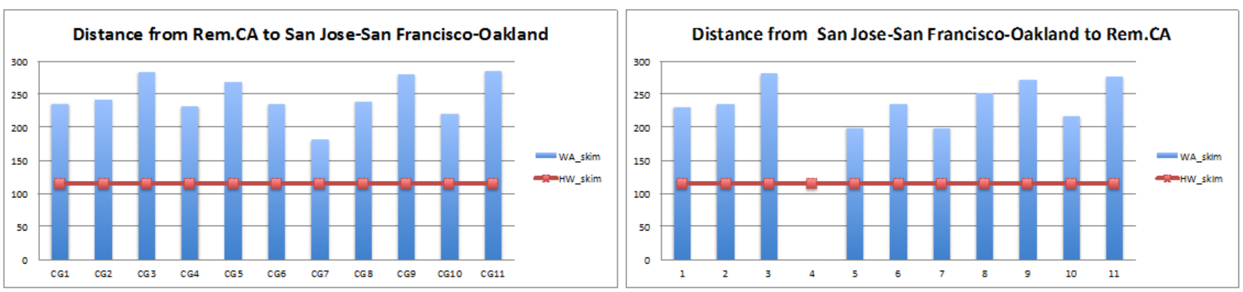
(a) Weighted Travel Distance between Remainder of CA and Los Angeles-Long Beach



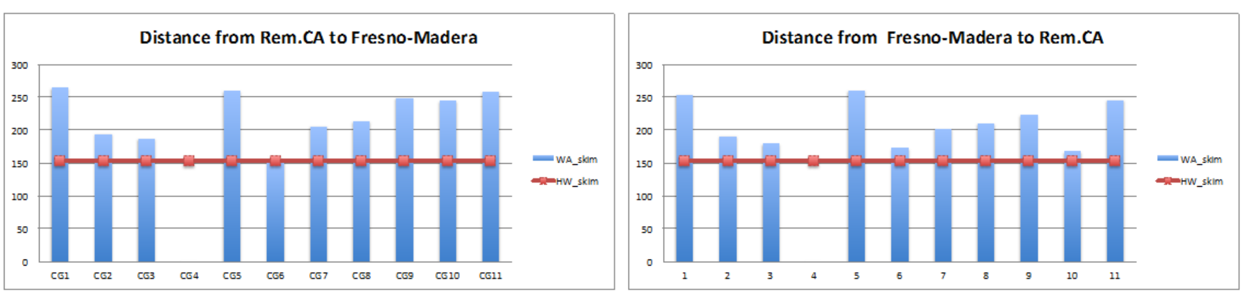
(b) Weighted Travel Distance between Remainder of CA and Sacramento-Roseville



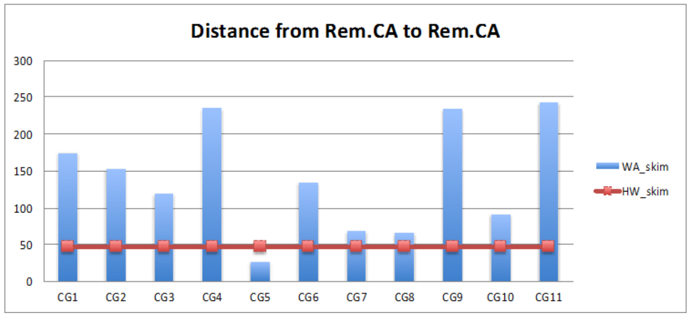
(c) Weighted Travel Distance between Remainder of CA and San Diego-Carlsbad-San Macros



(d) Weighted Travel Distance between Remainder of CA and San Jose- San Francisco-Oakland



(e) Weighted Travel Distance between Remainder of CA and Fresno-Madera



(f) Weighted Travel Distance between Remainder of CA and Remainder of CA

Figure 4-15 Weighted Travel Distance (miles) between Remainder of CA and California FAF Zones

In Figure 4-15 (a), the weighted travel distances are shorter than highway skims from the Remainder of California to Los Angeles-Long Beach. But in Figure 4-15 (b), the weighted travel distances are two to three times higher than the highway skim for most CGs. For intra-zonal movements as shown in Figure 4-15 (f), the weighted travel distances of CG4 and CG11 are still about five times higher than the highway skim. It shows that there are big differences between the highway skim and weighted travel distance by CG in each O-D pair related to Remainder of California, and the difference patterns are not fixed.

4.5 Calibration of CSFFM Total Generation Model

The Structural Regression Modeling Total Generation Model (SRMGM) of CSFFM is estimated based on FAF4 data. The freight production and consumption models are estimated for each CG using the QML method in Stata. Using the available FAF4 data with a 132 sample size, three levels of SRM models are developed in the final results: linear regression model, path analysis model and full SRM model.

4.5.1 Confirmatory Factor Analysis of Latent Variables

Two latent variables were measured in the total generation model. Latent variable CG8P represents the production level of CG8 (Logs, wood products, and textile and leather), and CG9C represents the consumption level of CG9 (Base metal and machinery). The path diagrams are shown below in Figure 4-16 and 4-17.

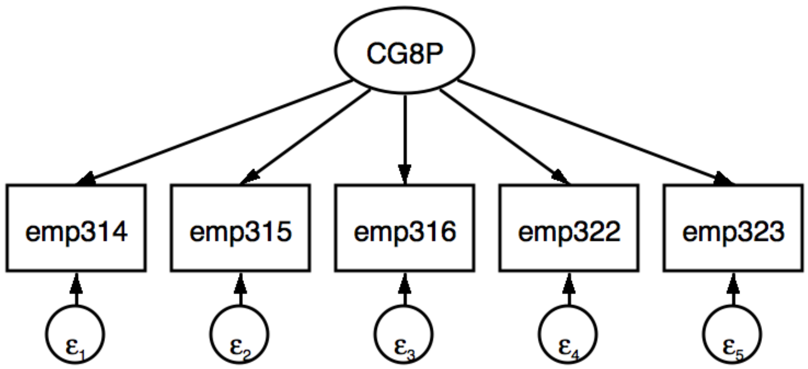


Figure 4-16 Path Diagram of Latent Variable CG8P

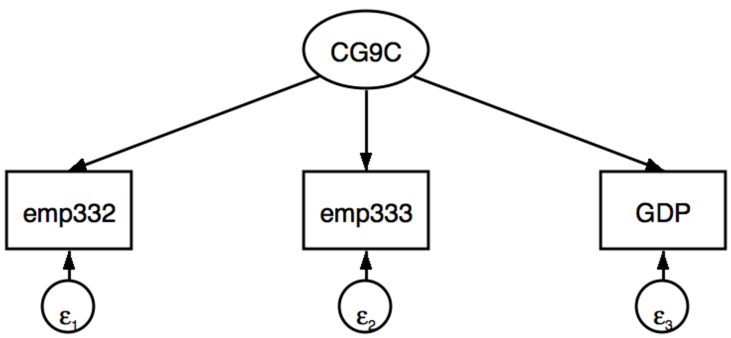


Figure 4-17 Path Diagram of Latent Variable CG9C

To estimate the measurement model, one of the factor loadings should be set to one to provide a scale for the latent variable (Bollen, 1989). In this study, the number of employees of Textile Product Mills (emp314) and Fabricated Metal Product Manufacturing (emp332) were both set to one to estimate the production level of CG8 and the consumption level of CG9 respectively for each FAF4 zone. Table 4-15 shows the unstandardized results and Standardized Root Mean Square Residual (SRMR) for the measurement models. All the coefficients are significant at $p<0.001$, and the SRMR for both measurement models are less than 0.08 which represents acceptable results (Hu and Benler, 1999).

Table 4-15 Measurement Model Results for SRMGM

Variable	Coefficient	p-value	SRMR
CG8P			
Emp314 (Textile Prod. Mftg. Emp.)	1.000	<0.001	0.069
Emp315 (Apparel Mftg. Emp.)	3.343	<0.001	
Emp316 (Leather Prod. Mftg. Emp)	0.378	<0.001	
Emp322(Paper Mftg. Emp.)	2.970	<0.001	
Emp323(Printing Emp.)	4.696	<0.001	
CG9C			
Emp332(Fabricated Metal Prod. Mftg.)	1.000	<0.001	0.000
Emp333(Machinery Mftg. Emp)	0.688	<0.001	
GDP	1.347	<0.001	

4.5.2 Model Results and Fitness

The total generation model contains twenty equations, eight path diagrams, two latent variables and covers ten CGs. Table 4-16 and 4-17 present the production and consumption model results respectively. The first column shows the path diagram ID for the each model. Column two is dependent (endogenous) variables. And column three is the corresponding explanatory or latent variables. The last three columns present the coefficient value, p-value, and coefficient of determination for each equation.

CG4 (Coal and metallic mineral ores) is not included because no proper variables can explain the production and consumption at the current stage. In the original CSFFM, the corresponding G6 was also excluded from the final model.

Table 4-16 SRMGM Results – Production

Path Diagram	Dependent Variables	Independent Variables	Coefficient	Std.Err.	p-value	R-square
1	G1_P	G1_C	0.532	0.150	<0.001	0.947
		Harvested Land (acreage)	0.003	0.001	<0.001	
1	G2_P	G1_C	0.073	0.025	0.003	0.817

		Emp311 (Food Mftg. Emp.)	0.280	0.083	0.001	
		Emp312 (Beverage & Tobacco Prod. Mftg. Emp.)	1.033	0.258	<0.001	
		Emp445 (Food and Beverage Stores Emp.)	0.058	0.026	0.027	
2	G3_P	G3_C	0.916	0.198	<0.001	0.898
		Est212 (No. Mining Establishments)	28.565	0.036	0.019	
3	G5_P	G5_C	0.675	0.097	<0.001	0.736
		Emp211 (Oil and Gas Extraction Emp.)	4.810	0.116	<0.001	
4	G6_P	G6_C	0.310	0.095	0.001	0.950
		Emp324 (Petroleum & Coal Mftg. Emp.)	12.588	1.828	<0.001	
4	G7_P	G6_C	0.271	0.020	<0.001	0.720
		Emp325 (No. of Chemical Mftg. Emp.)	0.338	0.067	<0.001	
5	G8_P	Emp113(No. Forestry and Logging Emp.)	8.073	1.110	<0.001	0.942
		CG8P	3.853	1.145	<0.001	
		constant	3502.886	481.900	<0.001	
7	G9_P	G9_C	0.871	0.032	<0.001	0.960
		Emp331(No. Primary Metal Mftg. Emp.)	0.423	0.092	<0.001	
8	G10_P	G10_C	0.794	0.112	<0.001	0.892
		Emp334(Computer and Electronic Mftg. Emp.)	-0.098	0.053	0.034	
		Emp339(Miscellaneous Mftg. Emp.)	-0.154	0.046	0.036	
		GDP	0.148	0.108	0.017	
6	G11_P	G11_C	0.535	0.158	0.001	0.881
		Emp493(No Warehousing and Storage Emp.)	0.356	0.124	0.004	

Table 4-17 SRMGM Results – Consumption

Path Diagram	Dependent Variables	Independent Variables	Coefficient	Std.Err.	p-value	R-square
1	G1_C	G1_P	0.775	0.026	<0.001	0.949
		Emp311 (Food Mftg. Emp.)	0.247	0.037	<0.001	
1	G2_C	Emp311 (Food Mftg. Emp.)	0.275	0.056	<0.001	0.880
		Emp493(No Warehousing and Storage Emp.)	0.318	0.064	<0.001	
		Emp445	0.078	0.016	<0.001	

2	G3_C	G3_P	0.846	0.228	<0.001	0.922
		Emp23 (Construction Emp.)	0.063	0.045	0.004	
		Emp213(No. Support Activities for Mining Emp.)	0.066	0.017	0.046	
3	G5_C	G5_P	0.284	0.164	<0.001	0.802
		Emp324 (Petroleum & Coal Mftg. Emp.)	9.980	0.105	<0.001	
		Emp447 (No. Gasoline Stations Emp.)	0.878	0.034	<0.001	
4	G6_C	G6_P	0.986	0.030	<0.001	0.934
		Emp447 (No. Gasoline Stations Emp.)	0.080	0.098	>0.05	
4	G7_C	G7_P	0.723	0.034	<0.001	0.869
		mp326 (No. of Plastics and Rubber Products Mftg. Emp.)	0.216	0.054	<0.001	
		pop(Population)	0.002	0.001	0.002	
		Emptot(Total Emp.)	-0.005	0.001	0.001	
		constant	819.771	338.343	0.015	
6	G8_C	G11_P	0.585	0.072	<0.001	0.871
		Emp113(No. Forestry and Logging Emp.)	0.579	0.061	<0.001	
		Emp313(Textile Mills Emp)	0.081	0.030	0.002	
		Emp321(Wood Product Manufacturing Emp.)	0.197	0.080	<0.001	
7	G9_C	Emp327(Nonmetallic Mineral Product Mftg. Emp.)	1.593	0.439	<0.001	0.881
		CG9C (CG9 Consumption level)	0.432	0.089	<0.001	
		constant	2098.008	564.904	0.001	
8	G10_C	G10_P	0.832	0.042	<0.001	0.977
		Emptot(Total Emp.)	0.001	0.019	<0.001	
6	G11_C	G11_P	0.785	0.082	<0.001	0.950
		Emp492(Messengers)	0.429	0.154	<0.001	

Note:
Gx_P: The production of commodity group x
Gx_C: The consumption of commodity group x
CGxP: Latent variable measured the production level of commodity group x
CGxC: Latent variable measured the consumption level of commodity group x
Std.Err: Standard Error

All the coefficients of above models are significant at the 0.05 level. The model fitness indices are presented in Table 4-18.

Table 4-18 SRMGM Fitness

Path Diagram ID	SRMR	Path Diagram ID	SRMR
1	0.023	5	0.094
2	0.051	6	0.096
3	0.063	7	0.054
4	0.030	8	0.065

For the equation goodness of fit, the coefficient of determination R^2 for all of the equations is greater than 0.72. Most of the R^2 of the production model are greater than 0.8. R^2 ranges from 0.802 to 0.977 in the consumption models. Regarding the Structural Equation Model (SEM) fitness, the SRMR of eight path diagrams are smaller than 0.1 and six of them are smaller than 0.08. The overall predictive accuracy of the model is acceptable.

4.5.3 Result Comparison

To illustrate the advantages of SRMGM, a result comparison between new model and the path analysis generation model (PAGM) in the original CSFFM is discussed in this section. In ideal conditions, the 2015 condition should be forecast using both PAGM and SRMGM to test the model performance. However, the official socioeconomic data in 2015 is not available during this research. Although backcasting to 2007 using SRMGM is another option, it is unfair for SRMGM since the PAGM applied adjustment factors to match 2007 real condition. Hence, the 2012 forecast of PAGM is decided to compare with SRMGM 2012 base results. Regarding to the different CG definition, the total production and consumption of each zone are aggregated for comparison. Root Mean Square Error (RMSE) is selected to represent model performance. Table 4-19 presents the RMSE values of PAGM and SRMGM. The RMSE of SRMGM production and consumption models are smaller than PAGM model.

Table 4-19 RMSE of PAGM and SRMGM in 2012

RMSE	Production Model	Consumption Model
SRMTGM	19884	18743
PAGM	34092	41288

4.5.4 Result Analysis

Two cases are discussed to illustrate the advantages of SRMGM in representing indirect effect, effect of latent variable and reflecting supply chain elasticity.

Figure 4-18 shows the path diagram of the case one model. There are two parts in the SRM: Structural model and Measurement model. The measurement model measures the latent variable CG8P which is determined by five indictors: emp314 (Textile Product Mills), emp315 (Apparel Manufacturing), emp316 (Leather and Allied Product Manufacturing), emp322 (Paper Manufacturing) and emp323 (Printing and Related Support Activities). The coefficients of the above five indicators are not significant when forming the regression model. But through the latent variable CG8P, the indirect effects are clearly presented.

The structural model predicts the production flow of CG8 which has two indictors: the number of employees of Logging (dem113) and the production level of CG8 (PLCG8).

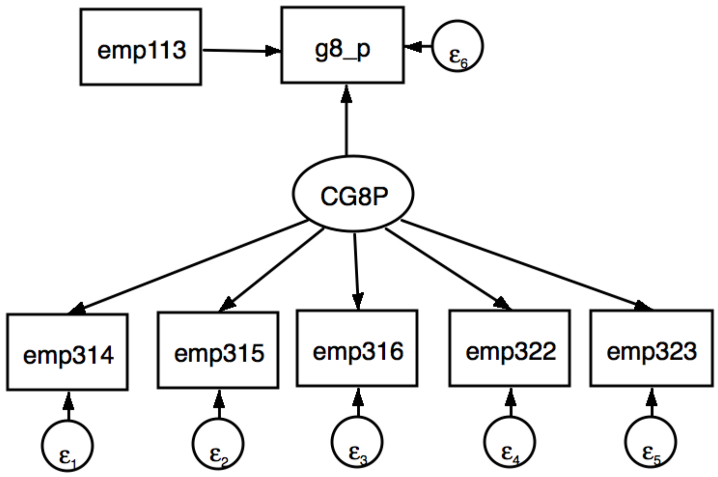


Figure 4-18 Path Diagram of Case 1(Path Diagram 5)

Compared this SRMGM with the independent model, we found that:

SRMGM: $g8\_p = 8.073 * Emp113 + 3.853 * CG8P + 3502.886$

Independent model: $g8\_p = 9.139 * Emp113 + 3071.335$

In this example, the R^2 of SRMGM is 0.942 which is greater than 0.745 of the independent model. The latent variable not only strengthens the model's explanation power, but also improves the model accuracy.

Figure 4-19 is the path diagram of case two, which contains both production and consumption models of the CG1 (Agriculture products and fish) and CG2 (Food, alcohol, and tobacco products).

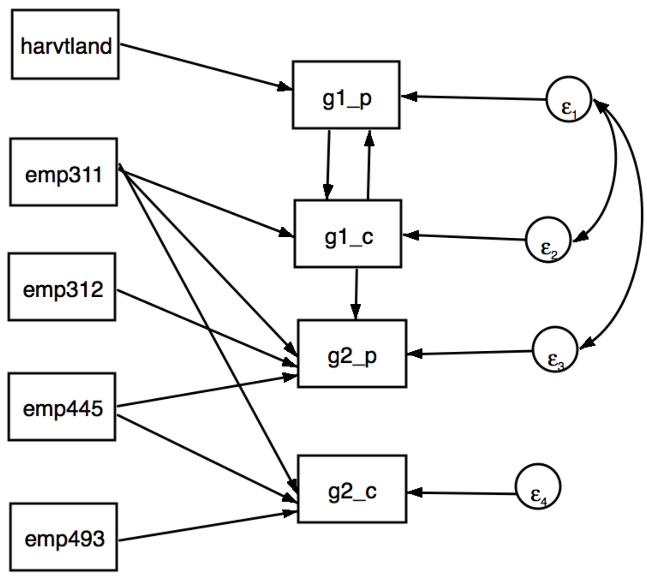


Figure 4-19 Path Diagram of Case 2 (Path Diagram 1)

There are five independent variables in this path diagram: the acreage of harvested land (harvtland), Food Manufacturing employments (emp311), Beverage and Tobacco Product Manufacturing employments (emp312), Food and Beverage Stores employments (emp445) and Warehousing and Storage employments (emp493). Emp311 has both direct effect and indirect effects on the production of CG2 (G2_P) as shown below:

Direct Effect: $Emp311 \xrightarrow{0.280} G2\_P$

Indirect Effect: $Emp311 \xrightarrow{0.247} G1\_C \xrightarrow{0.073} G2\_P$

The total effect of indicator Emp311 on G2_P equals to 0.280+0.247*0.073=0.298. This case also shows the supply chain elasticity between CG1 consumption and CG2 production in the same zone. One unit increase in Emp311 is associated with a 0.247 units increase in G1_C and totally 0.298 units increase in G2_P.

This path diagram also captures the intra-zonal freight flow movement which refers to the relationship between production and consumption of CG1.

$$Harvested\ Land \xrightarrow{0.003} G1\_P \underset{0.532}{\overset{0.775}{\rightleftarrows}} G1\_C$$

As shown above, one acre increase in Harvested Land is associated with three tons increase in G1_P and $\frac{0.003}{1-0.775\times0.532} \approx 0.005$ kilotons increase in G1_C within the same zone.

4.6 Calibration of CSFFM Domestic Direct Demand Model

There are 132 domestic zones and therefore 132×132 O-D pairs in FAF4 data. As discussed in Chapter three, the intra-zonal model and inter-zonal model should be estimated respectively to avoid the repetitive consideration of zonal variables for intra-zonal movements. The intra-zonal model measures the trips moving within the same zone and contains 132 O-D pairs, while the inter-zonal model has 132×131 O-D pairs.

4.6.1 Conform Factor Analysis of Latent Variables

One latent variable CG2C was measured in the Intra-zonal Structural Regression Modeling Direct Demand Model (SRMDDM) which represents the FAF4 zone consumption level of CG2 (Food, alcohol, and tobacco products). The path diagram is shown below (Figure 4-20):

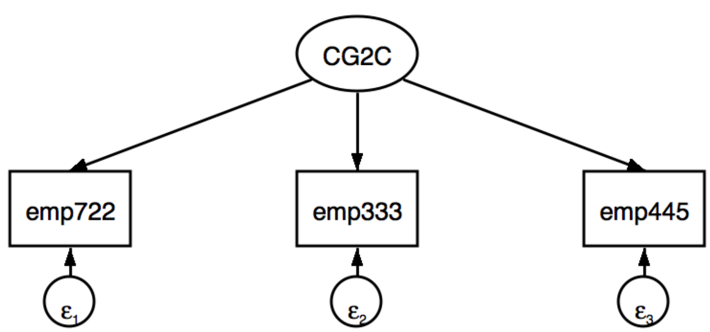


Figure 4-20 Path Diagram of Latent Variable CG9C

In this model, the number of employees of Food Services and Drinking Places (emp722) was set to one to estimate the consumption level of CG2. Table 4-20 shows the unstandardized results and the SRMR for this measurement model. All the coefficients are significant at p<0.001 and the SRMR is less than 0.08 within acceptable level.

Table 4-20 Measurement Model Results for Intra-zonal SRMDDM

Variable	Coefficient	p-value	SRMR
CG2C (for Intra-zonal Model)			
Emp722(Food Services Emp.)	1	<0.001	0.000
Emp493(Warehousing and Storage Emp.)	0.07	<0.001	
Emp445(Food and Beverage Stores Emp,)	0.298	<0.001	

Four latent variables were measured in the Intra-zonal SRMDDM. The path diagrams of them are shown below in Figure 4-21.

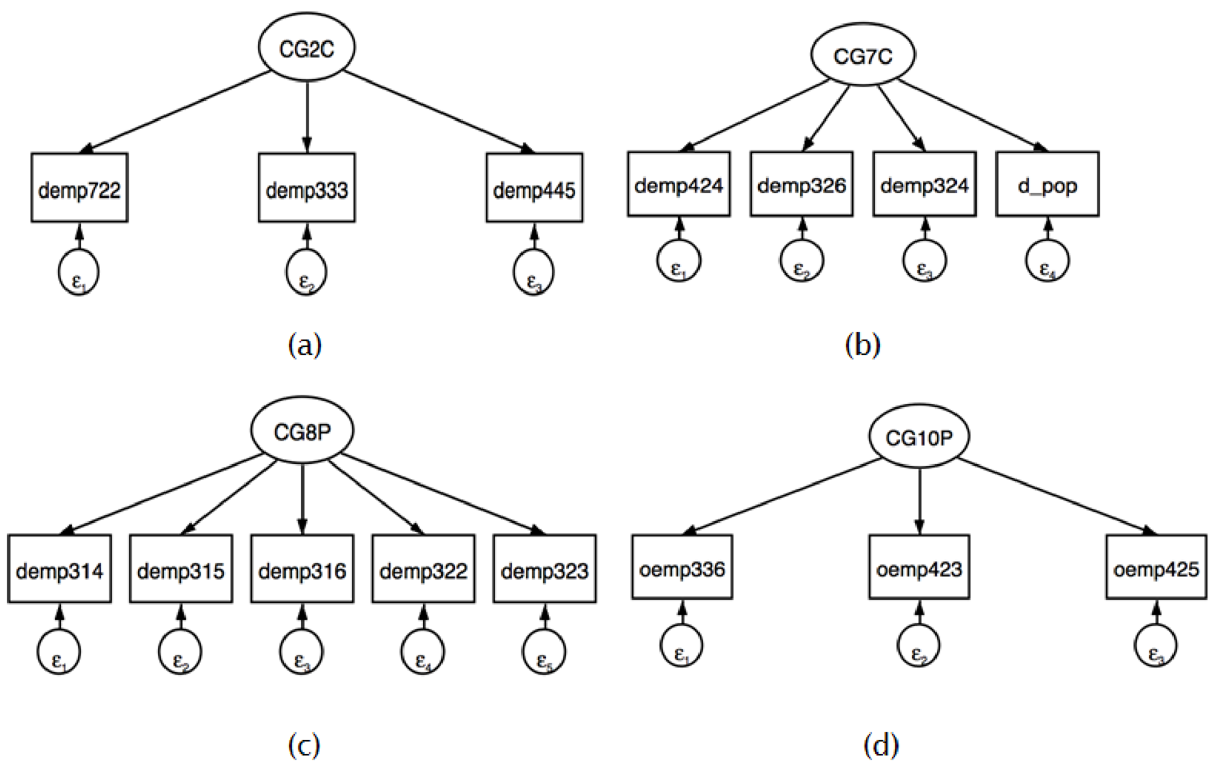


Figure 4-21 Path Diagrams of Latent Variable CG2C, CG7C, CG8P and CG10P

Latent variable CG2C and CG7C were used to measure the consumption level of CG2 and CG7 (Pharmaceutical and chemical products). CG8P and CG10P represent the production level of CG8 (Logs, wood products, and textile and leather) and CG10 (Electronic, motorized vehicles, and precision instruments).

The log transformation in the intra-zonal model can redress the skewness in socioeconomic data (Ranaiefar, 2013). So the Maximum Likelihood (ML) method can be used to estimate the model with Root Mean Square Error of Approximation (RMSEA) and Comparative-fit Index (CFI) as the model fitness measures. Table 4-21 shows the unstandardized results and the SRMR for those four measurement models. The entire coefficients are significant at $p<0.001$. The SRMR and RMSEA for each measurement model are less than 0.08 and all the CFI are greater than 0.95, which represents good fitness (Hooper, 2008; Hu and Bentler, 1999).

Table 4-21 Measurement Model Results for Inter-zonal SRMDDM

Variable	Coefficient	p-value	SRMR	RMSEA	CFI
CG2C (for Inter-zonal Model)					
demp722(Food Services Emp.)	1	<0.001	0	0	1
demp493(Warehousing and Storage Emp.)	1.216	<0.001			
demp445(Food and Beverage Stores Emp,)	1.009	<0.001			
CG7C					
demp424 (Merchant Wholesalers Emp.)	1	<0.001	0.011	0.077	0.996
demp326 (Plastics and Rubber Mftg. Emp.)	1.279	<0.001			
dEmp324 (Petroleum and Coal Prod. Mftg.)	1.168	<0.001			
d_pop(Population)	0.826	<0.001			
CG8P					
demp314 (Textile Prod. Mftg. Emp.)	1	<0.001	0.018	0.064	0.989
demp315 (Apparel Mftg. Emp.)	1.558	<0.001			
demp316 (Leather Prod. Mftg. Emp)	2.132	<0.001			
demp322(Paper Mftg. Emp.)	2.052	<0.001			
demp323(Printing Emp.)	0.905	<0.001			
CG10P					
oemp336 (Transportation Equipment Mftg.)	1	<0.001	0	0	1
oemp423 (Merchant Wholesalers Emp.)	0.741	<0.001			
oemp425 (Wholesale Electronic Agents)	1.062	<0.001			

4.6.2 Model Results and Fitness

The Intra-zonal SRMDDM contains eleven equations, three path diagrams, one latent variable and covers the entire eleven CGs. Table 4-22 presents the model estimation results.

Table 4-22 Results of Intra-zonal SRMDDM

Path Diagram	Dependent Variables	Independent Variables	Intra-Zonal Model			
			Coefficient	Std.Err.	p-value	R-square
1	CG1	Harvested Land (acreage)	0.003	0.000	<0.001	0.859
		Sold live stock (KTons)	0.004	0.001	<0.001	
2	CG2	EMP311(Food Mftg. Emp.)	0.144	0.020	<0.001	0.792
		CG2_ave_distance	13.473	3.909	<0.001	
		CG2C(CG2 Consumption level)	0.024	0.003	<0.001	
		constant	699.591	283.270	<0.001	
1	CG3	EST212(No. Mining Establishments)	51.690	11.858	<0.001	0.671
		EMP23(No. Construction Emp.)	0.293	0.015	<0.001	
1	CG4	EMP212 (No. Mining Emp.)	1.196	0.134	<0.001	0.600
		EMP213(No. Support Activities for Mining Emp.)	0.337	0.045	<0.001	
		EMP331(No. Primary Metal Mftg. Emp.)	0.510	0.114	<0.001	
		CG9	-0.270	0.063	<0.001	
1	CG5	EMP447(No. Gasoline Stations Emp.)	0.428	0.077	<0.001	0.722
		EMP486(No. Pipeline Transportation Emp.)	8.732	0.776	<0.001	
		CG7	-0.593	0.180	0.001	
1	CG6	GDP	0.814	0.219	<0.001	0.674
		EMP323(No. Printing Emp.)	-1.504	0.530	0.005	
		CG1	-0.227	0.081	0.005	
		CG7	2.205	0.919	0.016	
1	CG7	EMP325(No. of Chemical Mftg. Emp.)	0.136	0.063	0.031	0.703
		CG1	0.069	0.012	<0.001	
		CG6	0.108	0.031	<0.001	
1	CG8	EMP113(No. Forestry and Logging Emp.)	5.107	0.245	<0.001	0.916
		EMP321(No. Wood Product Mftg. Emp.)	0.394	0.072	<0.001	
		CG8_ave_distance	-11.354	3.975	0.004	
		CG9	0.230	0.027	<0.001	
1	CG9	GDP	0.125	0.025	<0.001	0.721

		CG3	0.319	0.032	<0.001	
3	CG10	EMP423(No. Merchant Wholesalers Emp.)	0.019	0.004	<0.001	0.689
		EMPTOT (Total Emp.)	-0.001	0.000	<0.001	
		EST336 (Transportation Equipment Mftg. Establishments)	5.634	0.501	<0.001	
3	CG11	EMP493(No Warehousing and Storage Emp.)	0.168	0.057	0.003	0.846
		ESTTOT(Total Establishments)	-0.051	0.024	0.036	
		Population	0.003	0.001	<0.001	
		GDP	0.053	0.027	0.045	
		constant	-737.586	271.679	0.007	

All the coefficients of above models are significant at the 0.05 level. The model fitness indices are presented in Table 4-23. The R^2 of all of the equations range from 0.6 to 0.916. Regarding SEM fitness, all the SRMR of the three path diagrams are smaller than 0.1 and two of them are smaller than 0.08. The overall predictive accuracy of the model is acceptable.

Table 4-23 Intra-zonal SRMDDM Fitness

Path Diagram ID	SRMR
1	0.086
2	0.048
3	0.014

The Inter-zonal model includes eleven equations, five path diagrams, four latent variables and covers the entire eleven commodity groups. Table 4-24 presents the model estimation results. The table attributes are the same as the Intra-zonal model.

Table 4-24 Results of Inter-zonal SRMDDM

Path Diagram	Dependent Variables	Independent Variables	Model			
			Coefficient	Std.Err.	p-value	R-square
1	CG1	demp311(Food Mftg. Emp.)	0.568	0.015	<0.001	0.863
		o_Harvested Land (acreage)	0.086	0.007	<0.001	
		log_distance_CG1	-4.030	0.015	<0.001	
		oemp311(Food Mftg. Emp.)	0.636	0.017	<0.001	

		demp114（Fishing Emp.)	0.018	0.003	<0.001	
		CG2	-0.108	0.004	<0.001	
		constant	15.888	0.243	<0.001	
1	CG2	demp311(Food Mftg. Emp.)	0.388	0.020	<0.001	0.852
		CG2C (CG2 Consumption level)	0.635	0.029	<0.001	
		log_distance_CG2	-3.920	0.016	<0.001	
		oemp311(Food Mftg. Emp.)	0.728	0.018	<0.001	
		oemp312(Beverage & Tobacco Prod. Mftg. Emp.)	0.058	0.007	<0.001	
		CG1	-0.089	0.004	<0.001	
		constant	16.872	0.267	<0.001	
2	CG3	oemp212(No. Mining Emp.)	0.079	0.005	<0.001	0.902
		demptot(Total Emp.)	-0.352	0.065	<0.001	
		d_pop(Population)	0.576	0.067	<0.001	
		log_distance_CG3	-3.893	0.010	<0.001	
		constant	21.984	0.236	<0.001	
3	CG4	oemp212 (No. Mining Emp.)	0.037	0.004	<0.001	0.849
		oemp213(No. Support Activities for Mining Emp.)	0.012	0.002	<0.001	
		demptot(Total Emp.)	0.039	0.009	<0.001	
		CG5	-0.016	0.002	<0.001	
		log_distance_G4	-4.071	0.014	<0.001	
		constant	26.633	0.180	<0.001	
3	CG5	oemp486(No. Pipeline Transportation Emp.)	0.139	0.006	<0.001	0.869
		demp324 (Petroleum & Coal Mftg. Emp.)	0.062	0.006	<0.001	
		demp447 (No. Gasoline Stations Emp.)	0.262	0.029	<0.001	
		demp325 (No. of Chemical Mftg. Emp.)	0.077	0.018	<0.001	
		demp326 (No. of Plastics and Rubber Products Mftg. Emp.)	-0.075	0.012	<0.001	
		log_distance_CG5	-3.815	0.012	<0.001	
		constant	22.270	0.223	<0.001	
*	CG6	oemp447 (No. Gasoline Stations Emp.)	0.106	0.010	<0.001	0.909
		demp324 (Petroleum & Coal Mftg. Emp.)	0.007	0.003	0.012	
		d_pop(Population)	0.062	0.010	<0.001	
		log_distance_CG6	-3.818	0.009	<0.001	
		constant	23.395	0.186	<0.001	
4	CG7	demp325 (No. of Chemical Mftg. Emp.)	0.597	0.019	<0.001	0.781
		oemp424 (Merchant Wholesalers Emp.)	0.336	0.023	<0.001	
		oemp331 (Primary Metal Mftg. Emp.)	0.031	0.006	<0.001	
		oemp324 (Petroleum & Coal Mftg. Emp.)	0.056	0.006	<0.001	
		CG7C (CG7 Consumption level)	1.135	0.017	<0.001	

		log_distance_CG7	-2.821	0.016	<0.001	
		constant	11.455	0.216	<0.001	
2	CG8	oemp113(No. Forestry and Logging Emp.)	0.091	0.003	<0.001	0.768
		o_gdp(GDP)	0.332	0.037	<0.001	
		demp321(Wood Product Manufacturing Emp.)	0.058	0.009	<0.001	
		demptot(Total Emp.)	-0.320	0.100	0.001	
		d_pop (Population)	1.862	0.107	<0.001	
		CG8P(CG8 Production level)	0.981	0.035	<0.001	
		CG3	0.019	0.004	<0.001	
		CG9	-0.267	0.007	<0.001	
		log_distance_CG8	-3.095	0.021	<0.001	
		constant	-4.435	0.488	<0.001	
2	CG9	oemp331(No. Primary Metal Mftg. Emp.)	0.157	0.006	<0.001	0.800
		oemp327(Nonmetallic Mineral Product Mftg. Emp.)	0.057	0.009	<0.001	
		o_gdp(GDP)	0.942	0.020	<0.001	
		demp331(No. Primary Metal Mftg. Emp.)	-0.018	0.006	0.001	
		demp332 (Fabricated Metal Product Mftg. Emp.)	0.111	0.019	<0.001	
		demptot(Total Emp.)	-0.900	0.090	<0.001	
		d_pop (Population)	2.240	0.092	<0.001	
		G3	0.016	0.004	<0.001	
		G8	-0.174	0.006	<0.001	
		log_distance_CG9	-3.145	0.019	<0.001	
		constant	-8.700	0.392	<0.001	
5	CG10	o_gdp(GDP)	0.371	0.026	<0.001	0.676
		desttot(Total Establishments)	0.896	0.106	<0.001	
		demptot(Total Emp.)	-1.482	0.111	<0.001	
		d_pop (Population)	1.246	0.094	<0.001	
		demp335(Electrical Equipment Mftg. Emp.)	-0.122	0.007	<0.001	
		demp423(Merchant Wholesalers Emp.)	0.718	0.044	<0.001	
		CG10P(CG10 Production level)	0.688	0.023	<0.001	
		log_distance_CG10	-1.649	0.016	<0.001	
		constant	-6.041	0.551	<0.001	
*	CG11	oemp337 (Furniture and Related Product Mftg. Emp)	0.488	0.016	<0.001	0.711
		oemp493 (No Warehousing and Storage Emp.)	0.212	0.015	<0.001	
		o_gdp(GDP)	0.194	0.019	<0.001	
		demp337(Furniture and Related Product Mftg. Emp)	-0.041	0.016	0.010	
		demp492(Messengers)	0.218	0.027	<0.001	
		demp493(No Warehousing and Storage Emp.)	-0.096	0.019	<0.001	

		demptot(Total Emp.)	-0.886	0.084	<0.001
		d_pop (Population)	2.023	0.081	<0.001
		log_distance_CG11	-2.032	0.015	<0.001
		constant	-10.214	0.356	<0.001

*Estimated independently using linear regression model

O_x: origin zone variable;

D_x: destination zone variable

All the coefficients of above models are significant at the 0.05 levels. The model fitness indies are presented in Table 4-25.

Table 4-25 Inter-zonal SRMDDM Fitness

Path Diagram ID	RMSEA	CFI	SRMR
1	0.078	0.981	0.022
2	0.087	0.909	0.003
3	0.059	0.992	0.006
4	0.044	0.991	0.017
5	0.065	0.968	0.019

The values of R^2 are from 0.676 to 0.909. Nine of eleven equations have R^2 greater than 0.75. The RMSEA of those five path diagrams are smaller than 0.1. Four of the five RMSEA are smaller than 0.08. The SRMR of the entire diagrams are less than 0.05. The CFI for all the path diagrams are greater than 0.90 and four of them are greater 0.95. The overall predictive accuracy of the model is good.

4.6.3 Result Comparison

In this section, the flow length frequency distribution and domestic production and consumption in the State of California will be compared among SRMDDM, Structural Equation Multi-Commodity OD Distribution (SEMCOD) model and FAF4 2012 observations. As reviewed previously, the original CSFFM aggregated 43 SCTG two-digit commodities to fifteen CGs. It should be noted that SEMCOD didn't estimate the model of CG3 (SCTG 16 - Crude petroleum) and CG6 (SCTG 14 and 15 - Coal and metallic ore) in the original CSFFM. When comparing SRMDDM and SEMCOD, the CGs containing SCTG 14, 15 and 16 were all excluded. As a result, CG4, 5 and 6 are excluded in SRMDDM, and CG3, 4 and 6 from original CSFFM were excluded from SEMCOD.

SEMCOD has one indicator "LogSum Utility" which is not used in SRMDDM. The LogSum Utility is a generalized cost of transportation between origin and destination and estimated in FAF3 zone level (Ranaiefar, 2013). So the 2012 forecasting result of SEMCOD is a 123×123 flow matrix. In order to compare the flow length frequency distribution of SEMCOD with SRMDDM, the 132×132 weighted travel distance matrix of FAF4 was transferred to FAF3 zone level (123 × 123) by removing the ten new-zone-related O-D pairs. The zone 342 (Philadelphia-Reading-Camden) in FAF4 is treated as zone 349 (Remainder of New Jersey) in FAF3 since they are using the same geographic area.

Figure 4-22 shows the flow length distributions of FAF4, SEMCOD and SRMDDM. The overall performance of SRMDDM well matches the FAF4 pattern. The forecasts of SEMCOD are less accurate at both 0 to 100 miles for underestimating flows over 10 percent, and 1000 miles and longer for overestimating flows by 15 percent to the maximum.

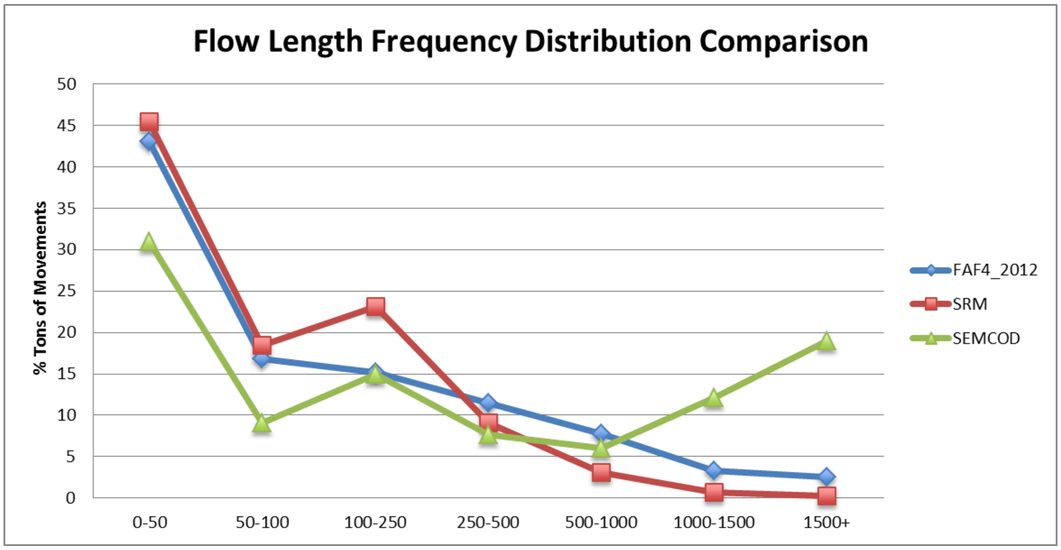
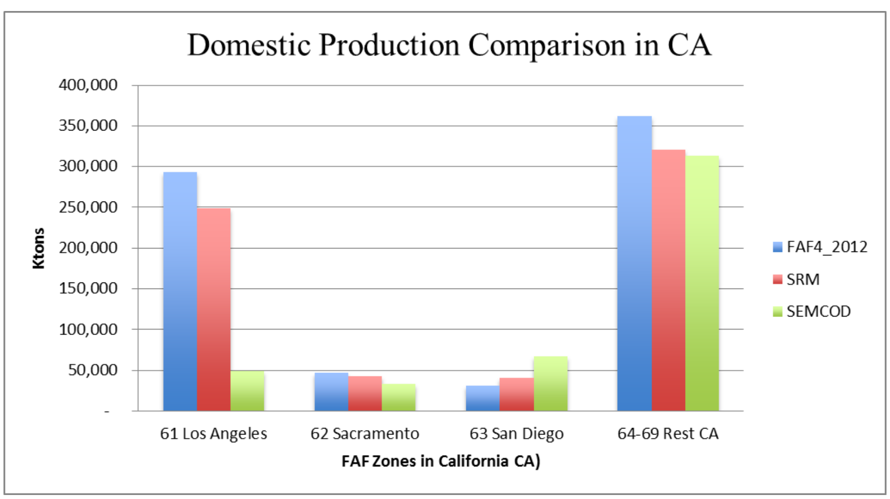
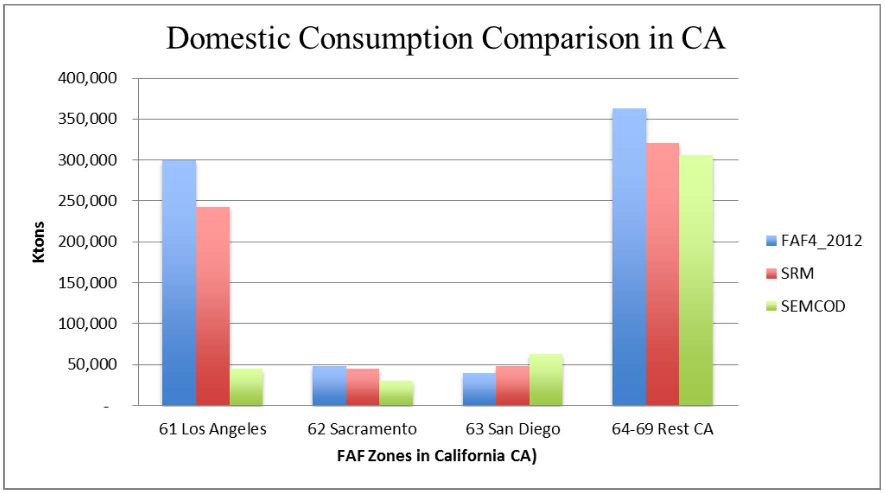


Figure 4-22 Flow Length Frequency Distribution of SRMDDM, SEMCOD and FAF4

The domestic production and consumption of California FAF4 zones were calculated to compare with FAF4 data. Due to the geographic changes from FAF3 to FAF4, the San Jose-San Francisco-Oakland zone (zone 64) and Remainder of California zone (zone 69) in FAF3 are combined together as new "Rest of California" zone, which is equal to the combination of zone 64 (San Jose-San Francisco-Oakland), 65 (Fresno-Madera) and 69 (Remainder of California) in FAF4. The comparison result is presented in Figure 4-23.



(a)



(b)

Figure 4-23 Domestic Total Freight Generation of SRMDDM, SEMCOD and FAF4

The above figure shows that SEMCOD underestimated both production and consumption in zone 61, 62 and the combination of 64 to 69. Especially for zone 61, the underestimation of SEMCOD is above 80 percent. In addition, SEMCOD overestimated both production and consumption in zone 63. Compared with SEMCOD, SRMDDM shows a better match of FAF4 condition in 2012.

4.6.4 Result Analysis

In this section, the CG10 will be analyzed to show the difference between the Intra-zonal model and the Inter-zonal model.

The emp423 (Merchant Wholesalers) is the indicator of CG10 for both intra-zonal model and inter-zonal model. Below is the relationship between emp423 and CG10:

Intra-Zonal Model: $emp423 \xrightarrow{0.019} G10$

Inter-Zonal Model: $demp423 \xrightarrow{0.718} \log\_G10$

In the intra-zonal model the coefficient of emp423 is equal to 0.019 with the standard error 0.004. One unit increased by emp423 is associated with 190 tons increased in G10 flows. In the inter-zonal model, the coefficient of demp423 is 0.718 and associated standard error is 0.044. Therefore, one unit increased in demp423 will lead to a 0.718 units increase in log_G10. So the emp423 has different influence on CG10 in intra-zonal flow and inter-zonal flow. The "Distance" indicator also has different effects on intra-zonal and inter-zonal model. For all CGs, "Distance" has a significant effect for inter-zonal flow. However, the influence of "Distance" is not significant for the most of CGs in intra-zonal model. Therefore, it is necessary to separate the intra-zonal and inter-zonal model in SRMDDM.

5 CONTRIBUTIONS AND FUTURE RESEARCH

This dissertation explored a framework for inter-regional commodity-based freight demand forecasting using Structural Regression Modeling (SRM). This framework was applied to update the original California Statewide Freight Forecasting Model (CSFFM) based on the latest Freight Analysis Framework Version 4 (FAF4) data for a 2012 base year. The strengths of this framework were further illustrated by comparison with original CSFFM forecast results in 2012. In this chapter, the contributions of this dissertation are summarized, followed by future research topics.

5.1 Contributions

The specific contributions of this research are concluded below:

- This research first introduces an optimization mathematical expression of a freight economic centroid to represent the origin and destination of each freight flow instead of a geometric centroid. At a FAF4 zone level, the freight economic centroids are generally closer to the highway system than geometric centroids and provide potential convenience for freight movements.
- The framework first applies a Fuzzy C-means (FCM) clustering algorithm to aggregate 43 SCTG two-digit commodities into eleven commodity groups (CG) which keep consistency with SCTG nine groups in FAF4. The new eleven CGs reduce the dimension and complexity of the freight demand forecasting model.
- This framework employs the weighted travel distance for multiple travel modes between each origin-destination (O-D) pair which acts a significant role in estimating freight flows.
- This dissertation first adopts SRM with latent variables into freight demand forecasting to

capture the direct effect, indirect effect and underlying relations between exogenous and endogenous variables to reflect supply chain relationships among CGs.

- This research is the first to use the latest FAF4 data in inter-regional freight demand forecasting.
- This framework is not limited to apply to the CSFFM with FAF data, and can also contribute to the development of other statewide and regional freight forecasting models with different zoning systems and freight data sources.

5.2 Future Research

The framework developed in this research can be further improved in the following ways:

- A comparison between Structural Regression Modeling Generation Model (SRMGM), Structural Regression Modeling Direct Demand Model (SRMDDM) and original CSFFM's Path Analysis Generation Model (PAGM) and Structural Equation Multi-Commodity OD Distribution (SEMCOD) model should be processed to forecast to 2015, 2020 and 2040 conditions when the socioeconomic and industrial data are available.
- Limited by the available socioeconomic and industrial variables, there is no proper total generation model measured for CG4 (Coal and petroleum products). Future study should explore new explanatory variables for CG4.
- In CSFFM, a comparable gravity distribution model can be developed as a comparison for SRMDDM. One thing should be noted in that the use of a gravity model should be based on a separate generation model. In order to make the comparison, a domestic generation model should be built first. The future year forecasts of the gravity model and SRMDDM can be compared to test the model performance.
- A future application of this research is to integrate CSFFM with the Heavy Duty Vehicle

Module of the Vision emission model of the California Air Resources Board (ARB) to evaluate the statewide total emissions and the effect of alternative fuel commercial vehicles in the ongoing California Natural Gas Vehicle Incentive Program (NGVIP).

6 REFERENCES

[1] Abdelwahab, W., & Sargious, M. (1992). Modelling the Demand for Freight Transport: A New Approach. Journal of Transport Economics and Policy, 49-70.

[2] Adie Tomer and Joseph Kane (2015). The Top 10 Metropolitan Port Complexes in the U.S.. Accessed online May 2017, https://www.brookings.edu/blog/the-avenue/2015/07/01/the-top-10-metropolitan-port-complexes-in-the-u-s/

[3] Adolf, J., Balzer, C., & Haase, F. (2016). Shell Commercial Vehicle Study. Diesel VS. Alternative Drive-Trains: Which Drive-Trains and Fuels Will Commercial Vehicles Use in the Future.

[4] Alabama Department of Transportation (2010). Alabama Statewide Freight Study and Action Plan: Final Report.

[5] Bachman, K. (2011). 6 Alternatives to Diesel for Freight Transport. Sustainable Manufacturing. Fabricatots & Manufacturers Association, International.

[6] Bagley, M. N., & Mokhtarian, P. L. (2002). The Impact of Residential Neighborhood Type on Travel Behavior: A Structural Equations Modeling Approach. The Annals of Regional Science, 36(2), 279-297.

[7] Bastida, C., & Holguin-Veras, J. (2009). Freight Generation Models: Comparative Analysis of Regression Models and Multiple Classification Analysis. Transportation Research Record: Journal of the Transportation Research Board, (2097), 51-61.

[8] Bentler, P. M., & Chou, C. P. (1987). Practical Issues in Structural Modeling. Sociological Methods & Research, 16(1), 78-117.

[9] Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The Fuzzy C-means Clustering Algorithm. Computers & Geosciences, 10(2-3), 191-203.

[10] Bollen, K. A. (1989). Structural Equations with Latent Variables. Wiley Series in Probability and Mathematical Statistics Section, ISSN 0271-6356.

[11] Brown, T. A. (2014). Confirmatory Factor Analysis for Applied Research. Guilford Publications.

[12] Brown, T. A., & Moore, M. T. (2012). Confirmatory Factor Analysis. Handbook of structural equation modeling, 361-379.

[13] Bureau of the Census (2004). Vehicle Inventory and Use Survey. US Department of Commerce.

[14] Bureau of Transportation Analysis (2015). Freight Analysis Framework Version 4 User's Guide for Release 4.0. Prepared by Center for Transportation Analysis, Oak Ridge National Laboratory.

[15] Bureau of Transportation Statistics, Foreign Trade: 11.4 .1 Disaggregating A Commodity Group to Associated two-digit SCTG Codes. Accessed online May 2017, https://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/subject_areas/freight_transportation/faf/faf4/trade

[16] Bryan, J., Weibrod, G., Martland, C.D., Wilbur Smith Associates, Inc. (2007). National Cooperative Highway Research Program (NCHRP) Report 586: Rail Freight Solutions to Roadway Congestion – Final Report and Guidebook.

[17] California Department of Transportation. (2010). 2010 Annual Average Daily Truck Traffic on the California State Highway System. The Transportation Agency.

[18] California Department of Transportation, Data Weigh-In-Motion, "Map of WIM locations in California", Accessed online May 2016, http://www.dot.ca.gov/trafficops/wim/locations.html

[19] California Emission Commission. Investments in California's Alternative and Renewable Fuel and Vehicle Technology Markets. Accessed online November 2017, http://www.energy.ca.gov/contracts/transportation.html

[20] California Statewide Freight Forecasting Model Draft Report (2015). University of California, Irvine.

[21] Caltrans Data WIM. Accessed July 2016. http://www.dot.ca.gov/trafficops/wim/datawim.html

[22] Cambridge Systematics, Inc., Global Insight, Cohen, H., Horowitz, A., Pendyala, R. (2008). National Cooperative Highway Research Program (NCHRP) Report 606: Forecasting Statewide Freight Toolkit.

[23] Cambridge Systematics.(1997). Quick Response Freight Manual, Report DOT-T-97-10, U.S. Department of Transportation and U.S. Environmental Protection Agency, Washington, D.C..

[24] CDM Smith(2014). Existing and Future Freight Goods Movements Assessment: Technical Memorandum. Coastal Region Metropolitan Planning Organization (CORE MPO) Freight Transportation Plan.

[25] CDM Smith (2015)., Review of Fright Data Sources for the Development of A Behavior-based Freight Model. U.S. Department of Transportation, Ferderal Highway Administration. Maricopa Association of Governments (MAG) System Analysis Program. MAG Project No. 0600-0631-15-E001-0639-0E.000B01.

[26] Chase, K. M., Anater, P., & Phelan, T. J. (2013). Freight Demand Modeling and Data Improvement. Transportation Research Board.

[27] Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. Structural Equation Modeling, 14(3), 464-504.

[28] Chen-Fu, L. (2014). Using truck GPS data for freight performance analysis in the Twin Cities metro area.

[29] Chiang, Y. S., Roberts Jr, P. O., & Ben-Akiva, M. (1981). Short-run Freight-Demand Model: Joint Choice of Mode and Shipment Size (abridgment) (No. 828).

[30] Chou, T. Y., Hsu, C. L., & Chen, M. C. (2008). A Fuzzy Multi-criteria Decision Model for International Tourist Hotels Location Selection. International Journal of Hospitality Management, 27(2), 293-301.

[31] Chuang, K. S., Tzeng, H. L., Chen, S., Wu, J., & Chen, T. J. (2006). Fuzzy C-means Clustering with Spatial Information for Image Segmentation. Computerized Medical Imaging and Graphics, 30(1), 9-15.

[32] Chun, Y., Kim, H., & Kim, C. (2012). Modeling Interregional Commodity Flows with Incorporating Network Autocorrelation in Spatial Interaction Models: An Application of the US Interstate Commodity Flows. Computers, Environment and Urban Systems, 36(6), 583-591.

[33] Commodity Flow Survey (2015). 2012 CFS Data Source: Table 1b. Shipment Characteristics by Mode of Transportation for the United SFtates: 2012 and 2007. Bureau of Transportation Statistics, U.S. Department of Transportation.

[34] Coops Are Open, Texas Weigh Station and DOT Information, Accessed online June 2016. http://www.coopsareopen.com/texas-weigh-stations.html

[35] Crockett, S. A. (2012). A Five-step Guide to Conducting SEM Analysis in Counseling Research. Counseling Outcome Research and Evaluation, 3(1), 30-47.

[36] Crossman, A. (2017). Understanding Path Analysis: A Brief Introduction. Accessed online 2017, https://www.thoughtco.com/path-analysis-3026444
[37] Donnelly, R. (2010). Best Practices for Incorporating Commodity Flow Survey and Related Data into the MPO and Statewide Planning Processes.
[38] Doustmohammadi, E., Sisiopiku, V. P., Anderson, M. D., Doustmohammadi, M., & Sullivan, A. (2016). Comparison of Freight Demand Forecasting Models. International Journal of Traffic and Transportation Engineering, 5(1), 19-26.
[39] Eaton. (2011). Driver Instructions: Hydraulic Launch Assist (HLA) TRDR1200. Accessed online January 2018, http://www.eaton.com/ecm/groups/public/@pub/@eaton/@roadranger/documents/content/ct_257101.pdf
[40] EPA (2017). United States Environmental Protection Agency: U.S. Transportation Sector Greenhouse Gas Emissions 1990-2015. Office of Transportation and Air Quality. EPA-420-F-17-013.
[41] FHWA(2005). Assessing the Effects of Freight Movement on Air Quality at the National and Regional Level: Final Report. Prepared by ICF Consulting.
[42] FHWA(2009). FAF2 Data and Documentation: 2002-2035.
[43] FHWA(2002). Freight Analysis Framework. FHWA-OP-03-006(R).
[44] FHWA(2012). Freight Analysis Framework 3: User Guide.
[45] FHWA(2015). Freight Analysis Framework Version 4: User Guide for Release 4.0.
[46] FHWA (2011). The Freight Analysis Framework Version 3 (FAF3): A Description of the FAF 3 Regional Database and How It Is Constructed.
[47] Fraer, R., Dinh, H., McCormick, R. L., Chandler, K., & Buchholz, B. (2005). Operating Experience and Teardown Analysis for Engines Operated on Biodiesel Blends (B20) (No. 2005-01-3641). SAE Technical Paper.
[48] Garson, G. D. (2013). Path analysis. Statistical Associates Publishing.
[49] Golob, T. F. (2003). Structural Equation Modeling for Travel Behavior Research. Transportation Research Part B: Methodological, 37(1), 1-25.
[50] Hernandez, S (2014). Integration of Weigh-In-Motion and Inductive Signature for Truck Body Classification, University of California, Irvine.

[51] Holguin-Veras, J., & Thorson, E. (2000). Trip Length Distributions in Commodity-based and Trip-based Freight Demand Modeling: Investigation of Relationships. Transportation Research Record: Journal of the Transportation Research Board, (1707), 37-48.
[52] Holguín-Veras, J., Jaller, M., Sanchez-Diaz, I., Wojtowicz, J., Campbell, S., Levinson, H., & Tavasszy, L. (2012). Freight trip generation and land use (No. NCHRP Project 08-80).
[53] Holguín-Veras, J., & Jaller, M. (2014). Comprehensive Freight Demand Data Collection Framework for Large Urban Areas. In Sustainable Urban Logistics: Concepts, Methods and Information Systems (pp. 91-112). Springer, Berlin, Heidelberg.
[54] Holguin-Veras, J., et al. (2015). National Cooperative Freight Research Program (NCFRP) Report 33: Improving Freight System Performance in Metropolitan Areas: A Planning Guide.
[55] Holguin-Veras, J., et al. (2017). National Cooperative Freight Research Program (NCFRP) Report 37: Using Commodity Flow Survey Microdata and Other Establishment Data to Estimate the Generation of Freight, Freight Trips, and Service Trips: Guidebook.
[56] Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. Articles, 2.
[57] Hox, J. J., & Bechger, T. M. (2007). An Introduction to Structural Equation Modeling.
[58] Hu, L. T., & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. Structural Equation Modeling: A Multidisciplinary Journal, 6(1), 1-55.
[59] Hunt, J. D., Donnelly, R., Abraham, J. E., Batten, C., Freedman, J., Hicks, J., ... & Upton, W. J. (2001, July). Design of A Statewide Land Use Transport Interaction Model for Oregon. In Proceedings of the 9th World Conference for Transport Research, Seoul, South Korea (p. 19).
[60] Hwang, H. L., Hargrove, S., Chin, S. M., Wilson, D. W., & Davidson, D. (2016). Freight Analysis Framework Version 4-Building the FAF4 Regional Database: Data Sources and Estimation Methodologies (No. ORNL/TM-2016/489). Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States).
[61] IHS Markit. Transearch: Prepare for Future Transportation Demand, Accessed online February 2017, https://ihsmarkit.com/products/transearch-freight-transportation-research.html
[62] Institute of Transportation Studies (ITS), University of California, Irvine. Natural Gas Vehicle Incentive Project (NGVIP). Accessed online November 2017, https://ngvip.its.uci.edu/

[63] Iowa Department of Transportation and Iowa State University Center for Transporting Research and Education. Iowa Statewide Freight commodity Model. Accessed 2014, https://tmip.org/content/iowa-statewide-freight-commodity-model

[64] Jeong, K., Tok, A., Ritchie, S. G., & Park, J. (2016). California Vehicle Inventory and Use Survey: Pilot Study Insights. Transportation Research Record: Journal of the Transportation Research Board, (2547), 32-40.

[65] Jiang, F., Johnson, P., & Calzada, C. (1999). Freight Demand Characteristics and Mode Choice: An Analysis of the Results of Modeling with Disaggregate Revealed Preference Data. Journal of Transportation and Statistics, 2(2), 149-158.

[66] Jonnavithula, S. S. (2004). Development of Structural Equations Models of Statewide Freight Flows.

[67] Kenny, D.A. (2015). Measuring Model Fit. Accessed online October 2017, http://davidakenny.net/cm/fit.htm

[68] Kenny, D.A. (2011). Terminology and Basis of SEM. Accessed online October 2017, http://davidakenny.net/cm/basics.htm

[69] Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The Performance of RMSEA in Models with Small Degrees of Freedom. Sociological Methods & Research, 44(3), 486-507.

[70] Kline, R. B., (2011). Principles and Practice of Structural Equation Modelling (Third Edition). ISBN 978-1-60623-877-6.

[71] Kruse, J.C., Protopapas, A., Bierling, D., Wang, B. (2014). National Cooperative Freight Research Program (NCFRP) Report 32: Integrating MTS Commerce Data with Multimodal Freight Transportation Performance Measures to Support MTS Maintenance Investment Decision Making.

[72] Kuzmyak, J. R. (2008). Forecasting metropolitan commercial and freight travel (Vol. 384). Transportation Research Board.

[73] Lecca, P. (Ed.). (2011). Systemic Approaches in Bioinformatics and Computational Systems Biology: Recent Advances: Recent Advances. IGI Global.

[74] Levinson, D. M., & Kumar, A. (1995). A Multi-modal Trip Distribution Model.

[75] Lischke, A. (2017). Drive Trains, Fuels And Technologies For Heavy Duty Vehicles In 2030 And Beyond.

[76] Lu, Q., Harvey, J., Le, T., Lea, J., Quinley, R., Redo, D., & Avis, J. (2002). Truck traffic analysis using weigh-in-motion (WIM) data in California. Report produced under the auspices of the California Partnered Pavement Research Program for the California Department of Transportation Pavement Research Center, Institute of Transportation Studies, University of California, Berkeley.

[77] Lupa, M.R(2008). Understanding FHWA's Freight Analysis Framework (FAF2) in Travel Models. Wilbur Smith Associates. TRB Conference Session 1: Statewide Planning Applications.

[78] LSA, (2008). Travel Demand Model Improvements for the Washtenaw Area Transportation Study. Technical Memorandum 4: Centroid Connector Placement.

[79] Malthouse, E. C. (2001). Checking Assumptions of Normality Before Conducting Factor Analyses. Journal Consumer Psychology, 10, 81.

[80] Maks Inc. (2016). FAF4 Freight Traffic Assignment: Final Report. Submitted to Oak Ridge National Laboratory.

[81] Mani, A., & Prozzi, J. (2004). State-of-the-practice in Freight Data: A Review of Available Freight Data in the US (No. 0-4713-P2). Center for Transportation Research, the University of Texas at Austin.

[82] Masoud, N., Hyun, K., (2012). 4-Step Commodity Model Freight Forecasting. University of California, Irvine.

[83] Moss, S., Fit Indices for Structural Equation Modeling. Accessed online May 2017, https://www.sicotests.com/psyarticle.asp?id=277

[84] Norton, P., Vertin, K., Bailey, B., Clark, N. N., Lyons, D. W., Goguen, S., & Eberhardt, J. (1998). Emissions from Trucks using Fischer-Tropsch Diesel Fuel (No. 982526). SAE Technical Paper.

[85] Ortúzar, J.D., Willumsen, L.G (2011). :Modeling Transport, 4th Ed., John Wiley & Sons Inc.

[86] Anderson Park, J., Cho, J., Gordon, P., Moore II, J. E., Richardson, H. W., & Yoon, S. (2011). Adding A Freight Network to A National Interstate Input–Output Model: A TransNIEMO Application for California. Journal of Transport Geography, 19(6), 1410-1422.

[87] Pinjari, A., Zanjani, A. B., Thakur, A., Irmania, A. N., Kamali, M., Short, J., ... & Park, L. (2014). Using Truck Fleet Data in Combination with Other Data Sources for Freight Modeling and Planning.

[88] Proussaloglou, K., Popuri, Y., Tempesta, D., Kasturirangan, K., & Cipra, D. (2007).
[89] Novak, D. C., Hodgdon, C., Guo, F., & Aultman-Hall, L. (2011). Nationwide Freight Generation Models: A Spatial Regression Approach. Networks and Spatial Economics, 11(1), 23-41.
[90] Parsons Brinkerhoff (2012). Oklahoma Freight Flows. Oklahoma Department of Transportation Planning & Research Division.
[91] Proussaloglou, K., Popuri, Y., Tempesta, D., Kasturirangan, K., & Cipra, D. (2007). Wisconsin Passenger and Freight Statewide Model: Case Study in Statewide Model Validation. Transportation Research Record: Journal of the Transportation Research Board, (2003), 120-129.
[92] Ranaiefar, F. (2013). Interregional Commodity Flow Model Using Structural Equation Modeling: Application to California Statewide Freight Forecasting Model. University of California, Irvine.
[93] Regan, A. C., & Garrido, R. A. (2002). Modeling Freight Demand and Shipper Behavior: State of the Art, Future Directions (No. UCI-ITS-LI-WP-02-2). The Institute.
[94] Resource System Group, Inc. (2012). Statewide Multi-Modal Freight Model. Presented to MTF Freight Modeling Session.
[95] Resource System Group, Inc. (2013). Florida Multimodal Statewide Freight Model: Model Inputs and Structure, and Integration with Regional Truck-touring Models. Presented to State of Florida Department of Transportation.
[96] Resource Systems Group, Inc., Florida Freight Supply-Chain Intermodal Model (FreightSim): Documentation. Accessed 2015, https://rsginc.atlassian.net/wiki/spaces/FREIGTSIM/pages/2457616/Model+Documentation
[97] Rushton, G., Armstrong, M. P., Gittler, J., Greene, B. R., Pavlik, C. E., West, M. M., & Zimmerman, D. L. (Eds.). (2007). Geocoding Health Data: The Use of Geographic Codes in Cancer Prevention and Control, Research and Practice. CRC Press.
[98] Rwakarehe, E. E., Zhong, M., & Christie, J. (2014). Development of a Freight Demand Model for the Province of Alberta Using Public Sources of Data. Procedia-Social and Behavioral Sciences, 138, 695-705.
[99] Sarkar, S., Vick, R. (2011). Using Population Weighted County Centroids to Calculate Migration Distance for Historical Linked Census Data: A Case Study of Male Migrants in the

New York and New Jersey Area, 1880-1910. Accessed online http://paa2011.princeton.edu/papers/111817

[100] SCAG (2012). Southern California Association of Governments (SCAG) Regional Model Chapter 7: Heavy Duty Truck Model.

[101] Schmitt, R.R(2008). Initial Thoughts on FAF2 Experience and FAF3 Design. Planning Team, FHWA Resource Center.

[102] Shihab, A. I. (2001). Fuzzy Clustering Algorithms and Their Application to Medical Image Analysis (Doctoral dissertation, University of London).

[103] Southworth, F. (2003). Freight Transportation Planning: Models and Methods. In: Transportation System Planning. Methods and Applications.

[104] Steiger, J. H. (1990). Structural Model Evaluation and Modification: An Interval Estimation Approach. Multivariate Behavioral Research, 25(2), 173-180.

[105] Suhr, D. (2006). The Basics of Structural Equation Modeling. Presented: Irvine, CA, SAS User Group of the Western Region of the United States (WUSS).

[106] Talvitie, A. (1973). A Direct Demand Model for Downtown Work Trips. Transportation, 2(2), 121-152.

[107] Tok, A., Hyun, K. K., Hernandez, S., Jeong, K., Sun, Y. E., Rindt, C., & Ritchie, S. G. (2017). Truck Activity Monitoring System (TAMS) for Freight Transportation Analysis (No. 17-06807).

[108] Tok, A., Zhao, M., Chow, J., Ritchie, S., & Arkhipov, D. (2011). Online Data Repository for Statewide Freight Planning and Analysis. Transportation Research Record: Journal of the Transportation Research Board, (2246), 121-129.

[109] Turnquist, M. A. (2006). Characteristics of Effective Freight Models. Freight Demand Modeling, 11.

[110] Transportation Research Circular: Statewide Travel Demand Modeling: A Peer Exchange (2005). Number E-C075.

[111] Transportation Research Circular (2011), Commodity Flow Survey Workshop. Number E-C158.

[112] Truck Activity Monitoring System website, Accessed online July 2016. http://freight.its.uci.edu/tams/

[113] UCI-ITS. (2013). California Vehicle Inventory and Use Survey – Framework and Pilot Study.
[114] United States. Department of Transportation. Research, Innovative Technology Administration, National Cooperative Freight Research Program, Cambridge Systematics, & GeoStats, LLP. (2010). Freight-demand Modeling to Support Public-sector Decision Making (Vol. 8). Transportation Research Board.
[115] Van Acker, V., & Witlox, F. (2010). Car Ownership as A Mediating Variable in Car Travel Behaviour Research Using A Structural Equation Modelling Approach to Identify Its Dual Relationship. Journal of Transport Geography, 18(1), 65-74.
[116] Van Acker, V., Witlox, F., & Van Wee, B. (2007). The Effects of the Land Use System on Travel Behavior: A Structural Equation Modeling Approach. Transportation Planning and Technology, 30(4), 331-353.
[117] Van Bergeijk, P. A., & Brakman, S. (Eds.). (2010). The Gravity Model in International Trade: Advances and Applications. Cambridge University Press.
[118] White, W. L., Anater, P., & Wall, B. (2007). Development of the Pennsylvania Statewide Commodity-Based Freight Model. In 11th TRB National Transportation Planning Application Conference.
[119] Wisetjindawat, W., Sano, K., & Matsumoto, S. (2005). Supply Chain Simulation for Modeling the Interactions in Freight Movement. Journal of the Eastern Asia Society for Transportation Studies, 6, 2991-3004.
[120] Wisetjindawat, W., Sano, K., & Matsumoto, S. (2006). Commodity Distribution Model Incorporating Spatial Interactions for Urban Freight Movement. Transportation Research Record: Journal of the Transportation Research Board, (1966), 41-50.
[121] Wolfram, Fuzzy Logic Manual: 1.12 Fuzzy Clustering. Accessed online August 2018, http://reference.wolfram.com/legacy/applications/fuzzylogic/Manual/12.html
[122] Wurfel, E., Bai, Y., Huan, L., & Buhr, V. (2009, August). Development of a Freight Analysis Framework for the Kansas City Metropolitan Area. In 2009 Mid-Continent Transportation Research SymposiumIowa Department of TransportationIowa State University, AmesUniversity of Northern Iowa, Cedar FallsNational Center for Freight and Infrastructure Research and Education (CFIRE) Wisconsin Department of Transportation.

[123] Wurel, E., Bai, Y., Huan, L., & Buhr, V. (2009). Freight Analysis Framework for Major Metropolitan Areas in Kansas. University of Kansas Center for Research, Inc..

[124] Xue, Q.L., Introduction to Path Analysis. Statistics for Psychosocial Research II: Structural Models. Johns Hopkins Bloomberg School of Public Health. Accessed online January 2018, http://ocw.jhsph.edu/courses/StructuralModels/PDFs/Lecture3.pdf

[125] Yang, C. H., Chow, J. Y., & Regan, A. (2009). State of The Art of Freight Forecasting Modeling: Lessons Learned and The Road Ahead (No. 09-3384).

[126] You, S. I. (2012). Methodology for tour-based Truck Demand Modeling. University of California, Irvine.

7 APPENDIXES

Appendix A: Path Diagrams of Total Generation Model

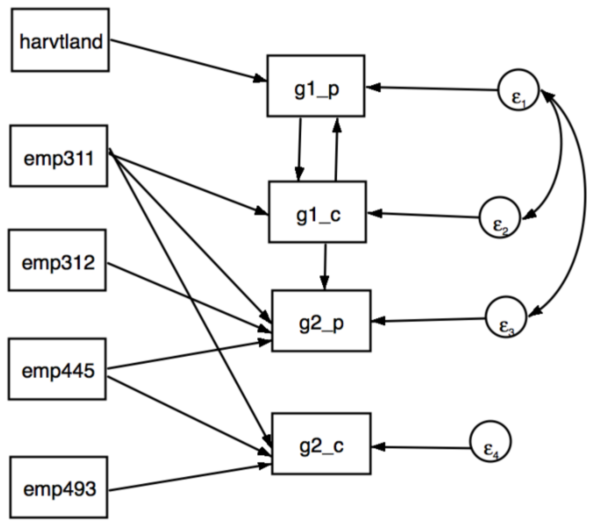


Figure A-1 Path Diagram 1 of Total Generation Model

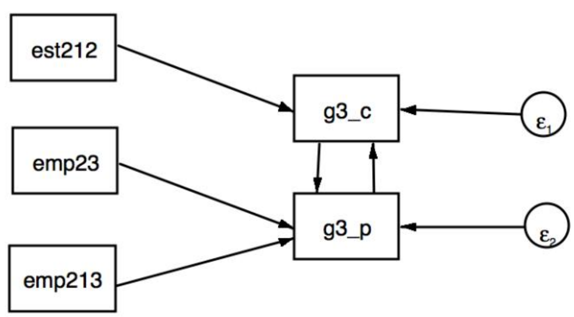


Figure A-2 Path Diagram 2 of Total Generation Model

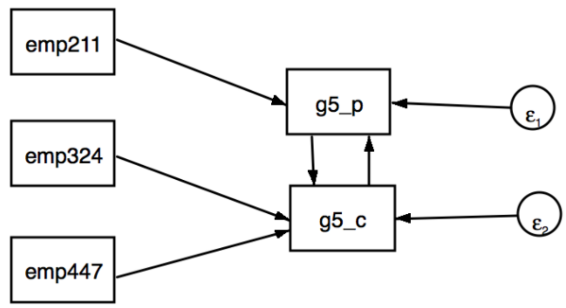


Figure A-3 Path Diagram 3 of Total Generation Model

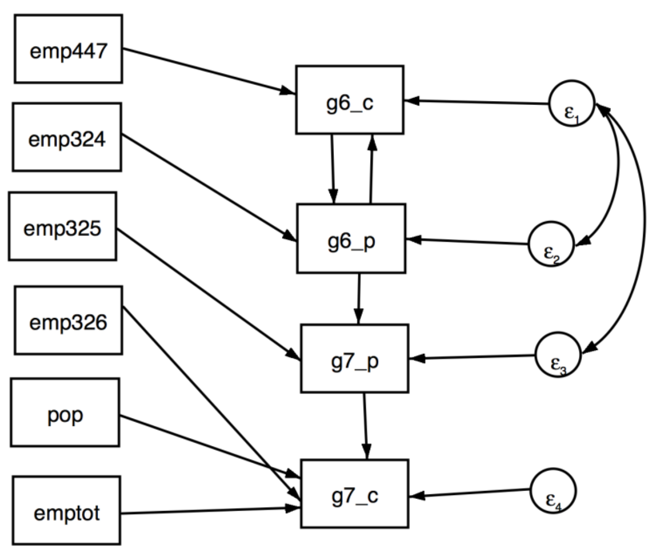


Figure A-4 Path Diagram 4 of Total Generation Model

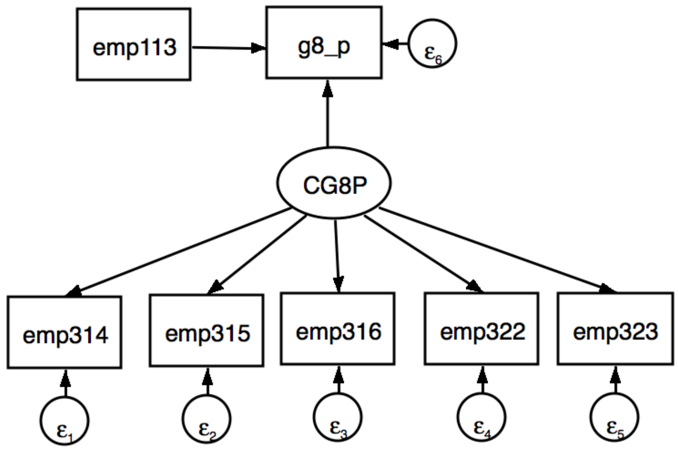


Figure A-5 Path Diagram 5 of Total Generation Model

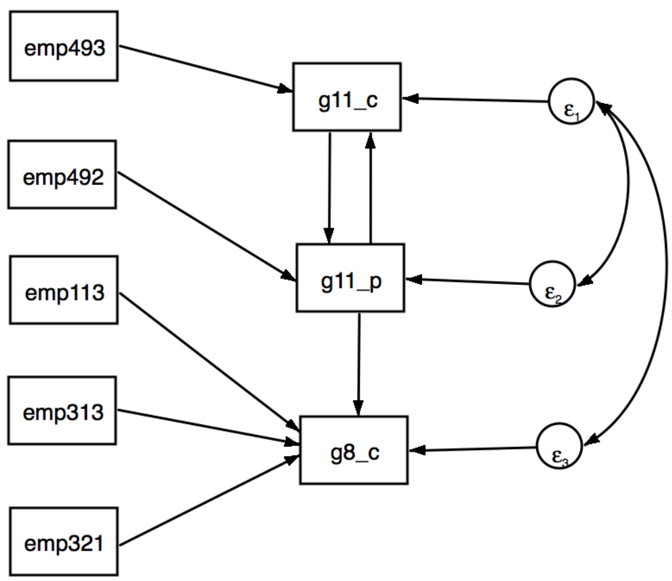


Figure A-6 Path Diagram 6 of Total Generation Model

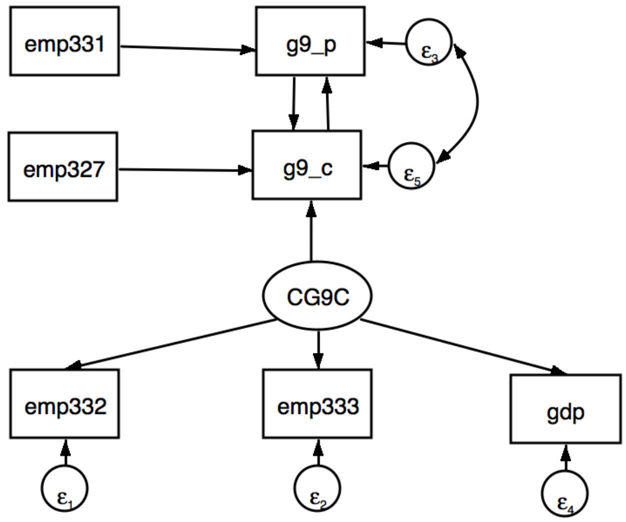


Figure A-7 Path Diagram 7 of Total Generation Model

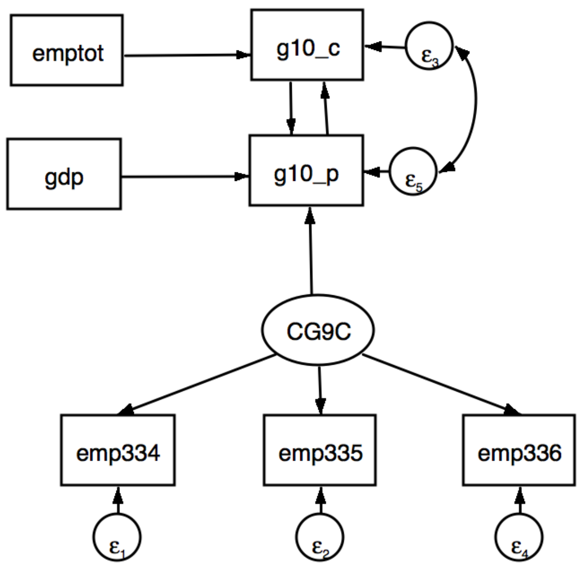


Figure A-8 Path Diagram 8 of Total Generation Model