Final Report

COMMERCIAL SERVICES VEHICLE MODEL AND SURVEY

CHICAGO METROPOLITAN AGENCY FOR PLANNING (CMAP)

Joanne Barry, Vice President - Research
R.A. Malatest & Associates Ltd.
858 Pandora Avenue
Victoria, BC V8W 1P4
Ph: 250-384-2770 ext 406 | Fax: 250-384-2774

Web: www.malatest.com

MALATEST

TABLE OF CONTENTS

1. BACKGROUND AND OBJECTIVES

1.1 Study Objectives

The overall purpose of the Commercial Services Vehicle Model and Survey was to develop a commercial services vehicle model (CSVM) for the Chicago Metropolitan Agency for Planning (CMAP). There were two foundational activities within this project to support the overall purpose:

- Implementing a commercial vehicle survey (CVS) to collect numerous data points about non-freight commercial vehicle activities in seven counties that can be used in the model, as well as collecting data about the firms, companies and organizations within the counties that own and deploy such vehicles; and
- Developing a commercial vehicle tour estimation process for northeastern Illinois to be used to generate a set of weekday commercial services vehicle tour trips that can be applied under current conditions, as well as used as a forecasting tool by adjusting various model parameters.

1.2 Background

While the outputs produced by CMAP's freight truck-touring model and the regional commercial vehicle touring model implemented alongside the freight truck touring model by RSG in other areas are similar, there are several important differences to each model's underlying assumptions and approach. The freight truck-touring model focuses on accounting for the delivery and pick-up of shipments at businesses. The demand for truck travel is explicitly connected to the production, consumption, and transfer of those shipments. The commercial vehicle touring model takes a different approach. It estimates the demand for service stops and local goods delivery stops at businesses as well as at residences using locally estimated stop generation models that are a function of the characteristics of the businesses in the region that operate commercial vehicles, the land use in the region, and the proximity of potential customers to businesses. This means that, while the freight truck-touring model is connected to external demand (i.e., the national supply chain model), the commercial vehicle touring model is not influenced by that factor. Instead, it covers a broader set of trip purposes and is more appropriate for the wide range of uses of light commercial vehicles that go beyond the pick-up and delivery of freight shipments.

1.3 Team Accountabilities

R. A. Malatest & Associates Ltd. (Malatest) acted as the prime contractor for this study and had responsibility for the following:

- Overall project management;
- Client liaison;
- Scheduling all activities and ensuring timelines were met;
- Designing and selecting the survey sample and final methodology;
- Drafting all survey instruments, including communications materials;
- Providing weekly progress reports on data collection;
- Reviewing and validating the collected survey data;

- Developing a sound data weighting methodology (if necessary) to allow the survey data to be expanded to be representative of the target population universe (commercial businesses generating/receiving shipments/services and commercial drivers reporting stops on full-day tours); and
- Drafting and finalizing reports that detail the data collection process and results.

Resource Systems Group, Inc. (RSG) served as a subcontractor and associate to Malatest for this study. RSG had the responsibility for the following:

- Data processing (various data sources for model development purposes);
- Estimation and calibration of the model;
- All processing scripts and open source modeling procedures;
- Delivery of all datasets and data files; and
- Drafting a technical memorandum which includes recommended updates, future improvements, and maintenance.

2. SURVEY METHODOLOGY

The following section details the activities that were undertaken in order to implement the CVS, including reviewing the data already available for model development to identify the gaps that survey data would fill, and then the process of designing and administering the CVS.

2.1 Data Review Summary

RSG worked with CMAP to obtain the data required for the estimation, calibration, validation and application of the CSVM. In the initial phase of the project, RSG reviewed data provided by CMAP and additional public and transferred data to determine what data currently existed and what data needed to be provided by the establishment and driver surveys. Chapter 3 of this report describes the processing and analysis of these data, including the newly collected establishment and driver surveys.

The data review technical memorandum summarized this data review. Each of the reviewed data items is shown in Figure 1, including the anticipated use in the development of the new commercial vehicle model. The combination of the available data and the proposed establishment and commercial vehicle survey was found to be adequate to support the intended model design.

Figure 1: DATA STATUS SUMMARY

Data Item	Year	Source	Data Type
Establishment Data	2017	Census County Business Patterns	Input
CMAP Regional Model Land Use Data	2019, 2050	CMAP	Input
CMAP Regional Model Travel Time, Distance, and Toll Skims	2019, 2050	CMAP	Input
Warehouses, Distribution Centers, and Intermodal Terminals Locations	2017	CMAP	Input
Establishment Survey	2021	CMAP (collected in this project)	Estimation
Commercial Vehicle Survey	2021	CMAP (collected in this project)	Estimation
Establishment and Commercial Vehicle Survey	2017	SEMCOG	Estimation, Calibration
Truck Origin-Destination GPS Data	2018	INRIX (processed by CMAP)	Calibration, Validation
Vehicle Classification Counts	2015	CMAP	Validation
Vehicle Miles Traveled Data	2019	FHWA	Validation

2.2 Data Requirements for Commercial Vehicle Survey

The data review identified a need for two types of data, at the establishment level to support understanding which types of businesses operate commercial service vehicles and for what purposes those vehicles make trips, and at the commercial vehicle level to understand vehicle travel behaviour (such as the number, types, and lengths of tours and trips). This need suggested a typical two stage commercial vehicle survey including an establishment survey and a commercial vehicle drive diary survey. The general parameters of those two surveys are introduced here and the administration of the surveys are described in the remainder of this section of the report.

Establishment Survey

To understand the number and type of commercial vehicles that are operated by business establishments in the region, Malatest developed an establishment survey. This survey targeted delivery and service-providing establishment types, using the following North American Industrial Classification System (NAICS) groupings:

- Retail and restaurants, (NAICS 44-45, 72)
- Government & Education (NAICS 61, 92)
- Utilities & Construction (NAICS 22, 23)
- Wholesale and distribution (NAICS 42)
- Non-construction building services/waste & remediation (NAICS 5617, 562)
- General services (Home health care & diagnostic labs, insurance adjusters, etc.) (NAICS 51-56, 62, 72, 81 other than 5617/562)

Commercial vehicle survey

To understand commercial vehicle travel behavior, Malatest developed a commercial vehicle driver diary survey, administered to a subset of the business establishments who participated in the establishment survey. It collected information on the origins and destinations of 24 hours of trips made by vehicles operated by the establishment, including details such as trip purpose, the duration and location type of stops, and information about the vehicle. This survey also targeted delivery and service-providing trips.

2.3 Survey Instrument Design

The design of the two survey instruments accounted for missing information or data gaps identified at this data review stage. The surveys included questions designed to collect the information outlined in Figure 2.

Figure 2: SURVEY DESIGN

Survey 1: Establishment Survey - Completed by establishments whether or not they employ their own drivers	
Form A	**Establishment Information:** Establishment size and type details, annual shipping information (value, quantity and location), trip generation information (e.g., # of vehicles arriving and departing business on a sampled day), frequency of shipments by mode, extent of vehicle fleet, # of employees who are drivers (if any)
Form B	**Outbound Commodities**: Quantity and size of goods or service shipments coming from the establishment on a typical day, its origin, type, weight and value
Form C	**Inbound Commodities:** Quantity and size of goods or service shipments received by the establishment on a typical day, its destination, type, weight and value
Survey 2: Drivers – Completed by drivers	
Form D	**Vehicle Information and Stop Information:** description of the vehicle used, ownership, capacity and fuel type, availability of GPS; stops made on the survey date, stop descriptions, locations, purposes and times, commodities dropped off/picked up
Form E	**Tour Information**: Whether pickup delivery stops, service stops, locations, order and routes were typical of other days

2.4 Sampling and List Sources

Malatest obtained randomly generated lists from Dun & Bradstreet, a broker that has business listings in the Chicago area. This decision was based on the ability to customize samples by both geography and industry sector, the ability to directly choose filters, and the ability to draw additional sample if required at a later date.

The initial sample drawn on August 6, 2021 consisted of 4,014 records. Additional samples were drawn on later dates due to the initial sample having many non-qualifiers upon surveying, and due to sample records being exhausted from calling. Additional samples were drawn, with all sample draw dates and sizes in Figure 3:

Figure 3: SURVEY SAMPLE SIZES

Sample Date	n
August 6, 2021	4,014
October 22, 2021	804
January 20, 2022	2,870
February 1, 2022	5,012
Various (non-D&B)	10
Total	**12,710**

The geographic area of selection included the following groupings of counties in Figure 4 :

Figure 4: SAMPLES BY COUNTY

County	Unique Records Available (D&B)	Sample Draw
Cook County – City of Chicago	17,941	3,253
Cook County - Suburban	14,396	3,380
Du Page / Lake	13,109	3,274
Kane / Kendall / McHenry	4,961	1,609
Will	3,193	1,194
Study Area Total	**53,600**	**12,710**

* The Cook County figure represents all of Cook County and is not divided between Chicago and the remainder of Cook County.

Within each of the six industry groups in the sample universe, a random sample of businesses in each area, moderated by the geographic distributions above was used.

Figure 5: SAMPLE BY INDUSTRY

NAICS Codes and/or specific businesses (most likely to be associated with a commercial service delivery)	Unique Records Available (D&B)	Sample Draw
Retail and restaurants, (NAICS 44-45, 72)	6,649	2,545
Government & education (NAICS 61, 92)	4,531	808
Utilities & construction (NAICS 22, 23)	3,651	1,902
Wholesale and distribution (NAICS 42)	4,297	1,962
Non-construction building services/waste & remediation (NAICS 5617, 562)	753	760
General services (home health care & diagnostic labs, insurance adjusters, etc. NAICS 51-56, 62, 81 other than 5617/562)	33,719	4,734

County Business Patterns (CBP) data were used for initial reference information on the survey universe. The industry stratification outlined in this section was based on a sample frame that targets commercial services movement.

2.5 Commercial Vehicle Survey Completion Targets

Malatest targeted the completion of 270 establishment surveys and 360 driver surveys. In anticipation of a spoilage rate of approximately 5% among establishments and 15% among drivers, a total of 285 establishment surveys and 415 driver surveys were anticipated. Establishment and driver survey completions were divided among the six industry subsectors listed in Figure 6, with driver surveys expected to follow similar patterns, depending on the number of drivers available from participating establishments.

Figure 6: COMPLETION TARGETS BY INDUSTRY

NAICS Codes and/or specific businesses (most likely to be associated with a commercial service delivery)	Sample Draw	Soft Completion Targets
Retail and restaurants, (NAICS 44-45, 72)	2,545	45
Government & education (NAICS 61, 92)	808	40
Utilities & construction (NAICS 22, 23)	1,902	40
Wholesale and distribution (NAICS 42)	1,962	40
Non-construction building services/waste & remediation (NAICS 5617, 562)	760	25
General services (home health care & diagnostic labs, insurance adjusters, etc. NAICS 51-56, 62, 81 other than 5617/562)	4,131	95
TOTALS	**12,710**	**285**

Soft completions targets were established by geographic region, though these targets were a lesser priority compared to industry targets. These targets are shown in Figure 7:

Figure 7: COMPLETION TARGETS BY COUNTY

County	Sample Draw	Soft Completion Targets
Cook County – City of Chicago	3,253	80
Cook County - Suburban	3,380	70
Du Page / Lake	3,274	70
Kane / Kendall / McHenry	1,609	35
Will	1,194	30
Study Area Total	**12,710**	**285**

2.6 Survey Implementation Process

The survey implementation process that was followed is outlined below.

- Step 1: Send establishments/employers a 'soft-notice' hard copy letter via mail to notify them about the survey and why it is important for them to participate.
- Step 2: Send a formal invitation email to establishments/employers with a link to the survey (accessible via www.CMAP.malatest.net). This website included access to Frequently Asked Questions (FAQ), instructions on how to participate, and details on how the establishments can invite drivers to participate in the driver survey. Malatest also conducted recruitment telephone calls to establishments who had not participated in the online survey.
- Step 3: Establishments were screened by NAICS code prior to sample draws from Dun and Bradstreet, as well as through the survey or during recruitment telephone calls to ensure they meet the eligibility criteria established for the study. Establishment representatives were asked if they have drivers or not. Those with drivers were asked to specify the number of drivers they had.
    - Note that all telephone recruitment efforts for each establishment was tracked. Call disposition codes and company call sheets were regularly reviewed to ensure correct tracking of call outcomes – especially those relevant to understanding the composition of sample universe (e.g., non-qualifiers, not-in-service, business reported as closed).

- Step 4: Eligible establishments/employers were invited to complete the survey online in English. They also had the option to call in to complete the survey over the phone in English or a different language (e.g., Spanish).
- Step 5: Establishments/employers with drivers were asked how they prefer to recruit drivers to participate in the driver survey: Each establishment/employer was assigned an ORG# to link drivers to employers. Employers could either send a recruitment email to their drivers with the ORG#, or provide Malatest with their driver contact information (e.g., their smartphone number or email) so Malatest could invite drivers to participate. Malatest only contacted those drivers whose contact information was provided by the establishments/employers. Beginning in December, a training manual was developed, and training was conducted with Chicago-based surveyors to conduct in-person visits to employers. Eligible establishments/employers were asked through the survey and during recruitment calls if they would allow an in-person visit to their business from surveyors at Quercus Consulting (sub-contracted through Malatest). During in-person visits, the representatives from Quercus informed the drivers directly about the survey by responding to questions and handing out postcards about the survey, left postcards for drivers who were not present, and provided the employer with recruitment materials for the driver survey (if requested).
- Step 6: Drivers were asked to register for the survey using their ORG# (at www.CMAP.malatest.net) to access the survey. The Driver Survey asked drivers to complete an activity log for one day of the week. The activity log ideally should have been completed on the same day as the data logged or within a day. Drivers were asked to record all the stops they made on the selected survey day. Malatest followed-up with drivers via email who registered for the survey or partially completed the survey. If the establishment recruited the drivers, Malatest asked the establishment to follow-up with drivers who had not participated.
- Step 7: Sent thank-you letters and incentives to participating establishments and drivers. Establishments received a $50 incentive (by check) and drivers received $100 (by being provided an online gift card). A few additional gift cards were provided when the respondent was inconvenienced (e.g., Internet issue, missed appointment).

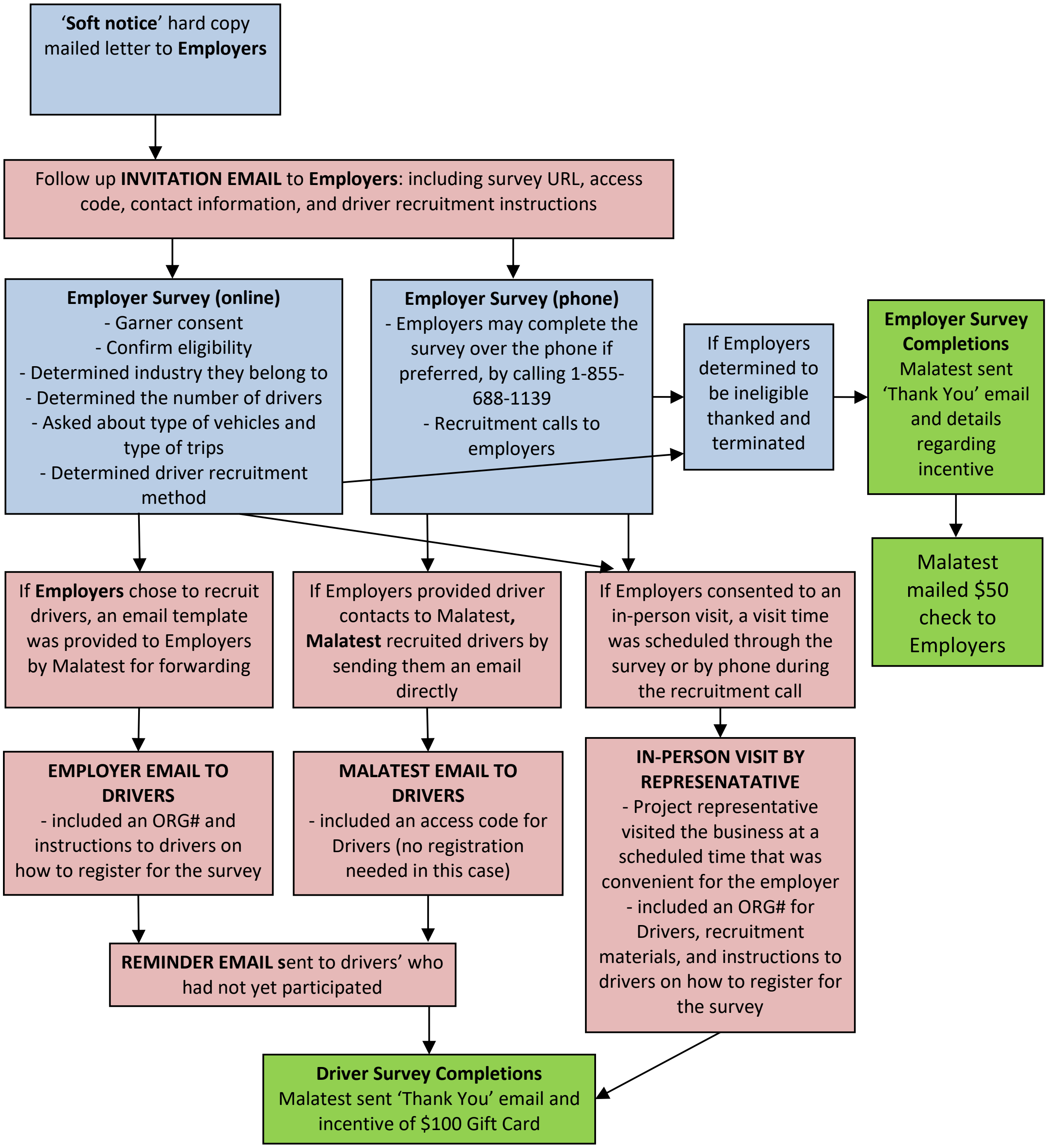
A pre-test of both the establishment survey and the driver survey was conducted in advance of the full survey launch. The purpose of the pre-test was to assess the functionality of the survey on different mobile devices and operating systems, and to ensure the questions were clear to respondents.

For the full survey launch, sample was released in batches rather than all at once to reduce non-response bias and retain the random nature of the sample. For earlier samples (drawn on August 6, 2021 and October 22, 2021), a total of 4-6 call attempts were made before a record was considered exhausted and retired. For later samples (drawn on January 20, 2022 and February 1, 2022), all records received at least one initial phone call. Due to the volume of records that needed to be contacted, follow-up call attempts were focused on contacts that either had either indicated they had eligible drivers, or had expressed interest in participating (e.g., scheduled a time for a surveyor to call back).

While the survey was in field, weekly meetings were held with CMAP, Malatest, and RSG representatives in attendance. The meetings reported on survey progress and discussed strategies that could be implemented to improve survey response rates. From these meetings, several methodological changes were devised and implemented, including:

- Adjusting or adding survey questions:
    - An additional employer eligibility question was added (October 7, 2021)
    - Two employer survey questions were changed from open-text response format to a multiple-choice, response-limited format (November 5, 2021)
    - A question asking employers if a surveyor could visit their business in-person was added (February 4, 2022)
- Refining the list of NAICS codes used to draw samples from Dun & Bradstreet to target businesses that were more likely to employ commercial drivers (October 21, 2021);
- Offering employers a copy of the driver survey to alleviate concerns about the questions their drivers would be asked (October 29,2021)
- Hiring Quercus Consulting to conduct in-person visits to businesses that consented to a visit during a phone call with a surveyor (December 9, 2021)
- Requesting to schedule an in-person visit to the business when speaking with employers over the phone in order to recruit drivers, rather than requesting that the employer recruit drivers or provide a list of drivers for Malatest to contact (December 9, 2021)
- Designing and distributing postcards and posters during in-person visits to businesses, which contained information on how to participate in the driver survey (December 9, 2021 to end of data collection)
- Adding an incentive of $25 for drivers who referred other drivers to participate in the survey (December 15, 2022)
- Conducting phone calls to employers who participated in the survey online but declined participation in the driver survey to ask if a surveyor could visit their business in-person to inform their drivers of the driver survey (February 17, 2022)

Figure 8: SURVEY PROCESS MAP



3. DATA PROCESSING AND ANALYSIS

The following section presents a summary of the final CVS data and then describes the cleaning and weighting of the CVS data and the processing and analysis of the other data that were used to supplement the survey data and produce model input data and model estimation and calibration datasets.

3.1 Final Survey Data

An overview of survey completions and employer and driver outcomes is provided in Figure 9:

Figure 9: SURVEY RESULTS OVERVIEW

Outcome	Total
Completed employer surveys	**207**
Employer surveys completed by phone	107
Employer surveys completed online	100
Employers consented to driver participation	74
Employers consented to an in-person visit	18
Drivers registered for the survey	43
Completed driver surveys	**35**

A breakdown of the amount and recipients of incentives paid is displayed in Figure 10. Please note that one employer refused compensation and returned their cheque, therefore this has been removed from the total incentives. All drivers were paid $100 in their choice of gift card (Amazon, Apple, Google Play, Office Max/Depot, Bass Pro Shops) for their one-day trip diary, and three drivers were paid $125 as they were compensated $25 extra after experiencing registration issues. After verifying the quality of drivers' trip diaries, five drivers who were paid were removed from the final driver sample after verifying the quality of their trip diary data during data cleaning.

Figure 10: INCENTIVES PAID

Recipient	n	$
Employers	206	$10,300
Drivers	40	$4,075
Driver Referrals	0	$0
Other	-	*$150
Total	**-**	**$14,525**

***** Includes a $100 incentive that was provided to a driver dispatcher and an extra $50 incentive to an employer for a rescheduled visit.

Valid employer surveys were completed by 207 employers. Employer survey completions are shown in Figure 11 by county and in Figure 12 industry.

Figure 11: SURVEY TARGETS & RESULTS (BY COUNTY)

County	Sample	Target Quotes	Completes	Quota Achieved (%)	Response Rate (%)
Cook County–City of Chicago	2,195	80	45	56.3%	2.1%
Cook County-Suburban	2,356	70	61	87.1%	2.6%
DuPage/Lake County	2,338	70	47	67.1%	2.0%
Kane/Kendall/McHenry County	1,149	35	32	91.4%	2.8%
Will County	819	30	22	73.3%	2.7%
Total	**8,857**	**285**	**207**	**72.6%**	**2.3%**

Table values reflect targeted samples only (i.e., excludes n = 2,132 records from the initial sample) and excludes 1,721 invalid cases (i.e., not in service numbers, wrong numbers, fax and modem lines, and non-qualifiers).

Figure 12: SURVEY TARGETS & RESULTS (BY INDUSTRY)

Industry	Sample	Target Quotes	Completes	Quota achieved (%)	Response Rate (%)
Retail and Restaurants	1,866	45	38	84.4%	2.0%
Government & Education	158	40	28	70.0%	17.7%
Utilities and Construction	1,511	40	43	107.5%	2.8%
Wholesale and Distribution	1,599	40	30	75.0%	1.9%
Non-building Services/waste and remediation	665	25	23	92.0%	3.5%
General Services	3,058	95	45	47.4%	1.5%
Total	**8,857**	**285**	**207**	**72.6%**	**2.3%**

Table values reflect targeted samples only (i.e., excludes n = 2,132 records from the initial sample) and excludes 1,721 invalid cases (i.e., not in service numbers, wrong numbers, fax and modem lines, and non-qualifiers).

As shown in Figure 13, of the 207 employers, 96% employed full-time and/or part-time drivers, and 4% outsourced their drivers.

Figure 13: DRIVERS EMPLOYED BY FIRM

Driver Employment	n	%of Employers
Full-time (30hoursormoreperweek)	171	65%
Part-time (lessthan30hoursperweek)	83	31%
None, drivers are outsourced	10	4%

The average number of full-time and part-time individuals employed at a business was 54 (n = 207), and most employers had fewer than 20 individuals employed at the location.

Figure 14: EMPLOYEES AT BUSINESS LOCATION

Number of Employees at Location	n	%of Employers
Less than 20	130	63%
20 to 100	57	28%
100 or more	20	10%
Total	**207**	**100%**

On average, an employer owned or leased 16.80 vehicles for business (n = 183), with most of these vehicles being passenger cars, sport utility vehicles, or pickup trucks. Many employers also possessed single unit trucks.

Figure 15: VEHICLES OWNED BY EMPLOYER BY VEHICLE TYPE

Vehicle Type	n	Mean	%of Employers
Passenger cars, sport utility vehicles for pickup trucks	154	12.7	74%
Single unit trucks	87	5.3	42%
Combo units (tractors/trailers)	23	35.0	11%
Other (open-end response)	30	5.2	14%

Valid driver surveys were completed by 35 drivers. Driver survey completions are shown below by county and industry.

Figure 16: COMPLETES BY COUNTY

County	Completes
DuPage/Lake County	7
Cook County–City of Chicago	6
Will County	1
Cook County-Suburban	24
Kane/Kendall/McHenry County	0
Total	**35**

Figure 17: COMPLETES BY INDUSTRY

Industry	Completes
Government & Education	5
Wholesale and Distribution	14
Utilities and Construction	2
Retail and Restaurants	1
Non-building Services/waste and remediation	0
General Services	13
Total	**35**

For 94% of drivers, their vehicle was owned by themselves or their employing organization (n = 35). Approximately half (51.4%) indicated that their vehicle was registered as a commercial vehicle in Illinois.

Figure 18: OWNER OF VEHICLE

Owner of the Vehicle	n	% of Drivers
The organization	19	54%
The driver	14	40%
Leasing company	1	3%
Other (open-end response)	1	3%
Total	**35**	**100%**

The majority of drivers' primary business use for the vehicle was as a commercial service vehicle.

Figure 19: VEHICLE PRIMARY USE

Primary Vehicle Use	n	%ofDrivers
Cargo/Freight Transport Vehicle(used PRIMARILY to transport cargo)	2	6%
Commercial Service Vehicle (used PRIMARILY for non-cargo transport purposes)	26	74%
Commercial Service and Cargo Delivery Vehicle(used for both service and cargo)	7	20%
Total	**35**	**100%**

3.2 Survey Data Cleaning and Weighting

Upon the completion of data collection, the data was cleaned and verified. Data were checked for consistency, that skip patterns had been followed correctly and that data was linked properly (e.g., drivers properly linked to their employers). In addition, additional trip logic checks were undertaken to verify that the geocoded location coordinates (geocoded post-survey based on the survey data) and reported trip tours appeared to be reasonable. Geocoded locations included the business establishment location, the location at which the vehicle was garaged, the starting location of the vehicle on the day of the reported trip, and trip stops. A few drivers did not report a final return-to-base trip after their final stop reported trip stops, and final trips returning to their starting location were imputed in these instances.

Data weighting allowed the data to be scaled up in such a way that it is generally representative of the universe of establishments within the industry group and geography. Data weights were calculated for two geographies (Cook County vs other counties) and six industry groups (retail and restaurants; government and education; utilities and construction; wholesale and distribution; non-construction building services; general services), for a total of 12 weighting strata. Weighting for each stratum were calculated by estimating the number of eligible business establishments in the stratum (i.e., those with vehicles and drivers that would undertake commercial vehicle trips). Estimates of the eligible number of businesses were developed by referring to a) CBP data (which provides the universe of all business establishments); b) the rate of non-qualifying businesses identified through NAICS codes prior to sampling (that is, NAICS codes were used to identify any businesses prior to sampling that would be unlikely to employ drivers), and; c) the rate of non-qualifying businesses determined during surveying (that is, the percentage of contacted businesses that were found to not have drivers). The nonqualification rates within each stratum were used to reduce the estimate of the eligible universe to produce a final estimate of the number of eligible businesses with drivers in each geography and industry group.

A set of expansion weights was developed to weight the employer surveys such that they represent the estimated total eligible businesses within each geography and industry group. A normalized set of expansion weights was also developed that sums to the total number of survey completions. No data weighting was performed for the driver survey data due to the limited number of driver survey completions. The cleaned and weighted dataset was then delivered to RSG for use in modeling.

Figure 20 shows the number of firms, expanded to the estimates of eligible businesses, and categorized by the purpose of the trips that their commercial vehicles make on a typical day (as reported in the establishment survey[1]).

Figure 20: EXPANDED FIRMS BY FIRM INDUSTRY & ACTIVITY

Employment Group	Goods	Goods and Service	Service	Other	Total
Admin_Support_Waste	218	1,692	1,412	-	3,322
Construction	533	4,285	670	267	5,754
Ed_Health_Social_Public	-	914	917	528	2,359
Office_Professional	507	1,424	2,341	-	4,272
Retail	1,415	1,750	289	365	3,819
Service_FoodDrink	875	730	-	183	1,788
Service_Other	302	1,219	507	-	2,028
Wholesale	554	498	249	46	1,348
Total	**4,404**	**12,513**	**6,385**	**1,388**	**24,689**

Figure 21 shows the total number of trips made by commercial vehicles owned by the establishments on a typical day for either goods pickup and delivery or to service purposes, based on responses to the same establishment survey question. This table also shows data expanded to the estimates of eligible businesses, meaning that the trip numbers are an estimate of the total number of commercial vehicle trips in the region made by vehicles owned by eligible businesses. Given the wording of the questions, it is likely that the responses fall somewhere between the number of tours started or ending at the business locations and the actual number of trips made by the vehicle including the stop-to-stop trips made during tours.

[1] Based on responses to questions AQ6A, “On a typical weekday, how many trips do owned/leased vehicles make to pickup cargo to bring to this location?”, AQ6B, “On a typical weekday, how many trips do owned/leased vehicles make to deliver cargo from this location to a different location?, and AQ6C, “On a typical weekday, how many trips do owned/leased vehicles make to travel from this location to provide a service at a different location?”

Figure 21: EXPANDED TRIPS BY INDUSTRY

Employment Group	Expanded Trips	% of Trips
Admin_Support_Waste	233,170	33%
Construction	206,284	29%
Ed_Health_Social_Public	59,153	8%
Office_Professional	50,385	7%
Retail	84,292	12%
Service_FoodDrink	14,273	2%
Service_Other	33,551	5%
Wholesale	22,588	3%
Total	**703,696**	**100%**

3.3 Data Analysis of Supplemental Data

RSG processed the other data sources supplemental to the new survey datasets collected during this project as required for the estimation, application, and calibration of the CSVM. This section of the report describes the processing and presents analysis of each dataset in turn. The datasets that were ultimately used in the development of the CSVM are as follows (including the survey data):

- **Spatial data**
  - TAZ system from the CMAP travel demand model
  - Correspondences between TAZs and other more disaggregate zone systems (subzones) and aggregate zone systems and geographical boundaries (mesozones, counties, City of Chicago boundary, CMAP MPO's boundary).
- **Correspondences**
  - Definitions of CMAP employment categories and correspondence with other employment data using standard NAICS codes
- **Land use data**
  - Business location data: business establishments by size, industry and location from 2017 CBP data
  - Employment data: number of jobs by industry (top level 2-digit NAICS code) by TAZ from the CMAP travel demand model for base year (2019) and future year (2050)
  - Household data: number of households by TAZ from the CMAP travel demand model for base year (2019) and future year (2050)
- **Business and commercial vehicle characteristics**
  - CMAP establishment survey data collected in 2021 and 2022 (described above)
  - CMAP commercial vehicle driver diary collected in 2021 and 2022 (described above)
  - Southeast Michigan Council of Governments (SEMCOG) commercial vehicle survey data collected in 2017 (the original data source used for estimation of the transferred models and also used to derive additional calibration tabulations)
- **Commercial vehicle origin-destination data**
  - Passively collected commercial vehicle GPS data sample from 2018 activity procured from INRIX by CMAP and provided to RSG after initial processing by CMAP
- **Transportation supply data**

- ○ Travel time, distance, and toll skims for congested conditions for vehicles by class (light vehicles, medium trucks, and heavy trucks) and time period from the CMAP travel demand model for base year (2019) and future year (2050).
- **Transportation network usage data**
  - ○ Vehicle classification counts for the CMAP region for 2015
  - ○ Vehicle miles traveled (VMT) data for urban areas in Illinois and for the Chicago IL-IN urbanized area from FHWA for 2019

Spatial Data

The TAZ system used in the CSVM is the Zone 17 system from the CMAP travel demand model. Along with the definition of the Zone 17 TAZs, spatial correspondences between TAZs and other more disaggregate zone systems (subzones) and aggregate zone systems and geographical boundaries (mesozones, counties, the City of Chicago boundary, and CMAP MPO boundary) were used to develop geographical correspondences.

Figure 22: TAZ CORRESPONDENCE

District	County	TAZ (Minimum Number)	TAZ (Maximum Number)
Chicago	COOK, IL	1	717
COOK, IL (Outside Chicago)		718	1732
DUPAGE, IL	DUPAGE, IL	1733	2111
KANE, IL	KANE, IL	2112	2304
KENDALL, IL	KENDALL, IL	2305	2325
LAKE, IL	LAKE, IL	2326	2583
MCHENRY, IL	MCHENRY, IL	2584	2702
WILL, IL	WILL, IL	2703	2926
GRUNDY, IL (CMAP Part)	GRUNDY, IL	2949	2949
DEKALB, IL (CMAP Part)	DEKALB, IL	2977	2977
Non-CMAP Part of Model Region	DEKALB, IL	2976	3021
	GRUNDY, IL	2927	2950
	BOONE, IL	2951	2975
	KANKAKEE, IL	3022	3073
	LASALLE, IL	3074	3145
	LEE, IL	3146	3151
	OGLE, IL	3152	3168
	WINNEBAGO, IL	3169	3247
	LAKE, IN	3248	3344
	LAPORTE, IN	3345	3400
	PORTER, IN	3401	3467
	KENOSHA, WI	3468	3512
	RACINE, WI	3513	3568
	WALWORTH, WI	3569	3632

The CMAP model region covers 21 counties in northeastern Illinois and adjoining portions of Indiana and Wisconsin. Within that larger model region, the CMAP MPO region encompasses seven counties along with townships in two adjacent counties. For the purposes of calibration, validation, and presentation of results, the model region was divided into 11 "districts", which were defined as each of the nine counties or portions of counties in the CMAP MPO region, with Cook County split into the City of Chicago and the rest of Cook County, and an 11th district that covers the part of the model region outside of the CMAP MPO boundary. The districts and their TAZ ranges shown in Figure 22, which summarizes the common data input[2] file *TAZ_System.csv*.

Correspondences

Several sources of employment data and survey data related to establishments were used in the development of model input data, estimation data, and calibration data. This included CMAP employment data which used top level 2-digit NAICS categories and CBP data with 6-digit NAICS codes. For the specification of employment groupings in the CSVM, a more concise set of employment categories was developed represented by groupings of the top level 2-digit NAICS categories. The correspondence between NAICS categories and CSVM employment groups, which is a common input data file called *corresp_naics2_empcats.csv*, is shown in Figure 23.

Figure 23: NAICS2-MODEL EMPLOYMENT CORRESPONDENCE

NAICS2	NAICS Description	CSVM Employment Groups
11	Agriculture, Forestry, Fishing and Hunting	Transport_Industry
21	Mining, Quarrying, and Oil and Gas Extraction	Transport_Industry
22	Utilities	Construction
23	Construction	Construction
31, 32, 33	Manufacturing	Transport_Industry
42	Wholesale Trade	Wholesale
44, 45	Retail Trade	Retail
48, 49	Transportation and Warehousing	Transport_Industry
51	Information	Office_Professional
52	Finance and Insurance	Office_Professional
53	Real Estate and Rental and Leasing	Office_Professional
54	Professional, Scientific, and Technical Services	Office_Professional
55	Management of Companies and Enterprises	Office_Professional
56	Administrative and Support and Waste Management and Remediation Services	Admin_Support_Waste
61	Educational Services	Ed_Health_Social_Public
62	Health Care and Social Assistance	Ed_Health_Social_Public
71	Arts, Entertainment, and Recreation	Service_FoodDrink
72	Accommodation and Food Services	Service_FoodDrink
81	Other Services (except Public Administration)	Service_Other
92	Public Administration	Ed_Health_Social_Public

[2] Common data inputs are CSVM input files that are not scenario specific and are used for every run of the model; they are located in the *lib/data* folder

Land Use Data

The CSVM includes a firm synthesis model that synthesizes set of business establishments in the region by scaling a business establishment location dataset to match with TAZ employment data. The TAZ employment data, along with TAZ household data, are also used as explanatory variables in several of the model components. To produce future year commercial vehicle forecasts using the model, a forecast of TAZ employment and households is required.

Business location data

The business location data required by the firm synthesis model is a list of business establishments by size, industry and location, which was obtained from 2017 CBP data and processed by CMAP (originally for the CMAP freight model) into the common input data file *data_emp_cbp.csv*. The CBP includes about 300,000 establishments in the CMAP model region. Information about establishment location, size (number of employees), detailed industry classification, FAF zone and “CBPZone” (a combined zoning system combining CMAP mesozone and national FAF zones used in the CMAP freight model) is included. A sample of the data is provided in Figure 24.

Figure 24: SAMPLE OF THE ESTABLISHMENT DATA

NAICS6	FAFZONE	CBPZONE	employment	establishment	e1	e2	e3	e4	e5	e6	e7	e8
113110	11	1	23	3	0	0	0	0	0	0	0	0
113310	11	1	231	46	33	0	0	0	0	0	0	0
115210	11	1	102	18	10	0	0	0	0	0	0	0
115310	11	1	0	12	5	0	0	0	0	0	0	0
212111	11	1	383	7	0	0	0	0	0	0	0	0
212112	11	1	37	3	0	0	0	0	0	0	0	0
212312	11	1	192	12	3	0	0	0	0	0	0	0
213112	11	1	50	5	3	0	0	0	0	0	0	0
221112	11	1	0	6	0	0	0	0	0	0	0	0

The 2017 CBP data requires some corrections in the firm synthesis model to account for censoring to avoid disclosure issues. For example, cells in the firm size distribution (the e1 to e8 columns) are censored where their values are small, which means that they do not sum to the number of establishments (the establishment column). The approach to accounting for censoring in the CBP data in the firm synthesis model is described in the model design section.

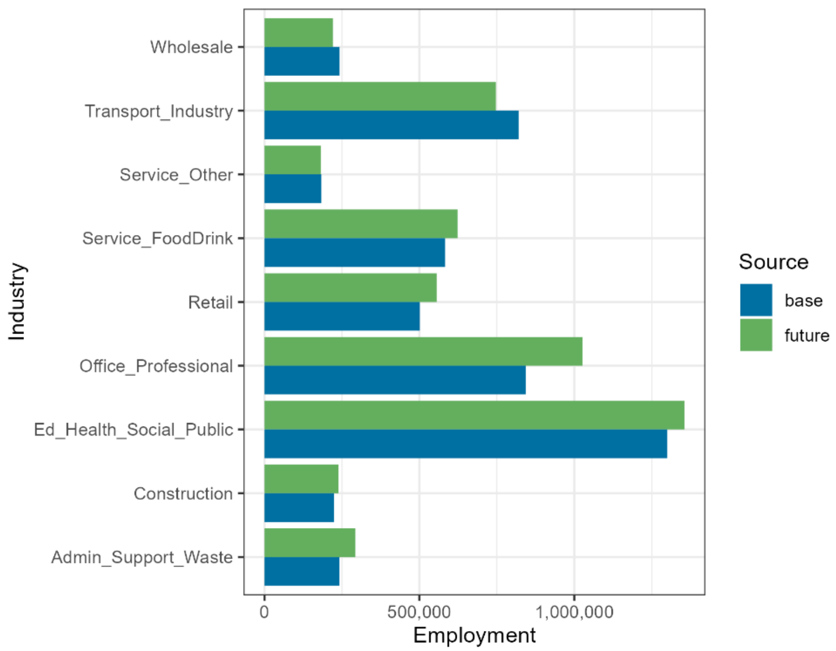
In addition to the CBP data, which does not cover either agricultural employment or public administration employment, a similarly formatted set of establishment data for farms derived from the USDA Census of Agriculture was used as a common input data file *data_emp_cbp_ag.csv*. As with the CBP data, this file was transferred from the CMAP freight model. The approach to synthesizing public administration establishments in the firm synthesis model is described in the model design section.

Employment Data

The employment data used by the CSVM is the number of jobs by industry by TAZ. It was summarized from the CMAP travel demand model's subzone employment data (jobs by top-level 2-digit NAICS code in each subzone) for the base year (2019) and future year (2050). The data were processed in the script *dev/Process_control_emp.R* into the scenario input files *data_emp_control_taz.csv*.

There are 4,941,000 jobs in 2019 and 5,242,000 in 2050, an increase of roughly 301,000 (6.1%). As shown in Figure 25, much of this growth is concentrated in the office and professional industry, while some categories such as wholesale and transport and industry are forecasted to decrease.

Figure 25: EMPLOYMENT BY INDUSTRY (2019 & 2050)

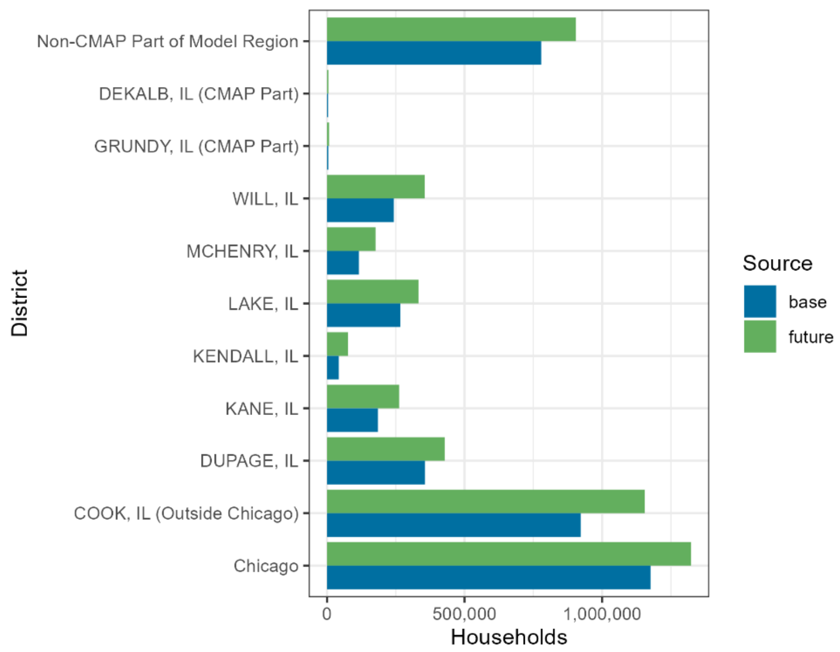


Household Data

The household data used by the CSVM is the number of households by TAZ. It was summarized from the CMAP travel demand model's subzone household data for 2019 and 2050. The data were processed in the script *dev/HH_TAZ_Count.R* into the scenario inputs *data_hh.csv*.

There are 4,094,000 household in the model region in 2019 and 5,026,000 in 2050, an increase of 22.8%. A significant portion of this growth in absolute terms if forecasted to place in the suburban portion of Cook County.

Figure 26: HOUSEHOLDS BY DISTRICT (2019 & 2050)



Business and commercial vehicle characteristics

In addition to the CVS conducted during this project in the CMAP region, the development of the CSVM made use of the SEMCOG CVS data collected in 2017. This was the original data source used for estimation of the transferred models that formed the starting point for model development of the CSVM and was also a source of additional calibration tabulations. The survey design, administration, and dataset are described in the project report “SEMCOG 2017 Commercial Vehicle Survey: Final Report”[3].

The SEMCOG CVS was comprised of three parts:

- An Establishment survey that was designed to understand the number and type of commercial vehicles that are operated by business establishments in the region.
- A Truck Travel Diary survey administered to a subset of the business establishments who participated in the establishment survey. It collected information on the origins and destinations of 24 hours of trips made by vehicles operated by the establishment, including details such as trip purpose, the types of freight moved, and information about the vehicle.
- A Passive GPS survey that collected GPS data for the trips made by a subset of the vehicles that participated in the truck travel diary.

The SEMCOG CVS effort is well described in the project’s final report. The data were provided to RSG in a spreadsheet with tabs containing establishment data, vehicle data, trip data, tour data, GPS data, and a data dictionary.

[3] SEMCOG 2017 CVS: Final Report, by ETC Institute, PDF document SEMCOG_CVS_FinalReport_20190319_Final.pdf

The dataset contains:

- 2,848 responses from establishments to the establishment survey.
- 1,959 responses to the truck diary survey, which includes trip information for the 1,769 trucks who traveled on the survey day.
- Data on 10,456 truck stops were collected from the 1,769 vehicles that completed the truck diary survey.
- GPS data from 500 vehicles.

The SEMCOG CVS project's final report describes the survey's sampling approach, administration, and the methodology applied to develop the expansion weights that are contained in the dataset. It then goes on to present results from the survey including summaries of results at the establishment level and at the truck level.

Commercial vehicle origin-destination data

CMAP provided a 1-week sample of passively collected GPS activity from 2018 for light and medium commercial vehicles. The data were procured by CMAP from INRIX and processed into a tour and trip database. The data were used to estimate, calibrate, or validate aspects of the CSVM, including characteristics such as the time of day of tours and trips, the number of stops on a tour, trip and tour lengths, and stop durations. The data includes roughly 693,000 trips and 297,000 tours. After processing, the data used to support calibration included roughly 101,000 trips, 77,000 stops, and 34,500 tours.

GPS Data Preparation

The script used to process the CMAP GPS data is *dev/CVGPS_01Skims_Prepare.R*. The script applies the following processing steps to the data:

Assigning Vehicle Base

Certain steps of the model consider the spatial relationship between a vehicle's base (e.g., its typical overnight parking location) and the stop locations that vehicle will travel to. To produce a calibration target for these model steps, a base subzone was inferred for each unique vehicle. To do this, tour origin subzones were tallied for each unique vehicle. Except in the case where the tally was one, the subzone with the highest tally was assigned as that vehicle's base subzone. For vehicles that were recorded as having only one tour, the origin of the tour was assigned as that vehicle's base. If a base subzone could not be inferred by either of the preceding methods, that vehicle was not assigned a base subzone and was excluded from the calibration sample.

Classifying Tour and Trip Characteristics

Trip and tour characteristics were summarized at the aggregate level. This included the following steps:

- Zone, county, and district were attached for the trip origin, destination, and vehicle base.
- Timestamps were formatted according to their eventual use as an arrival or departure time, or as a trip or tour duration and were converted from Coordinated Universal Time (UTC) to Central time (CT) where appropriate.

- Stop durations were set as the lead of “previous layover time”.
- Distances between trip origin and destination, and vehicle base and trip destination were attached using daily average network distance calculated from the CMAP time period skims.

Trips were classified by whether they started at the vehicle base or not and tours were classified as having one stop or multiple stops. The number of stops is a function of a tour’s final destination zone and the number of trips on that tour. If the final destination of a tour was that vehicle’s base subzone, the number of stops on that tour is the number of trips minus one. If the final destination on a tour is not that vehicle’s base zone, the number of stops on that tour is the number of trips.

Defining Completeness in the Data Sample
Finally, quality checks were performed to identify any trips and tours that could not be used in the analysis. Tours were excluded from summary tables for reasons such as:

- Missing stop or travel duration
- Missing departure times
- Missing stop distances
- Missing assigned vehicle base subzones
- Flagged as having unaccounted-for travel
- Having only one trip

Tours with unaccounted-for travel were defined as having mismatch between one trip’s destination zone and the next trip’s origin zone. Where a discontinuity existed, daily average values were attached to measure the unaccounted-for travel distance. Where this distance was greater than four miles, the tour was flagged as having a large discontinuity. Additionally, stop durations were sometimes marked as “NA” when they were likely an overnight stop. In general, this was done when trips lasted several hours and started in the evening or overnight.

Sample Versions
Once the sample was defined, two versions of the sample were used. One version retained return-to-base trips, and one included only stops as previously defined. Tabulations such as total tour distance and the travelling salesman solution consider the entire tour while individual stop and trip characteristics (distance from base to stop, stop duration, etc.) tend to consider only stops made away from the base during the tour and omit the final return-to-base trips.

Selected GPS Data Summaries

Once the GPS data had been filtered to only tours with complete data, summary tables were. The remainder of this section includes an overview of the results of this analysis and some selected tables that were later used during the calibration process. These summary tables are produced in the script *dev/CVGPS_02Skims_TablesForCalibration.R*.

Base-Stop Distances
Using a sample not containing return-to-base trips, direct base to trip destination distances were classified into distance bins, where the base zone is the most common tour starting zone for that unique vehicle and the distance are the daily average of network distances from skims. This allows for the

calibration of the model with respect to how far generated stops tend to be from their vehicle's base location. Figure 27 shows that 23.7% of trips ventured 5-10 miles away from their inferred base location. Roughly half of trips (51.6%) ventured 10 miles or fewer from their inferred base location. 48.4% of trips venture more than 10 miles from their base, with 28.3% of trips venturing more than 20 miles from their base.

Figure 27: BASE TO STOP DISTANCE

Bin	Trips	Proportion
0-2 Miles	5,983	7.8%
2-5 Miles	15,482	20.1%
5-10 Miles	18,246	23.7%
10-20 Miles	15,400	20.0%
20+ Miles	21,791	28.3%

Stop Durations

Using the sample not containing return-to-base trips, stop duration were classified into groups. This summary assists in the calibration of the stop duration model. 71.6% of all stops lasted 60 minutes or less, and 47.5% lasted 30 minutes or less. Frequencies decrease beyond 60 minutes.

Figure 28: STOP DURATION

Stop Duration (Minutes)	Trips	Proportion
0-15	14,513	19.8%
16-30	20,319	27.7%
31-45	11,129	15.2%
46-60	6,606	9.0%
61-75	4,677	6.4%
76-90	3,190	4.3%
91-150	6,720	9.2%
151-210	3,042	4.1%
211-270	1,704	2.3%
271-360	1,484	2.0%

Tour Destination Clustering

For each tour, given all the stop locations on a tour, a matrix of the distances between them was produced and the average distance value from that matrix was calculated. This distribution conveys how closely stops on a tour tend to be clustered. For this measure of the average distance between stops, 50.3% of tours were observed as having a value between 0 and 8 miles and 80.9% of tours had a value less than or equal to 20 miles.

Figure 29: TOUR STOP CLUSTERING

Miles	Tours	Proportion
0	344	1.7%
1	812	4.0%
2	1,409	7.0%
3	1,537	7.6%
4	1,443	7.1%
5	1,391	6.9%
6	1,310	6.5%
7	1,029	5.1%
8	902	4.5%
9	792	3.9%
10	706	3.5%
11	623	3.1%
12	565	2.8%
13	597	3.0%
14	544	2.7%
15	516	2.6%
16	440	2.2%
17	384	1.9%
18	366	1.8%
19	341	1.7%
20	314	1.6%

Travelling Salesman Analysis

Taking the tour start location and the stops on a tour, a matrix of the distances between them using the skims data provided by CMAP was produced. Then, the asymmetric travelling salesman algorithm used in the CSVM's tour sequencing model was applied to these matrices to determine the optimal routing order and the resulting tour distance for each tour. Comparing the actual distance travelled on a tour in the GPS data to the travelling salesman solution shows the difference in routing efficiency between actual routing and the travelling salesman solution. This analysis was done to understand whether the travelling salesman algorithm was an appropriate approach for modeling the stop sequencing or whether businesses and commercial vehicles are bound by schedules and other considerations that results in tours that are not well modeled using that algorithm.

In the analysis, 30% of tours were sequenced in the same way in both the GPS data and by the algorithm. While the remainder of the tours were sequencing differently, either resulting in shorter or longer tours, the overall ratio of vehicle miles traveled between the algorithm sequenced data (919,874 miles) and the GPS data (955,631 miles) was 96%, meaning that the results from the travelling salesman algorithm produced a reasonable approximation of the actual tour routing's optimization with respect to distance traveled.

Trip Departure Times and First Stop Arrival Time

Two tabulations related to time of day were developed. First, for all trips, departure hours were tallied to determine the distribution of trips start times in the CMAP GPS data. These results are utilized in the validation of the model with respect to trip departure times. In the GPS survey data, we observe that 56% of trips start between 7AM and Noon, with departures picking up rapidly between 5AM and 7AM, and little activity between the hours of 11PM and 4 AM (Figure 30, Trip Departure column). Commercial service trip departures taper off gradually after noon, reaching a minimum at midnight.

Similarly, the arrival time at the first stop of a tour was tallied. This target distribution is used to calibrate the tour arrival time model (Figure 30, Tour First Stop Arrival column). Comparing the distribution of all trip departures to the distribution of the arrival time at the first stop on tours, trip departures are more evenly distributed relative to trip start hours. On 42.4% of tours, arrival time at the first destination took place between 6AM and 10AM while only 34.7% of trips departed during this same period. Arrival times at the first stop on a tour are concentrated between the hours of 6AM and 9AM, and in general, are skewed towards the start of the day.

Figure 30: TRIP DEPARTURE TIME AND TOUR FIRST ARRIVAL TIME

Hour	Trip Departure	Tour First Stop Arrival
0	0.8%	0.8%
1	0.8%	0.7%
2	0.8%	0.7%
3	1.0%	1.0%
4	1.6%	1.6%
5	3.6%	3.9%
6	6.8%	8.8%
7	9.4%	12.7%
8	9.3%	12.0%
9	9.3%	9.1%
10	9.6%	7.9%
11	9.7%	7.7%
12	8.9%	7.2%
13	7.2%	5.6%
14	5.1%	4.2%
15	4.1%	3.8%
16	3.0%	3.0%
17	2.4%	2.3%
18	1.8%	1.7%
19	1.4%	1.4%
20	1.1%	1.1%
21	0.9%	1.0%
22	0.8%	0.8%
23	0.8%	0.9%

Transportation supply data
Travel time, distance, and toll skims were provided by CMAP from the CMAP travel demand model. The skims covered the base year (2019) and future year (2050) for both free-flow and congested conditions, and were segmented by vehicle class (light vehicles, medium trucks, and heavy trucks) and time period (eight time periods including AM and PM peak, pre and post peak shoulder periods and two off peak periods).

Average speed for the congested and free flow conditions were compared for the 2019 year. Free flow speed was roughly 45 miles per hour across all time periods and vehicles while average speeds in the congested data were lowest during P6 (2pm-4pm) at around 30-31 miles per hour and highest during P1 (8pm-6am) at around 40-42 miles per hour depending on vehicle class (see Figure 31). The congest skims were used as the inputs to the CSVM.

Figure 31: BASE YEAR SKIMS, AVERAGE SPEEDS

Vehicle	TOD	Speed (congested)	Speed (free flow)	Difference
Light	P1	41.9	45.5	-3.6
	P2	35.9	45.3	-9.3
	P3	31.0	45.3	-14.2
	P4	35.7	45.3	-9.6
	P5	33.2	45.3	-12.2
	P6	30.7	45.3	-14.6
	P7	32.5	45.3	-12.8
	P8	35.9	45.3	-9.4
Medium	P1	42.1	45.6	-3.5
	P2	36.6	45.4	-8.8
	P3	31.5	45.4	-13.9
	P4	36.4	45.4	-9.0
	P5	33.8	45.4	-11.6
	P6	31.2	45.4	-14.2
	P7	33.1	45.4	-12.3
	P8	36.7	45.4	-8.7
Heavy	P1	40.7	45.3	-4.6
	P2	35.4	45.2	-9.8
	P3	30.5	45.2	-14.7
	P4	35.0	45.2	-10.2
	P5	32.5	45.2	-12.7
	P6	30.1	45.2	-15.1
	P7	31.9	45.2	-13.3
	P8	35.2	45.2	-10.0

Figure 32 shows the future year congested skims. In comparison with the base year congested skims, generally, future year speeds were lower across all periods and vehicle class.

Figure 32: FUTURE YEAR CONGESTED SKIMS, AVERAGE SPEEDS

TOD	Light Vehicles	Medium Vehicles	Heavy Vehicles
P1	40.7	41.0	39.6
P2	34.3	34.9	33.7
P3	28.1	28.5	27.7
P4	33.5	34.0	32.8
P5	30.7	31.1	30.0
P6	27.9	28.2	27.3
P7	30.0	30.4	29.4
P8	33.7	34.3	33.0

Transportation network usage data

Two types of transportation network usage data were considered for use in validating the base year outputs from the CSVM: vehicle classification counts and estimates of VMT by functional class and vehicle type. As described in the base year validation section of the report, the VMT data were used for comparisons with the trip tables produced by the model, while the count data were not used in model validation but could be used in assignment validation on network assignments done with the CSVM trip tables combined with freight truck trip tables.

Vehicle Classification Counts

CMAP provided two sets of count data. One count file provided AADT for various vehicle classes for a thorough set of roads in the CMAP region. These data include information for heavy commercial vehicles (defined as six tire and larger including buses) and for multiple unit vehicles (defined as tractor-semitrailer combinations, large truck and trailer combinations, and two-trailer combinations). Single unit vehicle (defined as 2-axle 6-tire single frame, 3- and 4-axle single frame trucks, and buses) volumes can be calculated by subtracting the multiple unit volumes from the heavy commercial volumes. These data are largely gathered from 2015 and adjusted to 2019 levels. Figure 33 and Figure 34 illustrate the multiple unit truck and single unit truck AADTs, respectively, from this data set. A second set of data provided count data with temporal distribution from 2015, but this data does not distinguish between vehicle classes.

Figure 33: MULTIPLE UNIT TRUCK AADT

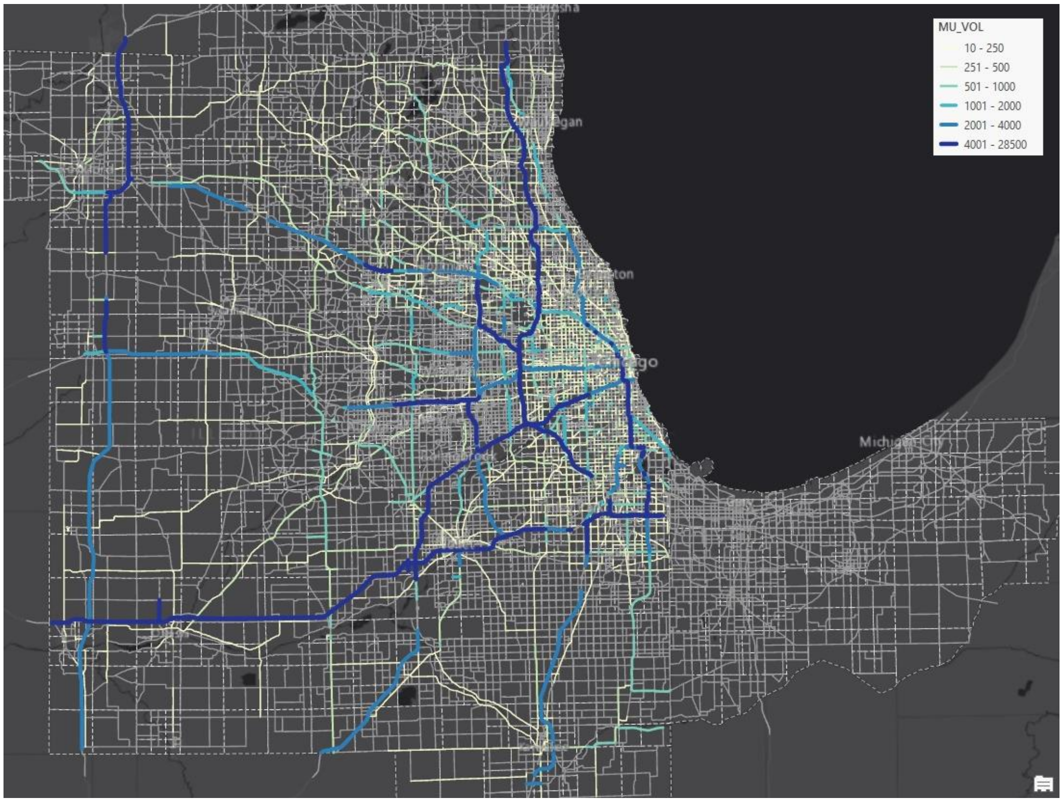
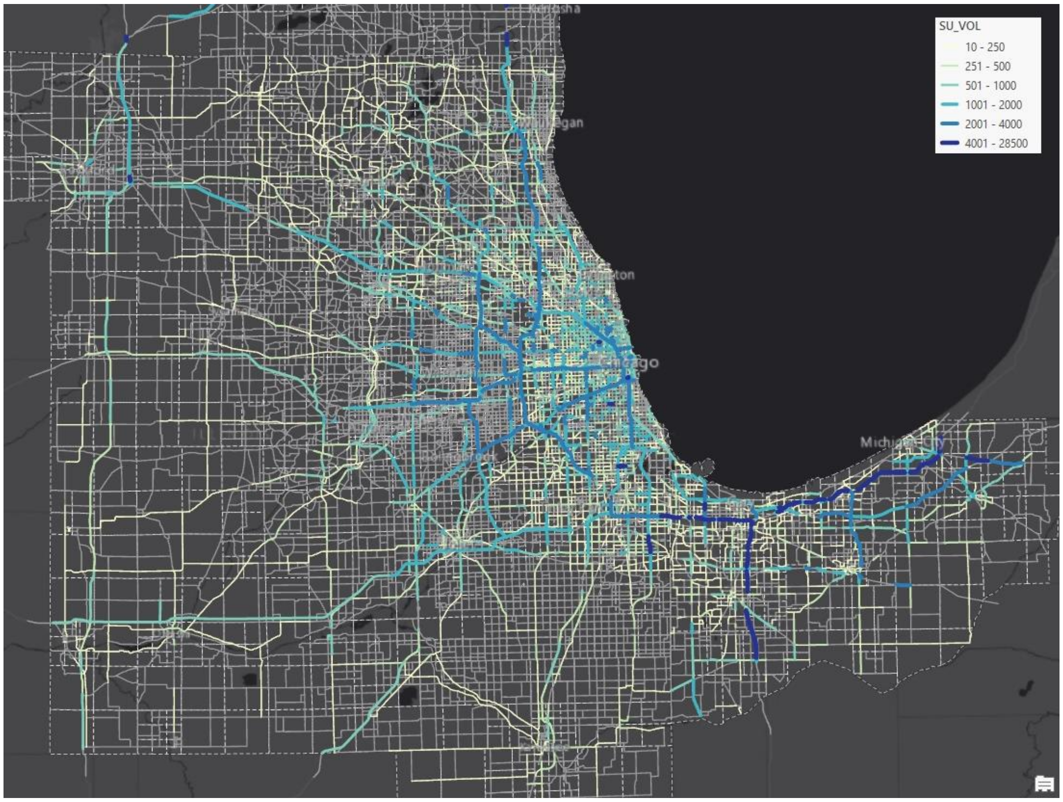


Figure 34: SINGLE UNIT TRUCK AADT



Vehicle miles traveled data

VMT data for urban areas in Illinois and for the Chicago IL-IN urbanized area for 2019 were downloaded from the FHWA's Highway Statistics website[4]. The data were processed into estimates of commercial vehicle VMT by vehicle class using the script *dev/calibration_targets_tt.R*. The approach to developing the VMT estimates includes the following steps:

- The expanded SEMCOG CVS data were tabulated to estimate the proportion of commercial VMT by vehicle class that is from internal-to-internal trips within the region and made by vehicles owned by the employment groups covered by the CSVM (i.e., excluding transportation, manufacturing, etc.). The proportions calculated were 69% of light commercial vehicle VMT, 55% of medium truck VMT, and 10% of heavy truck VMT.
- The expanded SEMCOG CVS data have been used to calibrate the SEMCOG Commercial Vehicle Model. For the SEMCOG region, the portion of <u>light vehicle</u> VMT that is attributable to light commercial vehicles (as opposed to passenger vehicles) is 9.5% and the portion of <u>total</u> VMT that is attributable to light commercial vehicles is 8.9%.
- An AADT to average weekday factor of 1.05 was calculated by comparing the 2017 FHWA VMT data for Detroit, Michigan with the VMT produced by the SEMCOG model's 2017 base year which was validated against 2017 average weekday traffic counts.
- The FHWA's Illinois state tabulations for urban areas for annual vehicle miles by functional class and distribution of annual vehicle distance traveled by vehicle type were used to allocate the Chicago IL-IN urbanized area daily VMT by functional class to vehicle types.
- The VMT by vehicle types estimates for the Chicago IL-IN urbanized area were factored up to account for the parts of the model region that are outside of the Chicago IL-IN urbanized area. The model region VMT was estimated to be 122% of the Chicago IL-IN urbanized area VMT.
- The factors estimated from the SEMCOG CVS data and model were then used to convert the VMT to average weekday VMT, to factor the light vehicle VMT to estimate light commercial VMT, and then finally to estimate the part of VMT that is covered by the CSVM, i.e., internal to internal trips made by vehicles operated by the covered employment groups.

Figure 35 shows a summary of the estimated VMT values for all commercial vehicles and for the commercial service vehicle VMT that is covered by the CSVM. Overall, approximately half of commercial vehicle VMT should be covered by the CSVM, with the majority being light commercial vehicles.

Figure 35: CMAP MODEL REGION ESTIMATED VMT FOR COMMERCIAL SERVICE VEHICLES

Commercial Vehicle Class	Total VMT	CSVM VMT	Percent CSVM
Light Commercial Vehicle	20,930,588	14,348,729	69%
Medium Truck	7,278,432	4,023,632	55%
Heavy Truck	12,961,598	1,313,915	10%
All Commercial Vehicles	**41,170,618**	**19,686,276**	**48%**

[4] https://www.fhwa.dot.gov/policyinformation/statistics/2019/

4. CSVM MODEL DEVELOPMENT

This section of the report describes the design of the CSVM, presents the results of statistical model estimation and calibration to target data of each model component, and then discusses the validation of the base year model against observed data. Finally, the results from a future year scenario and scenario tests are documented.

The CSVM was developed using the R programming language, an open-source language commonly used for statistical and data analysis. The functions and scripts that support the CMAP freight model are contained in a R package (function library) called "rFreight." The rFreight package is also the basis of the SEMCOG commercial vehicle model (SEMCOG CVM). RSG transferred the SEMCOG model's commercial vehicle touring model (CVTM) component as the starting point for the development of the CVTM component of the CSVM.

The transferred CVTM component was updated in this project to work with the spatial, land use, and travel model data for the CMAP region. RSG used the CVS data, augmented by the GPS data, local land use data, and travel time skims, to update and calibrate the parameters of the model components.

CMAP's existing firm synthesis model was updated to reflect the TAZ-level establishment locations needed for the CVTM and was also simplified to work just for the model region (as opposed to the entire US for the version in the freight model) and to omit the portions of the firm synthesis model that estimate commodity production and consumption for each business establishment as this information is not used in the CVTM.

4.1 Model Design

The CSVM contains two main model components followed by two additional steps of the model that process outputs:

1. The **firm synthesis model** that synthesizes a list of business establishments in the region. In this case, the existing CMAP firm synthesis model was adapted to add TAZ level controls (compared to the earlier mesozone level controls) and to add firm growth and decline to support the development of future year sets of business establishments.
2. The **CVTM** generates demand in each TAZ for commercial vehicle tour starts and commercial vehicle stops, connects those locations into tours, and produces a detailed commercial vehicle trip roster of all trip and tours by light commercial vehicles, and medium and heavy trucks, making internal to internal trips within the CMAP region and that are operated by the employment groups covered by the CSVM.
3. The **trip table** step processes the commercial vehicle trip roster into a set of vehicle trip matrices by time of day and vehicle class in the format required by CMAP's travel model assignment step. The output matrices are saved in an open matric (OMX) format file that can be read by Inro's EMME transportation planning software.
4. The **dashboard and summary spreadsheet** step which processes the model outputs to produce a stand-alone, interactive dashboard that is viewable in a web browser, and a spreadsheet containing larger tabulations of model outputs.

Firm Synthesis

The existing CMAP firm synthesis model (a component of the CMAP freight model) was adapted from its current mesozone structure to reflect the more detailed zone structure used in the passenger model. The firm synthesis model was further adapted for use in the CSVM by simplifying it to just cover the CMAP model region and to stop after employment scaling (where the version in the freight model continues to simulate each business establishment's commodity production and consumption).

The version of the firm synthesis model included in the CSVM has the following steps.

1. **Creating Firm Records** -- The firm synthesis model begins by enumerating business establishments using pre-processed aggregate establishment data from the CBP data, creating a list of firms.
2. **Placing Businesses within Zones** -- Until this point, the geographic identifier for all business establishments is county. The next step locates business establishment within the higher-resolution geographical units, first mesozones and then TAZs.
3. **Employment Scaling** – In the base year, the model's socioeconomic data that quantifies the employment by industry in each TAZ is used to scale the employment in the business establishments in each TAZ so it matches the model's socioeconomic data. The scaling involves adding or removing business establishments as well as growing and shrinking business establishments. In a future year run, the base year establishments are updated to match the future year employment data. This means that the future year establishment list is reasonably consistent with the base year list where there are only small changes in a particular TAZ.

At the end of firm synthesis, the model saves a database of business establishments.

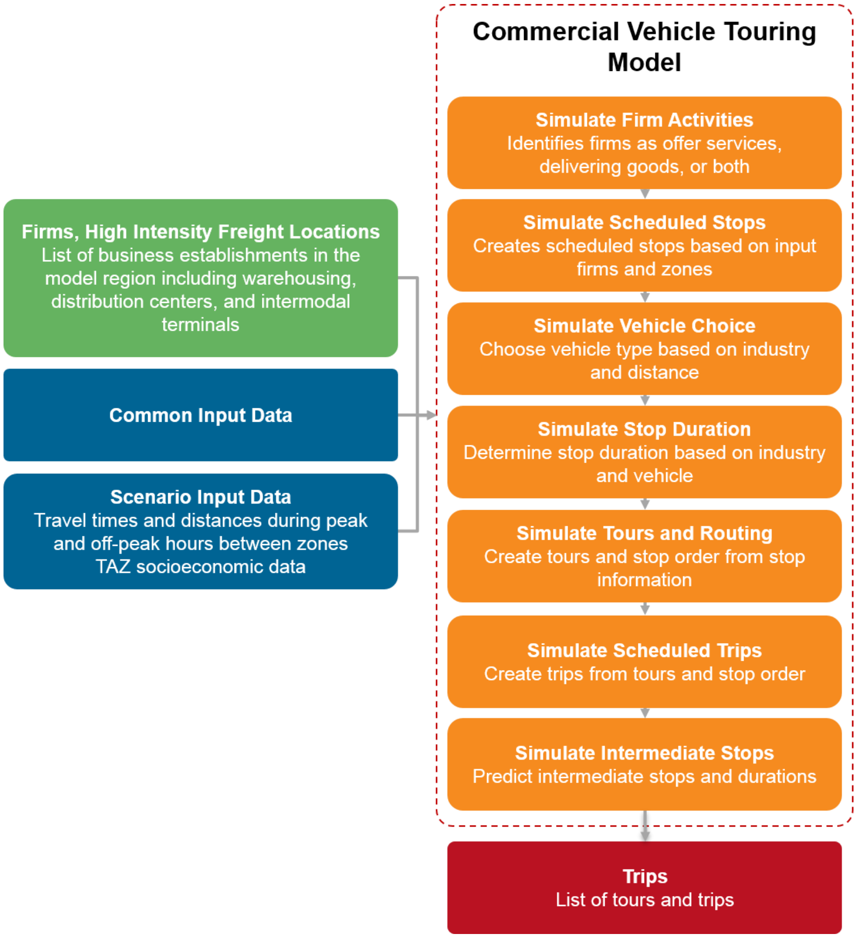
The firm synthesis model deals with some specific data issues that are apparent in the input data:

- The CBP data are censored to zero out the establishment size fields where there are fewer than three establishments. The size records are filled in based on the overall size distribution for each NAICS category for cases where the number of establishments is more than the total across the establishment size fields.
- The CBP data are missing both employment on farms and in some other agricultural businesses (NAICS category 11), and in public administration (NAICS category 92). As noted above in the data section of the report, the agricultural business records are added from US Census of Agriculture data. Public administration business establishments are synthesized based on the TAZ socio-economic data which describes the quantity of public administration employment as well as employment in other industries. Additional business establishments are added to each TAZ at a rate proportional to the share of employment in the TAZ that is public administration.

Commercial Vehicle Touring Model

The CVTM was transferred from the recently implemented SEMCOG CVM. The implementation in the SEMCOG region benefitted from a large commercial vehicle survey and provided a good start point for updated and calibration for the CMAP region. Figure 36 shows the CVTM's inputs, steps, and outputs. Each step in the model is briefly described here (with more details provided in the Base Year Model Estimation and Calibration section).

Figure 36: COMMERCIAL VEHICLE TOURING MODEL INPUTS, STEPS, AND OUTPUTS



Simulate Firm Activities

The firm activities model tags each synthetic establishment from the firm synthesis model with an industry type and a label that indicates whether the establishment's commercial vehicles make goods deliveries (or pickups), provide service calls, do both, or either the business does not operate commercial vehicles or on a typical day they do not make goods or service trips. The model applies a Monte Carlo simulation method to draw from observed distributions of establishments by industry type and stop type, constructed using truck diary data. Businesses in the transport and industry employment group are all allocated to the no trips category in the CSVM, so they will not generate any commercial vehicle trips (with the assumption that any commercial vehicle trips for these businesses are covered by the freight truck touring model portion of the CMAP freight model).

Simulate Scheduled Stops

The internal stop generation model predicts one day's worth of scheduled stops for each establishment by TAZ, covering all internal TAZs, using a hurdle-count model formulation (i.e., a binary model to predict whether a TAZ has a stop or not, and a count model predicted the number of stops given that a stop takes place). Scheduled stops from this model are grouped into two-market segments, good stops and service stops, and also the model also indicates whether a stop is at a residence or non-residential land use..

Simulate Vehicle Choice
For each stop, the vehicle choice model assigns one of three commercial vehicle types. These correspond to light (i.e., car, van, and pickup), medium/single-unit, and heavy/multi-unit truck types. The model is a multinomial logit model and predicts vehicle type as a function of the establishment's industry type, distance between establishment and stops to be served, and the stop's purpose—to deliver goods or provide services.

Simulate Stop Duration
The stop duration model is applied to scheduled stops generated by the stop generation model as well as any intermediate stops on the tour (the intermediate stop generation is discussed below and the stop duration is applied to those stops at the stage in the model). For each stop, an expected stop duration range is estimated using a multinomial logit model. Then an actual value in minutes is drawn from the range. The duration model is a function of vehicle type and stop purpose.

Simulate Tours and Routing
For each establishment, the stop clustering model groups scheduled goods and service stops into feasible commercial vehicle tours, based on vehicle type, spatial proximity of the stops, and expected stop duration. This is followed by a tour type model, which is a multinomial logit model that identifies whether each cluster of stops is part of a tour that starts and ends at the same location, or one that starts, and ends in different locations. Finally, a tour sequencing model is applied that uses a traveling salesman algorithm to order the stops within each tour into a route.

Simulate Scheduled Trips
The scheduled trips component identifies the timing for each of the tours created in the previous model component. For each tour, the arrival time at the first stop on the tour is predicted using a multinomial logit model.

Intermediate Stops
The intermediate stop choice model, a multinomial logit model, predicts whether there are intermediate stops between scheduled stops on each tour. The model simulates whether the driver makes one or more intermediate stops prior to each scheduled goods or service stop, or prior to returning to the establishment to complete the tour. Purposes for intermediate stops are breaks/meals, vehicle service/refueling, and personal business/other. The intermediate stop destination model predicts a destination TAZ for each intermediate stop. Specifically, for each intermediate stop, a multinomial logit model is used to select from a set of eligible TAZs (those which do not require excessive deviation from the tour route) based on attraction factor(s), such as food and drink service employment for break/meal stops.

4.2 Future Year Forecasts

To produce future year commercial vehicle forecast using the CSVM, the model requires future land use and network skims.

For future year land use, a forecast of TAZ level employment and households are required, but the model does not need a future year establishment dataset. This forecast of TAZ employment allows the simulation of a future year set of business establishments by the firm synthesis model, which updates the base year business establishments to match the future year employment by adding, removing, and growing or shrinking business establishments. The future year employment and households are also used to forecast future commercial vehicle stops in each TAZ.

The future year congested travel time and distance skims are applied in the steps of the CVTM that include time and distance variables.

For the standard future scenario (the results of which are described later in this report) no adjustments are made to model parameters to assert changes in commercial vehicle behavior over time. Several model parameters can be adjusted in the scenario configuration file to allow for scenario testing of assumptions about changes in commercial vehicle behavior.

4.3 Software Architecture

CSVM File Structure

The general software architecture of the CSVM is consistent with CMAP's freight model, and each of the CSVM components conforms to this structure. The R code components are in two sections:

- The CSVM includes a function library in the form of an R package called rFreight that contains functions used in the model.
- A set of scripts call functions from the rFreight library to implement each of the model components. A master controller file, __Master.R acts as the main interface for the models, initiating the model run and sequentially running the script files that apply each of the individual model components. The scripts are consistently structured to apply a single model component (such as firm synthesis): they load inputs, carry out any required data processing, apply the model, and save outputs, before passing control back to the controller file. In this way, the model code is modularized.

The CSVM's R code includes functions to load model inputs such as tabular inputs describing the number of business establishments or travel time skims, from the file system. The CSVM's R code also includes functions to produce summary tabulations of results and to build an interactive dashboard to display results in tables, charts, and maps. The CSVM model relies on the functionality of the EMME portions of the CMAP travel demand model to skim transportation networks and to assign commercial vehicle trip tables onto the highway network.

The CSVM's application code, model parameters, and scenario inputs and outputs are included entirely within a root model directory, *cmap_csvm*. Inside the *cmap_csvm* directory, the model has the file structure shown in Figure 37. The model application contains fours directories, only two of which are required to actually run the model, and several files that are used to initiate a model run:

- The *lib* folder which contains the model's scripts, parameters, inputs that do not vary across scenarios, and the R function packages used to run the model's scripts.
- The *scenarios* folder which contains scenarios and their inputs and outputs.
- The *dev* folder which contains model development scripts and other files. This directory is useful for making updates to model inputs but is not required for model application runs.
- The *docs* folders which provides integration with GitHub's IO pages. Web viewable files (e.g., HTML format files) placed in this folder and pushed to the CSVM's repository will be viewable via the GitHub IO site for the repository.

The model scripts, data inputs (both common and scenario specific) and the model outputs are described below.

Figure 37: CSVM FILE STRUCTURE

```
Cmap_csvm_vX.X.X
* dev                         Model development directory
* docs                        GitHub IO folder for web documents
* lib                         Model library directory
  + data                      Common input data directory
  + pkgs                      R application and R packages directory
    - library                 R packages directory
    - Pandoc                  Application directory for pandoc
  + scripts                   R scripts directory
    - db_markdown             Dashboard script directory
  + rFreight_0.1-34.zip       rFreight package
Scenarios                     Scenarios directory
  + base                      Base scenario directory
    - inputs                  Scenario inputs directory
    - outputs                 Scenario outputs directory
  + future                    Future scenario directory
    - inputs                  Scenario inputs directory
    - outputs                 Scenario outputs directory
* README.md                   Readme file, installation and run instructions
* run_cmap_csvm.bat           Batch file called to run the CSVM
* run_cmap_csvm.r             R script that starts the CSVM
```

CSVM Scripts

The R scripts that implement the CSVM model are included in the *lib/scripts* folder of the application (with the one exception being *run_cmap_csvm.r* which is in the root directory of the model). The scripts are listed and described in

Figure 38. There are several types of script which are named using the following naming convention:

- **MASTER:** A master controller file, __*Master.R* acts as the main interface for the models, initiating the model run and sequentially running the script files that apply each of the individual model components.
- **INITIALIZATION:** Scripts names prefixed with “init”, e.g., *init_start_rfreight_model.R*, contain code to start the model application and support functions such as loading the required R package function libraries.

- **VARIABLES:** There are several scripts that contain global variables used through the CSVM. The script naming convention is upper case and starting with an underscore, e.g., *_BASE_VARIABLES.R*.
- **CONTROL**: Each model component is run by a control script that either runs a model simulation or builds an output. In the case of a model simulation, the script is named after the model component suffixed with "_sim", e.g., *cv_sim.R* for the CVTM control script. For the model steps that build an output, the script is named after the model component suffixed with "_build", e.g., *tt_build.R* for the trip tables control script.
- **INPUT PROCESSING**: Each model component contains a script that loads the input files used by the model component into an environment in memory and does some initial processing. This script is run as the first step of each model component. These scripts have the same name as the control script for the model component with the additional suffix of "_process_inputs", e.g., *cv_sim_process_inputs.R* for the CVTM's input processing function.
- **MODEL STEP**: The control script for each model component can call additional scripts that contain functions implementing individual model steps. These scripts have the same name as the control script for the model component with an additional suffix describing the model step, e.g., cv_sim_activities.R which contains the function to implement the CVTM's activities model.

Figure 38: LIST OF MODEL SCRIPTS

Step	Step Name	Filename	Description	Script Type
0	Initialization	_BASE_VARIABLES.R	List of base variables	Variables
0	Initialization	_SCENARIO_VARIABLES.R	List of scenario variables	Variables
0	Initialization	_SYSTEM_VARIABLES.R	List of system variables	Variables
0	Initialization	_USER_VARIABLES.R	List of user variables	Variables
0	Initialization	__Master.R	Master script to control CSVM flow	Master
0	Initialization	init_install_special_packages.R	Initialization install R packages	Initialization
0	Initialization	init_start_rfreight_model.R	Initialization start model application	Initialization
1	Firm Syn.	firm_sim.R	Firm synthesis control function	Control
1	Firm Syn.	firm_sim_process_inputs.R	Firm synthesis input processing function	Input Processing
1	Firm Syn.	firm_sim_enumerate.R	Firm synthesis enumerates the CBP	Model Step
2	CV Touring	cv_sim.R	CVTM control function	Control
2	CV Touring	cv_sim_activities.R	CVTM activities model	Model Step
2	CV Touring	cv_sim_intermediatestops.R	CVTM intermediate stops model	Model Step
2	CV Touring	cv_sim_process_inputs.R	CVTM input processing function	Input Processing

Step	Step Name	Filename	Description	Script Type
2	CV Touring	cv_sim_scheduledstops.R	CVTM stop generation model	Model Step
2	CV Touring	cv_sim_scheduledtrips.R	CVTM trip scheduling model	Model Step
2	CV Touring	cv_sim_stopduration.R	CVTM stop duration model	Model Step
2	CV Touring	cv_sim_tours.R	CVTM tour model (clustering and sequencing)	Model Step
2	CV Touring	cv_sim_vehicle.R	CVTM vehicle choice model	Model Step
3	Trip Tables	tt_build.R	Trip table building function	Control and Model Step
3	Trip Tables	tt_build_process_inputs.R	Trip table input processing function	Input Processing
4	Dashboard	db_build.R	Dashboard control function	Control
4	Dashboard	db_build_graphics_functions.R	Dashboard helper functions	Support Functions
4	Dashboard	db_build_process_inputs.R	Dashboard input processing function	Input Processing
4	Dashboard	db_build_render.R	Dashboard rendering and post-processing	Model Step
4	Dashboard	db_build_spreadsheet.R	Spreadsheet summary of model outputs	Model Step
4	Dashboard	db_build_spreadsheet_functions.R	Spreadsheet summary helper functions	Support Functions
4	Dashboard	db_markdown	Contains dashboard Rmarkdown template and supporting files	Directory

CSVM Data Inputs

The tabular and parameter inputs to the CSVM are in two locations:

- Inputs that are common across all model scenarios are in the *lib/data* directory.
- Inputs for a particular scenario are in the *scenarios/[scenario name]/inputs* directory.

The first group, the common inputs, are listed in Figure 39, while the scenario specific inputs are listed in the Figure 40. The files are mainly comma separated variables (.csv) format files or encoded into one of R's binary formats for either single objects (.rds) or a workspace capable of holding multiple objects (.Rdata), but there are also ESRI shapefiles (.shp) files of TAZ layers and open matrix format (.omx) matrices.

Figure 39: LIST OF INPUTS THAT ARE COMMON ACROSS SCENARIOS

Step	Step Name	Filename	Description	Input Type
1	Firm Syn.	corresp_naics2_empcats.csv	Correspondence between NAICS and employment categories	Categories/ correspondence
1	Firm Syn.	data_emp_cbp.csv	List of establishments	Tabular data
1	Firm Syn.	data_emp_cbp_ag.csv	List of agricultural establishments	Tabular data
1	Firm Syn.	data_est_size_categories.csv	Establishment size categories	Categories/ correspondence
1	Firm Syn.	data_mesozone_emprankings.csv	Employment rankings data for mesozone allocation	Tabular data
1	Firm Syn.	TAZ_System.csv	tabular form of TAZ shapefile attribute data	Tabular data
2	CV Touring	cv_activities_model.RDS	Activities model coefficients	Model parameters
2	CV Touring	cv_arrival_model.RDS	Arrival model coefficients	Model parameters
2	CV Touring	cv_goods_res_model.RDS	Residential goods stops model coeffs.	Model parameters
2	CV Touring	cv_goods_non_res_model.RDS	Non-res goods stops model coefficients	Model parameters
2	CV Touring	cv_intermediate_deviations.rds	Intermediate stop deviation thresholds	Model parameters
2	CV Touring	cv_intermediate_model.rds	Intermediate stop generation model	Model parameters
2	CV Touring	cv_intermediate_model_attraction.rds	Intermediate stop location model	Model parameters
2	CV Touring	cv_service_res_model.RDS	Residential service stops model coeffs.	Model parameters
2	CV Touring	cv_service_non_res_model.RDS	Non-res service stops model coeffs.	Model parameters
2	CV Touring	cv_settings.RData	CVTM general parameters	Model parameters
2	CV Touring	cv_stopduration_model.RDS	Stop duration model coefficients	Model parameters
2	CV Touring	cv_tours_model.rds	Tour type model coefficients	Model parameters

Step	Step Name	Filename	Description	Input Type
2	CV Touring	cv_vehicle_model.RDS	Vehicle choice model coefficients	Model parameters
3	Trip Tables	data_skim_names.csv	Skim naming definitions	Tabular data
4	Dashboard	data_equity_zones.csv	Equity zone proportions	Tabular data
4	Dashboard	TAZ_System_Shape.dbf	Shapefile component for TAZ shapefile	Spatial data
4	Dashboard	TAZ_System_Shape.prj	Shapefile component for TAZ shapefile	Spatial data
4	Dashboard	TAZ_System_Shape.shp	Shapefile component for TAZ shapefile	Spatial data
4	Dashboard	TAZ_System_Shape.shx	Shapefile component for TAZ shapefile	Spatial data

Figure 40: LIST OF INPUTS THAT ARE SCENARIO SPECIFIC

Step	Step Name	Filename	Description	Input Type
1	Firm Syn.	data_emp_control_taz.csv	TAZ employment control data	Tabular data
2	CV Touring	data_hh.csv	TAZ household data	Tabular data
2	CV Touring	htruck_congested_skims.omx	Heavy truck congested skim matrices	Matrix data
2	CV Touring	ltruck_congested_skims.omx	Light truck congested skim matrices	Matrix data
2	CV Touring	mtruck_congested_skims.omx	Medium truck congested skim matrices	Matrix data
2	CV Touring	scenario_adjustments.R	Scenario parameter adjustments	Configuration script

CSVM Scenario Outputs

The main tabular outputs from the CSVM for a particular scenario are included in the *scenarios/[scenario name]/outputs* directory for the scenario. The model produces R binary format files (.Rdata) that contain compressed tables from each model step. The CSVM model adds trip tables to an .omx file for the scenario. These outputs are described in Figure 41. In addition to the main tabular and matrix outputs, the scenario outputs include a visualization dashboard (.HTML file), a summary spreadsheet (.xlsx file) and many summary tabulations (.csv files) and image files (.png files) which are produced during the dashboard step of the model.

Figure 41: LIST OF MODEL OUTPUTS

Step	Step Name	Filename	Description	Output Type
1	Firm Syn.	1.Firms.RData	Database of inputs and outputs from Firm Synthesis model	R workspace
2	CV Touring	2.CommercialVehicleTrips.RData	Database of inputs and outputs from CVTM	R workspace
2	CV Touring	skims_tod.rds	Processed time of day and daily average skim data	Tabular data
2	CV Touring	skims_tod_vehicle.rds	Processed time of day and daily average skim data by vehicle type	Tabular data
3	Trip Tables	3.TripTables.RData	Database of inputs and outputs from Trip Table step	R workspace
3	Trip Tables	CV_Trip_Tables.omx	CSVM Trip tables	Matrix data
4	Dashboard	4.DashboardTables.RData	Database of inputs to the dashboard step	R workspace
4	Dashboard	ReportDashboard.html	Dashboard showing scenario results and comparisons	HTML dashboard
4	Dashboard	Validation_base_[time_stamp].xlsx	Spreadsheet summary from base scenario comparing to validation data	Spreadsheet summary
4	Dashboard	Comparison_base_future_[time_stamp].xlsx	Spreadsheet summary from future scenario comparing to base scenario	Spreadsheet summary
4	Dashboard	*.csv and *.png	Other outputs from dashboard step	Summary tables and chart images

The CSVM dashboard and summary spreadsheet step processes the model outputs to produce a stand-alone, interactive dashboard that is viewable in a web browser, and a spreadsheet containing larger tabulations of model outputs. The dashboard is comprehensive and provides both detailed and summary-level statistics on the commercial vehicle travel patterns and impacts to the transportation system.

The dashboard presents the results of a CSVM scenario, which represents a simulation of commercial vehicle trips within the 21 county CMAP model region for the commercial vehicles operated by businesses in the employment groups covered by the CSVM. By default, the base year scenario results are compared to observed data (in a set of calibration results tables and charts for each model component). Also, by default, alternative and future year scenarios are compared to the base year model results to understand growth and the impacts of transportation policies and investments in the future. The dashboard can be configured during a model run to use other scenarios as the reference scenario instead of the base scenario.

The dashboard includes an overview of the modeling results, along with modeling results for both of the main components of the model and then summaries of the trip tables. It includes the following pages:

- Overview
- Firm Synthesis Model
  - Firms
  - Employment
  - Calibration (base year)
  - Comparison (Future year)
- Commercial Vehicle Touring Model
  - Stops and Geography
  - Stops and Timing
  - Tours
  - Trips
  - Calibration Stops (base year)
  - Calibration Tours (base year)
  - Comparison Stops (Future year)
  - Comparison Tours (Future year)
- Trip Tables
  - Daily Trips
  - Validation (base year)
  - Comparison (Future year)

The spreadsheets produced by the dashboard and summary spreadsheet step are naming based on the scenarios compared and the time of the model run. When the reference scenario is set to validation, the summaries are a comparison between the current scenario and the validation data and the spreadsheet is named *Validation_base_[time_stamp].xlsx*. When the reference scenario is set to another scenario, the summaries are a comparison between the current scenario and the reference scenario and the spreadsheet is named *Comparison_base_future_[time_stamp].xlsx*.

4.4 Base Year Model Estimation and Calibration

This section of the report describes updates to model estimation and then calibration of the base year (2019) CSVM. The general process followed to update the transferred model was to review each estimated model or empirical distribution and identify updates that could be made either with locally sourced data or by modifying the original estimation using transferred data (such as the SEMCOG CVS) so that categorical variables matched with those used in the CSVM. The main example of this updating employment category specific variables so that they matched with the CMAP employment groups.

The estimation scripts, inputs and outputs are included in the model structure at *dev/Estimation.* The script *dev/estimation_control.R* is used to copy the final set of estimated models from the development area of the model to the application's common inputs folder.

Once the estimated model and distributions had been updated, RSG used a sequential approach to calibrate the model components. This process involved wrapping the CSVM in a second layer of R scripts that, after each iteration of the model, evaluated the performance of the model against targets, and adjusted specified model constants until the targets were met within an acceptable tolerance. The calibration scripts are included in the model structure at *dev/calibration_*.R*

- *calibration_targets_firm_sim.R:* develops targets for the firm synthesis model calibration and validation
- *calibration_targets_cv_sim.R:* develops targets for the CVTM calibration
- *calibration_targets_tt.R:* develops targets for the validation of the trip tables
- *calibration_control.R:* control script for the calibration process. This script contains settings for the calibration of each model in the CSVM structure and a control process to loop through the model system, calibrating each model in turn and saving calibrated parameters and summaries.
- *calibration_functions.R:* contains functions used by the control script to calibrate each model in turn by summarizing model results, comparing with targets, and making parameter adjustments.

In addition to the figures included in this section, the base year dashboard provides a comprehensive set of calibration summaries for each step of the model.

Firm Synthesis

The firm synthesis model does not contain estimated models and does not require calibration. The only validation checks on the firm synthesis model are confirmatory, ensuring that the scaling algorithm does produce a set of establishments that match with the TAZ control data in terms of employment by industry by location. Figure 42 confirms that the modeled employment matches the control data by district, while Figure 43 confirms that the employment by NAICS 2 digit industry is matched.

Figure 42: BASE YEAR EMPLOYMENT BY DISTRICT

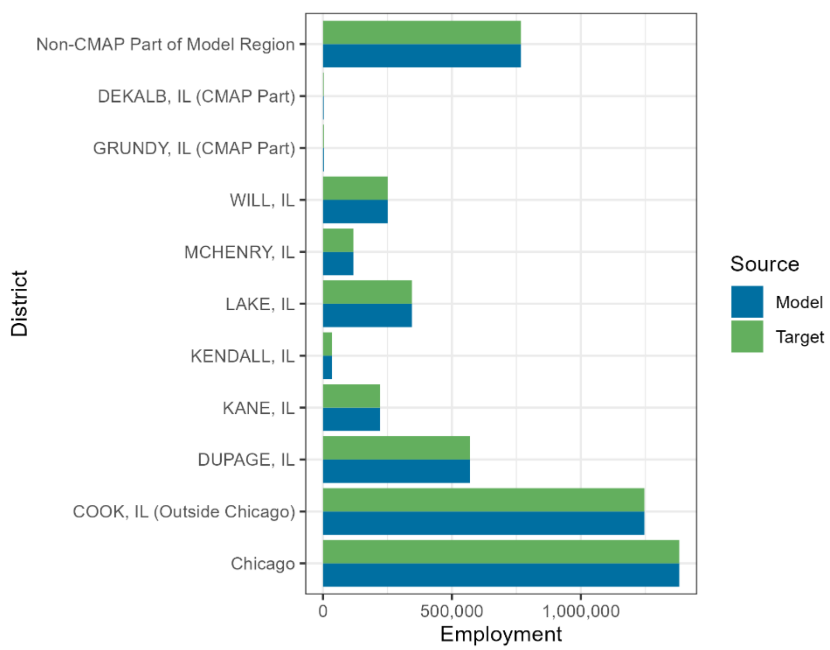
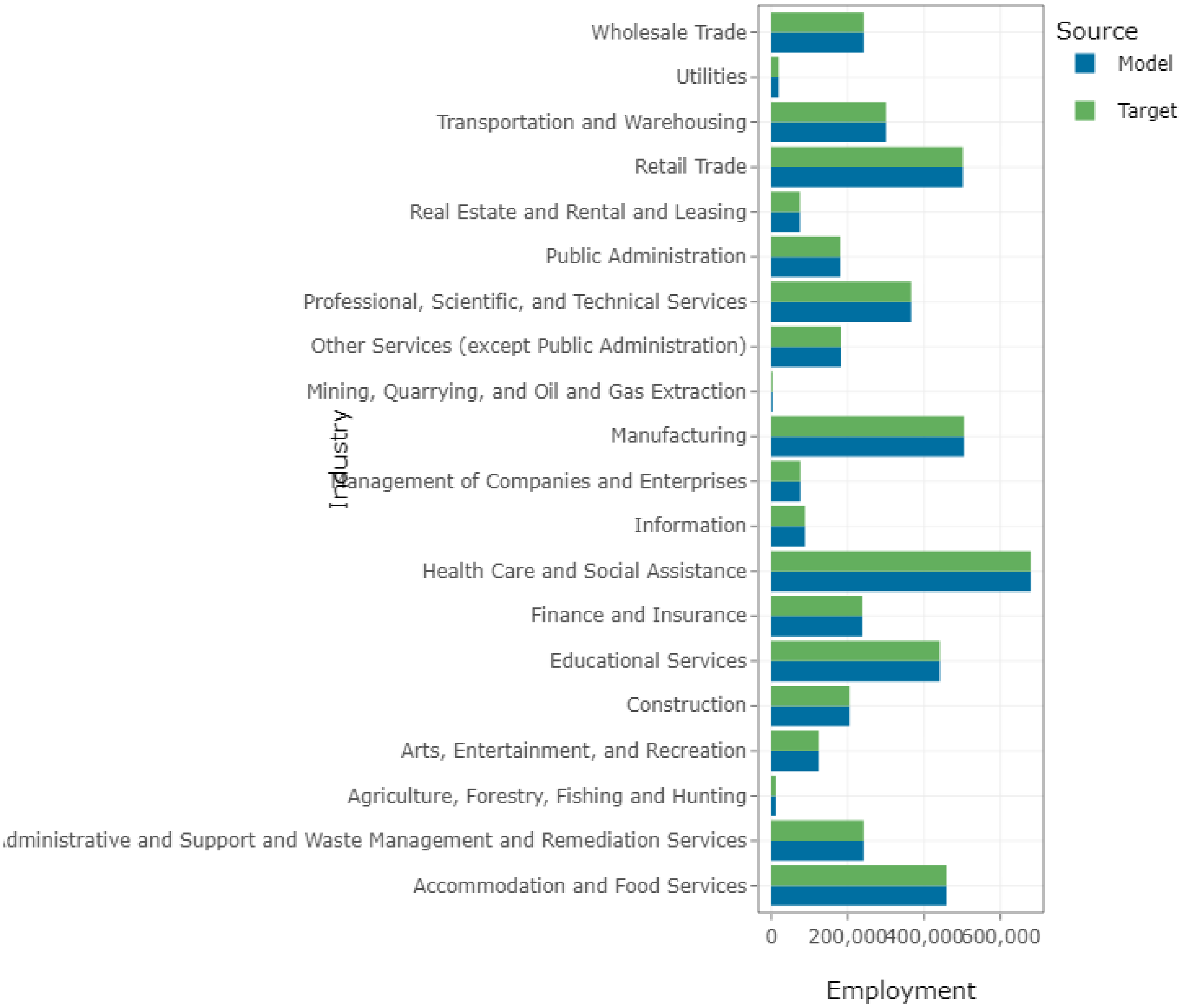


Figure 43: BASE YEAR EMPLOYMENT BY INDUSTRY



Commercial Vehicle Touring Model

As noted above, the model estimation phase included the development of an updated activities model distribution from the establishment survey, and re-estimation of the transferred hurdle and logit models. This was followed by calibration to various observed data tabulations. The calibration for the CVTM focused first on the stops models (scheduled stops, vehicle choice, and stop duration models) and then tour formation, tour timing and adding intermediate stops to tours. Further review of stops, trips, and tours is discussed in the section on base year validation.

Simulate Firm Activities

The CMAP establishment data collected in this project were used to develop the activity model, a crosstabulation of activity type (goods, services, or good & services) by industry group (Figure 44). NAICS

2 industry codes covered by the establishment survey were grouped into eight employment groups so that a reasonable sample size could be attained. A ninth employment group "transport_industry" accounts for the non-covered employment categories that are not modeled by the CSVM. The employment groups are used throughout the model system. The data were processed in the script *dev/Estimation/cv_activities/Estimation_CV_Activities.R*.

Figure 44: ACTIVITIES MODEL

EmpCatGroupedName	Goods	GoodsAndService	Other	Service
Admin_Support_Waste	6.6%	50.9%	0.0%	42.5%
Construction	9.3%	74.5%	4.6%	11.6%
Ed_Health_Social_Public	0.0%	35.7%	28.6%	35.8%
Office_Professional	11.9%	33.3%	0.0%	54.8%
Retail	35.4%	43.7%	13.7%	7.2%
Service_FoodDrink	44.4%	37.1%	18.5%	0.0%
Service_Other	13.0%	52.3%	13.0%	21.8%
Wholesale	39.5%	35.5%	7.2%	17.8%

Figure 45 through Figure 47 show the percentage of firms in each industry making stops for goods delivery, providing services, or both, comparing the model input tabulation (labeled target in the figures) with the model outputs. Overall, assigned activities for generated firms closely match what has been observed in the control data. Firms in the construction industry were the most flexible in the trips they made, roughly 75% making both goods and service stops and roughly 20% making either goods or service stops. Slightly more than half of firms in the "Office_Professional" group made only service stops, while a third made both goods and service stops, and only around 10% made goods stops exclusively.

Figure 45: FIRMS MAKING STOPS FOR 2019 (GOOD & SERVICES)

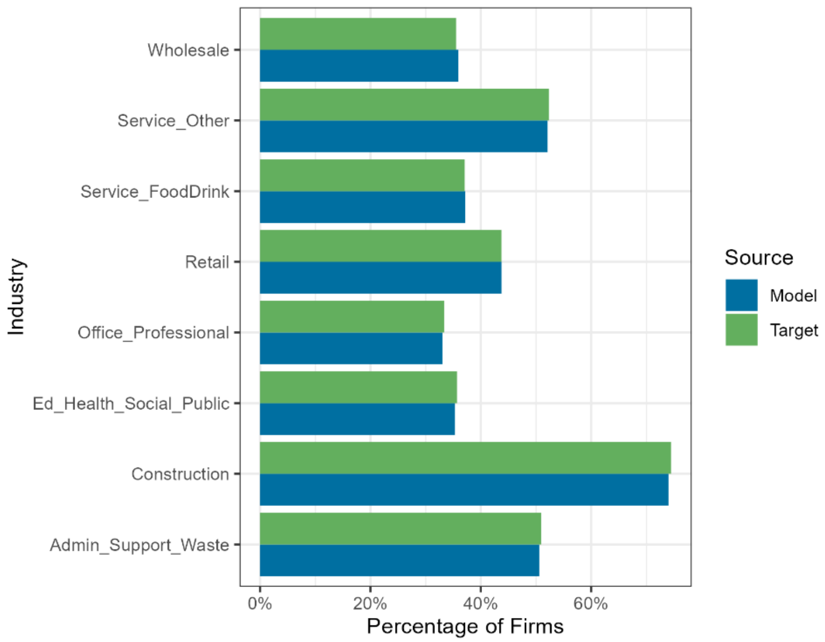


Figure 46: FIRMS MAKING STOPS FOR 2019 (SERVICES ONLY)

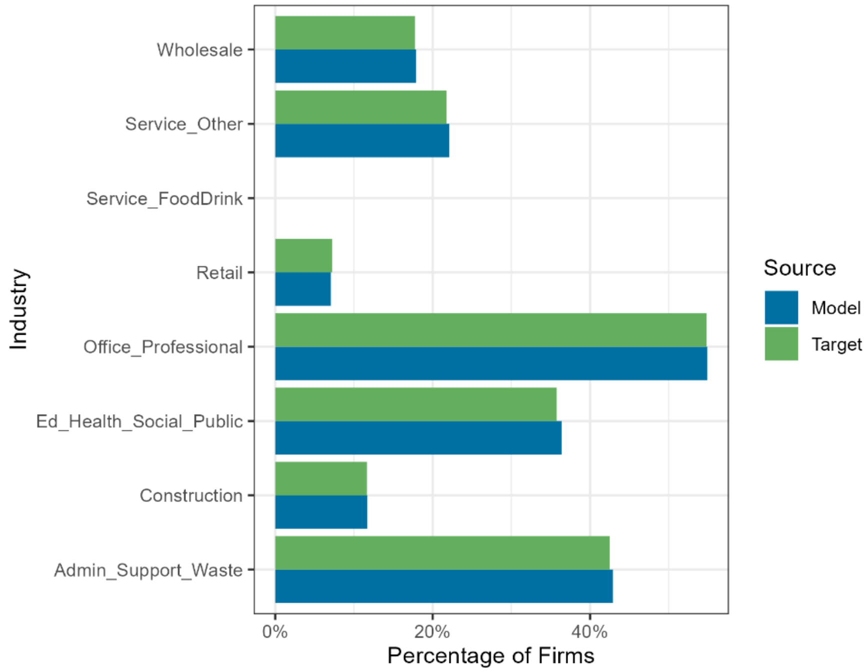
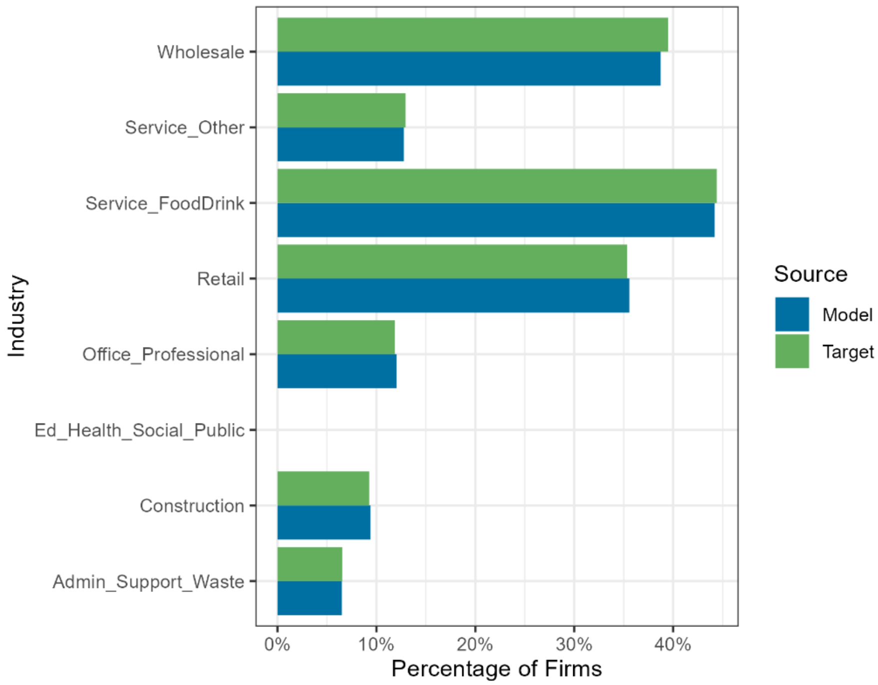


Figure 47: FIRMS MAKING STOPS FOR 2019 (GOODS ONLY)



Simulate Scheduled Stops

The internal stop generation model predicts one day's worth of scheduled stops for each establishment by TAZ, covering all internal TAZs, using a hurdle-count model formulation (i.e., a binary model to predict whether a TAZ has a stop or not, and a count model to predict the number of stops given that a stop takes place). Scheduled stops from this model were grouped into two-market segments, good stops and service stops, and the model also identifies whether the stop locations are at residential or non-residential land uses. The model was estimated using the SEMCOG CVS data, with updates made in the project to the estimation to reflect the employment groupings defined in the activities model based on groupings of the CMAP employment categories.

Figure 48 and Figure 49 show the calibrated parameters of the goods model, showing the zero hurdle model first followed by the count model for each of the residential and non-residential segments of the model. The parameters include those related to the industry of the business operating the commercial vehicle (i.e., the parameters named employment categories such as "Construction", the land use in the TAZ where the stop might be made (i.e., the parameters named with a prefix "NEmp" for number of employees in the TAZ in a particular industry, and the "HH" parameter for the number of households in a TAZ), and then impedance parameters such as the distance or travel time from the business to the TAZ where the stop might be made. Figure 50 and Figure 51 show the calibrated parameters for the service model, which is structured similarly to the good model.

Figure 48: GOODS MODEL (ZERO HURDLE)

Coefficient Name	Residential	Non Residential
(Intercept)	-2.851	-1.680
Admin_Support_Waste	-0.219	-1.269
Construction	-0.348	0.711
Ed_Health_Social_Public	-3.562	-2.689
Office_Professional	-1.556	-1.659
Service_FoodDrink	-0.763	-2.062
Service_Other	0.863	-0.565
Transport_Industry	-0.435	-0.485
Wholesale	0.949	1.140
log(TOTAL_EMPLOYEES)	0.103	0.244
log(dist)	-1.767	-1.678
log(dist):Construction	0.044	0.142
log(dist):Office_Professional	-0.066	0.160
log(dist):Transport_Industry	0.204	0.294
log1p(HH)	0.297	0.000
log1p(NEmp_Construction)	0.000	0.008
log1p(NEmp_Retail)	0.000	0.318
log1p(NEmp_Service_Other)	0.000	-0.150
log1p(NEmp_Transport_Industry)	0.000	0.270

Figure 49: GOODS MODEL (COUNT)

Coefficient Name	Residential	Non Residential
(Intercept)	-9.043	-3.128
log(TOTAL_EMPLOYEES)	1.023	0.130
log(time)	-1.207	-0.312
log(time):Transport_Industry	0.091	0.244
log1p(HH)	0.619	0.000
log1p(NEmp_Ed_Health_Social_Public)	0.000	0.045
log1p(NEmp_Office_Professional)	0.000	-0.221
log1p(NEmp_Retail)	0.000	-0.109
log1p(NEmp_Transport_Industry)	0.000	0.145

Figure 50: SERVICE MODEL (ZERO HURDLE)

Coefficient Name	Residential	Non Residential
(Intercept)	-1.282	-1.353
Admin_Support_Waste	2.427	3.516
Admin_Support_Waste:log(TOTAL_EMPLOYEES)	-0.518	-1.161
Construction	3.164	2.226
Ed_Health_Social_Public	0.748	-0.035
Office_Professional	0.678	-0.678
Office_Professional:log(TOTAL_EMPLOYEES)	-0.932	-0.225
Service_FoodDrink	1.739	0.701
Service_FoodDrink:log(TOTAL_EMPLOYEES)	-14.936	-0.463
Service_Other	1.672	-0.500
Service_Other:log(TOTAL_EMPLOYEES)	-0.306	0.301
Transport_Industry	-0.025	0.024
Wholesale	0.864	1.910
log(TOTAL_EMPLOYEES)	0.260	0.342
log(time)	-1.781	-1.608
log1p(HH)	0.335	0.000
log1p(NEmp_Admin_Support_Waste)	0.000	0.096
log1p(NEmp_Ed_Health_Social_Public)	0.000	0.071
log1p(NEmp_Office_Professional)	0.000	0.086
log1p(NEmp_Retail)	0.000	0.097
log1p(NEmp_Service_FoodDrink)	0.000	0.061
log1p(NEmp_Service_Other)	0.000	-0.155
log1p(NEmp_Transport_Industry)	0.000	0.118

Figure 51: SERVICE MODEL (COUNT)

Coefficient Name	Residential	Non Residential
(Intercept)	-9.043	-3.128
log(TOTAL_EMPLOYEES)	1.023	0.130
log(time)	-1.207	-0.312
log(time):Transport_Industry	0.091	0.244
log1p(HH)	0.619	0.000
log1p(NEmp_Ed_Health_Social_Public)	0.000	0.045
log1p(NEmp_Office_Professional)	0.000	-0.221
log1p(NEmp_Retail)	0.000	-0.109
log1p(NEmp_Transport_Industry)	0.000	0.145

The models were calibrated by iteratively adjusting the zero model's intercepts and the variables for the industry operating the commercial vehicle in the zero model until the number of stops made by each industry matched with the industry specific estimates and the overall total number of stops matched the estimated total number of stops. The estimates were made by applying industry specific stop rates from the SEMCOG CVS to the firm synthesis outputs from the CMAP CSVM. The transport and industry employment group was removed from the target totals for calibration. As well as adjustments to the stop numbers, the iterative calibration process also adjusted the impedance parameters (distance or travel time from the business to the TAZ where the stop might be) so that the average base to stop distance matched the value observed in the CMAP GPS data.

Figure 52 shows the number of stops generated by industry, compared to the targets derived from the SEMCOG CVS data applied to the firm synthesis outputs from the CMAP CSVM. Stops refers to stops made to deliver or pick up goods or to provide business services and does not include any intermediate stops related to driver needs or vehicle service needs. Overall and by industry the generated stops match the target data. The construction industry group produced the most stops, roughly 350,000. The wholesale industry group followed with around 110,000 generated stops. The "Service_FoodDrink" industry group produced the fewest stops, in line with the observed data, fewer than 25,000 stops.

Figure 53 shows the number of goods stops generated by industry. We see that the construction, wholesale, and retail industry groups generated the most goods stops. Figure 54 shows the number of generated service stops by industry group. The construction industry was the single largest producer of service stops by a large margin, producing around 250,000 stops. The Ed_Health_Social_Public industry group and the Admin_Support_Waste groups followed with just under 75,000 trips each.

Figure 52: TOTAL STOPS BY INDUSTRY FOR 2019

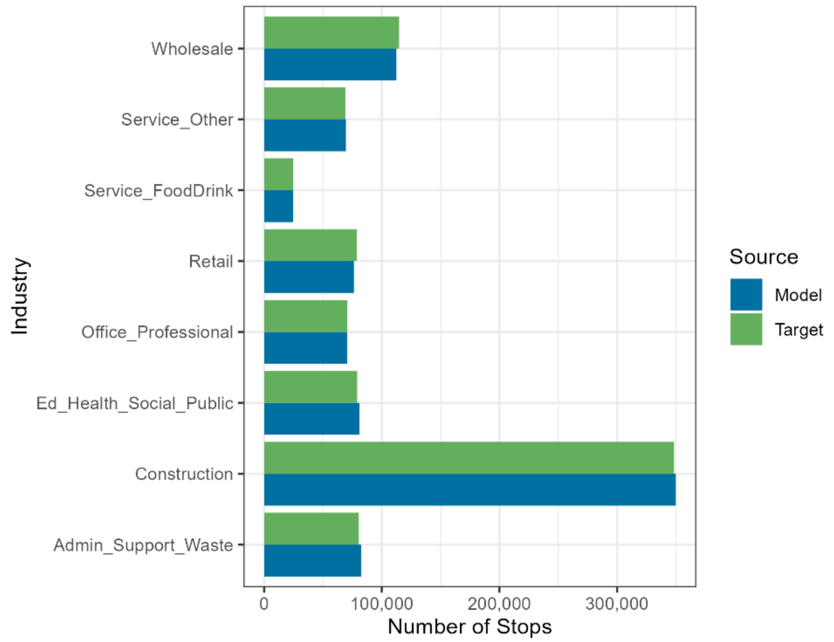


Figure 53: TOTAL GOODS STOPS (BY INDUSTRY)

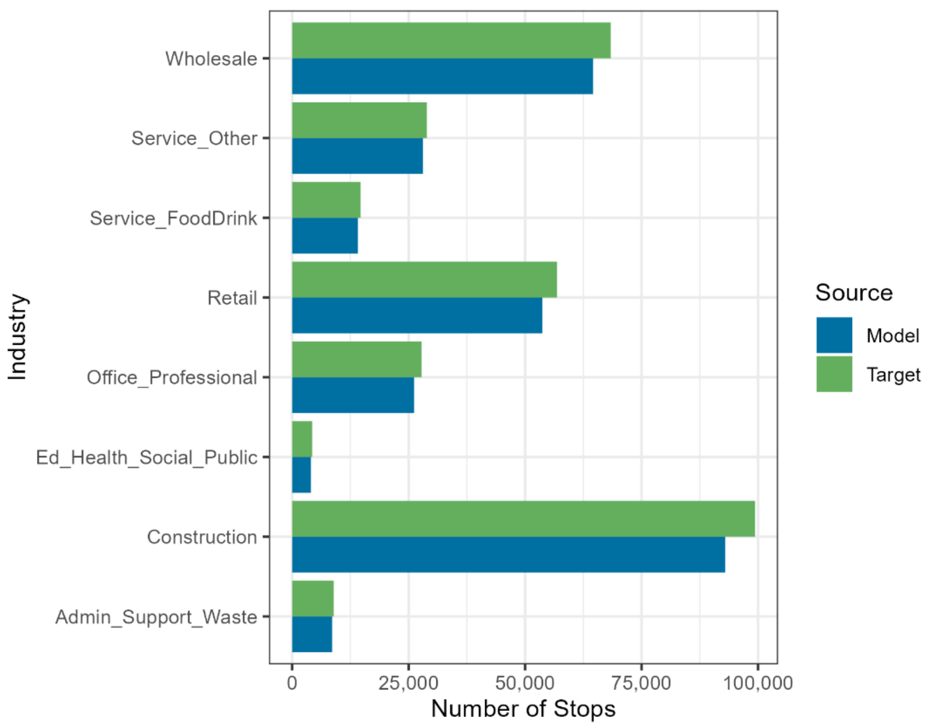
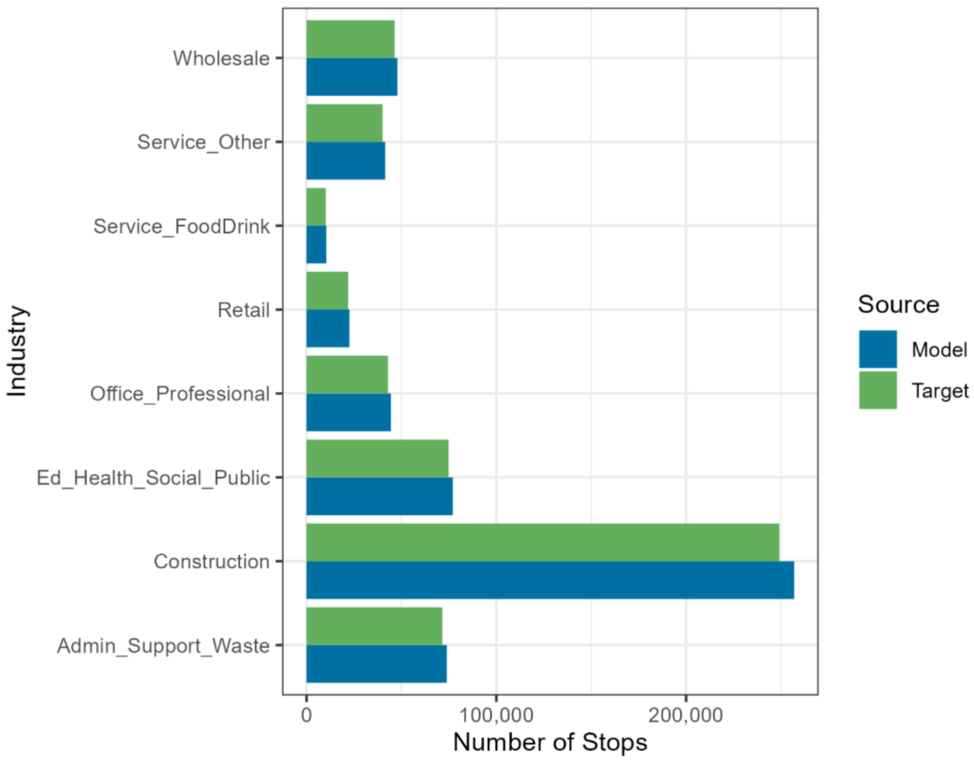


Figure 54: TOTAL SERVICE STOPS (BY INDUSTRY)



Simulate Vehicle Choice

For each stop, the vehicle choice model assigns one of three commercial vehicle types. These correspond to light (i.e., car, van, and pickup), medium/single-unit, and heavy/multi-unit truck types. The model is a multinomial logit model and predicts vehicle type as a function of the establishment's industry type, distance between establishment and stops to be served, and the stop's purpose—to deliver goods or provide services.

Figure 55 shows the calibrated parameters of the vehicle choice model. The three alternatives in the choice model are light commercial vehicles, medium trucks and heavy trucks. Parameters prefixed "beta_v1" are included in the utility equation for light commercial vehicles in addition to the "asc_light" alternative specific constant. The "beta_v2" parameters and the "asc_medium" are included in the utility equation for medium trucks. Heavy trucks are set at the base alternative in the model and the only parameter is "asc_heavy" which is fixed to zero. The parameters relate to the industry operating the commercial vehicle, the stop activity (either goods or service stops) and the distance between the business and the stop.

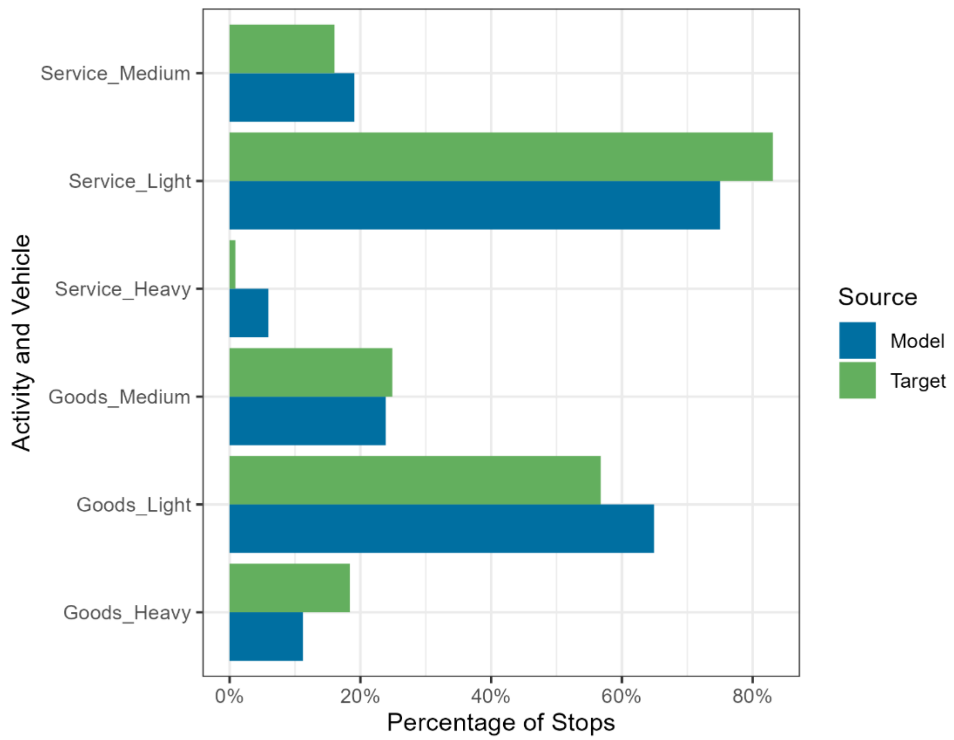
Figure 55: VEHICLE CHOICE MODEL

Coefficient Name	Coefficient
asc_light	2.214
asc_medium	-0.657
asc_heavy	0.000
beta_v1_industry_retail	0.000
beta_v1_industry_wholesale	-0.618
beta_v1_industry_construction	0.042
beta_v1_industry_transport_industry	0.000
beta_v1_industry_admin_support_waste	1.869
beta_v1_industry_ed_health_social_public	2.520
beta_v1_industry_service_other	1.496
beta_v1_industry_office_professional	-0.200
beta_v1_industry_service_fooddrink	3.429
beta_v2_industry_retail	0.000
beta_v2_industry_wholesale	-0.111
beta_v2_industry_construction	0.417
beta_v2_industry_transport_industry	0.000
beta_v2_industry_admin_support_waste	-0.585
beta_v2_industry_ed_health_social_public	5.094
beta_v2_industry_service_other	1.721
beta_v2_industry_office_professional	-0.133
beta_v2_industry_service_fooddrink	2.633
beta_v1_activity_deliver_pickup	0.000
beta_v1_activity_service	11.955
beta_v2_activity_deliver_pickup	0.000
beta_v2_activity_service	9.820
beta_v1_dist_00_02	0.000
beta_v1_dist_02_05	0.681
beta_v1_dist_05_10	0.365
beta_v1_dist_10_20	-0.033
beta_v1_dist_20_p	-0.819
beta_v2_dist_00_02	0.000
beta_v2_dist_02_05	1.539
beta_v2_dist_05_10	1.879
beta_v2_dist_10_20	1.150
beta_v2_dist_20_p	1.102

The alternative specific constants in the vehicle choice model were iteratively adjusted to match the observed shares of vehicles usage for goods stops and for service stops. Figure 56 provides the overall rate of vehicle choice for goods and service stops in the model compared to the shares from the expanded SEMCOG CVS data. Service stops in the control data tend to be made overwhelmingly with light vehicles, the model output reflects this overall, though the model slightly shifts utilization away from light vehicles in favor of medium and heavy vehicles for service stops. Whereas heavy vehicle service trips in the SEMCOG CVS data was near 0%, the rate in the model results was closer to 6%.

Similarly, a majority of goods stops in the SEMCOG CVS data utilized light vehicles, though the use of medium and heavy vehicle was more prevalent. In the case of vehicle choice for goods stops, the model shifts vehicle choice slightly in favor of light vehicles. Whereas the observed rate of light vehicle goods trips was roughly 55% in the control data, this rate is closer to 65% in the model. See Figure 55 for vehicle choice model coefficients.

Figure 56: VEHICLE CHOICE BY ACTIVITY FOR 2019



Simulate Stop Duration

The stop duration model is applied to scheduled stops generated by the stop generation model as well as any intermediate stops on the tour. For each stop, an expected stop duration range is estimated using a multinomial logit model. Then an actual value in minutes is drawn from the range.

Figure 57 shows parameters of the stop duration logit model, which is also applied to intermediate stops later in the model sequence to estimate the durations of those stops. The duration model is a function of vehicle type and stop purpose as well as having alternative specific constants for each of the choices, which are represented as time bins starting from 1-15 minutes, and continuing to 600 or more minutes.

Figure 57: STOP DURATION MODEL

Coefficient Name	Coefficient
asc_15	0.000
asc_30	-0.771
asc_45	-1.505
asc_60	-2.104
asc_75	-2.883
asc_90	-3.616
asc_150	-2.724
asc_210	-3.487
asc_270	-4.002
asc_390	-4.034
asc_600	-3.487
b_is_med_veh_30_75	-0.566
b_is_med_veh_90_plus	-1.126
b_is_hvy_veh_30_90	0.471
b_is_hvy_veh_150_plus	-0.829
b_activity_service_60_75	0.729
b_activity_service_90_plus	1.798
b_activity_other_30	0.100
b_activity_other_45	0.175
b_activity_other_60_90	0.317
b_activity_other_150_plus	0.511
b_activity_driver_needs_30	0.329
b_activity_driver_needs_45	0.525
b_activity_driver_needs_60_90	0.526
b_activity_driver_needs_150_plus	0.377

The stop duration model was calibrated by iteratively adjusting the alternative specific constants, the activity specific variables, and the vehicle specific variables to match with the distributions observed in the SEMCOG CVS data.

Figure 58 and Figure 59 show the distribution of stop durations in the model compared to SEMCOG CVS data by stop activity. The distributions match the observed data relatively well, with some small discrepancies. In the case of goods stops, generated stop durations in the longer duration bins tend to be less frequent and trips in the shortest duration group were inflated slightly. Stop durations between 1 -15 minutes were the most frequently generated, over 50% of all stops. Despite this shift away from longer stop durations, there is still the expected peak around the 91-150 minute bin.

In the case of service stops, generated stop durations fit the control data quite closely. It is still the case that 1-15 minute stop duration group is the most common, around 40%, and fewer stops are generated with longer durations. Relative to goods stops, we see a more uniform distribution of stop durations. For example, whereas the rate of goods stops lasting between 91-150 minutes was less than 5%, the proportion of service stops in this same bin was over 10%. See Figure 57 for stop duration model coefficients.

Figure 58: DISTRIBUTION OF STOP DURATION FOR 2019 (GOODS STOPS)

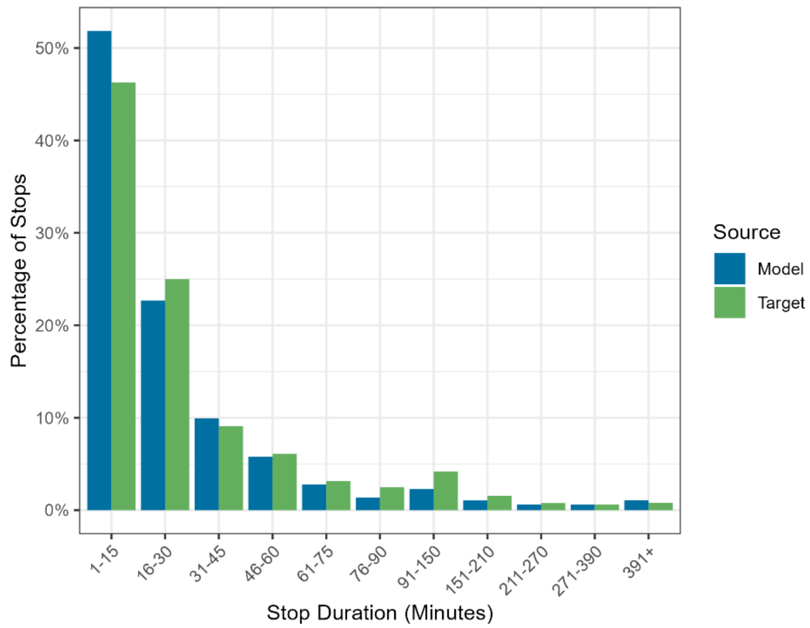
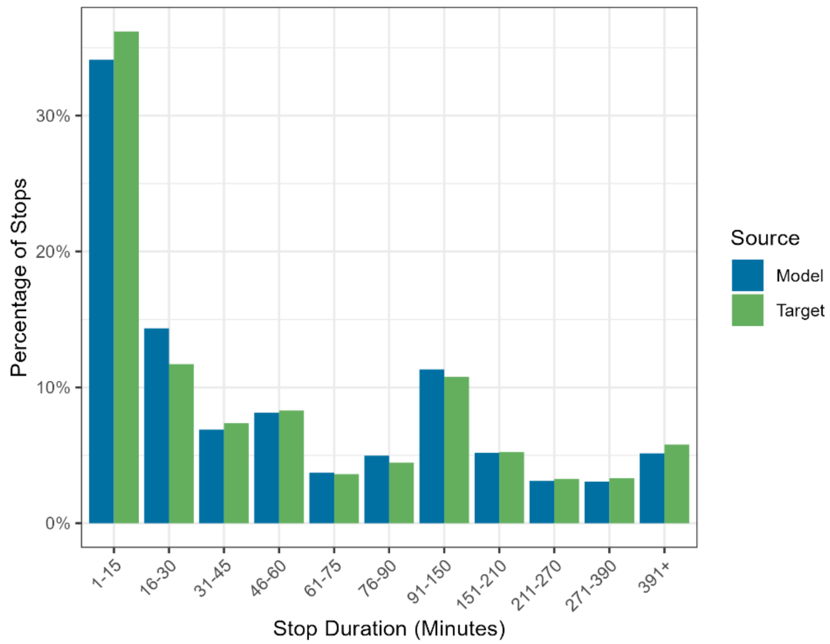


Figure 59: DISTRIBUTION OF STOP DURATION FOR 2019 (SERVICE STOPS)



Simulate Tours and Routing

For each establishment, the stop clustering model groups scheduled goods and service stops into feasible commercial vehicle tours, based on vehicle type, spatial proximity of the stops, and expected stop duration. This is followed by a tour type model, which is a multinomial logit model that identifies whether each cluster of stops is part of a tour that starts and ends at the same location, or one that starts, and ends in different locations. Finally, a tour sequencing model is applied that uses a traveling salesman algorithm to order the stops within each tour into a route.

Figure 60 shows the parameters of the tour type model which allocates a tour stop cluster to one of seven alternative tour types. If the tour cluster is only formed of one stop, then the model chooses between three tours types: “bbs”, a base to base single-stop tour, “bns”, a base to not-base (i.e., ending something other than the vehicle base) single-stop tour, or “nbs”, a not-base to base single-stop tour. If the cluster has multiple stops, then the model chooses between “bbm”, “bnm”, “nbm” which are similar to the single-stop tour types except with multiple stops, and also “nnm”, a not-base to not-base multiple stop tour.

Figure 60: TOUR MODEL

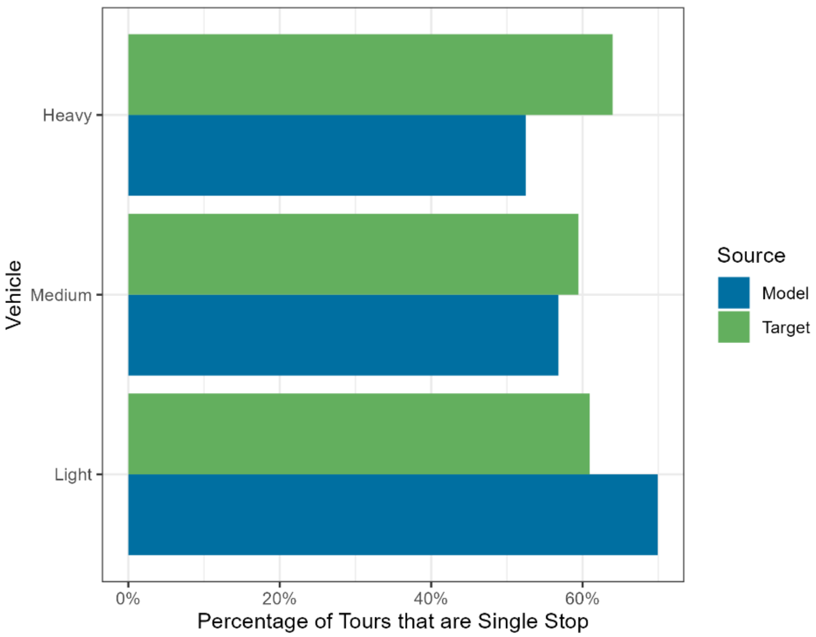
Coefficient Name	Coefficient
asc_bbm	0.000
asc_bbs	-0.069
asc_bnm	-2.071
asc_bns	-2.429
asc_nbm	-3.174
asc_nb0	-3.196
asc_nbs	-3.132
asc_nnm	-2.800
asc_bbm_is_med_veh	0.000
asc_bbs_is_med_veh	0.000
asc_bnm_is_med_veh	-0.702
asc_bns_is_med_veh	-0.780
asc_nbm_is_med_veh	-1.432
asc_nb0_is_med_veh	0.000
asc_nbs_is_med_veh	-1.025
asc_nnm_is_med_veh	-0.415
asc_bbm_is_hvy_veh	0.000
asc_bbs_is_hvy_veh	0.000
asc_bnm_is_hvy_veh	-0.228
asc_bns_is_hvy_veh	-1.369
asc_nbm_is_hvy_veh	0.633
asc_nb0_is_hvy_veh	0.000
asc_nbs_is_hvy_veh	-1.326
asc_nnm_is_hvy_veh	-0.838

The eighth tour type included in the tour model which was apparent in the data is a “nb0”, which is a non-base to base zero stop tour and is excluded from the application as it is understood to primarily be commuter to work type trips for commercial vehicle drivers. The model includes alternative specific constants, and then a set of vehicle type variables.

The proportion of single-stop tours vs. multi-stop tours, an outcome of the clustering step of the model, was calibrated by adjusting the clustering parameters in the CSVM to match the observed split between single-stop and multi-stop tours by vehicle type in the SEMCOG CSV data. The tour type model was calibrated by iteratively adjusting the alternative specific constants and the vehicle specific variables to match the shares of tours by tour type within single-stop and multi-stop tours as observed in the SEMCOG CVS data.

Figure 61 shows the rate of single-stop tours produced by the model, segmented by vehicle class. In the case of light vehicles, the model tends to favor single-stop tours slightly (roughly 70% vs 60%), while in the case of heavy vehicles, the model tends to favor multi-stop tours (roughly 49% vs 37%).

Figure 61: PROPORTION OF SINGLE-STOP TOURS FOR 2019



In Figure 62 through Figure 60, “N” refers to a tour start or end that is not the inferred vehicle base. “B” refers to a tour start or end that is the vehicle’s inferred base, and “S” or “M” refer to a single- or multi-stop tour. For example, “NBM” would refer to a tour that starts somewhere besides the vehicle’s base, ends at the vehicle’s base, and is composed of multiple stops. Overwhelmingly across all vehicle classes, most tours, whether single- or multi-stop, started and ended at that vehicle’s inferred base.

Figure 62: TOUR TYPOLOGY FOR 2019 (LIGHT VEHICLE TOURS)

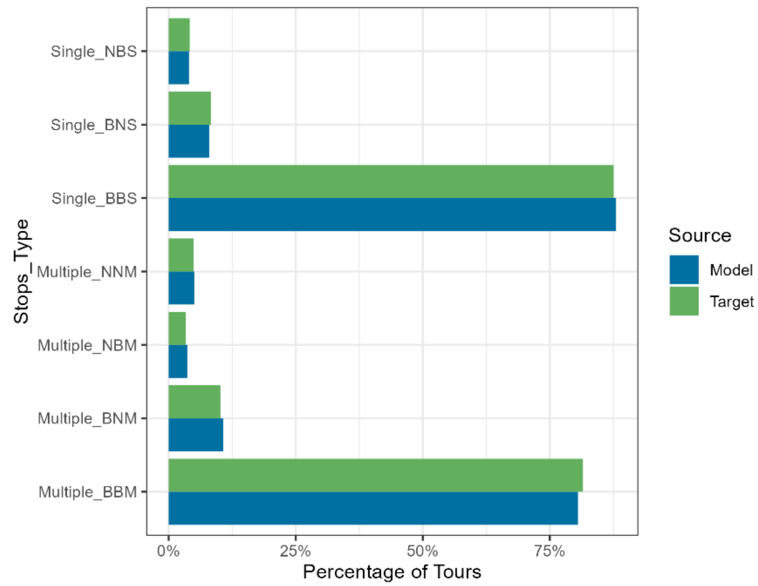


Figure 63: TOUR TYPOLOGY FOR 2019 (MEDIUM VEHICLE TOURS)

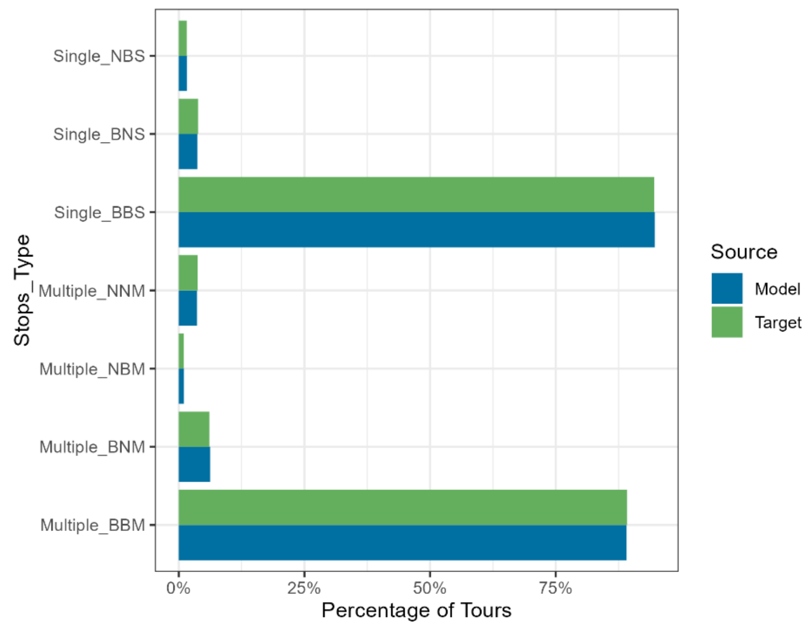
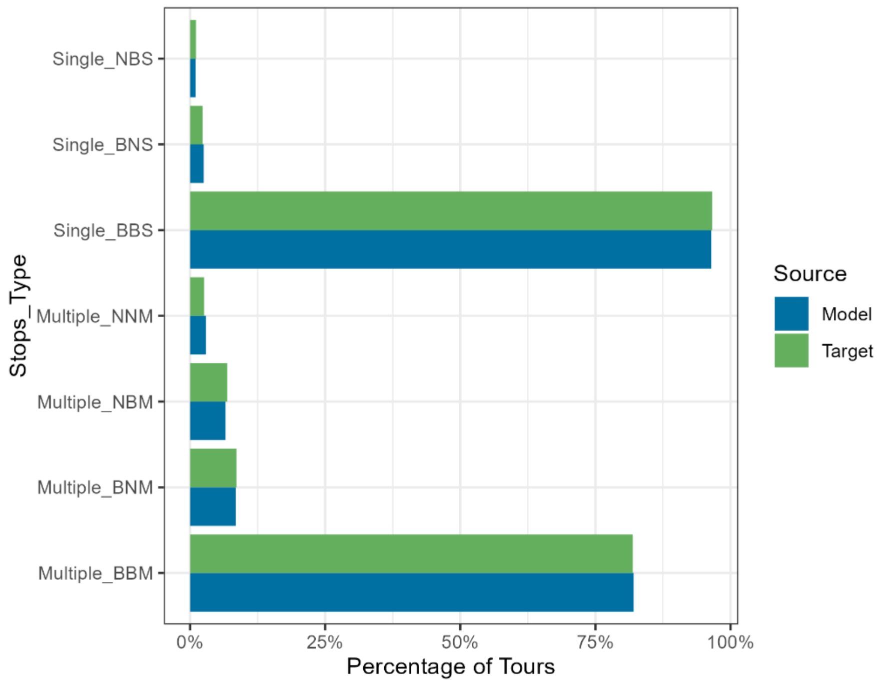


Figure 64: TOUR TYPOLOGY FOR 2019 (HEAVY VEHICLE TOURS)



Simulate Scheduled Trips

The scheduled trips component identifies the timing for each of the tours created in the previous model component. For each tour, the arrival time at the first stop on the tour is predicted using a multinomial logit model. Figure 65 shows the coefficients of the model, which includes alternative specific constants for each time bin (relative to the overnight time period from 10pm to 6am, as well as variables related to the type of tour and the vehicle making the tour.

Figure 65: ARRIVAL MODEL

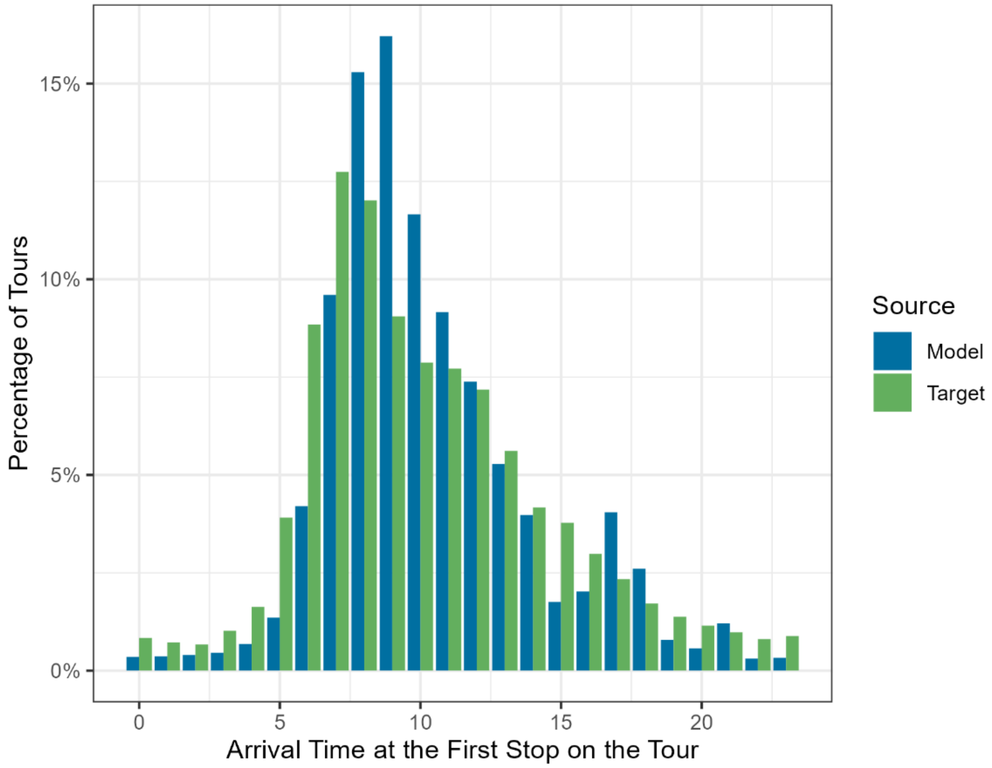
Coefficient Name	Coefficient
asc_overnight	0.000
asc_0600	0.782
asc_0700	0.950
asc_0730	1.430
asc_0800	1.900
asc_0830	2.026
asc_0900	2.293
asc_0930	2.080
asc_1000	2.012
asc_1030	1.379
asc_1100	1.357
asc_1130	1.379

Coefficient Name	Coefficient
asc_1200	1.357
asc_1230	1.007
asc_1300	1.184
asc_1330	0.792
asc_1400	1.069
asc_1430	0.406
asc_1500	0.792
asc_1530	0.463
asc_1600	0.518
asc_1630	-0.636
asc_1700	-0.111
asc_1730	-0.885
asc_1800	-0.111
asc_1900	-1.502
asc_2000	-2.064
asc_2100	-1.252
asc_begin_not_base_0900_1630	-1.102
asc_end_not_base_1700_2100	1.267
asc_single_stop_1030_1800	0.430
asc_is_no_stop_0600_0730	1.176
asc_is_med_hvy_veh_0600	-1.066
asc_is_med_hvy_veh_0700	-0.714
asc_is_med_hvy_veh_0730	-1.355
asc_is_med_hvy_veh_0800	-1.522
asc_is_med_hvy_veh_0830	-1.510
asc_is_med_hvy_veh_0900	-1.844
asc_is_med_hvy_veh_0930	-1.919
asc_is_med_hvy_veh_1000	-1.899
asc_is_med_hvy_veh_1030	-1.707
asc_is_med_hvy_veh_1100	-2.074
asc_is_med_hvy_veh_1130	-2.197
asc_is_med_hvy_veh_1200	-2.480
asc_is_med_hvy_veh_1230	-1.824
asc_is_med_hvy_veh_1300	-1.810
asc_is_med_hvy_veh_1330	-2.356
asc_is_med_hvy_veh_1400	-2.192
asc_is_med_hvy_veh_1430	-1.394
asc_is_med_hvy_veh_1500	-2.763
asc_is_med_hvy_veh_1530	-2.028
asc_is_med_hvy_veh_1600	-1.977

Coefficient Name	Coefficient
asc_is_med_hvy_veh_1630	-0.728
asc_is_med_hvy_veh_1700	-1.894
asc_is_med_hvy_veh_1730	-0.870
asc_is_med_hvy_veh_1800	-1.761
asc_is_med_hvy_veh_1900_2100	-0.855

The model was calibrated by iteratively adjusting the alternative specific constants of the vehicle specific variables to match with the target distribution by vehicle type derived from the SEMCOG CVS data. Figure 66 below shows the distribution of the first stop arrival times in the model compared to the target data. Generated tour start times tend to reflect the control data though we see more dramatic peaks during the AM peak period and the PM peak period.

Figure 66: DISTRIBUTION OF TOUR DEPARTURE TIME FOR 2019



Simulate Intermediate Stops

In the final model step, intermediate stops are generated for some tours . First, the intermediate stop choice model, a multinomial logit model, predicts whether there are intermediate stops between scheduled stops on each tour. The model simulates whether the driver makes one or more intermediate stops prior to each scheduled goods or service stop, or prior to returning to the establishment to complete the tour (see Figure 67). Purposes for intermediate stops are breaks/meals (driver needs, denoted “dn” in the model parameter names), vehicle service/refueling (“vs”), and personal business/other ("ot"). The alternatives in the model are set relative to no-stop ("ns"), i.e., no intermediate stop is added to the trip. The likelihood of making a stop is a function of the tour duration remaining at the point of the trip, the vehicle type, the time of day, and the distance back to the business establishment.

The intermediate stop destination model predicts a destination TAZ for each intermediate stop. Specifically, for each intermediate stop, a multinomial logit model is used to select from a set of eligible TAZs (those which do not require excessive deviation from the tour route) based on attraction factor(s), such as food and drink service employment for break/meal stops (see Figure 68).

Figure 67: INTERMEDIATE STOPS MODEL

Coefficient Name	Coefficient
asc_ns	0.000
asc_dn	-1.744
asc_vs	-4.025
asc_ot	-2.105
b_log_duration_dn	-0.097
b_log_duration_vs	0.207
b_log_duration_ot	0.269
b_med_veh_vs	-0.584
b_med_veh_ot	-1.301
b_hvy_veh_vs	-0.159
b_hvy_veh_ot	-2.372
b_is_lunch_dn	0.287
b_distance_1_2_dn	-0.414
b_distance_2_10_dn	-0.228
b_distance_10plus_dn	-0.333
b_distance_1_10_vs	-0.505
b_distance_10plus_vs	-0.218
b_distance_1_2_ot	-0.128
b_distance_2plus_ot	-0.333

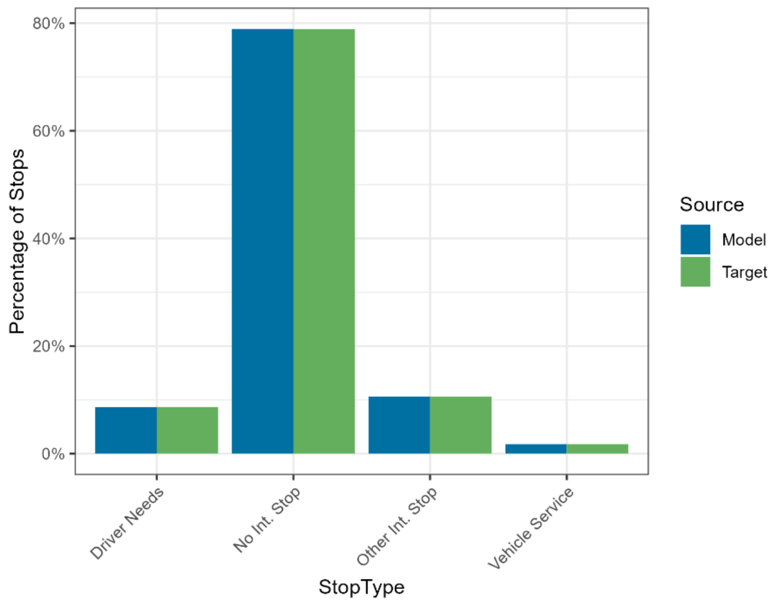
Figure 68: INTERMEDIATE STOP ATTRACTION MODEL

Coefficient Name	Coefficient
b_emp	0.187
b_dist	-0.133
b_retail	0.370
b_foodDrink_dn	0.164
b_dist_vs	-0.046

The intermediate stops model was calibrated by adjusting the alternative specific constants to match targets derived from the SEMCOG CSV data. The intermediate stop attraction model was not adjusted post estimation.

Figure 69 shows the share of intermediate stops by purpose from the model and in the SEMCOG CVS data. Reflective of what is observed in the SEMCOG CVS data, slightly over 20% of all stops generated by the model are intermediate stops, most being captured by the 'Other' category, followed closely by driver needs stops. Vehicle service stops made up the smallest portion of generated intermediate stops.

Figure 69: PREVALENCE OF INTERMEDIATE STOPS FOR 2019



4.5 Base Year Validation

Once each component of the base year CSVM had been calibrated as described in the section above, a further round of validation comparisons were made between the base year outputs (trip tables by time period and vehicle type, as well as the complete database of commercial vehicle trips and tours).

A summary of the base year scenario results are shown in Figure 70, which provides a set of statistics covering the main outputs from the CSVM, and Figure 71, which reports VMT by the eight time period and three vehicle type.

Figure 70: CSVM OVERVIEW STATISTICS AND VEHICLE MILES TRAVELED FOR 2019

Overview	Count
Synthesized Firms	291,724
Employment	4,940,712
Goods and Service Stops	877,523
Commercial Vehicle Tours	512,330
Total Trips	1,764,881
Total VMT	21,050,285

Figure 71: TOTAL VMT (BY PERIOD & VEHICLE)

TOD	Hours	Light	Medium	Heavy	Total
P1	8pm-6am	1,103,756	458,335	159,576	**1,721,667**
P2	6am-7am	657,053	223,468	79,462	**959,983**
P3	7am-9am	2,853,118	749,757	259,318	**3,862,193**
P4	9am-10am	1,933,924	446,398	152,988	**2,533,309**
P5	10am-2pm	6,332,466	1,273,336	440,943	**8,046,745**
P6	2pm-4pm	1,487,653	314,551	106,052	**1,908,256**
P7	4pm-6pm	1,052,624	190,226	71,795	**1,314,645**
P8	6pm-8pm	532,887	124,846	45,755	**703,487**
Total		**15,953,481**	**3,780,916**	**1,315,888**	**21,050,285**

The validation comparisons focus on comparing the CSVM outputs to estimates of actual VMT by vehicle type (see Figure 72) and trip and tour statistics for distance, number of stops, percentage of single stop tours, base-to-stop distance, cluster distance, stop and tour duration (see Figure 73).

The total VMT is estimated commercial vehicle VMT based on the FHWA reported average daily VMT by vehicle type for the Chicago IL-IN urban area in 2019, factored up to the 21-county model region with a portion of light vehicle VMT allocated to commercial vehicles based on the proportion from the SEMCOG CVS data. The observed CSVM VMT is the proportion of the VMT that is internal to internal and is from an industry covered by the CSVM (i.e., excluding freight model covered industries such as manufacturing and transportation and warehousing). The Transportation Network Usage section of this report describes the process of estimating the observed VMT in more detail.

Figure 72 shows that the model is reasonably close to matching the estimates of average weekday VMT by vehicle type for the region. All of the vehicle category estimates are within 11% of the observed data and overall, the model output is 6.9% higher. Given the inherent uncertainty in the estimated values themselves, which are based on FHWA VMT estimates that have been factored using factors derived from survey data, the match is satisfactory.

Figure 72: TRIP TABLE VALIDATION STATISTICS FOR 2019

Vehicle	Total VMT	Observed CSVM VMT	Model VMT	Difference	Ratio
Light	20,930,588	14,348,729	15,953,481	1,604,752	1.112
Medium	7,278,432	4,023,632	3,780,916	-242,716	0.940
Heavy	12,961,598	1,313,915	1,315,888	1,973	1.002
Total	**41,170,618**	**19,686,276**	**21,050,285**	**1,364,009**	**1.069**

Difference is Model – Observed, and Ratio is Model/Observed

The observed values in Figure 73 are all based on the CMAP GPS data that was processed by RSG to develop validation tabulations. Each measure relates to how well the CSVM matches with a tour characteristic observed in the GPS data.

- Average trip distance: This is the average of all trips. The model is producing shorter trips than observed in the GPS data, about 87% of the observed length.

- Average tour distance: this is the average of the total distance traveled during a tour, essentially the sum of all of the trip distances in each tour. The model is within 3% of the observed data, with 41.1 miles compared to 40.0 miles.
- Average number of stops in a tour (not counting the return to base stop at the end of the tour): The model is producing slightly more stops per tour, at 2.4 compared with 2.2. in the GPS data
- Proportion of single stop tours: While 41% of tours in the GPS are single stop, the model produces slightly fewer, at 37%.
- Average base to stop distance: this is the direct network distance between the commercial vehicle's based and each stop (as opposed to the actual distance traveled along the route of the tour). The model's stop are slightly closer to base, at 14.5 miles compared to 15.7 miles.
- The tour cluster distance is a measure of how close the stops on a tour (not including the base) are to each other. On average, the stops in the GPS data were 13 miles apart while the model's stops are spread on average 10 miles apart.
- Average stop duration is time the commercial vehicle is stopped at each stop during the tour. The CSVM's stop are 59 minutes on average, very close to the GPS data at 56 minutes.
- Average tour duration is the average of the sum of the time taken for a tour, including all travel time and stop time. The model is within 1.5% of the observed data, with 235 minutes compared to 237 minutes.

Overall, the tours produced by the CSVM are reasonably consistent with the CMAP GPS data. The results err very slightly towards shorter trips with slightly more stops on average in each tour resulting in overall tour distances and travel times that are almost identical to the GPS data. Given the slightly different make up of the GPS data, which is more heavily weighted towards medium trucks than light commercial vehicle compared to the population of commercial vehicles that the model is simulating, and may also include some of the freight moving medium trucks that are excluded from the CSVM, the model is producing a set of tours and trips that appear to be reasonable.

Figure 73: TRIP & TOUR VALIDATION STATISTICS FOR 2019

Field	Statistic	Observed	Model	Difference	Ratio
Trips: Distance	Mean (Miles)	13.71	11.93	-1.78	0.870
Tours: Distance	Mean (Miles)	40.04	41.09	1.05	1.026
Tours: Number of Stops (Excluding Return to Base)	Mean (# Stops)	2.23	2.44	0.21	1.094
Tours: Single Stop Tours (vs. Multi-Stop)	Proportion	0.41	0.37	-0.04	0.902
Trips: Base-to-Stop Distance	Mean (Miles)	15.73	14.49	-1.24	0.921
Tours: Cluster Distance	Mean (Miles)	13.26	10.46	-2.8	0.789
Trips: Stop Duration	Mean (Minutes)	56.12	58.85	2.73	1.049
Tours: Tour Duration	Mean (Minutes)	237.55	234.29	-3.26	0.986

Difference is Model – Observed, and Ratio is Model/Observed

Several larger validation tables, such as comparisons between the origin-destination movements in the CMAP GPS data and the base year results from the CSVM, can be found in the model dashboard on validation tab in the trip tables drop down menu.

4.6 Future Year Validation

This section of the report documents the results of the future year scenario for 2050 which has been run using the CSVM. The future year (2050) results are compared against the base year (2019) results to review how the changes in model inputs from the base year to the future year affect the estimate of VMT and other measures summarized from the model results.

Figure 74 compares VMT by vehicle segment for the base year and future year, and Figure 75 compares the key trip and tour characteristics between the base year and future year.

VMT in the future year grew by 6.6 % over the base year. Given the relatively small amount of employment growth, particularly in industries that are the highest users of commercial vehicles such as construction, and the increase in congestion making it harder for businesses to serve more distant customers, this level of growth from the model is reasonable under the baseline assumptions that commercial vehicle usage is the future is consistent with current behavior (sensitivity tests on changes from this assumption are described later in this report).

In general, tour and trip characteristics were relatively unchanged though the proportion of single-stop tours increased by three percentage points.

Figure 74: TRIP TABLE VALIDATION STATISTICS FOR 2050

Vehicle	Base	Future	Difference	Ratio
Light	15,953,481	16,999,451	1,045,970	1.066
Medium	3,780,916	4,044,141	263,226	1.070
Heavy	1,315,888	1,402,566	86,678	1.066
Total	**21,050,285**	**22,446,158**	**1,395,873**	**1.066**

Difference is Future – Base, and Ratio is Future/Base

Figure 75: TRIP & TOUR VALIDATION STATISTICS FOR 2050

Field	Statistic	Base	Future	Difference	Ratio
Trips: Distance	Mean (Miles)	11.93	11.64	-0.29	0.976
Tours: Distance	Mean (Miles)	41.09	40.06	-1.03	0.975
Tours: Number of Stops (Excluding Return to Base)	Mean (# Stops)	2.44	2.44	0	1.000
Tours: Single Stop Tours (vs. Multi-Stop)	Proportion	0.34	0.37	0.03	1.088
Trips: Base-to-Stop Distance	Mean (Miles)	14.49	14.14	-0.35	0.976
Tours: Cluster Distance	Mean (Miles)	10.46	10.18	-0.28	0.973
Trips: Stop Duration	Mean (Minutes)	58.85	58.18	-0.67	0.989
Tours: Tour Duration	Mean (Minutes)	234.29	237.07	2.78	1.012

Difference is Future – Base, and Ratio is Future/Base

Firm Synthesis

For the year 2050, most industries saw some level of growth except for the wholesale industry and the transport and industry employment group (which includes manufacturing), which saw some decline in the number of workers. Office and professional work saw the largest increase, upwards of 150,000 additional employees over the base year (see Figure 76). Figure 77 shows how this growth in employment is distributed across the entire model region.

Figure 76: EMPLOYMENT BY INDUSTRY FOR 2050

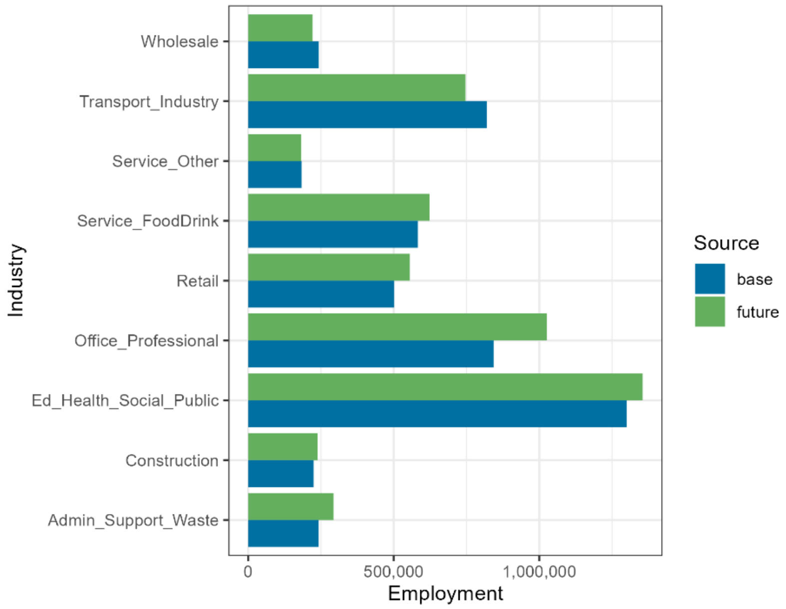
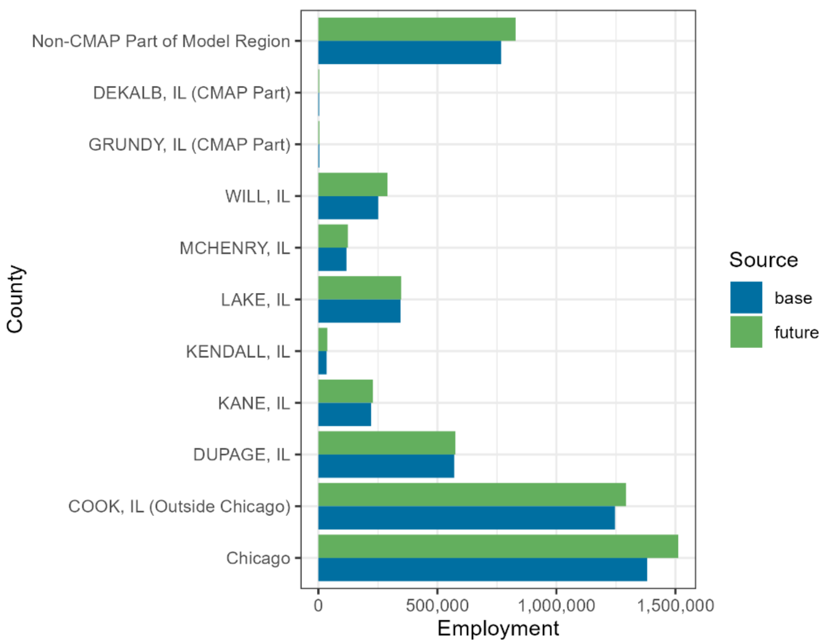


Figure 77: EMPLOYMENT BY DISTRICT FOR 2050



Activities

For the future year, the proportion of firms making either goods or service stops or both remains relatively unchanged. Retail and wholesale firms were the most likely to make only goods stops while firms in the public service or support service industry were most likely to make service stops exclusively. Construction firms were still the most likely to make both goods and service stops.

Figure 78: FIRMS MAKING STOPS FOR 2050 (GOODS)

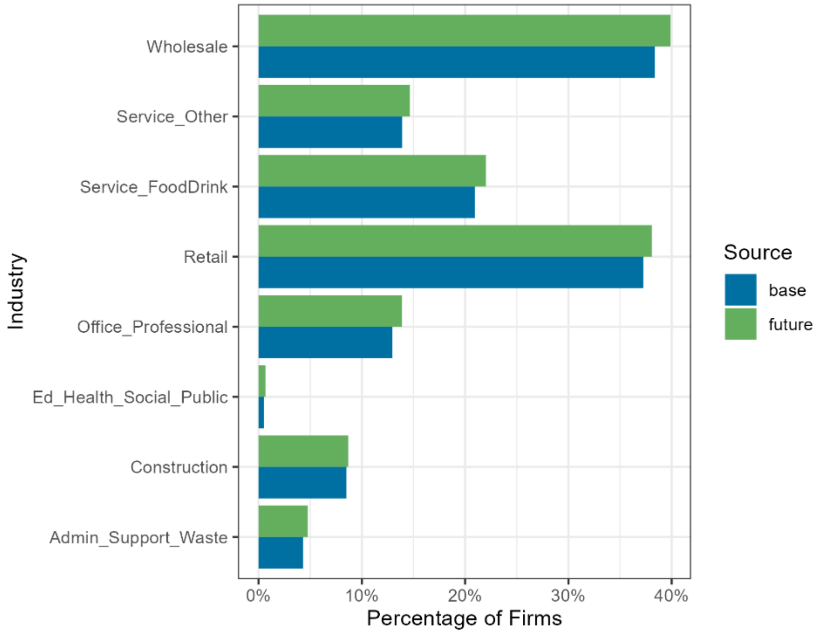


Figure 79: FIRMS MAKING STOPS FOR 2050 (SERVICES)

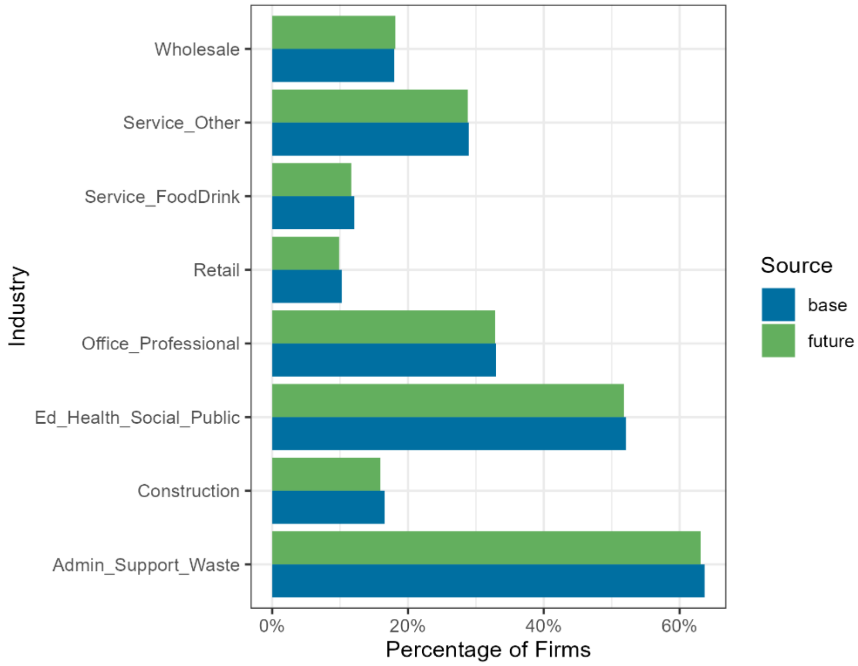
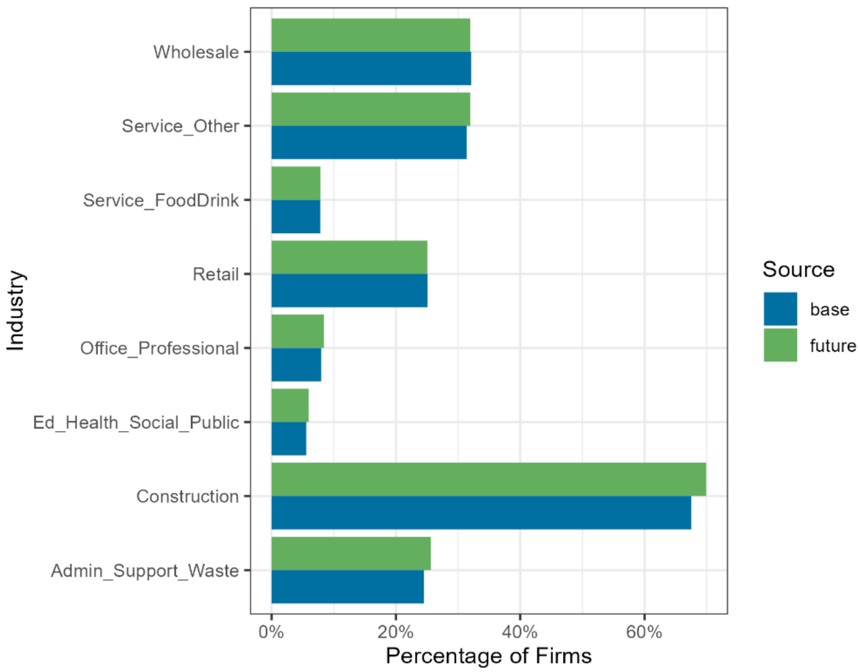


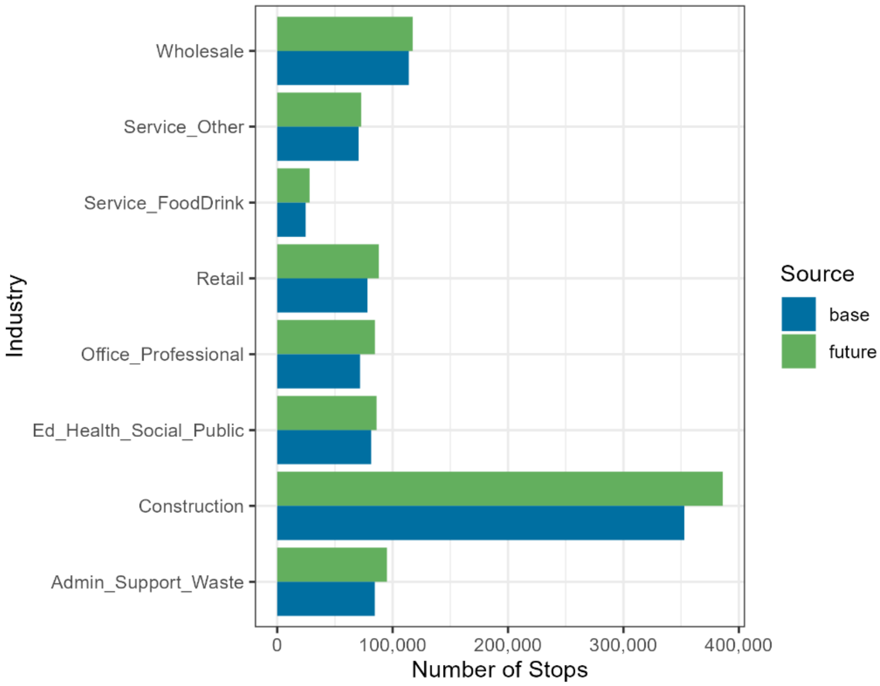
Figure 80: FIRMS MAKING STOPS FOR 2050 (BOTH GOODS AND SERVICES)



Scheduled Stops

Figure 81 shows the estimated growth in stops between 2019 and 2050. Generally, every industry saw some increase in the number of trips though the largest change is observed in the construction industry which will see upwards of 30,000 additional stops.

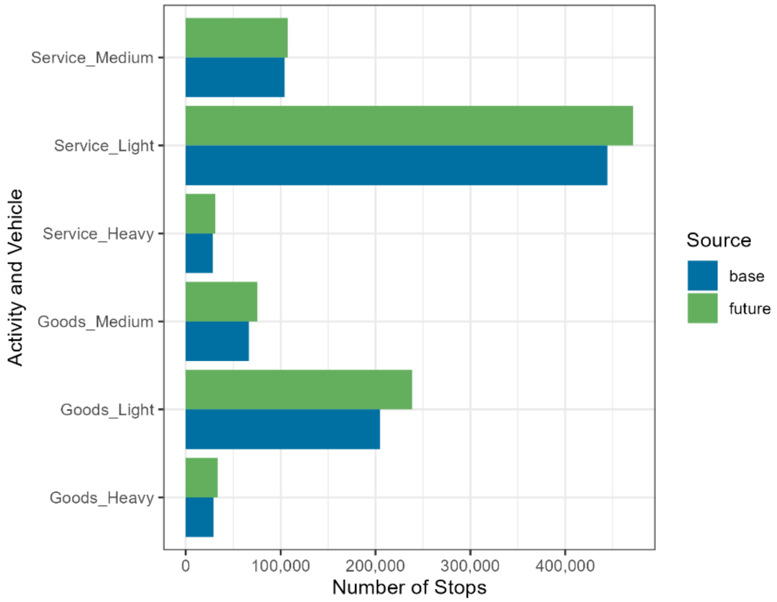
Figure 81: STOPS GENERATED BY INDUSTRY FOR 2050 (GOODS AND SERVICES)



Vehicle Choice

Figure 82 shows a breakdown of generated stop by vehicle class and activity for the 2050 scenario. With regards to service stops, trips assigned to medium and heavy vehicles have increased a small amount while trips assigned to light vehicles has increased by several thousand. Similarly, generated goods stops assigned to light vehicles also saw the largest increase while medium and heavy vehicle trips saw more modest increases.

Figure 82: VEHICLE CHOICE BY ACTIVITY FOR 2050



Stop Duration

Figure 83 and Figure 84 display the distribution of stop durations for goods and service stops, respectively. Generated stops were generally assigned shorter stop durations in keeping with the base year. Longer stop durations saw minimal growth while stop durations shorter than one hour saw more dramatic growth, especially stops between 1-15 minutes. With respect to service stops, the future year saw relatively uniform growth across all stop duration bins.

Figure 83: DISTRIBUTION OF STOP DURATION FOR 2050 (GOODS)

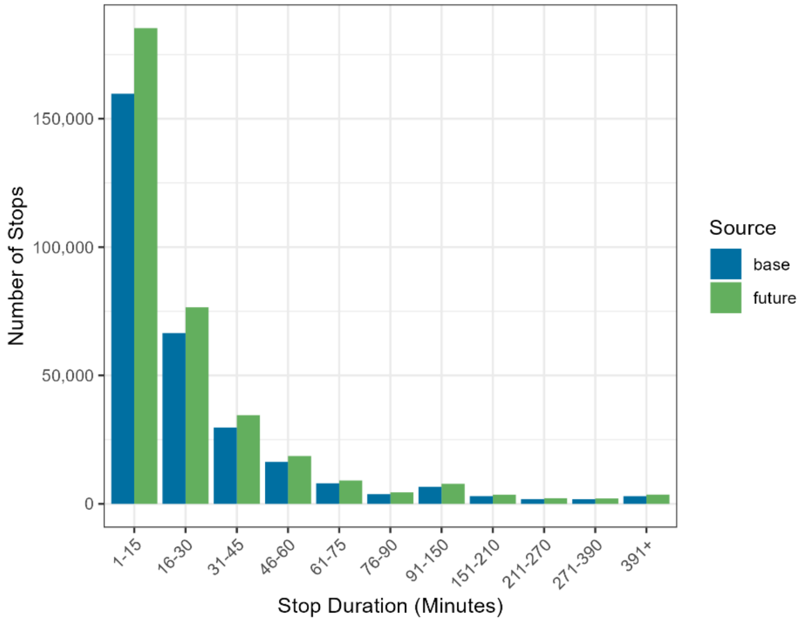
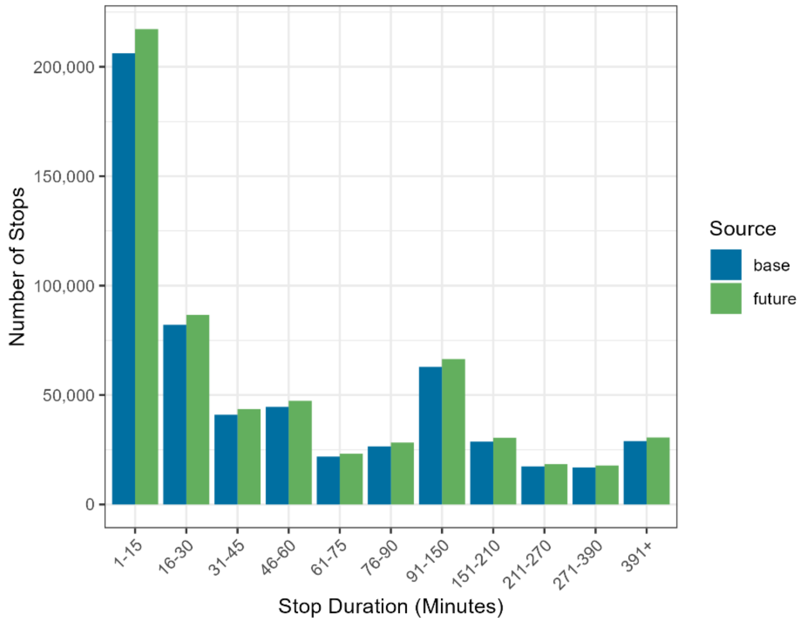


Figure 84: DISTRIBUTION OF STOP DURATION FOR 2050 (SERVICES)



Simulate Tours and Routing

Figure 85 through Figure 87 show the breakdown of tour types by light, medium, and heavy vehicles, respectively. For light and medium vehicles, we see the largest increase in assigned tour type in the Base-to-base single-stop tours followed by base-to-base multi-stop tours. Conversely, for heavy vehicles we saw a larger increase in base-to-base multi-stop tours followed by base-to-base single-stop tours though the magnitude of these increases was smaller than both light and medium vehicles.

Figure 85: TOUR TYPOLOGY FOR 2050 (LIGHT VEHICLES)

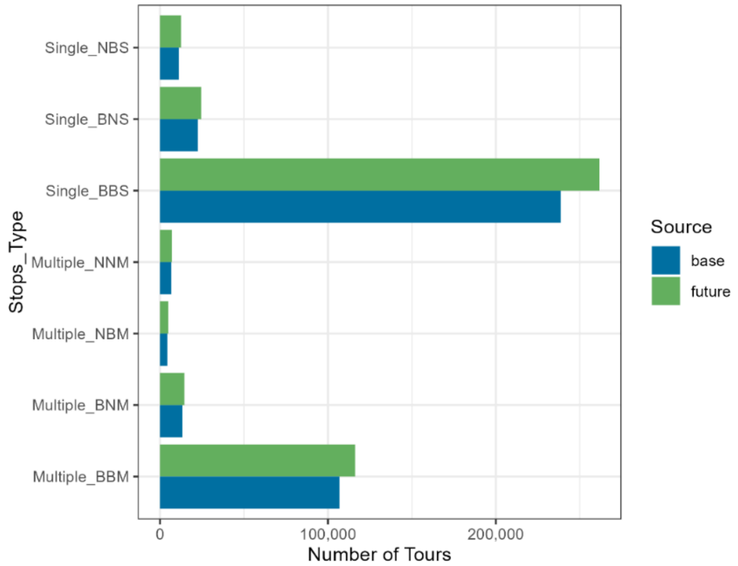


Figure 86: TOUR TYPOLOGY FOR 2050 (MEDIUM VEHICLES)

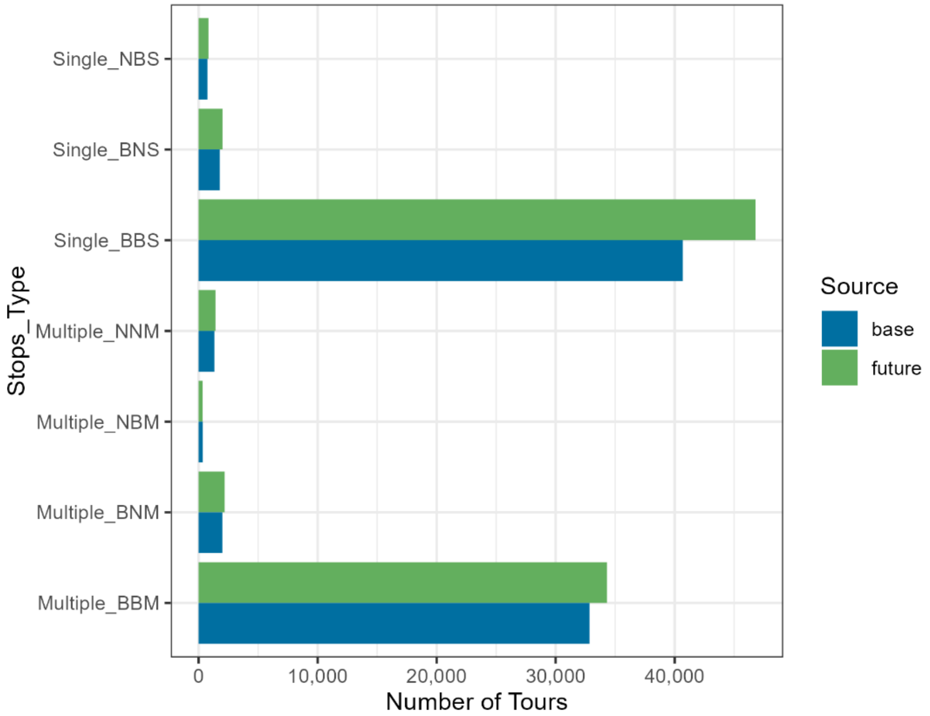
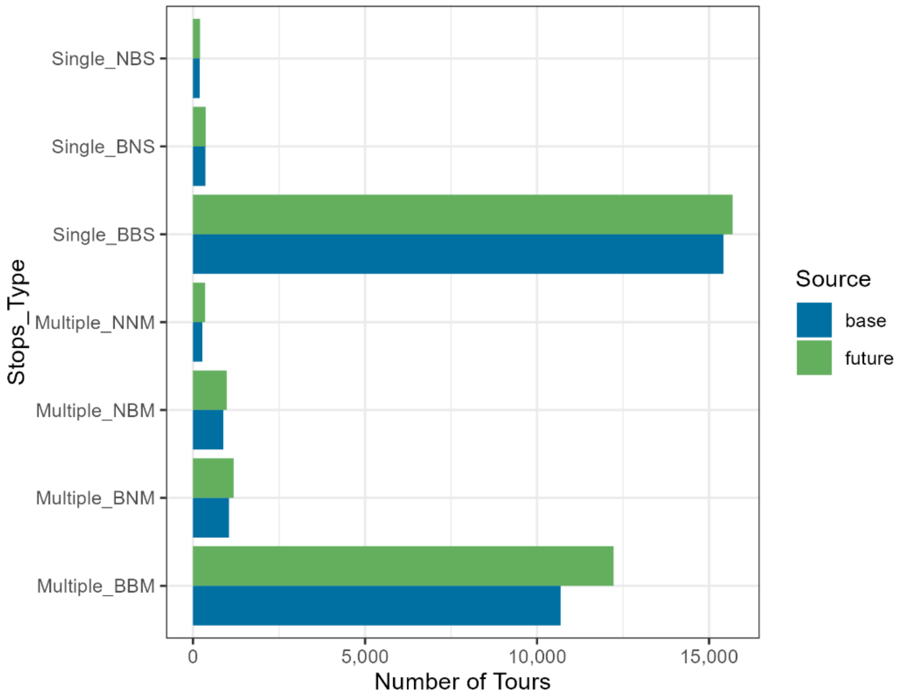


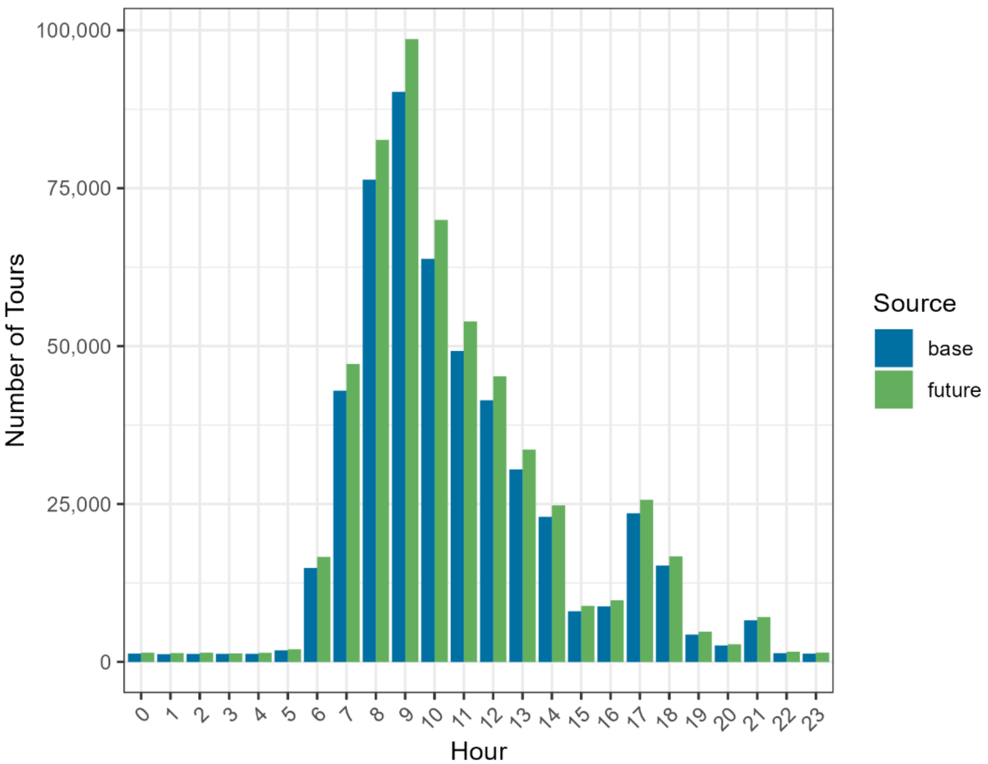
Figure 87: TOUR TYPOLOGY FOR 2050 (HEAVY VEHICLES)



Simulate Scheduled Trips

Figure 88 shows the distribution of tour start times in the future year (2050) compared against the base. The assignment of tour start times for the year 2050 appears proportionate to the rates at which they were assigned in the base year. The hours between 8AM and 11AM saw the highest level of activity.

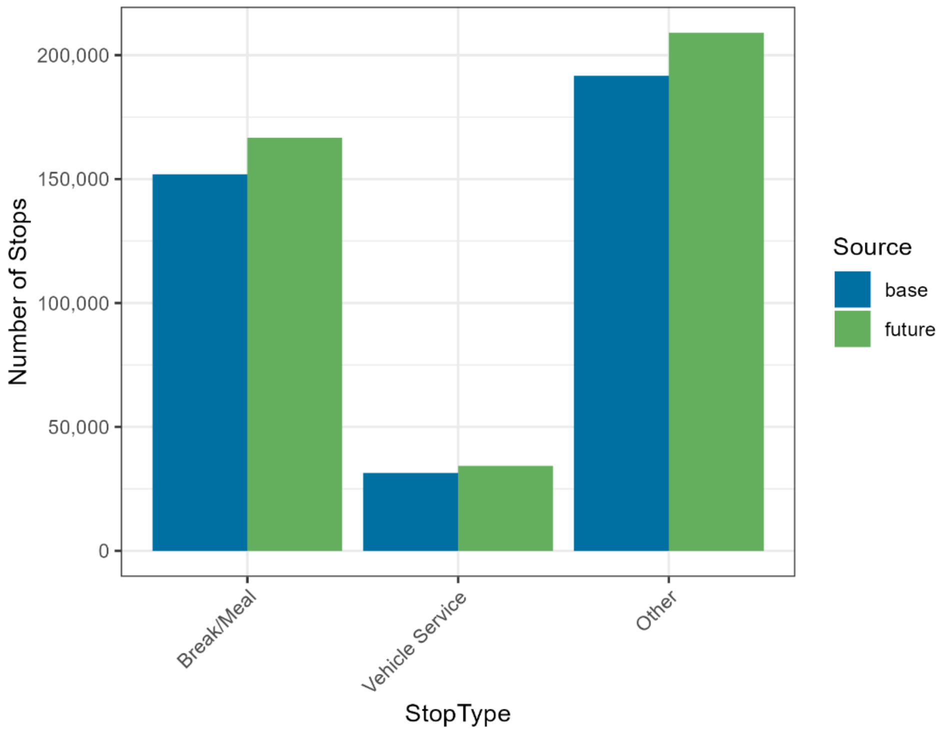
Figure 88: DISTRIBUTION OF START TIME FOR 2050



Simulate Intermediate Stops

Intermediate stops generated for the future year scenario are compared against the base year in Figure 89. The results from this step of the model also show a proportional increase in all categories.

Figure 89: INTERMEDIATE STOPS FOR 2050



The validation review of the future year in comparison to the base for each model component of the CSVM demonstrates stability in result and reasonably proportionate changes given the level of changes in the input data between the base and future scenarios.

4.7 Scenario Tests

CMAP tested the CSVM application and ran several sets of sensitivity tests to evaluate the response of model outputs to changes in input data and parameters. The descriptions of the sensitivity testing done is summarized in Figure 90Figure 90

Figure 90: SENSITIVITY TEST DEFINITIONS

Scenario	Definition
Home deliveries replacing shopping trips	To simulate a shift towards fewer shopping trips and more deliveries, CMAP asserted a shift towards additional goods stops by increasing the *asc_goods_adj* parameter in the scheduled stops model
Shift towards smaller vehicles	To simulate a shift towards smaller vehicles and smaller loads, CMAP asserted a shift towards lighter vehicles by increasing the *asc_vehcle_light_adj* or *asc_vehicle_medium_adj* parameters in the vehicle choice model
Warehouse and distribution centers closer to consumers	To simulate additional warehouses closer to consumers, CMAP asserted a shorter delivery shed via adjustments to the base to distance sampling parameter in the scheduled stops model
Higher Fuel Prices	To simulate changes to fuel taxes, CMAP asserted a higher travel cost via adjustments to the distance and travel time impedance variables in the scheduled stops model

More Deliveries

CMAP performed a test of model sensitivity to adjustment in the *asc_goods_adj* to model an expected shift towards delivery trips and away from shopping trips. In the 2019 scenario, we observe that in response to increases of 0.2 to *asc_goods_adj*, service trips remain constant while trips and VMT increase at a steadily increasing rate (see Figure 91). As would be expected, as the model produces additional goods trips, we see a corresponding increase in the share of trips assigned to medium and heavy vehicles (see Figure 92).

Figure 91: MORE DELIVERIES – STOPS & VMT (BASE YEAR)

	Scenario Adjustment		Base Year Response				
Test ID	asc_goods_adj	asc_service_adj	Goods Deliveries	Additional Goods Trips	Service Stops	Total VMT	Additional VMT
base	0	0	298,975	-	565,611	20,914,454	-
deliv1	**0.1**	0	322,262	23,287	565,611	21,380,235	465,781
deliv3	**0.3**	0	374,478	52,216	565,611	22,389,178	1,008,943
deliv5	**0.5**	0	433,528	59,050	565,611	23,577,572	1,188,394
deliv7	**0.7**	0	500,218	66,690	565,611	24,910,693	1,333,121
deliv9	**0.9**	0	575,745	75,527	565,611	26,449,846	1,539,153

Figure 92: MORE DELIVERIES – VEHICLE SHARE (BASE YEAR)

	Scenario Adjustment		Base Year Response		
Test ID	asc_goods_adj	asc_service_adj	Light Vehicle Share	Medium Vehicle Share	Heavy Vehicle Share
base	0	0	77.70%	16.20%	6.10%
deliv1	**0.1**	0	77.50%	16.50%	5.90%
deliv3	**0.3**	0	76.40%	17.10%	6.60%
deliv5	**0.5**	0	75.70%	17.70%	6.60%
deliv7	**0.7**	0	74.80%	18.00%	7.20%
deliv9	**0.9**	0	74.10%	18.40%	7.40%

A similar sensitivity test was performed for the future year with increases to *asc_goods_adj* of 0.1 to ensure the model behaved like the base year. A similar pattern is observed with respect to trip generation, vehicle assignment, and VMT (see Figure 93 & Figure 94).

Figure 93: MORE DELIVERIES - STOPS & VMT (FUTURE YEAR)

	Scenario Adjustment		Future Year Response				
Test ID	asc_goods_adj	asc_service_adj	Goods Deliveries	Additional Goods Trips	Service Stops	Total VMT	Additional VMT
future	0	0	334,422	-	593,911	21,632,796	-
Fdeliv1	**0.1**	0	360,232	25,810	593,911	22,092,616	459,820
Fdeliv2	**0.2**	0	388,042	27,810	593,911	22,655,331	562,715
Fdeliv3	**0.3**	0	417,852	29,810	593,911	23,197,993	542,662
Fdeliv4	**0.4**	0	449,142	31,290	593,911	23,842,504	644,511

Figure 94: MORE DELIVERIES - VEHICLE SHARE (FUTURE YEAR)

	Scenario Adjustment		Future Year Response		
Test ID	asc_goods_adj	asc_service_adj	Light Vehicle Share	Medium Vehicle Share	Heavy Vehicle Share
future	0	0	77.70%	16.50%	5.80%
Fdeliv1	**0.1**	0	77.30%	16.60%	6.10%
Fdeliv2	**0.2**	0	77.10%	16.70%	6.20%
Fdeliv3	**0.3**	0	76.60%	17.10%	6.30%
Fdeliv4	**0.4**	0	76.10%	17.40%	6.50%

Lighter Vehicles

CMAP performed a test of model sensitivity to adjustment in the *asc_vehicle_light_adj* and *asc_vehicle_medium_adj* to model an expected shift towards lighter vehicles and smaller loads. By adjusting *asc_vehicle_light_adj* and *asc_vehicle_medium_adj* independently, we observe the marginal effects of each parameter. As the medium vehicle specific constant is increased, vehicle choice shifts towards medium vehicles and away from light and heavy vehicles in a predictable manner, the same is true for the light vehicle specific constant. As expected, as vehicle share is shifted towards light vehicles, additional trips are generated. On the other hand, as vehicle choice is shifted towards medium vehicles, the number of trips decreases. Furthermore, changes to the constants seem to have a proportional effect as evidenced by similar shifts in vehicle share across the base year and future year.

Figure 95: LIGHTER VEHICLES - VEHICLE SHARE (BASE YEARS)

	Scenario Adjustments			Base Year Response		
Test ID	**asc_vehicle_light_adj**	**asc_vehicle_medium_adj**	**asc_vehicle_heavy_adj**	**Light Vehicle Trips**	**Medium Vehicle Trips**	**Heavy Vehicle Trips**
base	0	0	0	78%	16%	6%
t1	**0.1**	0	0	79%	15%	5%
t5	**0.5**	0	0	84%	12%	4%
t9	**0.9**	0	0	88%	9%	3%
m1	0	**0.1**	0	77%	18%	6%
m5	0	**0.5**	0	72%	23%	5%
m9	0	**0.9**	0	66%	30%	4%

Figure 96: LIGHTER VEHICLES - TRIPS & VMT (BASE YEAR)

	Scenario Adjustments			Base Year Response	
Test ID	**asc_vehicle_light_adj**	**asc_vehicle_medium_adj**	**asc_vehicle_heavy_adj**	**Total Trips**	**Total VMT**
base	0	0	0	1,743,286	20,639,295
t1	**0.1**	0	0	1,747,990	20,684,348
t5	**0.5**	0	0	1,765,975	20,807,978
t9	**0.9**	0	0	1,782,599	20,906,703
m1	0	**0.1**	0	1,739,700	20,616,112
m5	0	**0.5**	0	1,721,585	20,496,337
m9	0	**0.9**	0	1,700,475	20,343,073

Figure 97: Lighter Vehicles - Vehicle Share (Future Year)

	Scenario Adjustments			Future Year Response		
Test ID	**asc_vehicle_light_adj**	**asc_vehicle_medium_adj**	**asc_vehicle_heavy_adj**	**Light Vehicle Trips**	**Medium Vehicle Trips**	**Heavy Vehicle Trips**
future	0	0	0	78%	16%	6%
t1	**0.1**	0	0	79%	16%	5%
t5	**0.5**	0	0	84%	12%	4%
t9	**0.9**	0	0	88%	9%	3%
m1	0	**0.1**	0	77%	18%	6%
m5	0	**0.5**	0	72%	23%	5%
m9	0	**0.9**	0	66%	30%	4%

Figure 98: LIGHTER VEHICLES - TRIPS & VMT (FUTURE YEAR)

	Scenario Adjustments			Future Year Response	
Test ID	**asc_vehicle_light_adj**	**asc_vehicle_medium_adj**	**asc_vehicle_heavy_adj**	**Total Trips**	**Total VMT**
future	0	0	0	1,873,430	21,611,417
t1	**0.1**	0	0	1,878,223	21,657,133
t5	**0.5**	0	0	1,898,034	21,795,270
t9	**0.9**	0	0	1,915,341	21,896,556
m1	0	**0.1**	0	1,868,054	21,586,986
m5	0	**0.5**	0	1,849,135	21,444,526
m9	0	**0.9**	0	1,827,091	21,309,109

Stop Proximity

Interested in modeling the location of warehouses closer to consumers, CMAP performed a test on the sensitivity of base to stop distances to the adjustment of the *base_dist* factor for either goods or services. At the moment, adjustments to this factor alone produce a change in the total number of generated trips in addition to a change in the distribution of base to stop distances. Therefore, the goods and service constants were also adjusted to offset this change. Figure 99 shows the parameter adjustments for this test. Figure 100 and Figure 101 show how the number of stops changed in response. In all cases, distance factors were adjusted by .1 and goods and service constants were adjusted at 0.1 intervals to explore the balance between these two determinants of stop generation in the model.

Figure 99: Stop Proximity Test Specification

Test ID	base_dist_goods_factor	base_dist_service_factor	asc_goods_adj	asc_service_adj
Base/Future	1	1	0	0
1a	**0.1**	1	**-0.3**	**-0.3**
1b	**0.1**	1	**-0.4**	**-0.2**
5a	1	**0.1**	**-0.3**	**-0.3**
5b	1	**0.1**	**-0.2**	**-0.4**

Figure 100 and Figure 101 show the effect of the two kinds of scenario adjustments on the number of generated stops for the base year and the future year.

Figure 100: Stop Generation Effect (Base Year)

Test ID	Goods Deliveries	Change (Goods)	Service Stops	Change (Services)	Total Stops
Base	300,046	-	566,988	-	1,747,178
1a	307,603	3%	559,161	-1%	1,752,775
1b	285,219	-5%	598,506	6%	1,795,779
5a	295,836	-1%	606,764	7%	1,830,654
5b	319,242	6%	567,577	0%	1,789,937

Figure 101: Stop Generation Effect (Future Year)

Test ID	Goods Deliveries	Change (Goods)	Service Stops	Change (Services)	Total Stops
Future	337,062		594,584		1,878,114
1a	345,864	3%	590,447	-1%	1,890,800
1b	321,116	-5%	632,120	6%	1,937,185
5a	333,259	-1%	639,622	8%	1,985,427
5b	358,908	6%	597,876	1%	1,940,237

The resulting change in average base-to-stop distance (i.e., service area or delivery shed), the relevant statistics with respect to modeling more proximal warehouses and distribution centers, can be found in Figure 102 and Figure 103.

Figure 102: Service Shed Effect (Base Year)

Test ID	Base-Stop Distance (Goods)	Change (Goods)	Base-Stop Distance (Services)	Change (Services)	Mean Distance
Base	14.2		13.6		12
1a	10.9	-23%	9.8	-28%	10
1b	10.8	-24%	10	-26%	10
5a	12.9	-9%	11.3	-17%	10
5b	12.3	-13%	11.8	-13%	10

Figure 103: Service Shed Effect (Future Year)

Test ID	Base-Stop Distance (Goods)	Change (Goods)	Base-Stop Distance (Services)	Change (Services)	Mean Distance
Future	13.6		12.9		12
1a	10.7	-21%	9.2	-29%	10
1b	10.6	-22%	9.4	-27%	10
5a	11.9	-13%	10.9	-16%	9
5b	12.1	-11%	10.8	-16%	9

Figure 104 and Figure 105 show how changes in service shed and total trips translate into travel time and total VMT generated by the model in the base year and the future year.

Figure 104: VMT & Travel Time Effect (Base Year)

Test ID	Mean Travel Time	Change in Travel Time	Total VMT	Change in VMT
Base	26		20,692,182	
1a	23	-12%	17,435,881	-16%
1b	23	-12%	18,095,694	-13%
5a	22	-15%	17,632,963	-15%
5b	22	-15%	17,070,696	-18%

Figure 105: VMT & Travel Time Effect (Future Year)

Test ID	Mean Travel Time	Change in Travel Time	Total VMT	Change in VMT
Future	28		21,672,448	
1a	24	-14%	18,299,007	-16%
1b	24	-14%	18,971,485	-12%
5a	23	-18%	18,595,221	-14%
5b	23	-18%	18,014,733	-17%

While the model is sensitive to the distance parameters and the model also demonstrated the ability to balance out generated stop numbers with the effects of the distance reduction, the model specification (with a combined set of TAZs generated for consideration as stop locations for each business), the model is not currently demonstrating the ability to separately modify the stop distance distribution by service and goods stops. The scheduled stop model would require additional segmentation to separate these effects, adding segmentation by activity to the part of the model step that draws a sample of TAZs for consideration as stop locations.

Travel Cost Changes

Interested in modeling changes to travel costs in the region, CMAP performed a test of the sensitivity of the model with respect to adjustment in the travel impedance factor for goods and service stops. The goods and service impedance factors were adjusted independently of one another at increments of 0.05. Figure 106 and Figure 107 show how the .05 increments translate into fewer or additional stops. In the base year, impedance is reduced for goods stops and increased for service stops. In the future year, an increase in impedance for service stops and a decrease in impedance for goods stops are tested. In general, an increase or decrease in impedance led to somewhat similar increases or decreases in the total number of trips. With regard to the statistics of interest, increases in impedance had larger downward effects on VMT, service shed, and travel time than did a decrease in impedance in the opposite direction.

Figure 106: Stop Generation Effect (Base Year)

Test ID	impedance_goods_factor	impedance_service_factor	Goods Deliveries	Change (Goods)	Service Stops	Change (Services)	Total
Base	1	1	300,046		566,988		1,747,178
2a	**0.9**	1	427,282	42%	566,988	0%	1,960,605
2b	**0.95**	1	357,062	19%	566,988	0%	1,841,578
6c	1	**1.05**	300,046	0%	444,321	-22%	1,490,015
6d	1	**1.1**	300,046	0%	348,976	-38%	1,291,524

Figure 107: Stop Generation Effect (Future Year)

Test ID	impedence_goods_factor	impedence_service_factor	Goods Deliveries	Change (Goods)	Service Stops	Change (Services)	Total
Future	1	1	337,062		594,584		1,878,114
2c	**1.05**	1	285,514	-15%	594,584	0%	1,791,777
2d	**1.1**	1	242,927	-28%	594,584	0%	1,720,525
6a	1	**0.9**	337,062	0%	971,999	63%	2,669,739
6b	1	**0.95**	337,062	0%	760,477	28%	2,225,301

Figure 108 and Figure 109 show how changes to the impedance factors translate into changes in the average service shed. Increases in impedance lead to more significant decreases in average service shed while decreases in impedance lead to more measured increases in service shed.

Figure 108: Service Shed Effect (Base Year)

Test ID	impedence_goods_factor	impedence_service_factor	Base-to-Stop (Goods)	Change (Goods)	Base-to-Stop (Services)	Change (Services)
Base	1	1	14.2		13.6	
2a	**0.9**	1	15.9	12%	13.4	-1%
2b	**0.95**	1	14.8	4%	13.4	-1%
6c	1	**1.05**	13.8	-3%	12.2	-10%
6d	1	**1.1**	13.8	-3%	10.9	-20%

Figure 109: Service Shed Effect (Future Year)

Test ID	impedence_goods_factor	impedence_service_factor	Base-to-Stop (Goods)	Change (Goods)	Base-to-Stop (Services)	Change (Services)
Future	1	1	13.6		12.9	
2c	**1.05**	1	12.5	-8%	12.9	0%
2d	**1.1**	1	11.6	-15%	12.9	0%
6a	1	**0.9**	13.5	-1%	15.7	22%
6b	1	**0.95**	13.6	0%	14.3	11%

Figure 110 and Figure 111 show how changes to either the goods or service impedance factor affect travel times and overall VMT in the model. Like average service sheds, we observe that an increase in impedance results in more dramatic decreases in VMT while an equal decrease in impedance results in more measured increase in VMT.

Figure 110: VMT & Travel Time Effect (Base Year)

Test ID	impedence_goods_factor	impedence_service_factor	Mean Travel Time	Travel Time Change	Total VMT	VMT Change
Base	1	1	26		20,692,182	
2a	**0.9**	1	27	4%	23,764,933	15%
2b	**0.95**	1	26	0%	22,000,637	6%
6c	1	**1.05**	25	-4%	16,787,816	-19%
6d	1	**1.1**	24	-8%	13,930,028	-33%

Figure 111: VMT & Travel Time Effect (Future Year)

Test ID	impedence_goods_factor	impedence_service_factor	Mean Travel Time	Travel Time Change	Total VMT	VMT Change
Future	1	1	28		21,672,448	
2c	**1.05**	1	27	-4%	20,502,483	-5%
2d	**1.1**	1	27	-4%	19,577,419	-10%
6a	1	**0.9**	30	7%	34,608,008	60%
6b	1	**0.95**	29	4%	27,152,650	25%

5. MODEL MAINTENANCE, FUTURE DATA COLLECTION, AND MODEL ENHANCEMENTS

This section of the report presents recommendations for model maintenance, future data collection, and ideas to enhance the model moving forward. These recommendations include future survey efforts, ideas for representation of changes to vehicle technology, and other additional policy sensitivities, and for model system integration, and suggestions for data collection or data processing needs andresolving weaknesses in the model design.

5.1 Model Maintenance

There are two aspects of model maintenance to consider. The first is related to software, where it might be necessary or preferred to update the version of R used with the model to align the model system with a more current version of R and one that was being used within the agency for other purposes.

The second is a typical model base year update where, in the usual planning cycle, the base year is moved forward several years. In this case the CSVM is currently using a 2019 base year, which is the last full year prior to the COVID 19 pandemic. A move forward could be done in the relatively short term to current post pandemic travel conditions (say 2022) or, in a few years time to a more current year such as 2025 as part of a more comprehensive travel model update.

Software Versioning

The CSVM is implemented and tested in R version 4.1.2. This version was released in November 2021. The current version of R is version 4.2.2 which was released in October 2022. The type and number of changes in version is relatively small for the final increment (i.e., 4.1.2 to 4.1.3). There were some changes between R 4.1.3 and R 4.2.0 (including certain changes to how numerical vectors are handled) that may be sufficient to require code changes in the rFreight package and in the CSVM scripts, although that has not been fully researched by RSG at the time of writing of this report.

To update the version of R used for the model, the following general steps are necessary:

- Update the version of R installed on the computer(s) used to run the CSVM.
- Obtain a rebuilt version of rFreight for the new version of R. R packages are built to match with specific version of R. rFreight is currently working with up to R v4.1.3, with planned updates to the current version of R, version 4.2.2 in early 2023. While RSG does not release rFreight via CRAN, it does attempt to maintain the package in line with R version updates once any major changes in R are known to be stable.
- Reinstall the model and run it to install the new version of rFreight and updates to the other packages used by the model.
- Test the model to identify any issues introduced by the new version of R and new package versions. Even an error free run could be problematic, so results should be compared with a run with identical inputs and settings of the CSVM implemented in R version 4.1.2.
- It is possible that code updates may be necessary if the implementation of certain functions and other features of R have changed between versions. The likelihood of this occurring increases with time as software version changes accumulate.

Base year update

A base year update can typically involve changes to the spatial system in the model (zones, subzones), the socio-economic data (both in terms of numbers of jobs and households due to the base year change but also possibly to format, categorization and other similar changes), and network data (changes in skims values and also potentially due to model design changes such as the time periods that are modeled). There are several considerations:

- The current base year is 2019. A simple update where the only changes were to land use and skims without any TAZ system changes would require updated input data (employment and households), potentially new CBP data, and updated network skims. The CBP data used in the model is from 2017, and while it is scaled to match the employment control data, the data should probably be updated once the base year is further removed from 2017.
- If the base year update includes an update to the TAZ system, an update to input files that include TAZ numbering (and other associated geographies such as mesozones and associations between subzones and TAZs and TAZs and counties) would be required. If the level of detail in the TAZ system is changed considerably, then certain aspects of the model including the sampling approach in the stop generation model, would need to be adjusted in order to ensure that the model was still producing reasonable results.
- If the level of changes caused by the shift in model base year are sufficient, recalibration would be necessary. This requires running the calibration scripts. If more up to date calibration sources are available, then the target data used in calibration could also be updated using the target data processing scripts as a starting point.
- Revalidation of the model would require updating the validation data (including updated FHWA VMT data, and an updated source for local tour making characteristics such as updated local GPS data)

The processing scripts for data and the model calibration scripts (provided to CMAP in the “dev” folder of the CSVM), and the model dashboards to create comparisons with validation data can all support efforts to move the model base year forward.

5.2 Future Data Collection

This project benefitted from new locally collected data in the form of the establishment survey, a sample of truck diaries (although the ultimate sample size of the truck diary data was relatively small) and truck GPS data for the CMAP region. These data were supplemented by transferred data from the SEMCOG region where a larger truck diary survey was completed in 2017, as well as publicly available data from FHWA, the US Census Bureau, and the USDA.

For future survey and data collection efforts that would benefit the CSVM, there are three main avenues that CMAP could consider: a further round of truck diary data collection, updating and expanding truck GPS data, and additional truck counts.

Truck diary surveys

Collecting additional truck diary surveys in the CMAP region would have the benefit of collecting detailed data that could be directly used to re-estimate and recalibrate the model components but (as found in this survey effort) would likely require significant resources to collect sufficient data. Before considering this, it is likely that CMAP would want to revisit the data collection carried out during this project and review efforts by other agencies to look at budget requirements and methods for achieving higher sample sizes.

Expanded truck GPS data

The INRIX data used in the calibration and validation of the CSVM on this project were based on a sample of primarily medium trucks with some light commercial vehicles. The data were processed by CMAP but have not been expanded to match with classified truck counts. Further work on updated passive GPS data from various sources (INRIX, ATRI) could be done to create vehicle class specific truck trip tables that were expanded to match with classified truck counts.

For light commercial vehicles, techniques are available to take large datasets of location-based service (LBS) data and identify light commercial vehicle trips and separate them out from the predominant passenger trips. The LBS data space is currently changing quickly as vendors adjust to new privacy requirements and so precisely what data are available and what can be done with it is somewhat in flux. However, vehicle class specific and expanded datasets would be a valuable resources for more detailed model estimation, calibration, and validation.

Classified trucks counts and passenger/commercial splits

A robust set of classified truck counts with good representation of different functional classes and geographies are necessary to validate any network assignments done with the trip tables produced by the CSVM (and also the freight truck touring model component of the CMAP freight model).

One detail in the counts that is not currently available is an attempt to distinguish between commercial trips and passenger travel in the light vehicle counts. This has been attempted in some regions by carrying out at a subset of count location an observational survey (either during the count or by processing videos of the count location) to identify vehicles that appear to be commercial vehicles (for example, based on truck body configuration, vehicle logos, the materials carried in open pick up trucks, etc.). This can support the separate validation of light commercial vehicles and light passenger vehicles.

5.3 Future Model Enhancements

There are several themes that future model enhancements can be grouped into:

- Better connections between the CSVM and other parts of the CMAP modeling system (integration) to improve the sensitivity and feedback of all of the models
- Better representation of current truck travel behavior in the CSVM by adding detail (such as additional segmentation) and improving the specification of models
- Better policy sensitivity and ability to represent changes over time by adding detail to the models to include new and emerging aspects of the commercial vehicle model sector and by including additional variables and features to model new policies.

Model Integration

The model is currently a stand-alone application that is run with inputs that include congested travel time and distance skims from the CMAP travel demand model. The model produces a set of trip tables that can then be assigned to the CMAP highway network along with trip tables from other classes of vehicles.

In order to simplify workflow and to allow for feedback of the impacts of congestion on demand to all classes of vehicle, the CSVM could be integrated into the overall model system so that it runs along with the other demand models during each system iteration of the CMAP travel demand model.

The CSVM is designed to run in this way, with the ability to run the firm synthesis, CVTM, trip tables, and dashboard steps separately. Typically, the firm synthesis model would be run only once during the first iteration, the CVTM and trip tables steps would be run in each iteration, and the dashboard would only be run in the final iteration.

The model can be called from system commands with arguments (e.g., the scenario to run and the model steps to run) passed to the CSVM's batch file. The skims and trip tables can be passed to and from the CSVM in OMX format, which EMME can read and write and is also the format used by the ActivitySim activity based model framework.

The file structure of the CSVM is flexible and described in a set of system variables. It can be adjusted to match with the file structure of the overall travel demand model system. This allows, for example, the scenario input and output structure used by the CSVM to be fully integrated with the travel demand model's structure to simplify model set up and avoid issues such as duplicated copies of files. The trip tables function in the CSVM can be adjusted to directly write into a centralized set of trip table files, if this is desired.

Representation of current truck behavior

While the current version of the CSVM attempts to capture nuances and details of the characteristics of the travel of the commercial vehicle fleet, there are clearly limitations to what we understand based on the available data and research on the motivations of businesses, commercial vehicle drivers, and other decision makers.

There are several approaches to consider for improving how the model captures these characteristics (in addition to the data collection suggestions mentioned earlier):

- The firm synthesis model currently deals with public administration employment in a simple manner. It is missing from the CBP data and so the firm synthesis model simulates an appropriate set of establishments based on the TAZ employment. The current method is correct in terms of the number of employees by TAZ, but essentially treats all public employment as the same. So, for example, office based administrative government employment and public works employment have very different truck and commercial vehicle operating characteristics, but this is not reflected in the model. A more detailed treatment of government employment would involve identifying different classes of employment (and in particular those that involve commercial vehicle use) and sizing and locating them appropriately. At its most detailed this might involve collecting data on each local jurisdiction and its public works and other commercial vehicle intensive activities.

- The segmentation of the stop generation model is currently by service and goods activity purposes. For each business that operates commercial vehicles, it generate a set of stops based on the proximity of each TAZ and its land use. In the current CSVM and based on the estimation data, this tends to emphasize the impacts of TAZ employment by industry and the response to residential development in each TAZ is simplistic. An improvement would be to separately generate stops for deliveries and service calls at business establishments and at residences. This might allow for additional detail and policy sensitivity in the CSVM, such as the inclusion of additional household related variables (e.g., average TAZ household income) and built form variables (e.g., the proportions of different types of housing units such as multi-family buildings and single family homes).
- The vehicle type model is currently limited to three broad vehicle classes (light commercial vehicles which are FHWA class 3 or smaller and include all two axles four type vehicle, medium trucks which are FHWA class 5 to 7 and includes all single unit trucks from two axles size tyre pick up trucks to large 3 and4 four axle straight trucks, and heavy trucks which are FHWA class 8 and larger truck and includes all tractor trailer multi unit combinations). Within those broad classes, the vehicle choice model could be enhanced to associate businesses and the stops that their vehicles make with more detailed vehicles classes (for example, splitting light commercial vehicles into cars, SUVs, pick up trucks, and small box vans). This model could also be enhanced, or a separate component added to add the vehicle powertrains (e.g., internal combustion, hybrid, battery electric) to support the estimation of fuel usage and emissions, to more accurately represent costs faced for vehicle operations, and to support other analyses where travel by different power train types is useful.
- The current model design allows for difference in tour patterns by vehicle class, by activities on the tours, and by industry operating the vehicle. There are additional factors that may be important and introducing additional segmentation into the model may allow for a more detailed representation of tour types and other tour characteristics. This could include:
    - Segments for route-based operations (e.g., garbage and recycling collection, parcel and postal delivery, public works street cleaning and snow plowing)
    - Segmentation of construction related trips/establishments (since it is the largest segment) into trips to existing residential development, to existing business development, to new build residential, to new build commercial, and to public infrastructure construction. Each of these categories likely have different attraction rates, time of day profiles and other differences.
    - Different time of day constraints for different segments to capture different typical hours of business operation. For example, construction tends to skew early particularly for work on new development sites, while home goods deliveries tend to be later conforming to the standard work day, and food service delivery for ready to eat meals to homes skews later with peaks around lunch and then in the late afternoon and evening.
    - Seasonality is important for some industries, e.g., construction (with much less new build and some types of services to existing home stopped or close to stopped in winter) and retail delivery (with a peak in the fall approaching holidays).

Policy sensitivity and emerging changes to commercial vehicle travel

In addition to a more detailed representation of the current system of commercial vehicle travel, there are limitations to the policy sensitivity of the CSVM. There are also emerging changes to the behavior of businesses and consumers as well as emerging technologies that CMAP may need to evaluate and understand in their travel modeling and planning. These issues include:

- Sensitivity to cost changes. The CSVM does not explicitly represent costs of travel (e.g., vehicle operating costs such as fuel costs) in the decision-making process. This means that while the model's sensitivity to costs can be represented in a very simple way by adjusting the response to travel impedance, the model is not designed to support detailed modeling of cost related policy changes. Enhancing the model to do this would require two tasks. The first would be to gather research on empirical data that captures actual observed responses of commercial vehicle operators across different industries to vehicle operating cost changes, and the second would be to update the model design to incorporate that sensitivity.
- In addition to vehicle operating costs changes, the commercial vehicle trip demand generation in the CSVM has limited sensitivity to pricing policies such as VMT pricing and road pricing. The specification of models within the CSVM would need to be updated to incorporate this. Route choice and other detailed responses to pricing on specific routes is dependent on the model system's assignment model.
- Alternative modes to commercial vehicles for deliveries are being evaluated and could be included as alternatives in the CSVM. These mode include drones for delivery of small items such as meals and small packages.
- A segment of commercial vehicles that is rapidly growing is app based delivery such as UberEats, Amazon Prime, and Instacart. This is typically a flexible fleet of contract drivers who use personal vehicles for local delivery for items such as groceries, small packages and restaurant take out. Certain regions (such as San Diego) are beginning to collect survey data on the behavior of these fleets. The CSVM could be enhanced to represent app delivery as an option for certain types of stops and for certain industries.
- Interaction between passenger trip generation and commercial vehicle trip generation is an issue with changing consumer behavior. For example, increasing consumer use of e-commerce leads to increases in home delivery and potentially reduces the number of home-based shopping trips a household makes. This may reduce household travel or allow for substitution for travel with a different purpose such as leisure or recreational travel. The CSVM could be enhanced as part of broader model system changes to allow for the representation of interactions between currently separate demand models.