U.S. Department of Transportation
Office of the Secretary of Transportation

Bureau of Transportation Statistics

Freight Analysis Framework Version 5 (FAF5) Experimental County-Level Estimates Technical Report

About the Bureau of Transportation Statistics

Leadership

Patricia Hu, *Director*, *Office of Director*
Rolf R. Schmitt, *Deputy Director*, *Office of Director*

Publication Management

Stephanie Lawrence, *Director*, *Office of Statistical and Economic Analysis*

About This Report

Project Manager

Monique Stinson, *Freight Estimation, Forecasting, and Analysis Manager*, *Office of Statistical and Economic Analysis*

Major Contributors

Jingsi Shaw and Praveen Hettige

Editor

Mikki Stacey

Visual Information Specialists

Praveen Hettige, Jingsi Shaw, Monique Stinson, and Alpha Wingfield

Other Contributors

Hyeonsup Lim, Laura Dods, and Ed Strocko

Report DOI

https://doi.org/10.21949/b16k-zb09

Title

Freight Analysis Framework Version 5 (FAF5) Experimental County-Level Estimates: Technical Report

Publication Date

January 2025

Performing Organization

Bureau of Transportation Statistics
1200 New Jersey Avenue, SE
Washington, DC 20590

Key Words

Freight Analysis Framework; version 5; FAF5; county; estimates

Abstract

The Freight Analysis Framework (FAF) database provides estimates of the weight and value of shipments throughout the United States for all commodity types and forms of transportation using a geographic system of 132 FAF zones. The user community has expressed a need for more geographically granular commodity flow data. In response to this need, BTS developed an experimental county-to-county commodity flow product using publicly available data and transparent methods. The development involves creating three estimates of county-level flows using existing disaggregation methods, constructing validation targets, then using these inputs to create a composite estimate of county-to-county flows. This technical report documents the data, methods, uses, and limitations of this release. BTS welcomes users to email FAF@dot.gov with feedback on this experimental product.

Recommended Citation

United States Department of Transportation, Bureau of Transportation Statistics. *Freight Analysis Framework Version 5 (FAF5) Experimental County-Level Estimates: Technical Report*. Washington, DC: 2025. https://doi.org/10.21949/b16k-zb09.

BTS information service contact information:

Ask-A-Librarian https://transportation.libanswers.com/

Phone 202-366-DATA (3282)

Quality Assurance Statement
The Bureau of Transportation Statistics (BTS) provides high quality information to serve government, industry, and the public in a manner that promotes public understanding. Standards and policies are used to ensure and maximize the quality, objectivity, utility, and integrity of its information. BTS reviews quality issues on a regular basis and adjusts its programs and processes to ensure continuous quality improvement.

Acknowledgments

This research was supported, in part, by an appointment to the U.S. Department of Transportation (USDOT) Research Participation Program administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy (DOE) and USDOT. ORISE is managed by Oak Ridge Associated Universities (ORAU) under DOE contract number DE-SC0014664. All opinions expressed in this paper are the authors' and do not necessarily reflect the policies and views of USDOT, DOE, or ORAU/ORISE.

Table of Contents

List of Figures

List of Tables

Executive Summary

The Freight Analysis Framework (FAF) database provides estimates of the weights and values of shipments throughout the United States for all commodity types and forms of transportation using a geographic system of 132 FAF zones [BTS, FHWA 2024]. The smallest zone is a single county, and the largest zones are entire states. The large zone size limits the useability of FAF for many applications [National Academies of Sciences, Engineering, and Medicine 2013]. The user community has expressed a need for more geographically granular commodity flow data to support planning, policymaking, and operational decisions at the state and local levels.

In response to this need, the Bureau of Transportation Statistics (BTS) developed an experimental county-to-county commodity flow product. As part of this effort, BTS aimed to use publicly available data and transparent methods. BTS also used external data sources to help improve the quality of this experimental product. The main data inputs to this effort include FAF version 5.6.1 (FAF5.6.1) Origin-Destination-Commodity-Mode (ODCM) flows, NextGen truck trip estimates [FHWA 2023], Public Use Waybill sample (PUWS) rail shipment data [STB, RAILINC 2024a], waterborne freight volumes [USACE 2023a, USACE 2023b], County Business Patterns (CBP) employment data [Census 2024a], and other data sources that Section 2 describes.

The development process uses the following steps:

1. Create three estimates of county-level flows using existing disaggregation methods that are available to BTS.
2. Construct county-level validation targets using real-world data (or estimates thereof).
3. Develop a composite estimate of the results from the first step by regressing the validation data on the individual estimates, then applying the resulting model to the three estimates.

BTS tested ordinary least squares (OLS), log-log, and principal component regressions (PCRs), then selected OLS to blend the results of the three disaggregation processes, thereby forming a single, composite estimate of annual tonnage for each county-to-county pair, mode, and commodity group. This process applies to truck, rail, and water flows. Disaggregation factors for truck are then applied to air and to multiple modes and mail. The process for developing disaggregation factors for pipeline involves applying the disaggregation factors resulting from two of the third disaggregation methods to counties with access to the pipeline network.

BTS chose this methodology to build on previous research and take advantage of the strengths of existing disaggregation methods. This methodology also lets BTS create estimates for all counties, even in cases where industry-specific employment data are suppressed in public datasets. BTS chose these three existing disaggregation methods because they are transparent enough for BTS to replicate the methodology.

This initial release offers users the option to download state-specific files or the entire set of disaggregation factors for creating customized data queries. The experimental product includes estimates of county-level tons for year 2022. The release represents FAF modes as follows: rail, water, pipeline, and multiple modes and mail modes in the same way as the main FAF database. The data product bundles flows by truck-only and air together, and it excludes flows by Other, Unknown, or No Domestic Mode. This release aggregates Standard Classification of Transported Goods (SCTG) commodities into five groups:

- Agricultural products (SCTG 01–09)
- Gravel and mining products (SCTG 10–14)
- Coal and other energy products (SCTG 15–19)
- Chemicals, wood, and metals (SCTG 20–33)
- Manufactured goods, mixed freight, waste, and unknown (SCTG 34–99)

This technical report describes the data sources BTS used to develop the estimates, the process for creating the estimates, how to use the estimates, and limitations of this experimental product.

BTS welcomes users to email FAF@dot.gov with feedback on this experimental product.

Introduction

The FAF database provides estimates of the annual weights and values of shipments to, from, and within the United States for all forms of transportation [BTS, FHWA 2024]. FAF version 5 (FAF5) classifies freight into 42 commodity types and represents the origins and destinations of commodity flows, or volumes, using a geographic system of 132 FAF zones that aggregate to states. The smallest zones comprise a single county, and the largest zones comprise entire states.

The U.S. Department of Transportation (USDOT) originally developed FAF in the late 1990s as an internal tool for analyzing freight policies. FAF has become a national resource for the broader transportation community to understand freight movement [Berthaume, Morton 2015]. The large size of FAF zones, however, limits the useability of FAF for many applications [National Academies of Sciences, Engineering, and Medicine 2013]. Consequently, the user community has expressed a need for more geographically granular commodity flow data to support planning, policymaking, and operational decisions at the state and local levels.

In response to this need, BTS undertook an effort to create an experimental county-to-county commodity flow product. This experimental product includes estimates of county-level freight volumes in tons for the year 2022. This initial release does not include other years or other flow quantities, and it aggregates SCTG commodities into five groups:

- Agricultural products (SCTG 01–09)
- Gravel and mining products (SCTG 10–14)
- Coal and other energy products (SCTG 15–19)
- Chemicals, wood, and metals (SCTG 20–33)
- Manufactured goods, mixed freight, waste, and unknown (SCTG 34–99)

This experimental product excludes flows with other, unknown, or no domestic mode. Since this new product focuses on tons and a negligible portion of U.S. freight flows are transported via air on a tonnage basis (less than 0.1 percent), flows with the modes truck-only and air are bundled together. The other modes (rail, water, pipeline, and multiple modes and mail) have the same representation as in the main FAF database.

The initial release of this product allows users to download the following files:

- State-specific files (one for each state and one for Washington, DC): In each file, flows are represented at the county level for the state of interest and every adjacent state and at the FAF zone level for all other areas.
- The full set of county-level factors: Advanced users can download this file and merge it with the FAF data to create either a county-level database for a customized geographic area or full U.S. county-to-county flows.

This technical report describes the data sources BTS used to develop the estimates, the process for creating the estimates, how to use the estimates, and limitations of this experimental product. BTS welcomes feedback on all aspects of this product, especially regarding the following:

- Quality (e.g., how well do the estimates compare to local knowledge or other benchmark data? Can particular strengths or weaknesses be observed, and do users have suggestions to improve the fidelity of the estimates?)
- Features (e.g., should the product have more or fewer commodity groups?)
- Useability (e.g., are the files easy to use? Should BTS offer summary products—and, if so, what would be most useful to users? Is a visualization tool desirable?)

BTS invites users to email FAF@dot.gov with feedback.

1. Background

BTS develops the benchmark FAF [BTS 2024a] ODCM database using many data sources [Hwang et al. 2021]. Shipment data from the quinquennial Commodity Flow Survey (CFS) [BTS, Census 2020] are FAF's primary input. The CFS uses stratified sampling to survey approximately 100,000 U.S. shippers. Additional sources of FAF input data include the U.S. Department of Agriculture (USDA), the U.S. Geological Survey (USGS), and others. BTS uses these and other sources to develop annual estimates of flows in the years between the quinquennial CFS data-collection efforts [BTS 2025a].

The CFS is the main source of goods-movement data for about two-thirds of FAF tonnage and three-fourths of FAF value estimates. The use of CFS data impacts the level of geographic detail in the final FAF ODCM database. The CFS sampling approach uses a geographic layer with 132 zones, which include 84 metropolitan areas, 35 remainder-of-state areas, and 13 whole states. These zones allow BTS to construct statistically valid estimates of commodity flows for the origins, destinations, and industries that form the CFS sampling strata. Aggregated flows (e.g., state-to-state flows) are also statistically valid. The CFS flow estimates, however, are not statistically valid for areas that are smaller than the sample regions. As a result, USDOT provides FAF data at the CFS zone and state levels. In other words, FAF zones and CFS zones are equivalent.

The level of geography used in FAF has implications for FAF's useability. Because FAF uses relatively large zones, data users can readily use FAF for analyses that involve the entire United States or individual states. FAF also offers estimates of total flows originating or terminating in each zone. In a 2013 workshop, however, stakeholder discussions revealed that, without further transformation, "the FAF cannot be used to address concerns at a level of granularity more detailed than annual statewide flows and therefore cannot be used to address growing regional and local needs" [Berthaume, Morton 2015]. Consequently, it "could be useful to enhance the FAF to provide users with some level of understanding regarding freight movements at the regional level to inform local freight studies and projects" [Berthaume, Morton 2015].

1.1. NEEDS AND OBJECTIVES

The U.S. freight transportation system is essential to U.S. goods movement and the economy. As the National Academies of Sciences, Engineering, and Medicine [2013] notes, while freight transportation demand has risen substantially, existing infrastructure and operations are sometimes insufficient to meet this growing demand. All levels of government require sound data and technical tools to inform decisions regarding infastructure and operations. Obtaining subnational freight flow data is a challenge for local and state entities in particular. Methods to develop such data include establishment surveys, truck intercept surveys, economic data collection, and commodity flow disaggregation.

Many agencies choose to use a commodity flow disaggregation approach to develop subnational flows, and FAF is a widely used input for this process [Golias et al. 2021]. Disaggregated flow data then inform a variety of analyses involving trade flows, regional truck trip patterns, long-distance mode shares, economic impacts of freight, freight planning, corridor studies, and other topics [National Academies of Sciences, Engineering, and Medicine 2013].

The primary objective of the current effort is to meet the need for more geographically granular data to support state and local freight analyses, taking advantage of approaches to estimation using multiple methods. Throughout this effort, BTS' other objectives include the following:

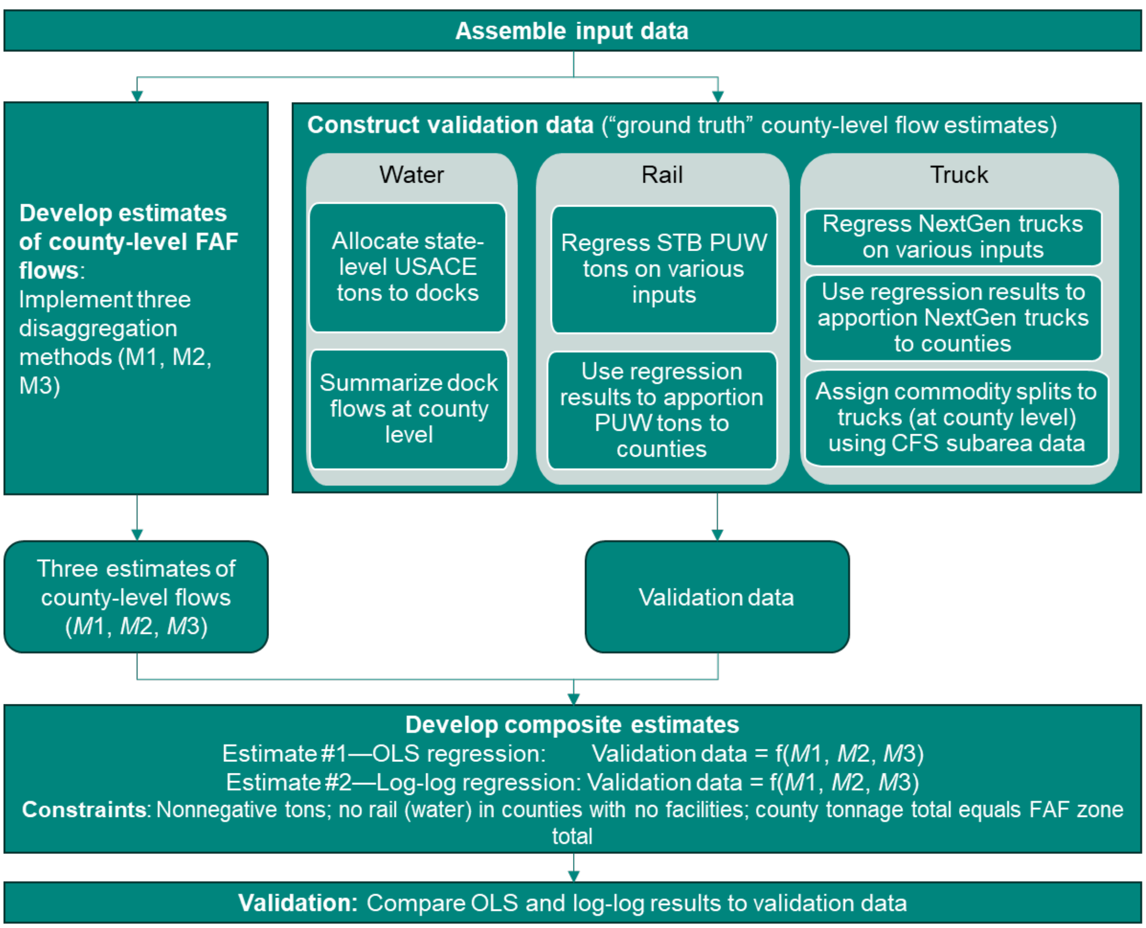
- Using a transparent approach
- Utilizing publicly available data
- Incorporating external data sources for validation and improved product quality

1.2. APPROACH AND DATA

BTS designed a disaggregation approach to meet the objectives noted in Section 1.1. Briefly, the approach involves the following steps (Figure 1):

1. Create three estimates of county-level flows using existing disaggregation methods that are readily available to BTS.
2. Construct county-level validation targets using real-world data.
3. Form a composite estimate of the results from the first step by regressing the validation targets on these individual estimates.

Figure 1. Approach Overview



Source: BTS.
Note: M1, M2, and M3 refer to the three disaggregation methods used in this effort. Section 3.1 contains more details.
USACE = U.S. Army Corps of Engineers; STB = Surface Transportation Board; PUW = Public Use Waybill.

This effort uses the following publicly available data (all data sources are specific to 2022 unless otherwise noted throughout the report):

- Flow data
  - FAF annual tons (BTS and the Federal Highway Administration [FHWA]) [BTS, FHWA 2024]
  - Public Use Waybill (PUW) (Surface Transportation Board [STB 2024b]) sample of rail shipments
  - NextGen medium- and heavy-duty truck flows (FHWA [2022] using data from the American Transportation Research Institute, INRIX, and FHWA)
  - CFS 2017 subarea tonnage estimates by commodity group (BTS and Census [2021])
  - State-to-state water flows and principal port volumes (U.S. Army Corps of Engineers [USACE 2023b])
- Freight network infrastructure data
  - Active docks and principal port freight docks ([BTS 2025b] and [USACE 2023a])
  - North American Rail Network (NARN) tracks (Federal Railroad Administration (FRA)) [BTS 2024d]
  - FAF5 highway network (BTS and FHWA) [BTS, FHWA 2022]
  - Border crossings (BTS National Transportation Atlas Database (NTAD) [BTS 2025b])
  - Pipeline and other energy-related data (Energy Information Administration [EIA 2022])
- Site-specific data
  - Commodity warehouse capacity (USDA) [USDA-WCMD 2024]
  - Global Energy Monitor for coal power plants [Global Energy Monitor 2024]
  - Buildings and their attributes (area, location, and land use type, e.g., industrial) (Federal Emergency Management Agency (FEMA)) [ORNL, FEMA 2024]
  - Surface and underground coal mines [Esri 2024]
- County-level data
  - CBP employment (Census) [Census 2024a]
  - County population (Census) [Census 2024b]
  - Number of private building permits (Census) [Census 2024c]
  - National Agricultural Statistics Service data (NASS) [USDA 2024]

Section 2 and Section 3, respectively, discuss the data sources and methodology in more detail.

1.3. LITERATURE SCAN

Several studies have conducted a thorough literature review on disaggregation methods for freight flows. Golias et al. [2021] found proportional weighting (also known as proportional allocation) is the most widely used approach among the reviewed studies, followed by regression and other methods. Socioeconomic data variables, such as employment, population, and payroll, are commonly used to allocate the aggregate-level zone values to a more granular level. Other variables, such as industry- or commodity-specific activity data (e.g., electricity generated, livestock sales, farm acreage), are also used in disaggregation.

Proportional weighting uses the following approach: For each FAF zone, analysts develop county-level disaggregation factors (or percentages) that sum to 100 percent for the FAF zone. Then, analysts can apply these factors to the counties in the FAF zone, thereby apportioning all the FAF flow from the FAF region to the counties that constitute the FAF region. Each FAF zone has two sets of disaggregation factors: production and attraction. Production factors allocate the flows that originate in the FAF zone, and attraction factors allocate the flows that are destined to the FAF zone.

Opie, Rowinski, and Spasovic [2009] disaggregated FAF data to the county level by developing several sets of disaggregation factors for different commodity types. For production factors, the study tested total employment, vehicle-miles traveled (VMT), counts of trucks, and North American Industry Classification System (NAICS) three- or six-digit commodity-specific industry employment. For attraction factors, data on population, age-specific and income-adjusted population, VMT, number of trucks, and commodity-specific employment were used. The study applied the various sets of factors to the 2002 FAF for the state of New Jersey. The authors compared the disaggregation results with Global Insight's Transearch data to find the best results for trip productions and attractions. Besides using employment and population as disaggregation factors, several studies used the outputs from Input–Output models with proportional weighting. Fischer, Ang-Olson, and La [2000] used Transearch data and Input–Output modeling data from IMPLAN to allocate the Transearch flows to traffic analysis zones. The effort uses information on employment, land use, and commerical facilities.

Besides proportional weighting, regression is often used to identify the relationship between freight flows and key socioeconomic, industry-specific, or business attributes of a region. Cambridge Systematics [2009] developed regression models to link the production and attraction of each commodity type to employment, population, energy production, and agriculture activities.

Other methods, such as iterative proportional fitting, cross-classification, econometric, structural equations, and behavior-based approaches, are used in freight flow disaggregation. Ranaiefar, Chow, and Ritchie [2013] used a structural equation modeling approach to identify the relationship between commodity flow and socioeconomic factors, including employment, number of establishments, population, farm acreage, gross domestic product, capacity of refineries, and electricity generation of power plants. Maricopa Association of Governments [2018] developed an agent-based supply chain and freight transport model. It includes disaggregate behavior-based logistics and transportation-choice models to simulate commodity flows at the firm level.

1.4. LIMITATIONS AND EXTENSIONS

The initial release of this experimental product will benefit from extensive user testing and feedback. The estimates in this release do not benefit from nationwide ground truth measures of tonnage by commodity for all modes. This initial release uses validation data that are incomplete or indirect measures of actual, ground truth ODCM data. Validation of the benchmark and annual estimates among FAF regions defies traditional methods of calculating estimation variability since multiple data sources are blended with techniques that can either offset or magnify sampling errors in the source data. To summarize, although FAF commodity flow estimates are widely accepted as the best publicly available source, the estimates are subject to continuous improvement.

Potential enhancements to this product could involve the following:

- Instead of a regression-based composite estimate, the composite estimation could utilize machine learning methods, such as an ensemble-based stacking approach.
- Instead of using the methods selected here, BTS could develop an enriched, new disaggregation formulation that leverages all the most recent data.

Another possible extension is using disaggregate data to estimate shipping activity (e.g., using an agent-based or other granular approach) and then summarizing these flows to the county level (rather than developing aggregate estimates then disaggregating them).

Data sources could be revisited. The water benchmarks could use USACE Manuscript data directly rather than state flows. The water validation dataset currently uses principal port volume from 2021; it should be updated to 2022 to be consistent with the other data sources. The STB PUW was the main input for the rail validation, but this dataset has major limitations, including significant suppression of flows. The Confidential Waybill sample would likely provide better information for the county-level estimates. Rail and water FAF modes reflect shipments that use rail or water only, with no other modes. Geographically detailed rail and water flow predictors—such as land use data for parcels that are adjacent to port or rail tracks—may generate better estimates than county-level data.

Pipeline flow estimates in this release use only network constraints as implied by open-source, energy related information. However, more detailed information could improve the accuracy of county-level estimates for pipeline. Multimodal estimates likewise use the basic assumption that the flow patterns are essentially the same as truck at origins and destinations. BTS is concurrently developing a multimodal assignment, which can be integrated with the county-level flow estimates to improve it in this respect.

2. Data

This section describes the data sources used in this effort. These sources include the FAF origin–destination (OD) flow data, data for creating disaggregation factors, and data for developing validation targets. All data sources are for 2022 unless otherwise noted.

2.1. DISAGGREGATION DATA

FAF estimates dometic and international commodity flows among states, substate regions, and major international gateways. The latest version of FAF (version 5.6.1) is an OD database of commodity flows among 132 domestic regions and 8 international regions. The full dataset includes tons, value, and ton-miles of commodity movements among regions by 7 modes of transportation and 42 types of commodity. For this experimental product, the tons of OD flows in 2022 from FAF5.6.1 are used as the aggregate-level input into the disaggregation process. More details about FAF5 are available from BTS and FHWA [2024] and Hwang et al. [2021].

The following data sources are inputs to processes that create disaggregation factors:

- CBP employment data for 2022 [Census 2024a] provide industry and county-level information on the number of establishments, counts of employees in mid-March, first quarter payroll, and annual payroll. Since the reference year of 2017, cells in the CBP data are only released if they contain three or more establishments. Otherwise, the values are suppressed. Using these data could lead to an underestimation of the number of employees in specific industries for some counties.
- County population data 2022 [Census 2024b] are mainly used to generate disaggregation factors at the destination end for the proportional allocation method and as explanatory variables in regression models.
- Global Energy Monitor data for coal-based power plants provides information on the location of active power plants that use coal as the primary energy source. The annual kilowatt hours (KWH) of electricity generated at the plants is used as a key factor for explaining the total tons of coal shipped to a specific destination [Global Energy Monitor 2022].
- NASS data from USDA [2022] include information on livestock estimates and farm acreage at the county level.

These factors, when applied, transform the FAF data from the FAF zone level to the county level.

Section 3 discusses the use of these data for each disaggregation method.

2.2. VALIDATION DATA

BTS constructed three validation datasets to use as ground truth data in the composite estimation process. As noted in Section 1.1, one objective of this experimental product development is transparency. BTS constructed the validation datasets using only publicly available data accordingly for this initial release. The validation data use the same commodity groups that the Introduction defines unless noted otherwise. The development of water, rail, and truck validation data is described throughout this subsection.

2.2.1. Water

The USACE Waterborne Commerce Statistics Center provides waterborne cargo flows by commodity type at the state level [USACE 2023a] and total flows for principal U.S. ports [USACE 2023b]. The state-level data contain 14 commodity types. Manuscript cargo data [USACE 2023c] provide more detailed commodity information for inbound and outbound port flows. This experimental product uses the Manuscript flows to create a correspondence between the state-level commodity types and the five commodity groups. The principal port volume data are from 2021 and may be updated to 2022.

The USACE state-level data use Public Domain Database Commodity (PDDC) codes, and FAF uses the SCTG system. As discussed in the Introduction, this experimental product uses five categories comprising SCTG groups. To facilitate the analysis, BTS created a crosswalk between the PDDC codes and the five-category system (Table 1). For most PDDC codes, the correspondence is clear. For example, PDDC code 1000 denotes coal, which belongs to SCTG group 15–19 (energy products). Other PDDC codes contain more than one SCTG commodity type (Table 2). Creating the correspondence for these PDDC codes involved summarizing the USACE Manuscript cargo data for years 2013 to 2022 using the detailed USACE commodity types shown in Table 2, then computing the shares of PDDC-based tons in each SCTG group.

Table 1. Crosswalk Between PDDC Codes and SCTG Groups

PDDC code	PDDC name	SCTG 01–09	SCTG 10–14	SCTG 15–19	SCTG 20–33	SCTG 34–99
1000	Coal, Lignite, and Coal Coke	-	-	100%	-	-
2100	Crude Petroleum	-	-	100%	-	-
2229	Petroleum Products	-	-	100%	-	-
3100	Chemical Fertilizers	-	-	-	100%	-
3200	Chemicals excl. Fertilizers	-	-	-	100%	-
4142	Lumber, Logs, Wood Chips, Pulp	-	-	-	100%	-
4349	Sand, Gravel, Shells, Clay, Salt, and Slag	-	90%	-	10%	-
4400	Iron Ore, Iron, & Steel Scrap	-	85%	-	-	15%
4600	Non-Ferrous Ores and Scrap	-	92%	-	-	8%
5155	Primary Non-Metal Products	-	-	-	100%	-
5354	Primary Metal Products	-	-	-	100%	-
6168	Food and Food Products	100%	-	-	-	-
7000	Manufactured Goods	-	-	-	15%	85%
8099	Unknown & Not Elsewhere Classified	-	-	-	-	100%

-Not applicable.

Table 2. PDDC Codes With Multiple SCTG Group Affiliations

PDDC code	PDDC name	Detailed commodity type (USACE Manuscript cargo data)	SCTG county-level commodity group	
4349	Sand, Gravel, Shells, Clay, Salt, and Slag	4310	Building stone	SCTG 10–14
		4322	Limestone	SCTG 10–14
		4323	Gypsum	SCTG 10–14
		4327	Phosphate rock	SCTG 10–14
		4331	Sand & gravel	SCTG 10–14
		4333	Dredged material	SCTG 34–99
		4338	Soil & fill dirt	SCTG 10–14
		4515	Marine shells	SCTG 01–09
		4741	Sulphur (dry)	SCTG 20–33
		4782	Clay & refractory materials	SCTG 20–33
		4860	Slag	SCTG 20–33
		4900	Nonmetallic minerals (NEC)	SCTG 10–14
4400	Iron Ore, Iron, & Steel Scrap	4410	Iron ore	SCTG 10–14
		4420	Iron & steel scrap	SCTG 34–99
4600	Non-Ferrous Ores and Scrap	4630	Copper ore	SCTG 10–14
		4650	Aluminum ore	SCTG 10–14
		4670	Manganese ore	SCTG 10–14
		4680	Non-ferrous scrap	SCTG 34–99
		4690	Non-ferrous ores (NEC)	SCTG 10–14
7000	Manufactured Goods	7110	Machinery (not electric)	SCTG 34–99
		7120	Electrical machinery	SCTG 34–99
		7210	Vehicles & parts	SCTG 34–99
		7220	Aircraft & parts	SCTG 34–99
		7230	Ships & boats	SCTG 34–99
		7300	Ordnance & accessories	SCTG 34–99
		7400	Manufactured wood products	SCTG 20–33
		7500	Textile products	SCTG 20–33
		7600	Rubber & plastic products	SCTG 20–33
		7800	Empty containers	SCTG 34–99
		7900	Manufactured products (NEC)	SCTG 34–99

NEC = not elsewhere classified.

2.2.2. Rail

The primary data source used for the rail validation data is the PUWS published by STB [STB, RAILINC 2024a]. In addition, the data from FRA NARN, USA Structures data published by FEMA, commodity storage warehouse data by USDA, and Bulding Permits Survey data published by the U.S. Census Bureau are used to develop regression equations. Furthermore, the Surface and Underground Coal Mines in the U.S. data published through the ArcGIS Hub is used to obtain the county-level coal productions. These data sources are described further in the following list:

- PUWS is the nonproprietary subset version of the Confidential Carloads Waybill Sample (CCWS) published by STB. To hide the sensitive information in the CCWS, such as station and carrier information, PUWS reports the origins and terminations of the commodity flows by Business Economic Areas (BEA) zones [STB, RAILINC 2024b]. The most recent PUWS data available were the 2022 data. PUWS uses the Standard Transportation Commodity Code (STCC) commodity classification system. Expanded billed volume of each commodity flowing between BEA zones was used for the analysis after converting the STCC classifications to SCTG classifications (Appendix A). For the purpose of rail validation data generation, the flow tons are aggregated into seven commodity groups: SCTG 01–09, SCTG 10–14, SCTG 15, SCTG 16–19, SCTG 20–33, SCTG 34–40 and 43, and finally SCTG 41 and 99 together. These seven groups are used instead of the five groups since county-level production data are available for SCTG 15 (coal), and splitting SCTG 34–99 into two groups led to improved goodness of fit in the regressions. After computing the estimated tons using regression models, they are again grouped into the five commodity groups.
- NARN data are publicly available through USDOT BTS' NTAD [BTS 2025c]. This database covers all 50 U.S. states, the District of Columbia, Mexico, and Canada and provides details of rail lines, such as ownership, type, and use by passengers vs. freight [BTS 2024d].
- USA Structures (FEMA) is a dataset of all structures in the United States that are larger than 440 sq ft. This dataset was created as a collaboration between the Department of Homeland Security, Federal Insurance and Mitigation Administration, FEMA's Response Geospatial Office, Oak Ridge National Laboratory (ORNL), and the USGS for use in Flood Insurance Mitigation, Emergency Preparedness and Response. Building occupancy classification, such as commercial, industrial, agricultural, residential, is the primary variable of interest. The dataset is available to download publicly through the ESRI Living Atlas [ORNL, FEMA 2024].
- Commodity Warehouses (USDA) contains data for approved storage warehouses and is available to download through the interactive dashboard at USDA [USDA-WCMD 2024]. This resource was launched by the Warehouse and Commodity Management Division (WCMD) of the Agricultural Marketing Service. It provides details on capacity and the number of working warehouse units for commodities, such as cotton (in bales), cotton seeds (in tons), dry edible beans (in hundredweight (CWT)), grains (in bushels), peanuts (in tons), and sugar (in CWT). The capacity variable is a summation of all the observed measures, and it does not represent the actual capacity of a warehouse.
- The Building Permits Survey (BPS) provides national, state, and local statistics on new privately owned residential construction [Census 2024c]. Data are available monthly, year-to-date, and annually at the national, state, and county levels. County-level annual data are downloaded from the Census website for the years 2020–2022.
- Surface and Underground Coal Mines in the United States contains the data collected by EIA using the form EIA-7A, which collects data from all coal mining companies that owned a mining operation that produced 50,000 or more short tons of coal during the reporting year. EIA county-level coal-production data for 2021 are available to download through the ArcGIS Hub [ESRI 2024].

2.2.3. Truck

Truck trip data from the NextGen data product [FHWA 2023] are the primary source of flow data for the truck validation dataset development. This national-level dataset provides annual estimates of truck trips throughout the United States. The product uses a geography comprising

583 zones. FHWA estimates the NextGen truck trips using Global Positioning System (GPS) data and other analysis from ATRI and INRIX, then calibrates the flows using traffic count stations. The data include trips made by freight trucks and light duty trucks that make intercity and local deliveries. Pickup trucks are excluded. Information on payload, including what commodity is carried and whether a truck is full or empty, is not included.

Commodity information for truck and parcel ground flows are available from the 2017 CFS Subarea estimates [BTS, Census 2021]. The 132 CFS zones are subdivided into 329 subareas in this dataset.

The FAF5 highway network [BTS, FHWA 2022] is a geospatial representation of the U.S. highway network, with a particular focus on roads that are used to transport freight. The network was developed using National Highway System (NHS) links and other links that are not in the NHS. Centroid connectors and ferry links are excluded from the current analysis. Links with Urban_Code of 99998 or 99999 are labeled “rural”; other links are labeled “urban.” The number of lanes is available for most links, but in some cases, it is zero in both directions. In these cases, BTS assumes that the link has one lane. The validation data use the sum of lane-miles in each county on urban NHS links, rural NHS links, and other links.

Transborder data [BTS 2024b] provide information on the volume of goods crossing the U.S.–Mexico and U.S.–Canada borders at individual border crossings. The TransBorder data are a subset of the Census FT900 U.S. International Trade in Goods and Services data.

A selected docks shapefile [BTS 2025b] contains the locations of docks that appear to be actively used for handling freight. Trucks often transport freight to or from such docks. Principal port designation information [USACE 2023b] supplements this information in the validation data development.

2.3. NETWORK CONSTRAINTS

To ensure the freight flow is not assigned to counties that do not have access to a specific mode, several network datasets are used to determine the availablity of the mode at the county level. The network contraints are applied to rail, waterborne, and pipeline. The following are the modal-specific data used to impose the network contraints:

- Rail: NARN Lines for non-passenger transport from NTAD.
- Waterborne: Active docks released based on the USACE Waterborne network.
- Pipeline: Terminals of various energy sources released by EIA [2022], namely, information on natural gas below-ground storage, light natural gas above-ground storage, peak shaving facility, light natural gas import and export terminals, natural gas processing plants, petroleum ports, terminals transporting petroleum products, NTAD intermodal pipeline terminals [BTS 2024c], petroleum terminals, petroleum refinery, power plants, natural gas pipeline endpoints, ethanol plants, and coal mines. BTS uses these data to determine whether a county has access to the pipeline network.

3. Methodology

This section describes the methods used throughout this study. Three methods generate individual estimates of disaggregated flows (Section 3.1). Additionally, BTS uses various methods (including regression and construction) to develop validation targets. BTS then tests three types of regression methods to estimate the composite flows. Finally, BTS uses a separate set of methods to develop air, multiple modes and mail, and pipeline flows.

3.1. DISAGGREGATION APPROACH

This section introduces three disaggregation methods (M1–M3) that are applied to the aggregated FAF zone level to generate the county-level origin and destination flows. Throughout this report, the three disaggregation methods are referenced as follows:

- M1: proportional allocation
- M2: updated Cambridge Systematics disaggregation procedure
- M3: ORNL disaggregation procedure

M1 is proportional allocation, which is widely used in practice. M2 is based on regression models that Cambridge Systematics, Inc. originally estimated in 2009 [Cambridge Systematics 2009]. BTS used 2022 data to reestimate the model parameters for the current effort. M3 is the disaggregation procedure developed by ORNL [n.d.].

3.1.1. Proportional Allocation

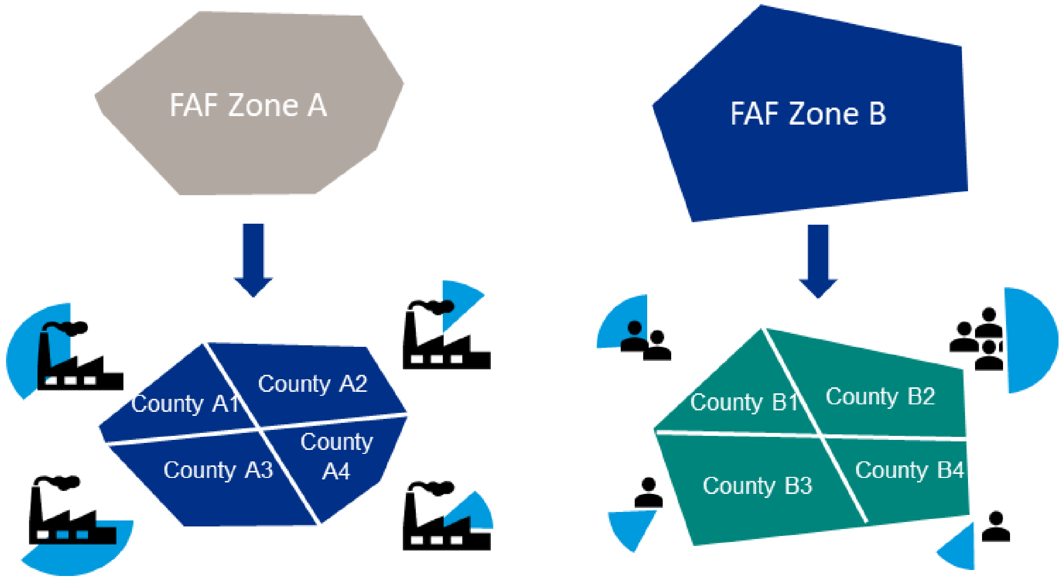
While the methods used by transportation planners to generate disaggregated commodity flows vary, most have involved using employment and population data as the primary factors to disaggregate the nationwide or statewide commodity flows down to the county level [Opie, Rowinski, Spasovic 2009]. On the one hand, employment is found as a key indicator in explaining the production and attraction of commodities [Cambridge Systematics 2009]. On the other hand, employment and population data are often available at the granular level, such as for counties or even traffic analysis zones [Sorratini, Smith 2000]. The procedures of applying proportional allocation follow:

1. Get 2022 data on employment and population at the county level.
2. For each FAF zone, calculate the ratios for employment and population for each county within that FAF zone. The results are the production and attraction ratios/factors (production and attraction factors can be used interchangeably with origin and destination factors).
3. Proportionally allocate the production and attraction flows to each county based on the production and attraction ratios.

The following is a hypothetical example to illustrate the process:

FAF zones A and B each comprise four counties (A1, A2, A3, A4, B1, B2, B3, and B4) as shown in Figure 2.

Figure 2. Counties by FAF Zone



Source: BTS.

From FAF data, the FAF zone level flow (W_{AB}) is known. The production factor is calculated based on the total employment in each county within the production FAF zone. Specifically, for county A1, the production factor (Prod_R_{A1}) can be calculated as $Prod\_R_{A1} = Emp_{A1} / Emp_A$, where Emp_{A1} is the total employment in county A1 and Emp_A is the total employment in the whole FAF zone A. Similarly, the attraction factor is calculated using the total population in the counties within a FAF zone. For county B1 in Zone B, the attraction factor ($Attr\_R_{B1}$) can be calculated as $Attr\_R_{B1} = Pop_{B1} / Pop_B$, where Pop_{B1} is the total population in county B1 and Pop_B is the total population in the whole FAF zone B.

In this example, there are 16 OD pairs at the county level (4 × 4). Each OD pair flow is obtained through the following formula:

$$W_{A1B1} = W_{AB} * Prod\_R_{A1} * Attr\_R_{B1} \tag{1}$$

Where:

- W_{A1B1} = commodity flow from county *A1* to county *B1*
- $Prod\_R_{A1}$ = production factor for county *A1*
- $Attr\_R_{B1}$ = attraction factor for county *B1*

For example, if the total flow between Zone A and Zone B is 600 tons, the production factor for A1 is 1/3, and the attraction factor for B1 is 1/4, then the total flow between A1 and B1 is calculated as 600 × 1/3 × 1/4 = 50 tons. To obtain the factors for specific modes, such as rail, water, and pipeline, only the counties that have access to the network for that mode are considered for developing the production and attraction factors.

3.1.2. Updated Cambridge Systematics Disaggregation Procedure

Cambridge Systematics [2009] applied linear regression to identify the relationship between employment by industries and the commodities those industries produce and consume. CBP data were the main data source from which to extract the employment information of a specific industry sector. For certain commodities, other variables, such as farm acreage, sales of livestock, and electricity generated, were used in model estimation. The procedure of disaggregating the FAF zone data down to the county level follows:

1. Determine employment in 2022 based on the three-digit NAICS codes at the county level.
2. Develop linear regression models for tonnage for each of the 42 commodity types by origin and destination. The model specifications are similar to those of the previous study, which includes explanatory variables such as industry-specific employment by three-digit NAICS, total employment, population, farm acreage and livestock estimates from NASS, and electricity generation by coal power plants released by Global Energy Monitor. For some commodities, the tons of commodities were aggregated to construct the model.
3. Apply the estimated coefficients to the county-level data to estimate the production and attraction of specific commodities at the county level. The regression models do not include an intercept following the assumption stated in the previous study that, when the employment or population is zero, there would be zero tons of commodities shipped out or to a region. This step is illustrated in Equations 2 and 3.

$$Prod\_W_{ij} = \beta_{i1} * V_1 + \cdots + \beta_{in} * V_n \tag{2}$$

   Where:
   - $Prod\_W_{ij}$ = production in county i for commodity j
   - β_{i1} = coefficient of explanatory variable 1 (V_1)
   - V_1 = explanatory variable 1
   - β_{in} = coefficient of explanatory variable n (V_n)
   - V_n = explanatory variable n

$$Attr\_W_{ij} = \beta_{i1} * V_1 + \cdots + \beta_{im} * V_m \tag{3}$$

   Where:
   - $Attr\_W_{ij}$ = attraction in county i for commodity j
   - β_{im} = coefficient of explanatory variable m (V_m)
   - V_m = explanatory variable m

   The model results are included in Appendix B.

4. After obtaining the estimated commodity flow in terms of tonnage for production and attraction at the county level, calculate the ratio of county production or attraction to the FAF zone. Similarly to the proportional allocation, only counties with access to a specific mode network are included to calculate the expansion factors to disaggregate to FAF flow to the county level. Then the final expansion factors are applied to the FAF regional matrix to obtain the counties matrix as described in Equation 4.

$$W_{A1B1} = W_{AB} * (\frac{Prod\_W_{A1}}{Prod\_W_A}) * (\frac{Attr\_W_{B1}}{Attr\_W_B}) \quad (4)$$

Where:

- W_{A1B1} = OD flow between county A1 and county B1
- $Prod\_W_{A1}$ = production in county A1
- $Prod\_W_A$ = production in FAF zone A
- $Attr\_W_{B1}$ = attraction in county B1
- $Attr\_W_B$ = attraction in FAF zone B

3.1.3. Oak Ridge National Laboratory Disaggregation Procedure

ORNL uses payroll as a key indicator to determine the disaggregation factors at the county-level. A set of 28 production and attraction equations were estimated by regressing total freight shipped or received by FAF zone based on Economic Census (EC) and CBP data by industry sector. The explanatory variables for the production and attraction estimation models include number of establishments, number of employees, annual payroll, and receipt totals from the EC and CBP data.

Freight productions and attractions at the county level were estimated by applying these models to the county-level data by industry sector. These estimates are converted to factors and are used as factors to expand each FAF OD flow to the corresponding county productions and attractions.

ORNL used a matrix-balancing technique to distribute the production and attraction to the county level. The matrix-balancing technique is based on information theory, with given prior matrix information. The prior information was a matrix of weights reflecting the resistance for travel or disutility for using a certain mode. The weights were developed based on a function of spatial correlation between production of origin and attraction of destination by mode. At the end, ORNL calculated the weights of county-level freight movement by seven modes: truck, rail, water, air, multiple modes and mail, pipeline, and other or unknown [Oak Ridge National Laboratory unpublished].

Since M1 uses total employment and population to generate disaggregation factors, while M2 and M3 rely on more specific industry-based employment, M1 includes the most county pairs in the disaggregation results, and M2 includes the fewest county pairs in the results.

3.2. CONSTRUCTION OF VALIDATION DATA

This section describes the procedures used to develop validation data for county-level water, rail, and truck flows.

3.2.1. Water

BTS developed water validation data using a combination of flow data and network data. State-level flow data by commodity type are the basis of the total volumes by commodity type originating in and destined to each state. The shares of commodity type at each port are assumed to be uniform throughout the state. Then, the main steps of developing water validation data are:

1. Compute the percentage of state-level volume that is handled by each principal port.
2. Allocate principal port volume equally among its docks [BTS 2025b].
3. Assign the remaining state-level volume to docks that do not belong to principal ports.
4. Allocate the remaining state-level volumes equally to the non-principal port active docks throughout the state.
5. Summarize dock-level flows to the county level.
6. Convert flows into percentages, with county-level percentages adding up to 100 percent within each FAF zone and commodity group combination.

The process for computing origin shares and destination shares are the same. The percentages are then joined to the FAF data at the origin and destination ends and are multiplied by the FAF OD volume to compute the validation flow target for each county-to-county flow by commodity type.

3.2.2. Rail

The main steps of developing rail-validation data are summarized as follows:

1. Use the OLS regression to estimate the parameters of the best-fit line that relates various explanatory variables to the number tons in origin and destination BEA zones reported in PUWS data.
2. Apply the estimated parameters to county-level data to estimate the number of tons originating in or destined to each county.
3. Find the domestic origin and domestic destination counties for import and export tons in PUWS data.
4. Sum the total number of estimated tons in each FAF zone, and then, compute the percentage of tons by county in each FAF zone.
5. Use the shares of volume by commodity group from the FAF data to estimate the share of rail tons by commodity group in each county.

The output of this process is the annual share of rail tons by county and commodity group in each FAF zone. These estimates constitute the rail-validation dataset. The next few paragraphs in this subsection provide the detailed steps of constructing the rail validation dataset.

The primary data source for the weight of commodities (measured in tons) to develop the rail validation data is the PUWS, which provides the amount of tons moved from one BEA zone to another by five-digit STCCs. The expanded billed weight is used as the dependent variable.

After converting the STCC classifications into SCTG classifications and making the seven commodity groups as described in Section 2.2.2, BTS splits the PUWS dataset into two parts, origins and terminations, based on the BEA zone and commodity group. The rebill codes 0 and 1 are selected for the origins data, and 0 and 3 are selected for the terminations. Here, the rebill code 0 is defined as a local shipment or normal through-rate, 1 is defined as originated–delivered Rule 11 shipment, and rebill code 3 is defined as received–terminated Rule 11 shipment. Although PUWS contains other variables (e.g., all rail/intermodal codes and type of move) that could be important, most of the tons amounts fall into the unknown category. Therefore, these variables are not used when aggregating the flow values into BEA zones.

The predictor variables for the regression models of rail validation data development come from various sources as described in Section 2.2. These predictor variables are selected so they do not overlap with the variables selected for the three methods (M1, M2, and M3) described in Section 3.1. Each of the data sources are preprocessed as follows.

Length of rail tracks is the first predictor variable of interest due to the hypothesis that having more rail tracks could imply more rail flow traffic. FRA NARN data are used to compute rail track length by applying the following preprocessing steps:

1. Filter the tracks for which network (NET) is either O (other track/minor industrial leads), Y (yard tracks), M (main sub network), or I (major industrial lead).
2. Drop all passenger (PASSNGR) categories B (Amtrak & Commuter), C (Commuter), T (Tourist, Museum, or Science Passenger Service), R (Rapid Transit), D (Alaska Railroad Passenger Service), O (Ontario Northland [Canada Network Only]), V (Via Rail Canada [Canada Network Only]), A (Amtrak), I (Intercity High-Speed Rail), and E (Intercity High-Speed Rail & Commuter).
3. Create a new binary variable to identify Class I carriers and Other carriers based on railroad ownership (RROWNER1). The Class I carriers listed on the NARN data are: Union Pacific (UP), Burlington Northern Santa Fe (BNSF) Railway, CSX Transportation (CSXT), Norfolk Southern (NS) Railway, Canadian National (CN) Railway, Kansas City Southern (KCS) Railway, and Canadian Pacific (CPRS) Railway.
4. Aggregate the data to the county level and compute the Class I and Other miles.

To capture the next predictor variable of interest, direct railroad users, the size of structures (in square meters) near rail tracks is calculated using FEMA USA structures data. The SQMETERS variable in the FEMA dataset provides the sizes of the various structures. To compute the size of structures predictor variable, the industrial, agriculture, and commercial layers are selected from the FEMA data. Then, a buffer of 0.1 mi is created around the rail tracks where the NET variable of the NARN data is either O or Y or I. This buffer is also applied to all the endpoints of the main rail network if it is not captured by O, Y, and I. Finally, the summation of the SQMETERS variable of the selected FEMA layers (within buffer) is computed for each county.

To improve the regression model for agricultural productions, a variable of interest to consider is the data related to grain elevators. Although farm acreage might be a better indicator of agricultural productions, this variable is already used for the method described in Section 3.1.2. Therefore, farm acreage is not used as a predictor variable in the construction of ground truth data. To capture grain elevator attributes, the number of functional warehouse units in a county and the capacity of warehouses data are obtained from the USDA WCMD dashboard. The capacity values in the downloaded data use a mixture of units. The capacity variables are converted to cubic meters as given in Table 3. Although this computation is not an exact measure of the capacity of a warehouse, it is assumed to be sufficient to get an approximate measure of warehouse volume.

Table 3. Conversion of Observed Warehouse Capacity to Volume

Commodity	Observed value (units)	Conversion factors	Approximate warehouse volume
Cotton	Bales	bale volume = 0.48 m^3	*Bales* × 0.48 m^3
Cotton seed	Seed mass (tons)	seed density = 0.4414 ton/m^3	*Seed mass*∕0.4414 ton/m^3
Dry edible bean	CWT	1 CWT = 45.36 kg Bean density = 789 kg/m^3	CWT × 45.36∕789 kg/m^3
Grain	Bushels	1 U.S. bushel = 0.035239 m^3	*Bushels* × 0.035239
Peanut	Peanut mass (tons)	Peanut density = 0.497 ton/m^3	*Peanut mass*∕0.497 ton/m^3
Sugar	CWTs	1 CWT = 45.36 kg Sugar density = 828 kg/m^3	*CWTs* × 45.36∕828 kg/m^3

Note: Bale volume is from NCC [n.d.], and the cotton seed density is obtained from Ashley et al. [2018] based on the required storage space per ton. As per USDA Commodity Specification Dry Edible Beans, Peas, and Lentils documentation [USDA 2014], dry edible beans are pinto beans, red beans (small red type only), black eye beans, light red kidney beans, dark red kidney beans, baby lima beans, pink beans, great northern beans, garbanzo beans, pea, lentils, split green, and pigeon peas. Dry Bean density is an average value computed based on dry pinto beans, dry lima beans, dry garbanzo beans, and dried peas as found in Machine and Process Design [n.d.]. Peanut density represents an avarage of shelled and unshelled peanuts. Sugar density is an average of brown, granulated, and powdered sugar [Machine and Process Design n.d.].

Next, using the data from BPS, the total number of permits, number of units, and values are computed at the county level for 2020 through 2022. The number of units variable is selected as another predictor for the regression models, specifically for SCTG 10–14 (gravel and other mining products), with the intention of capturing demand or attraction of gravel and mining products. The number of units variable represents the total number of units in a selected permit. For example, a two-story building has two units. In addition to the above-mentioned predictors, the number of active water docks [BTS 2025b] by county is also used as a predictor variable. Rail often carries cargo to or from water docks. Therefore, the hypothesis in this scenario is that having more water docks can impact the amount of tons that originates or terminates at these locations.The number of active docks is further broken into variables for the number of docks in West Coast, East Coast, Gulf, and Inland Water systems.

After preparing the predictor variable values at the county level, BTs aggregated them to the BEA-zone level to get the final predictors for the regression analysis. Then, multiple linear regression analysis is performed at the BEA-zone level with PUWS expanded billed weight (tons) as the dependent variable and other selected predictors (Class Ⅰ miles, Other miles, Size of Industrial building in square meters, etc.) aggregated to BEA-zone level totals for each commodity group. This approach is somewhat like the Cambridge Systematics method. To identify the best set of predictor variables in the regression models, leaps [Lumley 2024] and rFSA [Lambert et al. 2018] packages in R are utilized with best subset regression criteria. Although the best subset regression is not advised in practice when the number of predictor variables is too large, in this situation, there are only a handful of predictors. The rFSA allows identifying the best models with the interaction terms between predictor variables, which is a facility the leaps package does not offer by default. The adjusted R-squared is used as the model evaluation criteria.

This process identified 13 regression models, and they are summarized in Table 4 and Table 5.

Table 4. Rail Validation Regression Models for Origins by Commodity Group

SCTG	Variable (units)	Estimate	Std. error	*t* statistic	*p*-value	Adj R-sq
Agricultural products—01–09	*Intercept*	288473.32	403301.36	0.72	0.4760	0.5871
	Number of functional warehouse units	936.96	205.20	4.57	0.0000	
	Warehouse capacity (m^3)	3.971×10^{-3}	0.02	0.20	0.8431	
	Class I track miles	819.08	481.50	1.70	0.0918	
	Warehouse capacity (m^3) × *Class I track miles*	8.6×10^{-5}	1.6×10^{-5}	5.26	0.0000	
Gravel and mining products—10–14	*Intercept*	2.50	2.74	0.91	0.3654	0.1251
	log1p(*Class I track miles*)	0.68	0.37	1.84	0.0701	
	log1p(*Total number of private building units (2020–2022)*)	0.53	0.23	2.25	0.0277	
Other energy products—16–19	*Class I track miles*	240.53	86.06	2.79	0.0062	0.6234
	FEMA agriculture building area (m^3)	4.41	0.76	5.80	0.0000	
	Number of active docks in Gulf	7323.86	1682.94	4.35	0.0000	
Chemicals, wood, and metals—20–33	*Intercept*	3806.25	191905.59	0.02	0.9842	0.8008
	Class I track miles	863.29	242.99	3.55	0.0005	
	Total number of active docks	4279.77	3551.98	1.20	0.2302	
	Class I track miles × Total number of active docks	16.32	3.37	4.85	0.0000	
	FEMA agriculture building area (m^3)	-3.97	2.87	-1.38	0.1689	
	Class I track miles × FEMA agriculture building area (m^3)	0.01	1.4×10^{-3}	4.36	0.0000	
Manufactured goods, mixed freight—34–40, 43	*FEMA industrial and commercial building area* (m^3)	0.28	0.02	13.96	0.0000	0.6606
	Number of active docks in west coast	27093.78	10544.67	2.57	0.0112	
Waste and unknown—41 & 99	*Total number of active docks*	2438.24	787.68	3.10	0.0024	0.6279
	FEMA agriculture building area (m^3)	1.87	0.39	4.83	0.0000	
	FEMA industrial and commercial building area (m^3)	0.03	4.4×10^{-3}	5.82	0.0000	

Note: log1p() denotes log(value + 1). The symbol × denotes an interaction between two predictor variables. SCTG 15 (coal) does not require a regression model as production data are available at the county level.

Table 5. Rail Validation Regression Models for Destinations by Commodity Group

SCTG	Variable (units)	Estimate	Std. error	*t* statistic	*p*-value	Adj R-sq
Agricultural products—01–09	*Class I track miles*	1036.74	305.37	3.40	0.0009	0.6911
	FEMA industrial building area (m^3)	0.10	0.04	2.60	0.0103	
	Total number of active docks West coast	101620.82	10771.85	9.43	0.0000	
	Number of active docks in Gulf	22650.74	4535.07	4.99	0.0000	
Gravel and mining products—10–14	*Intercept*	261099.50	191985.60	1.36	0.1763	0.6137
	FEMA agriculture building area (m^3)	-0.46	2.75	-0.17	0.8672	
	Total number of private building units (2020–2022)	17.04	3.28	5.19	0.0000	
	FEMA agriculture building area (m^3) × *Total number of private building units (2020–2022)*	6.6×10^{-5}	1.4×10^{-5}	4.65	0.0000	
Coal—15	*Intercept*	1036077.00	1720691.00	0.60	0.5516	0.3842
	Class I track miles	5204.56	1498.45	3.47	0.0016	
	FEMA industrial building area (m^3)	-0.32	0.11	-2.85	0.0079	
	Total number of active docks	49038.01	19381.70	2.53	0.0169	
Other energy products—16–19	*FEMA agriculture building area* (m^3)	4.63	0.49	9.41	0.0000	0.6944
	FEMA industrial building area (m^3)	0.03	0.01	3.93	0.0001	
	Total number of active docks	3929.78	998.68	3.93	0.0001	
Chemicals, wood, and metals—20–33	*Class I track miles*	172.13	160.92	1.07	0.2864	0.8695
	FEMA industrial building area (m^3)	0.25	0.02	13.84	0.0000	
	FEMA agriculture building area (m^3)	10.57	1.09	9.71	0.0000	
	Total number of active docks	5678.25	2083.47	2.73	0.0072	
Manufactured goods, mixed freight—34–40, 43	*Intercept*	797672.97	187364.40	4.26	0.0000	0.8805
	Class I track miles	-2196.62	260.11	-8.45	0.0000	
	FEMA industrial and commercial building area (m^3)	0.03	0.03	0.97	0.3314	
	Class I track miles × *FEMA industrial and commercial building area* (m^3)	2.38×10^{-4}	1.6×10^{-5}	15.18	0.0000	
Waste and unknown—41 & 99	*FEMA agriculture building area* (m^3)	1.29	0.22	5.93	0.0000	0.7707
	FEMA industrial and commercial building area (m^3)	0.03	2.5×10^{-3}	11.60	0.0000	
	Total number of active docks	1022.76	442.77	2.31	0.0224	

Note: The symbol × denotes an interaction between two predictor variables.

Per Table 4 and Table 5, the following are some key observations:

- Class I miles have a positive impact on the number of tons flowed for all regression models except the SCTG 34–40,43 group for destinations. The negative coefficient of this particular group might be due to manufactured goods and mixed freight products mostly going to urban areas, where there are fewer rail tracks.
- In general, the size of the building (industrial, agricultural, commercial) and number of active docks also have a positive impact on the number of tons flowed.

Then, the selected best models are used predict the origin and destination flow volumes (in tons) at the county level by the commodity groups subject to the constraint that negative predicted values are not allowed. In addition, if the predicted volume is more than zero but there are no rail tracks in the county, the predicted volume is set to zero. Next, the predicted volumes are rescaled to match the observed PUWS volumes. If the total of predicted tons is zero but observed PUWS is more than zero, the observed volume is equally distributed for the counties within the BEA zone where rail tracks are available.

Note that the predicted volumes still do not give the total volume that originates or terminates from a given county in the United States as the imports and exports data must be factored in. Therefore, the final step is identifying the domestic origin and domestic destination FAF zones and their corresponding counties for the imported and exported commodities, respectively. The PUWS data obtained from STB can be split into several parts: Domestic origin BEA to domestic destination BEA, foreign origin (Canada or Mexico) to domestic destination (Imports), domestic origin BEA to foreign destination (Canada or Mexico only) (Exports), one foreign origin to another foreign origin, and unknown origin to known destination and vice-versa. While the PUWS contains another category in which the origin and termination BEA zones are identified as “000” to hide sensitive details, this category was not used to construct the ground truth data.

Computing the county-level origin and termination volume is straightforward for the category domestic origin to domestic destination as this is a direct output of the regression predictions. To find the domestic origin FAF zones for the imports and domestic destination FAF zones for the exports, FAF 5.6.1 data first are filtered with rail as the domestic mode, import as trade type, and rail as the foreign in-mode. Then, the dataset is grouped by the foreign origin, domestic origin, domestic destination, and five commodity groups. Then, the percentage of tons flowed through the domestic origin is computed for the five commodity groups. Another set of percentage of tons is computed ignoring the commodity group to account for any mismatches due to the commodity group being unavailable when merging with the PUWS data. This process outputs a set of origin FAF zones for the imports. The same approach can be extended to the exports to identify the domestic destination FAF zones. After origin and destination FAF zones are identified, identifing the corresponding counties within those FAF zones is necessary. This step involves identifying the port of entry counties within the port of entry FAF zones for the imports and identifying counties at ports of exits for the exports. Based on the data available through the TransBorder Freight program, the import and export values are used to identify the port of entry and exit counties.

Once all the missing county information is filled, the dataset is aggregated into origins and destinations by the county and commodity group. Finally, the origins and destinations are rescaled to match the FAF volumes. Here, rescaling means that intead of using the predicted values as is, these values are converted to percentages. These percentages become the final origin and destination factors. By applying these factors to the FAF data, the final total tonnage at origins and destinations are calculated.

3.2.3. Truck

The truck validation data development uses the following steps:

1. Estimate the parameters of the best-fit line that relates various explanatory variables to the number of truck trips.
2. Apply the estimated parameters to county-level data to estimate the number of truck trips originating in or destined to each county.
3. Sum the total number of estimated truck trips in each FAF zone, and then, compute the percentage of truck trips by county in each FAF zone.
4. Use the shares of volume by commodity group from the CFS Subarea data to estimate the share of truck tons by commodity group in each county.

The output of this process is the annual share of truck tons by county and commodity group in each FAF zone. These estimates constitute the truck validation dataset.

This procedure uses OLS to develop the best-fit line. The NextGen truck trip variable is regressed on several explanatory variables. The following explanatory variables were tested:

- Network supply: the number of highway centerline or lane miles
- Port and Transborder activity indicators:
    - Number of docks belonging to principal ports (on the West Coast, East Coast, and Gulf)
    - Inland and Great Lakes trade indicators: the number of principal port docks and other active docks
    - Number of land border crossings
    - Value of goods crossing the land border

BTS developed models for origins and destinations separately. However, the NextGen truck trip table is balanced (truck trip origins and destinations are roughly equal), so BTS also developed a model using a dependent variable that equals the mean of trip origins and destinations. The estimated parameters of this model were similar to the models that use trip origins or trip destinations only. Therefore, for simplicity, the single model that uses the mean of origins and destinations is selected. Additionally, since these models have negative intercept terms, which will create negative estimates of trips when applied, the final model does not have an intercept.

Table 6 presents the estimated parameters for the selected regression model. The explanatory variables in the final model include both centerline miles and indicators of water and Transborder trade activity.

Table 6. Truck Validation Estimates: Model Parameters

	Parameter	Estimate	Std. error	*t* statistic	*p*-value
Centerline miles	In FAF5 & NHS Networks: Urban	55,417	1,312	42.25	<2e-16
	In FAF5 & NHS Networks: Rural	4,434	2,444	1.81	0.0702
	Other (in FAF5 only)	6,784	1,887	3.60	0.0004
Port and Transborder activity indicators	Number of East Coast Principal Port Docks	521,582	94,318	5.53	4.86E-08
	Number of West Coast Principal Port Docks	374,445	88,728	4.22	2.84E-05
	Has U.S.–Mexico Border Crossing (Binary)	4,891,929	141,446	1.56	0.1200
Number of observations		*583*	-	-	-
R-squared		*0.897*	-	-	-
Adjusted R-squared		*0.896*	-	-	-

-Not applicable.

Applying the model from Table 6 to county-level data produces an estimate of total ground truth truck tons for each county. The final step is estimating tonnage by commodity group in each county. The CFS Subarea data provide truck tonnage estimates for 329 subareas in 4 commodity categories: SCTG 01–09, SCTG 10–19, SCTG 20–34, and SCTG 35–43. Since the first category contains all all commodities in SCTG 01–09 in the two datasets, the percentage of flows in this CFS Subarea category is used as-is to estimate the percentage of flows in this category in the county-level validation estimate (Table 7). For example, if a given CFS Subarea has 100 origin tons, and 50 of these tons are in the SCTG 01–09 category, BTS assumes half of the validation tons are SCTG 01–09 for each county that belongs to this subarea. Then, half of the SCTG 10–19 share is assigned to SCTG 10–14 and half to 15–19. The SCTG 20–34 share is assigned to the county-level SCTG 20–33 group. The CFS Subarea SCTG 35–43 share is then assigned to the county-level SCTG 34–99 group.

Table 7. Assigning CFS Subarea Commodity Shares to FAF County-Level Data

CFS subarea category	Assignment of CFS subarea commodity group share to county-level FAF commodity group share
SCTG 01–09	100% of the CFS Subarea share is assigned to FAF group SCTG 01–09
SCTG 10–19	50% of the CFS Subarea share is assigned to FAF group SCTG 10–14 50% of the CFS Subarea share is assigned to FAF group SCTG 15–19
SCTG 20–34	100% of the CFS Subarea share is assigned to FAF group SCTG 20–33
SCTG 35–43	100% of the CFS Subarea share is assigned to FAF group SCTG 34–99

3.3. DEVELOPMENT OF COMPOSITE ESTIMATES

Upon creating the ground truth estimates and computing the estimates from M1, M2, and M3 for origins and destinations at the county level, the three estimates can be used to construct county-level composite estimates by the five commodity groups and three modes (truck, rail, and water). The three methods that are considered include linear regression, log-log regression, and PCR. Use of a regression method allows getting the contribution of each method as a weighted sum toward explaining the ground truth data. Use of a regression method allows getting the contribution of each method as a weighted sum toward explaining the ground truth data.

3.3.1. Linear Regression

Multiple linear regression models are fitted with the ground truth data as the dependent variable, and M1, M2, and M3 are independent variables for origins and destinations by the commodity group. This approach results in a total of 10 multiple linear regression models for each mode. If the intercept term of any of the models is negative, the regression model is refitted without the intercept term.

3.3.2. Log-Log Regression

Log-log transformed regression models are used to check if the adjusted R-squared of composite estimate models can be improved. The predicted values of log-log regression models are transformed to tons by computing the exponential of the predicted value. One advantage of using log transformation of the dependent variable is that, after taking the inverse transformation of the prediction, the tons value would be always positive. However, it has been reported in literature that the log transformed regression models are biased and can under predict the target values [Beauchamp, Olson 1973 and Jia, Rathi 2008]. To rectify this bias, a correction is applied as described by Beauchamp and Olson [1973] and Jia and Rathi [2008].

3.3.3. Principal Component Regression

The estimates generated from M1, M2, and M3 are highly correlated since their explanatory variables are related. In the presence of correlations between predictor variables, multicollinearity can occur with linear regression. Therefore, PCR is also performed as a remedy. Principal component analysis (PCA) is applied to the M1, M2 and M3 data, and the number of principal components (PCs) required for the regression is determined by using the percentage of variance explained criteria with a 95 percent cutoff.

3.3.4. Steps to Develop Final Composite Estimates

Several high-level observations, based on the results of linear regression, log-log regression, and pricipal component regression, follow:

- The pairwise collection coefficients as shown in Section 4.1. indicate that all three disaggregation methods are highly correlated. This correlation confirms the intuition that these methods allocate the commodity flows in a similar way to the county level regardless of which indicators (total employment versus industry-specific employment versus payroll) are used.
- Because M2 and M3 rely on industry-specific employment and payroll data to allocate the OD flows to the county level, the estimated flow of a given commodity is zero for with either no employment or suppressed data in the specified industries. However, in reality, the county may generate or attract flows of the commodity. A composite process helps create estimates for counties with no CBP employment in specific industries.

The following steps were applied to generate the final composite measures:

1. If the multiple linear regression model predicted any values to be negative, but the volume estimates of M1, M2, M3 and ground truth are more than zero, then the average of all four values is used as the estimated volume.
2. If a particular county does not contain the required network facilities to have commodity flows for a specific mode, those counties should have zero tons for that mode. For example, if a given county does not have any rail tracks, then no rail flows should be assigned to it.
3. After computing the predicted values and adjusting them based on the constraints, the predicted values are rescaled so that the FAF zone totals are maintained. In cases where the final composite estimate does not capture any volume for a given FAF zone or commodity group, factors based on employment and population (same as M1) are applied. An alternative approach may be to instead use the mean result from all three estimation techniques (M1, M2, and M3).

3.4. METHODS FOR OTHER MODES

The preceding discussion in this section applies to truck, rail, and water modes. BTS applied different methods to air, multiple modes and mail, and pipeline flows.

Flows by air and multiple modes and mail use the same disaggregation factors as truck flows because air is a small percentage of tons (less than 0.1 percent), and therefore, this experimental product groups air flows together with truck-only flows for simplicity. For multiple modes and mail, validation data were not available; BTS applied the truck factors to this mode due to this data gap and in acknowledgement of the fact that many of these flows use trucking at one end of the trip.

BTS developed county-level pipeline flow estimates as follows:

1. The network constraints from Section 2.3 are applied to the disaggregation factors developed from M3 to ensure that pipeline flows begin and end in counties that appear to have pipeline origination or termination points (similar to the rail and water processing). The resulting flows are then rescaled to ensure that the county-based sums match the FAF zonal flows.
2. Some combinations of FAF zones and commodity groups have pipeline flows in the FAF5.6.1 data, but are not captured in the rescaled M3 results. When this happens, factors are developed based on the M1 method and are applied to the FAF5.6.1 data to ensure that that the FAF zone flows can be disaggregated.

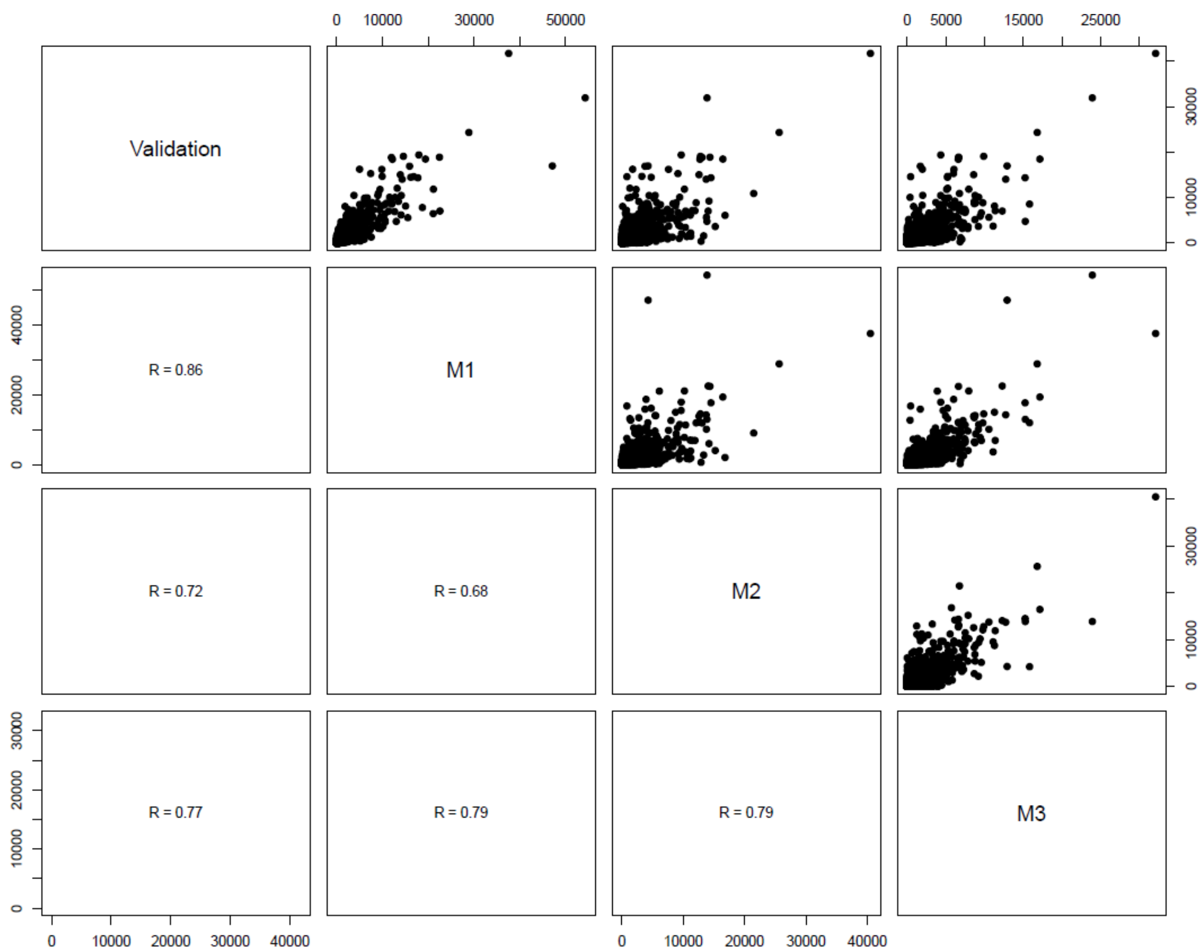
4. Results

This section summarizes the results of computing the composite estimate of county-to-county flows. M1, M2, and M3 refer to three estimation methods—proportional allocation, Cambridge Systematics method, and ORNL method, respectively, as detailed in Section 3.1.

4.1. SCATTERPLOTS BETWEEN DISAGGREGATION METHODS AND VALIDATION DATA

Figure 3 illustrates the relationship between the validation data and the estimates generated from M1, M2, and M3, as well as the relationships between the three methods for the truck origins for the SCTG 01–09 commodity group. Figure 3 can be interpreted by following the variable names and identifying their intersections. For example, reading horizontally along the Validation variable and vertically along the M2 variable leads to the third cell in the top row, which is a scatterplot of Validation versus M2. To find the correlation coefficient, follow the Validation variable vertically and the M2 variable horizontally to the third cell in the first column, showing a correlation coefficient (R) of 0.72.

Figure 3. Pairwise Matrix Scatterplot Between Validation Data, M1, M2, and M3 With Pairwise Correlations: SCTG 01–09 Truck Origin Data



Source: BTS.

After obtaining all the pairwise relationship between the variables as explained, the validation data have a high positive linear correlation coefficient (above 0.7) with all three methods of estimation. The linear correlation coefficient values are also high among M1, M2, and M3, which is expected as all three methods estimate the same variable of interest. Due to the strong linear relationship between validation data with M1, M2, and M3, linear regression analysis can be performed. However, although there is a high linear positive correlation for the truck modes origins and destinations for all commodity groups, the linear correlation coefficient can get as low as 0.3 for certain commodity groups within rail and water modes. Refer to Appendix D for the remaining pairwise scatterplots.

4.2. MULTIPLE LINEAR REGRESSION RESULTS

Per Table 8 (truck origins) and Table 9 (truck destinations), for the commodity group SCTG 01–09, the OLS regression model has the highest adjusted R-squared. All three methods are statistically significant in predicting the validation data. M1 often has a higher coefficient in predicting validation data as given in Table 8 and Table 9. However, some regression models indicate higher weights for M2 or M3 depending on the specific commodity groups and modes being analyzed. The remaining composite estimate regression results are presented in Appendix D.

Table 8. Composite Estimate Linear Regression Without Transformation (OLS) and With Log Transformation (Log-Log): SCTG 01–09 Truck Origin Data

Model	Parameter	Estimate	Std. error	*t* statistic	*p*-value	Adj R-sq
OLS	*Intercept*	133.27	19.84	6.72	0.0000	0.7813
	M1	0.54	0.01	47.69	0.0000	
	M2	0.22	0.01	15.16	0.0000	
	M3	0.10	0.02	5.08	0.0000	
Log-log	*Intercept*	2.11	0.06	34.13	0.0000	0.5662
	log1p(*M*1)	0.57	0.02	35.60	0.0000	
	log1p(*M*2)	0.17	0.01	17.59	0.0000	
	log1p(*M*3)	-0.07	0.01	-6.01	0.0000	

Note: log1p() denotes log(value + 1).

Table 9. Composite Estimate Linear Regression Without Transformation (OLS) and With Log Transformation (Log-Log): SCTG 01–09 Truck Destination Data

Model	Parameter	Estimate	Std. error	*t* statistic	*p*-value	Adj R-sq
OLS	*Intercept*	133.27	19.84	6.72	0.0000	0.7997
	M1	0.54	0.01	47.69	0.0000	
	M2	0.22	0.01	15.16	0.0000	
	M3	0.10	0.02	5.08	0.0000	
Log-log	*Intercept*	1.27	0.06	20.24	0.0000	0.6567
	log1p(*M*1)	0.62	0.02	34.01	0.0000	
	log1p(*M*2)	0.23	0.01	18.25	0.0000	
	log1p(*M*3)	-0.05	0.01	-4.77	0.0000	

Note: log1p() denotes log(value + 1).

To further investigate the effects of multicollinearity, BTS performed PCA. The PCA results indicated that the first PC captures most of the explained variance (at least 80–90 percent), and PC loadings for each variable are approximately equal across all commodity groups and modes for the first PC. Then, BTS performed PCR. However, it did not improve the adjusted R-squared value. Therefore, the PCA or PCR results are not included in this document.

Upon comparing all the regression analysis results for the composite estimates, the following conclusions are reached:

- For truck origins and destinations, the OLS regression model has the highest adjusted R-squared among all commodity groups.
- OLS regression is also the best model in terms of adjusted R-squared for the rail mode in all commodity groups except SCTG 01–09 rail origins.
- For water origins and destinations, the log-log regression has the highest adjusted R-squared for the commodity groups SCTG 01–09 and 10–14.

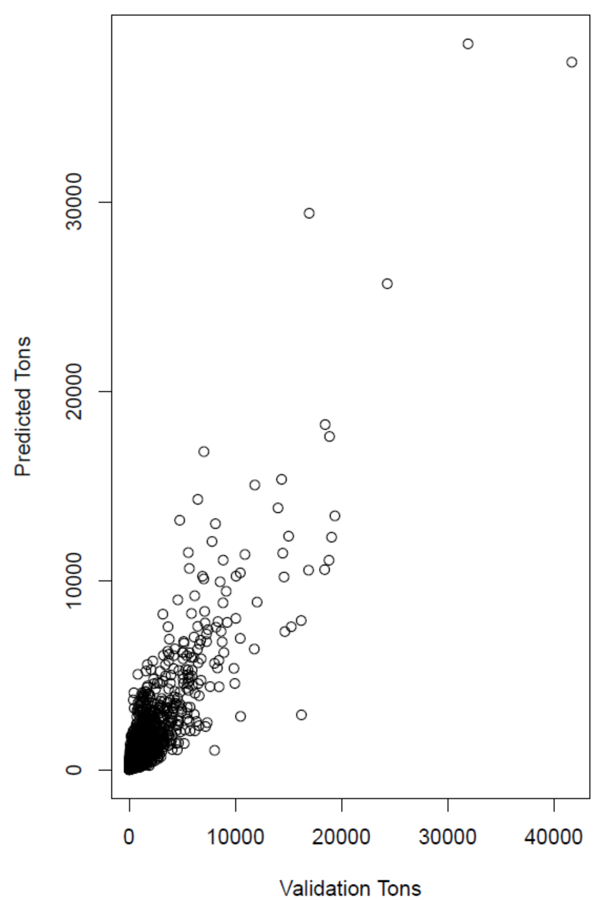
In general, the OLS regression model performed better than the other two models.

4.3. COMPARISON BETWEEN VALIDATION DATA AND COMPOSITE RESULTS

Figure 4 is a scatterplot illustrating the relationship between predicted and validated tons for truck origins and destinations for SCTG 01–09.

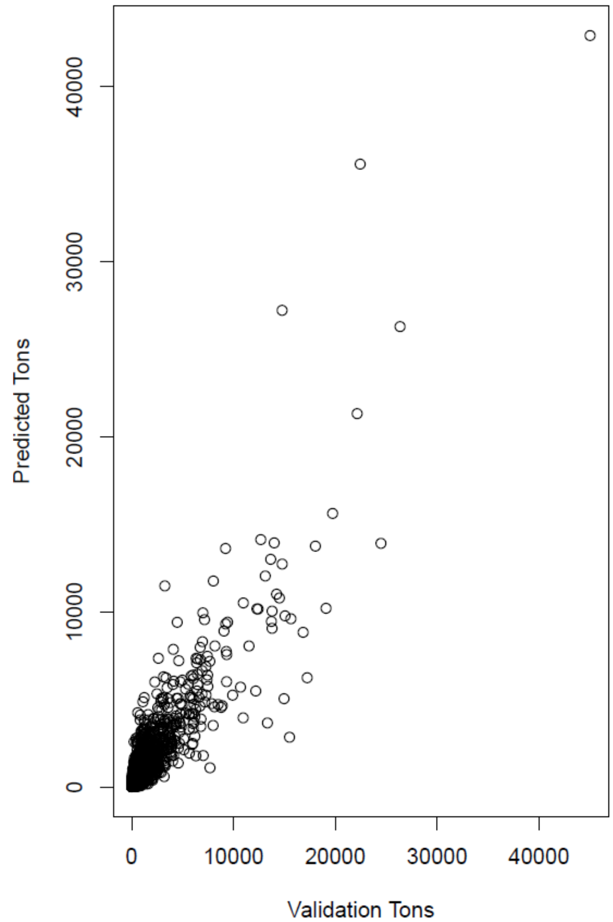
Figure 4. Predicted Versus Validated Tons for SCTG 01–09 Truck Origins and Destinations

A. Origins



Source: BTS.

B. Destinations



Source: BTS.

Figure 4 illustrates that, in general, the predicted and observed values have good agreement. Please refer to Appendix E for the remaining figures.

4.3.1. Statistical Summary of Validation Results and Composite Results From OLS and Log-Log (Tables for Min, Median, Max)

Table 10 and Table 11 show summary statistics of the validation results and composite estimates.

Table 10. Summary Statistics of Validation Data, M1, M2, M3, and Rescaled OLS and Log-Log Composite Estimates: SCTG 01–09 Truck Origins

Variable	Min	Median	Mean	Max	Sum
Validation	0.01	346.89	942.21	41,689.67	2,962,312
M1	0.00	249.24	942.21	54,273.70	2,962,312
M2	0.00	259.52	942.20	40,523.56	2,962,288
M3	0.00	355.75	942.09	32,086.90	2,961,917
OLS	36.17	371.15	942.21	38,353.79	2,962,312
Log-log	4.61	438.50	942.21	30,347.72	2,962,312

Table 11. Summary Statistics of Validation Data, M1, M2, M3, and Rescaled OLS and Log-Log composite estimates: SCTG 01–09 Truck Destinations

Variable	Min	Median	Mean	Max	Sum
Validation	0.01	331.64	942.21	45,021.19	2,962,312
M1	0.08	330.10	942.21	42,260.29	2,962,312
M2	0.00	244.55	942.20	48,192.15	2,962,288
M3	0.00	594.10	942.09	39,786.57	2,961,932
OLS	20.25	396.26	942.21	42,909.68	2,962,312
Log-log	2.31	396.79	942.21	38,041.52	2,962,312

The median and the maximum values of the rescaled OLS regression are much closer to the validation targets compared to the rescaled log-log regression esimates in both origins and destinations.

The minimum value of the validation data is much smaller than the rescaled OLS predicted value for the data in Table 10 and Table 11. However, this pattern is only observed for all commodity groups except SCTG 15–19 in truck origins and SCTG 01–09 in truck destinations. Rail and water modes' minimum values are closer to the validation data regardless of commodity group in both origins and destinations for both OLS and log-log rescaled estimates.

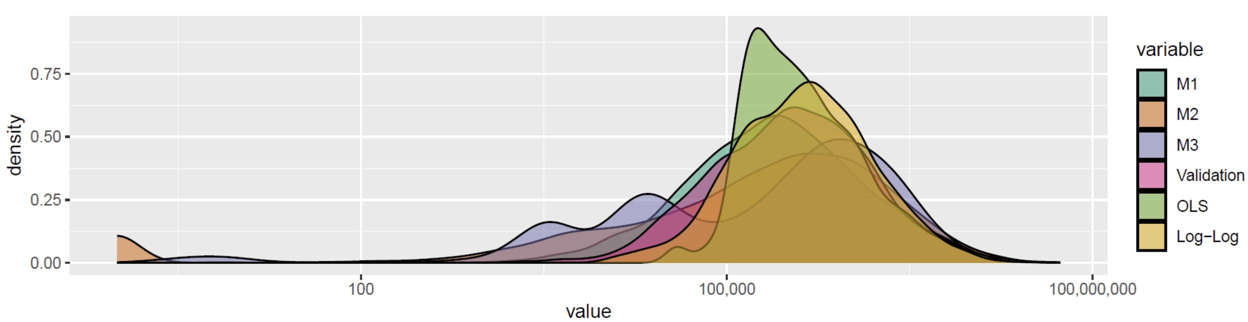
In general, the rescaled OLS values seem to perform better in terms of capturing the central tendency and the maximum value.

4.3.2. Area Under the Curve Comparison Among Validation Results, Composite Results From OLS and Log-Log (Graphics)

The overlaid density plots allow comparing the overall distribution (shape, spread, and central tendency) of the variables of interest. Some key observations to look for are alignment of the peaks, width of the curves (spread), shape of the curves, and overlapping areas. In Figure 5 and Figure 6, the x-axes in the overlaid density curve plots are in a log scale to illustrate the distributions clearly.

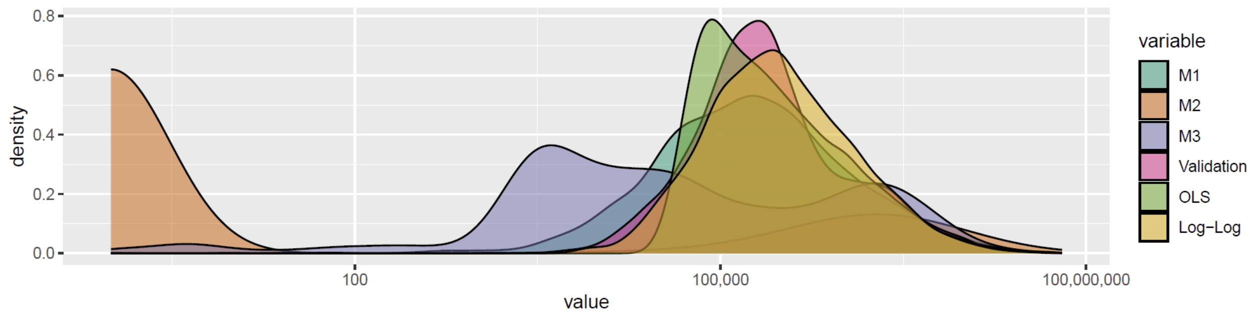
Figure 5. Overlaid Density Curves for M1, M2, M3, Validation Data, Rescaled Predicted Values of OLS, and Log-Log: Truck Origins

A. SCTG 01–09



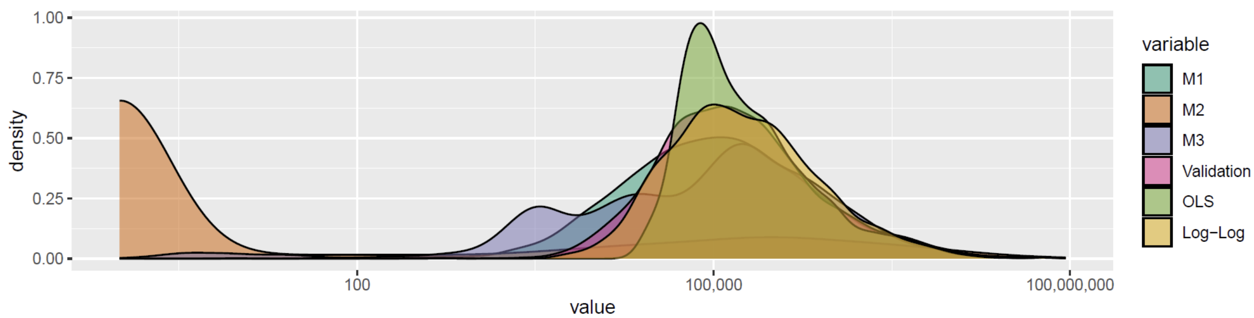
Source: BTS.

B. SCTG 10–14



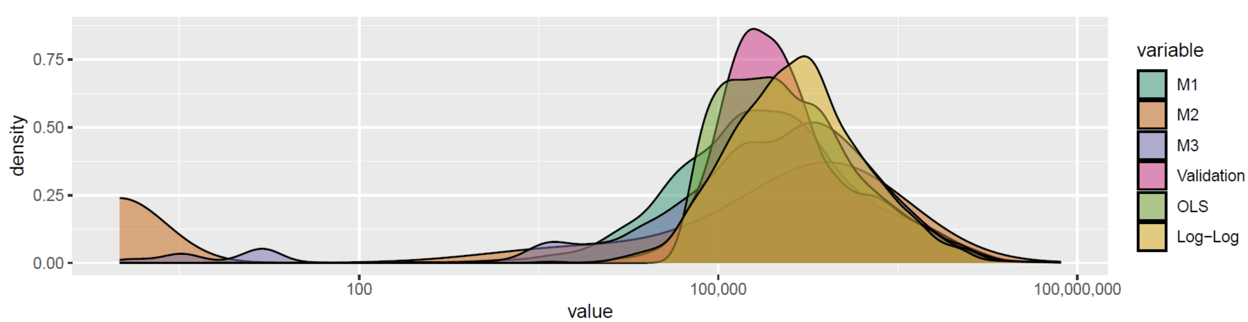
Source: BTS.

C. SCTG 15–19



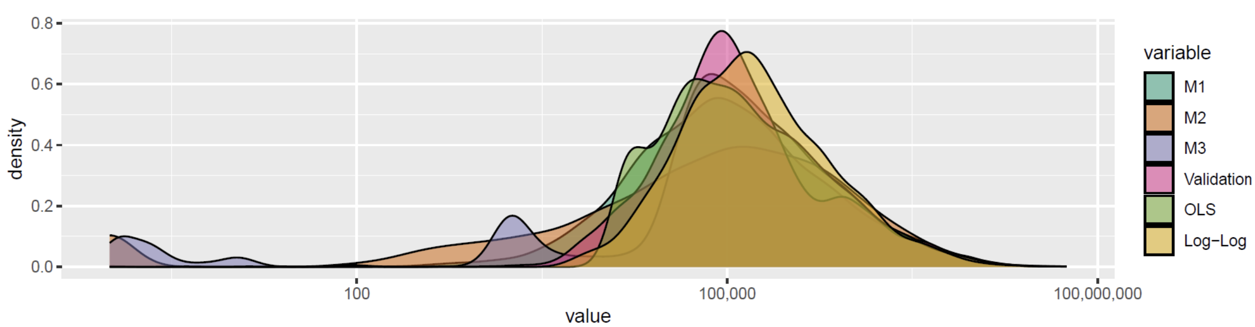
Source: BTS.

D. SCTG 20–33



Source: BTS.

E. SCTG 34–99

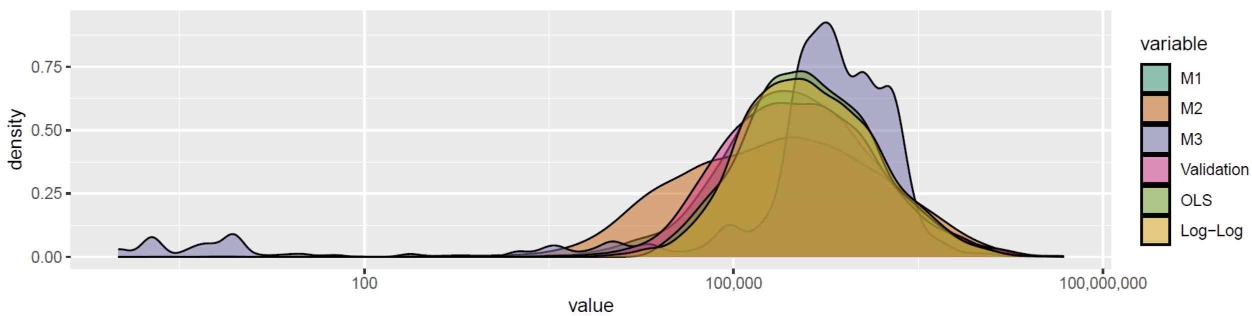


Source: BTS.

Based on Figure 5, Figure 6, and the figures in Appendix G, we can conclude that in general, all the computed volumes are centered near a similar value. There is a high degree of overlap between the curves, indicating that predicted-rescaled values closely match the observed validation data. The M2 values seem to concentrate around two regions for all the origins regardless of the mode. The concentration at zero is due to M2's use of industry-specific employment, which is zero for some counties in the CBP data. However, the destination values for M2 seem to closely follow the other estimated values across all modes and commodity groups.

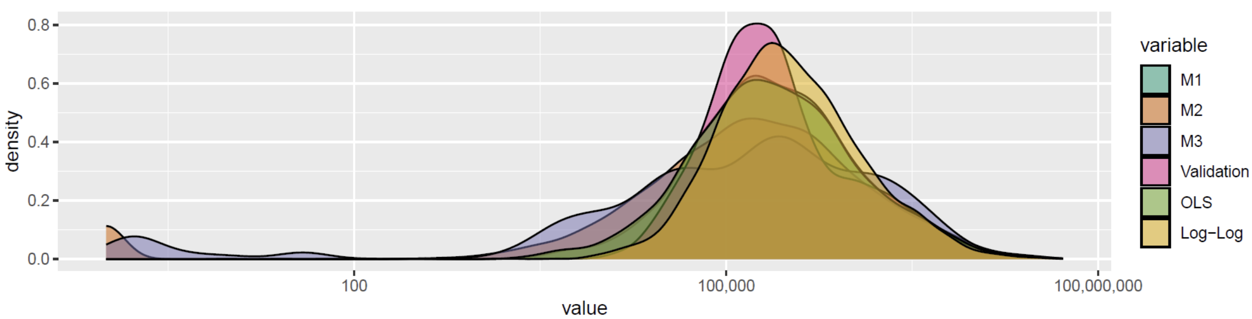
Figure 6. The Overlaid Density Curves for M1, M2, M3, Validation Data, Rescaled Predicted Values of OLS, and Log-Log: Truck Destinations

A. SCTG 01–09



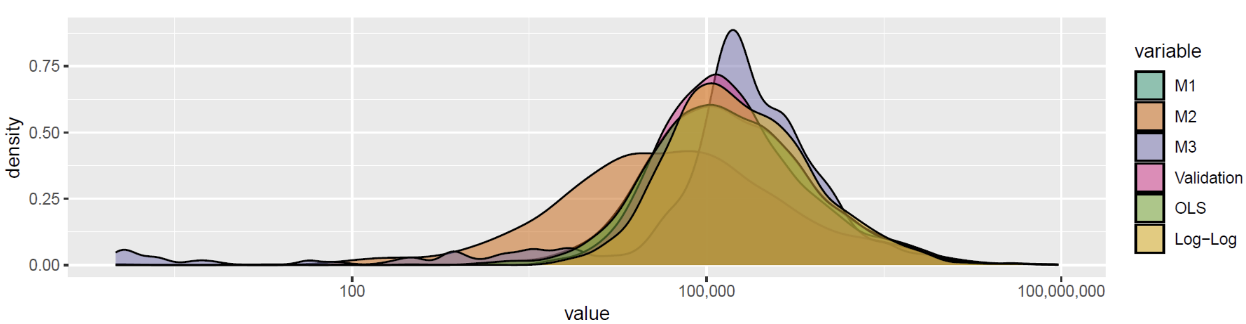
Source: BTS.

B. SCTG 10–14



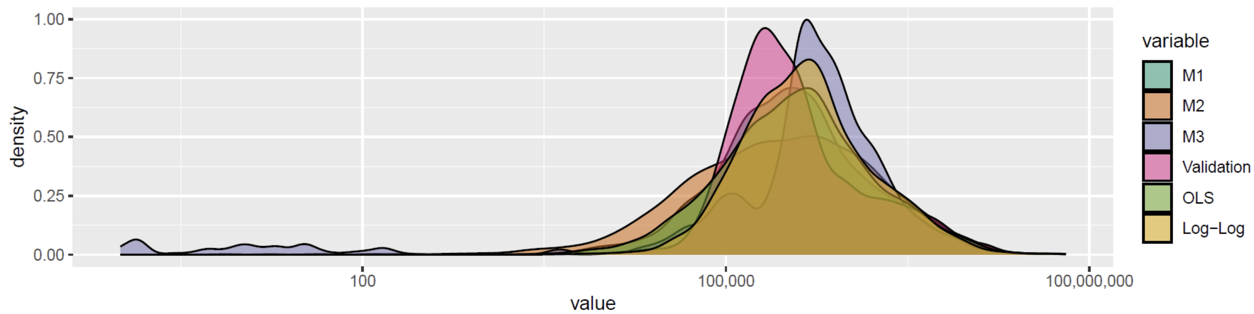
Source: BTS.

C. SCTG 15–19



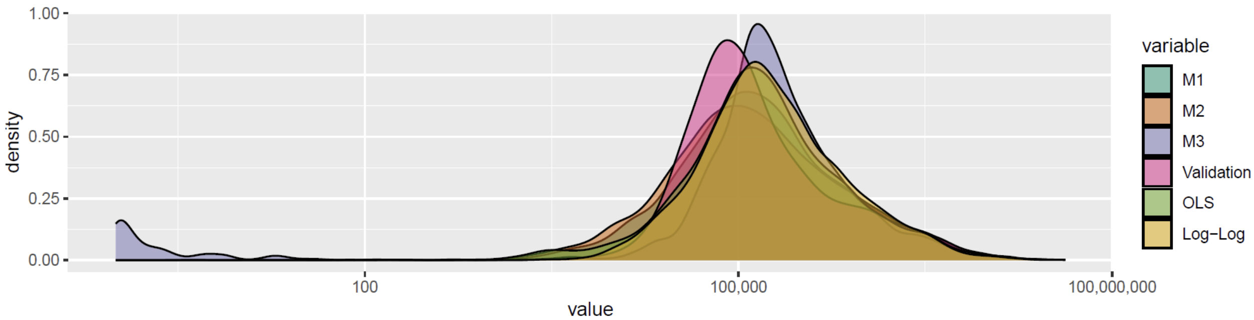
Source: BTS.

D. SCTG 20–33



Source: BTS.

E. SCTG 34–99



Source: BTS.

5. User Guide

Users can access files for this data release at https://www.bts.gov/faf/county. Data users can download two types of files: state-specific files and disaggregation factors for all counties in the United States. Table 12 lists the five commodity groups that the release files use to refer to commodity type. The *sctgG5* field contains this information. Table 13 lists the mode groups and their codes, which the the state-level files use in the *dms_mode* field. This webpage provides a detailed explanation of these release files. The release files exclude FAF modes 7 and 8.

Table 12. SCTG Group Codes in the Disaggregation Data Products

sctgG5 code	Definition	FAF SCTG code
sctg0109	Agricultural products	01–09
sctg1014	Gravel and mining products	10–14
sctg1519	Coal and other energy products	15–19
sctg2033	Chemical, wood, and metals	20–33
sctg3499	Manufactured goods, mixed freight, waste, and unknown	34–99

Table 13. New Mode Code in Disaggregation Factor Tables

New mode code	New mode definition	FAF mode code
11	Truck and air	1—Truck 4—Air
2	Rail	2—Rail
3	Water	3—Water
5	Multiple modes and mail	5—Multiple modes and mail
6	Pipeline	6—Pipeline

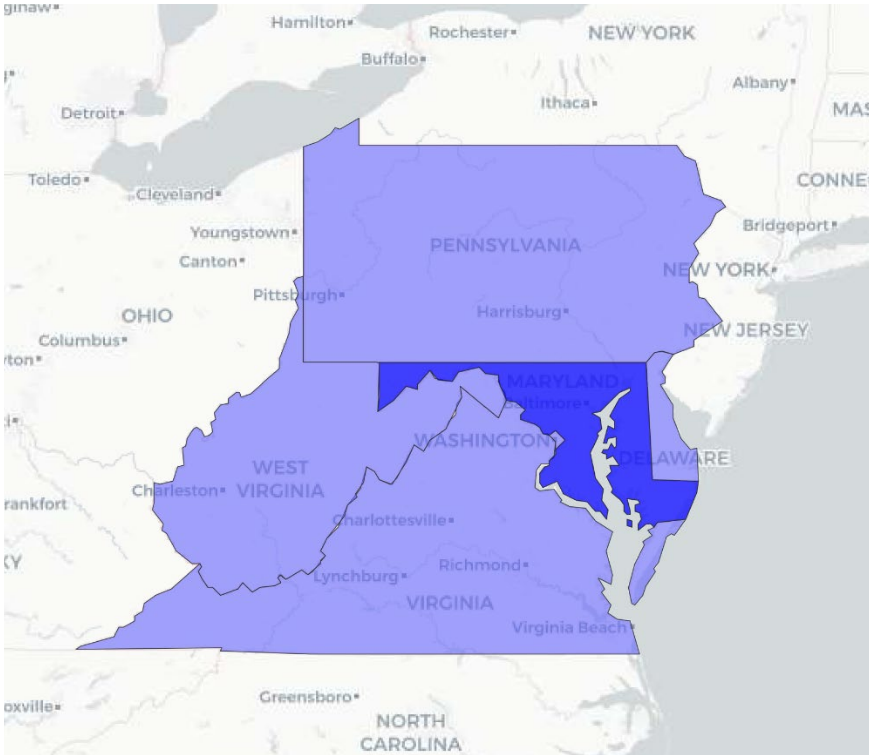
5.1. STATE-BASED DISAGGREGATION RESULTS

The release includes 51 state-specific zip files (including one zip file for Washington, DC). Each file represents flows using county-level geography for the main state and for states adjacent to the main state. Geographic representation of flows outside this area consists of FAF zones. Each state zip file contains four tables representing county-level OD flows for the state of interest and every adjacent state, county-to-FAF zone OD flows from the multistate area to all other FAF zones, FAF zone-to-county OD flows from all other FAF zones to the multistate area, and FAF zone-to-FAF zone OD flows from all other FAF zones to all other FAF zones.

As an example, Figure 7 shows Maryland and its multistate area, which includes four adjacent states (Pennsylvania, Delaware, Virginia, and West Virginia) and the District of Columbia. The zip file for this multistate area contains four tables with the following information:

- Table 1: county-to-county commodity flows between all counties in this multistate area
- Table 2: county-to-FAF zone commodity flows from counties in the multistate area to the FAF zones outside this area
- Table 3: FAF zone-to-county commodity flows from the FAF zones outside the multistate area to counties in the multistate area
- Table 4: FAF zone-to-FAF zone commodity flows between all other FAF zones outside the multistate area.

Figure 7. Maryland and Surrounding States



Source: BTS and OpenStreetMap 2024.

5.2. DISAGGREGATION FACTORS: DESCRIPTION AND APPLICATION

The full set of county-level factors (origin and destination factors) are contained in one zip file. Users can merge these factors with FAF regional databases to create a county-level database for a customized geographic area or for the entire United States. Users will need to download the FAF regional databases from www.bts.gov/faf. The FAF database download includes metadata with variable dictionaries. The factors are available for four modes and five commodity groups. The zip file contains files with disaggregation factors for the following four modes:

- Rail (rail_origin_factors.csv and rail_destination_factors.csv)
- Water (water_origin_factors.csv and water_destination_factors.csv)
- Truck (truck_origin_factors.csv and truck_destination_factors.csv)
- Pipeline (pipeline_origin_factors.csv and pipeline_destination_factors.csv)

Users can apply the truck factors to air and multiple modes and mail modes to generate county-to-county flows for these modes. This experimental product does not include methods to disaggregate flows by other or unknown mode, or flows that use no domestic mode.

Table 14 explains the variables in the origin factor files. The data structure in the destination factor file is the same. The factor files include the county FIPS code, the county's corresponding FAF zone code, commodity group, and the disaggregation factor for that combination of county and commodity group.

Table 14. Example of FAF Origin Factor Table

Variable name	Description
dms_orig	Origin FAF code
dms_orig_cnty	Origin county FIPS code
sctgG5	Commodity group code
f_orig	The proportion of FAF zone tons of sctgG5 that originate in this county

Users can apply the process that Figure 8 illustrates to disaggregate FAF zone flows to the county level. The steps are as follows:

1. (Optional) Select FAF trips for the area of interest. This step will reduce the size of the input dataset, which improves efficiency in the remaining disaggregation steps. For example, to generate county-to-county tonnage estimates only for Iowa to Idaho, the user can select flows with *dms_orig* = 190 and *dms_dest* = 160, filtering out all other flows.
2. Select FAF trips for the mode of interest. For example, to disaggregate rail flows, select flows with *dms_mode* = 2 (rail) from the previous step, filtering out all other flows.
3. Use the commodity correspondence from Table 1 to summarize the flows from the previous step (red box in Figure 1) using the five commodity groups. This step will reduce the 42 commodity categories to 5.
4. Join the resulting table from the previous step to the origin factor table based on the FAF origin zone (*dms_orig*) and commodity group (*sctgG5*) columns. The resulting file now contains the factor for the origin county and commodity group (*f_orig*).
5. Join the resulting table from the previous step to the destination factor table based on the FAF destination zone (*dms_dest*) and commodity group (*sctgG5*) columns. The resulting file now contains the factor for the destination county and commodity group (*f_dest*).

The resulting table contains all possible combinations of county pairs and commodity groups that are present in the original FAF estimates for the selected area and mode. Users can calculate the final county-to-county tonnage by commodity group by multiplying the total FAF zone-level tons by the origin factor and the destination factor. The Figure 8 note shows these formulas.

Figure 8. Example: Applying Disaggregation Factors

FAF Table with hypothetical numbers

fr_orig	dms_orig	dms_dest	fr_dest	sctg2	…	tons_2022
	A	B		1		18,000
	A	B		2		12,000
	A	B		5		3,000
	A	B		9		2,000

FAF table with aggregated results

fr_orig	dms_orig	dms_dest	fr_dest	sctgG5	..	tot_tons_2022
	A	B		sctg0109	···	35,000

dms_orig	dms_orig_cnty	sctgG5	f_orig
A	A1	sctg0109	0.333
A	A2	sctg0109	0.333
A	A3	sctg0109	0.167
A	A4	sctg0109	0.167

Origin Factor

dms_dest	dms_dest_cnty	sctgG5	f_dest
B	B1	sctg0109	0.250
B	B2	sctg0109	0.125
B	B3	sctg0109	0.125
B	B4	sctg0109	0.500

Destination Factor

fr_orig	dms_orig	dms_dest	fr_dest	dms_orig_cnty	dms_dest_cnty	sctgG5	f	tons_2022_c2c
	A	B		A1	B1	sctg0109	0.083	291.67
	A	B		A1	B2	sctg0109	0.042	145.83
	A	B		A1	B3	sctg0109	0.042	145.83
	A	B		A1	B4	sctg0109	0.167	583.33
	…	…		…	…	…	…	…
	A	B		A4	B4	sctg0109	0.083	291.67

Source: BTS.
Note: *f* = *f_orig* × *f_dest*, *tons_2022_c2c* = *tons_2022* × *f*.

6. References

Ashley, Harrison, Joe Thomas, Greg Holt, and Thomas Valco. 2018. "Engineering and Ginning: Cottonseed Air-Handling and Storage Requirements," *Journal of Cotton Science*, 22, no. 1. The Cotton Foundation.

Beauchamp, John J., and Jerry S. Olson. 1973. "Corrections for Bias in Regression Estimates After Logarithmic Transformation," *Ecology*, 54, no. 6.Washington, DC: Ecological Society of America. https://doi.org/10.2307/1934208. Last accessed January 2, 2025.

Berthaume, Andrew, and Tom Morton. 2015. *National Multimodal Freight Analysis Framework Research Workshop: Workshop Summary Report, December 11, 2013*. Washington, DC: U.S. Department of Transportation, Federal Highway Administration. https://www.fhwa.dot.gov/publications/research/ear/14054/14054.pdf. Last accessed January, 2, 2025.

BTS. 2025a, forthcoming. *Development of the Freight Analysis Framework Version 5 (FAF5) Annual Estimates*. Washington, DC: U.S. Department of Transportation, Bureau of Transportation Statistics.

BTS. 2025b, forthcoming. Data from "National Transportation Atlas Database." Multimodal Network datasets. Washington, DC: U.S. Department of Transportation.

BTS. 2025c. "National Transportation Atlas Database." Washington, DC: U.S. Department of Transportation. https://geodata.bts.gov/. Last accessed December 31, 2024.

BTS. 2024a. *Freight Analysis Framework (FAF) Versions*. Washington, DC: Bureau of Transportation Statistics. https://www.bts.gov/faf/versions. Last accessed December 18, 2024.

BTS. 2024b. *TransBorder Freight Data*. Washington, DC: Bureau of Transportation Statistics. https://data.bts.gov/stories/s/TransBorder-Freight-Data/myhq-rm6q/. Last accessed October 11, 2024.

BTS. 2024c. "National Transportation Atlas Database." Intermodal Freight Facilities Pipeline Terminals. Washington, DC: U.S. Department of Transportation. https://geodata.bts.gov/. Last accessed December 31, 2024.

BTS. 2024d. "National Transportation Atlas Database." North American Rail Network Lines. Washington, DC: U.S. Department of Transportation. https://geodata.bts.gov/. Last accessed August 13, 2024.

BTS and Census. 2021. *2017 Commodity Flow Survey Subarea Estimates*. Washington, DC: U.S. Census Bureau. https://www.census.gov/data/experimental-data-products/commodity-flow-survey-subarea-estimates.html. Last accessed October 11, 2024.

BTS and Census. 2020. *2017 Commodity Flow Survey Methodology*. Washington, DC: U.S. Census Bureau. https://www2.census.gov/programs-surveys/cfs/technical-documentation/methodology/2017cfsmethodology.pdf. Last accessed December 18, 2024.

BTS and FHWA. 2024. *Freight Analysis Framework*. Washington, DC: Bureau of Transportation Statistics. https://www.bts.gov/faf. Last accessed December 18, 2024.

BTS and FHWA. 2022. Data from "Freight Analysis Framework." FAF Model Highway Network dataset. Washington, DC: U.S. Department of Transportation, Federal Highway Administration. https://ops.fhwa.dot.gov/freight/freight_analysis/faf/. Last accessed October 7, 2024.

BTS and OpenStreetMap. 2024. *Maryland and Surrounding States*. Map made by BTS using R. Washington, DC: Bureau of Transportation Statistics.

Cambridge Systematics. 2009. *Development of a Computerized Method to Subdivide the FAF2 Regional Commodity OD Data to County Level OD Data*. Washington, DC: Federal Highway Administration.

Census. 2024a. *County Business Patterns*. Washington, DC: U.S. Census Bureau. https://www.census.gov/programs-surveys/cbp.html. Last accessed December 1, 2024.

Census. 2024b. *County Population Totals and Components of Change: 2020-2023*. Washington, DC: U.S. Census Bureau. https://www.census.gov/data/tables/time-series/demo/popest/2020s-counties-total.html. Last accessed December 1, 2024.

Census. 2024c. *Building Permits Survey (BPS)*. Washington, DC: U.S. Census Bureau. https://www.census.gov/construction/bps/index.html. Last accessed November 5 2024.

EIA. 2022. "API," *Open Data.* Washington, DC: U.S. Energy Information Administration. https://www.eia.gov/opendata/. Last accessed December 20, 2024.

Esri. 2024. *Surface and Underground Coal Mines in the U.S*. Redlands, CA: Esri, ArcGIS Hub. https://hub.arcgis.com/datasets/fedmaps::surface-and-underground-coal-mines-in-the-u-s--2/about. Last accessed November 23 2024

FHWA. 2023. "2022 Data Highlights," *NextGen NHTS National OD Data*. Washington, DC: U.S. Department of Transportation, Federal Highway Administration. https://nhts.ornl.gov/od/. Last accessed October 23, 2024.

Fischer, Michael, Jeffrey Ang-Olson, and Anthony La. 2000. "External Urban Truck Types Based on Commodity Flows: A Model," *Transportation Research Record: Journal of the Transportation Research Board*, 1707, no. 1. Thousand Oaks, CA: Sage Journals. https://doi.org/10.3141/1707-09. Last accessed January 2, 2025.

Global Energy Monitor. 2024. *Global Coal Plant Tracker*. Covina, CA: Global Energy Monitor. https://globalenergymonitor.org/projects/global-coal-plant-tracker/. Last accessed December 20, 2024.

Golias, Mihalis, Sabyasachee Mishra, Martin Lipinski, Karlis Pujats, Mitra Salehi Esfandarani, Neda Nazemi, and Hana Takhtfiroozeh. 2021. *Comprehensive Planning Guidebook for Commodity and Freight Movement in Tennessee*. Memphis, Tennessee: Tennessee Department of Transportation. https://www.tn.gov/content/dam/tn/tdot/research/final-reports/2019-final-reports-and-summaries/RES2019-14_Final_Report_Approved.pdf. Last accessed January 2, 2025.

Hwang, Ho-Ling, Hyeonsup Lim, Shih-Miao Chin, Majbah Uddin, Alec Biehl, Fei Xie, Stephanie Hargrove, Yuandong Liu, and Chieh (Ross) Wang. 2021. *Freight Analysis Framework Version 5 (FAF5) Base Year 2017 Data Development Technical Report*. Washington, DC: U.S. Department of Transportation. https://rosap.ntl.bts.gov/view/dot/68154. Last accessed January 2, 2025.

Jia, Siwei, and Sarika Rathi. 2008. "On Predicting Log-Transformed Linear Models with Heteroscedasticity," *Statistics and Data Analysis*. Cary, NC: SAS Global Forum. https://support.sas.com/resources/papers/proceedings/pdfs/sgf2008/370-2008.pdf. Last accessed January 2, 2025.

Lambert, Joshua, Liyu Gong, Corrine F. Elliott, Katherine Thompson, and Arnold Stromberg. 2018."rFSA: An R Package for Finding Best Subsets and Interactions," *R Journal*, 10, no. 2. Indianapolis, IN: The R Foundation. https://doi.org/10.32614/RJ-2018-059. Last accessed January 2, 2025.

Lumley, Thomas. 2024. *Leaps: Regression Subset Selection*. R Package Version 2.9. http://CRAN.R-project.org/package=leaps. Last accessed January 2, 2025.

Machine & Process Design. n.d. *Bulk Density*. Anoka, MN: Machine & Process Design. https://www.mpd-inc.com/bulk-density/. Last accessed December 29, 2024.

Maricopa Association of Governments. 2018. *Mega-Regional Multi-Modal Agent-based Behavioral Freight Model: Final Report*. Phoenix, AZ: Maricopa Association of Governments.

National Academies of Sciences, Engineering, and Medicine. 2013. *Guidebook for Developing Subnational Commodity Flow Data*. Washington, DC: The National Academies Press. https://doi.org/10.17226/22523. Last accessed January 2, 2025.

NCC. n.d. *U.S. Cotton Bale Dimensions*. Cordova, NT: National Cotton Council of America. https://www.cotton.org/tech/bale/bale-description.cfm. Last accessed December 20, 2024

Opie, Keir, Jakub Rowinski, and Lazar Spasovic. 2009. "Commodity-Specific Disaggregation of 2002 Freight Analysis Frameowrk Data to County Level in New Jersey," *Transportation Research Record: Journal of the Transportation Research Board*, 2121, no. 1. Thousand Oaks, CA: Sage Journals. https://doi.org/10.3141/2121-14. Last accessed January 2, 2025.

ORNL. n.d. *Processes for FAF4 Ton-Miles Estimation*. Unpublished.

ORNL and FEMA. 2024. *USA Structures*. Washignton, DC: FEMA Geospatial Resource Center. https://gis-fema.hub.arcgis.com/pages/usa-structures. Last accessed September 25, 2024

Ranaiefar, Fatemeh, Joseph Chow, and Stephen Ritchie. 2013. "Structural Commodity Generation Model that Uses Public Data: Geographic Scalability and Supply Chain Elasticity Analysis," *Transportation Research Record: Journal of the Transportation Research Board*, 2378, no. 1. Thousand Oaks, CA: Sage Journals. https://doi.org/10.3141/2378-08. Last accessed January 2, 2025.

Sorratini, Jose, and Robert Smith Jr. 2000. "Development of a Statewide Truck Trip Forecasting Model Based on Commodity Flows and Input-Output Coefficients," *Transportation Research Record: Journal of the Transportation Research Board*, 1707, no. 1. Thousand Oaks, CA: Sage Journals. https://journals.sagepub.com/doi/epdf/10.3141/1707-06. Last accessed January 2, 2025.

STB and RAILINC. 2024a. *2022 Public Use Waybill Sample*. Washington, DC: Surface Transportation Board. https://www.stb.gov/wp-content/uploads/PublicUseWaybillSample2022.zip. Last accessed January 2, 2025.

STB and RAILINC. 2024b. *2022 Surface Transportation Board Carload Waybill Sample Reference Guide*. Washington, DC: Surface Transportation Board.

USACE. 2023a. *Waterborne tonnage for principal U.S. ports and all 50 states and U.S. territories*. https://usace.contentdm.oclc.org/digital/collection/p16021coll2/id/1492. Last accessed September 23, 2024.

USACE. 2023b. *2022- State to State Commodity Tonnages Public Domain Database in Excel Format*. https://www.iwr.usace.army.mil/About/Technical-Centers/WCSC-Waterborne-Commerce-Statistics-Center/. Last accessed October 2, 2024.

USACE. 2023c. *[2000-2022 Trips] Manuscript cargo and trips data files, statistics on foreign and domestic waterborne commerce move on the United States waters*. Washington, DC: U.S. Army Corps of Engineers Digital Library. https://usace.contentdm.oclc.org/digital/collection/p16021coll2/id/1690/. Last accessed September 24, 2024

USDA. 2014. "Ⅰ. General," *Commodity Specifications Dry, Edible Beans, Peas, and Lentils*. Washington, DC: U.S. Department of Agriculture. https://www.ams.usda.gov/sites/default/files/media/Commodity%20Specification%20for%20Dry%20Edible%20Beans%2C%20Peas%20and%20Lentils%2C%20June%202014.pdf. Last accessed January 2, 2025.

USDA. 2024. National Agricultural Statistics Service: Data & Statistics. Washington, DC: U.S. Department of Agriculture. https://www.nass.usda.gov/Data_and_Statistics/. Last accsessed January 3, 2025.

USDA-WCMD. 2024. *Approved Storage Warehouses*. Washington, DC: U.S. Department of Agriculture Agricultural Marketing Service. https://www.ams.usda.gov/services/warehouse/approved-storage-warehouses. Last accessed October 4, 2024

Appendix A. SCTG Commodity Codes

Table 15 summarizes SCTG codes and their associated commodity descriptions [BTS 2025a].

Table 15. SCTG Commodity Descriptions by SCTG Code

Code	Commodity description
01	Animals and fish (live)
02	Cereal grains (includes seed)
03	Agricultural products (excludes animal feed, cereal grains, and forage products)
04	Animal feed, eggs, honey, and other products of animal origin
05	Meat, poultry, fish, seafood, and their preparations
06	Milled grain products and preparations, and bakery products
07	Other prepared foodstuffs, fats, and oils
08	Alcoholic beverages and denatured alcohol
09	Tobacco products
10	Monumental or building stone
11	Natural sands
12	Gravel and crushed stone (excludes dolomite and slate)
13	Other non-metallic minerals not elsewhere classified
14	Metallic ores and concentrates
15	Coal
16	Crude petroleum
17	Gasoline, aviation turbine fuel, and ethanol (includes kerosene, and fuel alcohols)
18	Fuel oils (includes diesel, bunker c, and biodiesel)
19	Natural gas and other fossil products, not elsewhere classified
20	Basic chemicals
21	Pharmaceutical products
22	Fertilizers
23	Other chemical products and preparations
24	Plastics and rubber
25	Logs and other wood in the rough
26	Wood products
27	Pulp, newsprint, paper, and paperboard
28	Paper or paperboard articles
29	Printed products
30	Textiles, leather, and articles of textiles or leather
31	Non-metallic mineral products
32	Base metal in primary or semi-finished forms and in finished basic shapes
33	Articles of base metal
34	Machinery
35	Electronic and other electrical equipment and components, and office equipment
36	Motorized and other vehicles (includes parts)
37	Transportation equipment, not elsewhere classified
38	Precision instruments and apparatus
39	Furniture, mattresses and mattress supports, lamps, lighting fittings, and illuminated signs
40	Miscellaneous manufactured products
41	Waste and scrap (excludes agriculture or food)
43	Mixed freight

Appendix B. Updated Cambridge Systematics Regression Models (M2)

Table 16 contains detailed results for the updated M2 regresssion models.

Table 16. Model Results of the Updated Cambridge Systematics Procedure

SCTG	NAICS3	Variable	Production model Coeff	Production model *t* stat	Attraction model Coeff	Attraction model *t* stat
1—Live animals/fish	115	Support Activities for Agriculture and Forestry	0.280	3.760	-	-
	311	Food Manufacturing	-	-	0.101	10.099
	-	Farm acres (in thousands)	-	-	0.057	2.433
2—Cereal grains	311	Food Manufacturing	0.515	4.892	0.531	5.840
	-	Farm acres (in thousands)	1.448	3.312	1.260	3.336
3—Other agriculture products	311	Food Manufacturing	0.282	8.014	0.342	5.883
	-	Farm acres (in thousands)	0.319	2.177	-	-
4—Animal feed	115	Support Activities for Agriculture and Forestry	0.957	3.247	-	-
	-	Farm acres (in thousands)	0.659	4.881	0.461	8.076
	-	Livestock sold (in million)	-	-	273.000	17.460
	-	Population 2000	-	-	0.00043	6.514
5—Meat/seafood	311	Food Manufacturing	0.063	19.210	0.033	8.228
	722	Food Services and Drinking Places	-	-	0.006	11.062
6—Milled grain products	311	Food Manufacturing	0.078	15.450	0.061	9.523
	722	Food Services and Drinking Places	-	-	0.004	5.090
7—Other foodstuffs	311	Food Manufacturing	0.392	14.514	0.276	14.833
	325	Chemical Manufacturing	0.021	0.442	-	-
	722	Food Services and Drinking Places	-	-	0.018	7.426
8—Alcoholic beverages	312	Beverage and Tobacco Product Manufacturing	0.240	8.218	-	-
	311	Food Manufacturing	0.029	5.072	0.018	5.063
	722	Accommodation and Food Services	-	-	0.008	16.090
9—Tobacco products	312	Beverage and Tobacco Product Manufacturing	0.009	2.085	0.007	1.007
	-	2000 Population	-	-	0.0000019	0.284
10—Building stone	212	Mining (except Oil and Gas)	2.610	9.443	-	-
	23	Construction	-	-	0.002	12.910
11—Natural sands	212	Mining (except Oil and Gas)	2.610	9.443	-	-
	23	Construction	-	-	0.094	13.390

SCTG	NAICS3	Variable	Production model		Attraction model	
			Coeff	*t* stat	Coeff	*t* stat
12—Gravel	212	Mining (except Oil and Gas)	2.610	9.443	-	-
	23	Construction	-	-	0.210	14.060
13—Nonmetallic minerals	212	Mining (except Oil and Gas)	2.610	9.443	-	-
	321–327	Nondurable goods	-	-	0.065	9.587
14—Metallic ores	212	Mining (except Oil and Gas)	2.610	9.443	-	-
	331	Primary Metal Manufacturing	-	-	0.239	5.074
15—Coal	212	Mining (except Oil and Gas)	2.610	9.443	0.738	2.633
	324	Petroleum and Coal Products Manufacturing	-	-	0.321	0.844
	-	Electricity Generation KWH	-	-	607.475	14.451
16—Crude petroleum	211	Oil and Gas Extraction	8.703	7.260	0.675	0.899
	324	Petroleum & Coal Products Manufacturing	-	-	9.817	8.374
17—Gasoline	324	Petroleum & Coal Products Manufacturing	10.831	11.380	6.457	6.702
	481–484, 486, 488	transportation	-	-	0.246	7.502
18—Fuel oils	324	Petroleum & Coal Products Manufacturing	7.549	9.630	5.876	6.418
	-	Total employment	-	-	0.002	3.040
19—Other coal & petroleum products	211	Oil and Gas Extraction	6.867	4.483	-	-
	324	Petroleum & Coal Products Manufacturing	7.319	3.042	5.474	3.218
	-	2000 Population	-	-	0.005	9.073
20—Basic chemicals	325	Chemical Manufacturing	0.160	11.150	0.422	9.503
21—Pharmaceuticals	325	Chemical Manufacturing	0.160	11.150	-	-
	-	2000 Population	-	-	0.000	22.290
22—Fertilizers	325	Chemical Manufacturing	0.160	11.150	-	-
	-	Farm acres (in thousands)	-	-	0.222	6.589
	-	2000 Population	-	-	0.000	5.218
23—Chemical products	325	Chemical Manufacturing	0.160	11.150	0.035	2.632
	-	2000 Population	-	-	0.000	6.771
24—Plastics/rubber	23	Construction	-	-	0.005	1.138
	325	Chemical Manufacturing	-	-	0.176	4.820
	326	Plastics and Rubber Products Manufacturing	0.245	6.783	0.047	1.532

SCTG	NAICS3	Variable	Production model Coeff	Production model *t* stat	Attraction model Coeff	Attraction model *t* stat
25—Logs	113	Forestry and Logging	7.439	17.084	8.161	26.327
	321	Wood Product Manufacturing	0.339	4.644	-	-
	322	Paper Manufacturing	-	-	0.339	4.683
	-	Farm acres (in thousand)	-	-	0.054	0.945
26—Wood products	23	Construction	-	-	0.022	10.403
	321	Wood Product Manufacturing	0.792	22.450	0.458	12.299
	337	Furniture and Related Product Manufacturing	-	-	0.022	0.557
27—Newsprint/paper	113	Forestry and Logging	1.109	9.563	-	-
	311	Food Manufacturing	-	-	0.020	3.383
	322	Paper Manufacturing	-	-	0.208	6.071
	323	Printing and Related Support Activities	0.187	8.692	0.111	4.099
28—Paper articles	311	Food Manufacturing	-	-	0.021	7.153
	322	Paper Manufacturing	0.219	7.861	0.082	5.013
	323	Printing and Related Support Activities	0.053	2.513	0.008	0.511
	-	Population 2000	-	-	0.000	5.431
29—Printed products	322	Paper Manufacturing	0.008	0.870	-	-
	323	Printing and Related Support Activities	0.066	8.986	0.038	5.891
	511	Publishing Industries (except Internet)	-	-	0.001	-1.453
	81	Other Services (except Public Administration)	-	-	0.003	6.135
30—Textiles/leather	313	Textile Mills	0.134	4.975	0.045	3.943
	314	Textile Product Mills	0.180	9.000	0.087	9.359
	-	Population 2000	-	-	0.000	20.064
31—Nonmetallic mineral products	23	Construction	-	-	0.126	14.739
	327	Nonmetallic Mineral Product Manufacturing	2.935	25.100	-	-
	321–327	Nondurable goods	-	-	0.067	3.243
32—Base metals	331	Primary Metal Manufacturing	0.808	13.136	0.452	8.384
	332	Fabricated Metal Product Manufacturing	-	-	0.052	2.231
	333	Machinery Manufacturing	0.024	1.239	0.046	1.503
33—Articles of base metals	23	Construction	-	-	0.003	2.252
	332	Fabricated Metal Product Manufacturing	0.080	13.700	0.057	8.057
34—Machinery	332	Fabricated Metal Product Manufacturing	0.064	6.517	-	-
	23	Construction	-	-	0.008	8.952

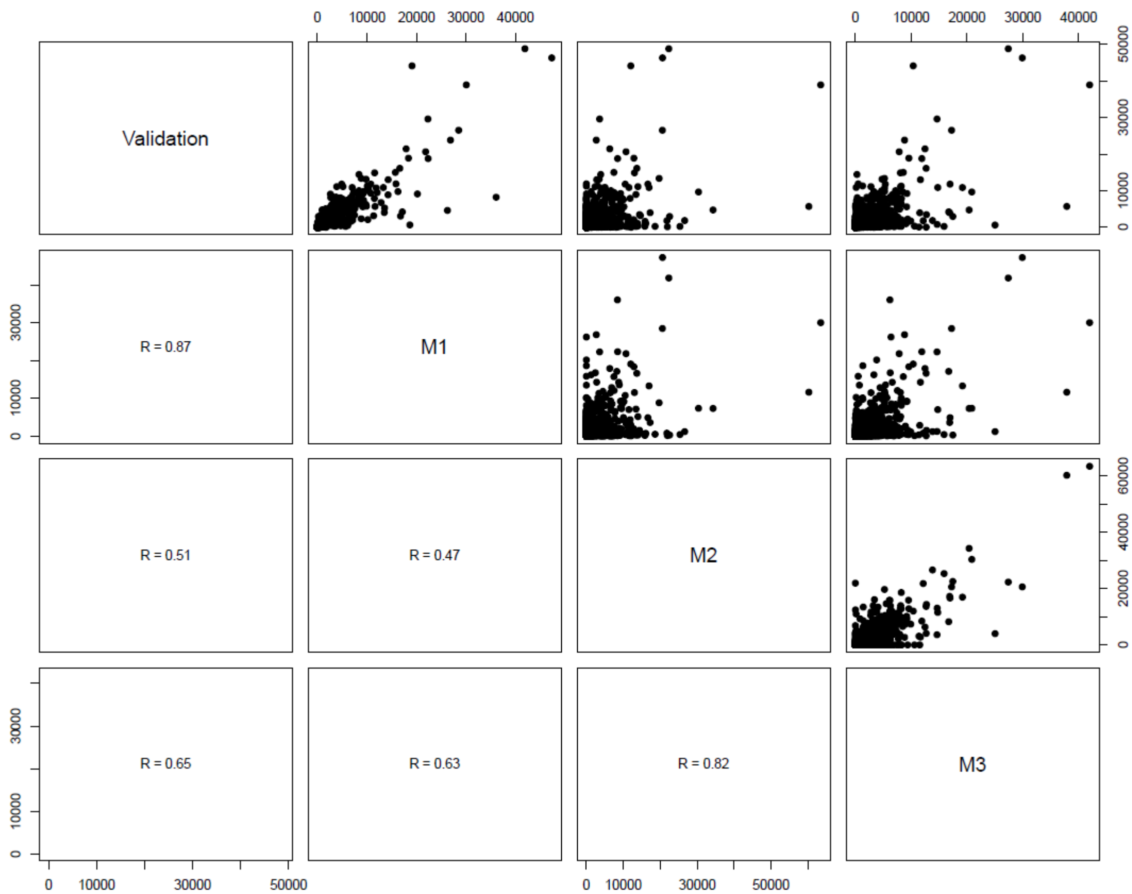
SCTG	NAICS3	Variable	Production model Coeff	Production model *t* stat	Attraction model Coeff	Attraction model *t* stat
35—Electronic & electrical	333	Machinery Manufacturing	0.019	2.672	-	-
	334	Computer and Electronic Product Manufacturing	0.029	4.859	-	-
	335	Electrical Equipment, Appliance, and Component Manufacturing	0.068	2.652	-	-
	-	Total employment	-	-	0.000	29.060
36—Motorized vehicles	336	Transportation Equipment Manufacturing	0.051	12.420	0.081	12.518
	-	Total employment	-	-	0.000	2.146
37—Transportation equipment	336	Transportation Equipment Manufacturing	0.051	12.420	-	-
	-	Total employment	-	-	0.000	5.813
38—Precision instruments	339	Miscellaneous Manufacturing	0.010	8.948	-	-
	-	Total employment	-	-	0.000	10.920
39—Furniture	337	Furniture and Related Product Manufacturing	0.092	17.340	-	-
	-	2000 Population	-	-	0.000	50.370
40—Misc. manufactured products	339	Miscellaneous Manufacturing	0.135	17.270		
	-	2000 Population	-	-	0.000	25.260
41—Waste and scrap	221	Oil and Gas Extraction	0.593	9.566	-	-
	321–327	nondurable	0.106	5.331	-	-
	-	2000 Population	-	-	0.002	23.660
43—Mixed freight	321–327	nondurable	0.066	7.103	-	-
	481–484, 486, 488	transportation	0.058	6.974	-	-
	-	2000 Population	-	-	0.001	7.296
	-	Total employment	-	-	0.000	-0.894

-Not applicable.

Appendix C. Pairwise Scatterplots of Validation Data and M1, M2, and M3

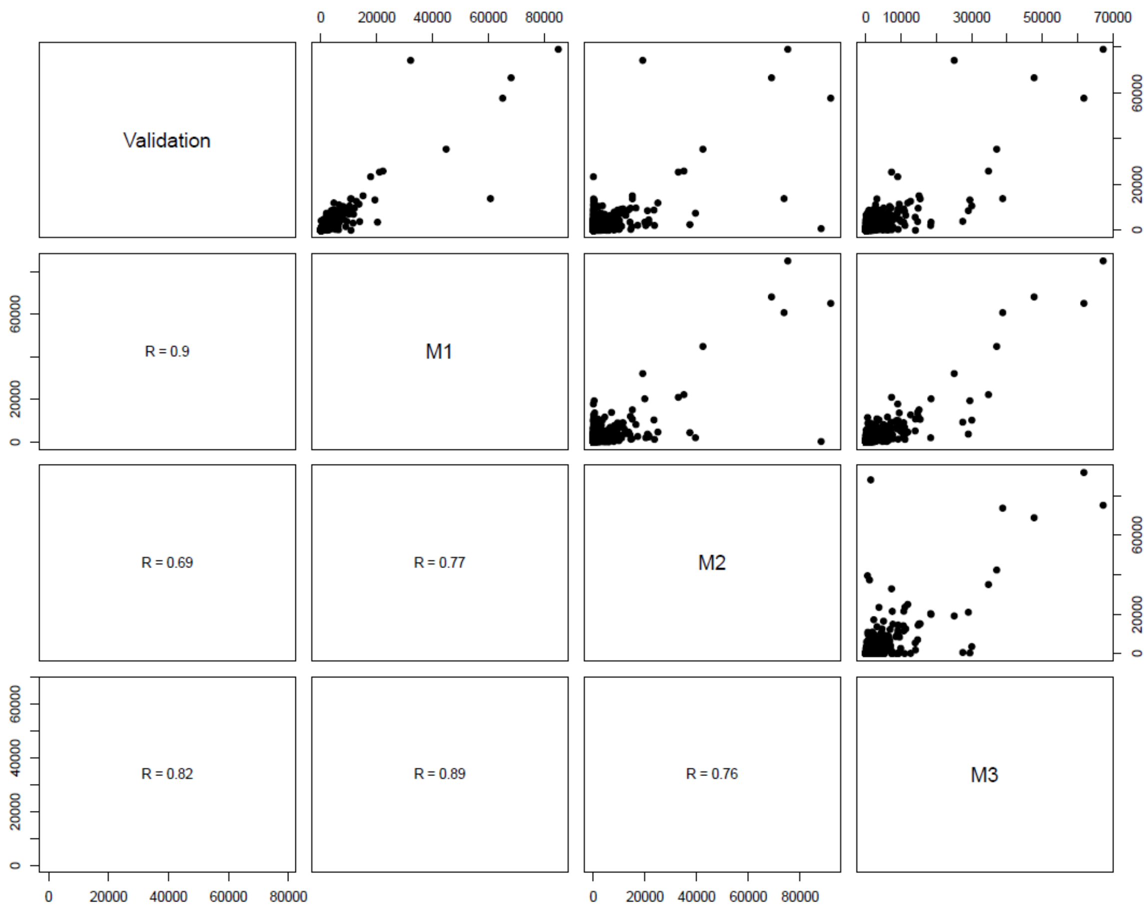
Figure 9–Figure 37 are pairwise scatter plots illustrating the relationship between validation data, M1, M2, and M3 with pairwise correlations. Truck origin pairwise scatter plot for SCTG 01–09 is shown and discussed in the main text.

Figure 9. Pairwise Matrix Scatterplot Between Validation Data, M1, M2, and M3 With Pairwise Correlations: SCTG 10–14 Truck Origin Data



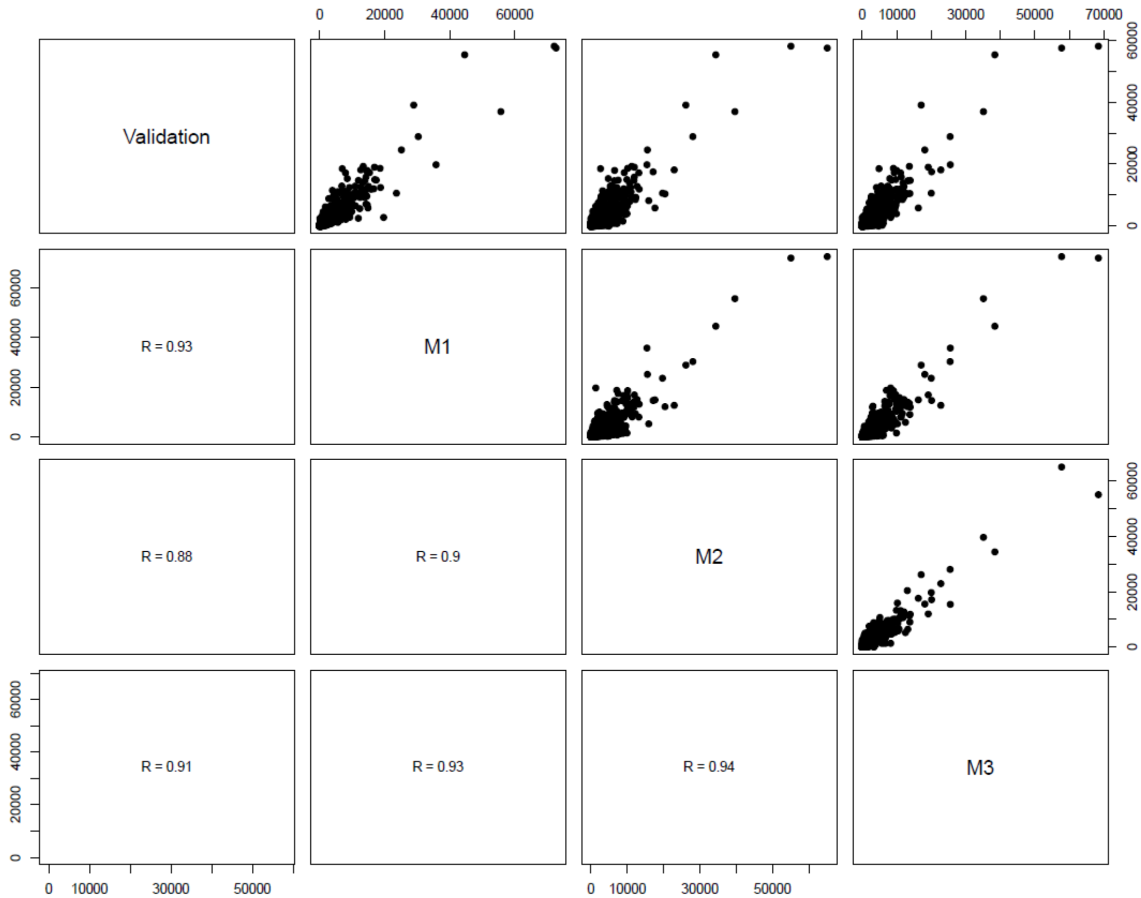
Source: BTS.

Figure 10. Pairwise Matrix Scatterplot Between Validation Data, M1, M2, and M3 With Pairwise Correlations: SCTG 15–19 Truck Origin Data



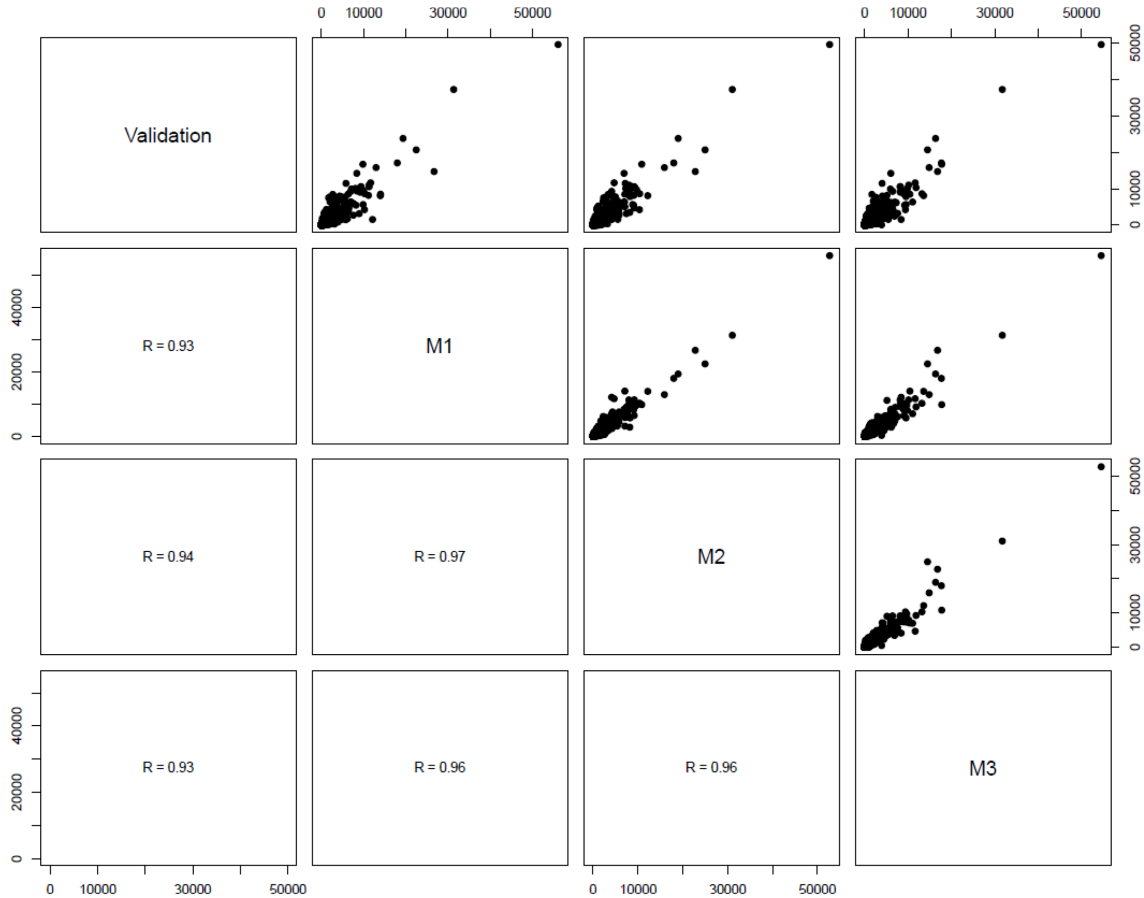
Source: BTS.

Figure 11. Pairwise Matrix Scatterplot Between Validation Data, M1, M2, and M3 With Pairwise Correlations: SCTG 20–33 Truck Origin Data



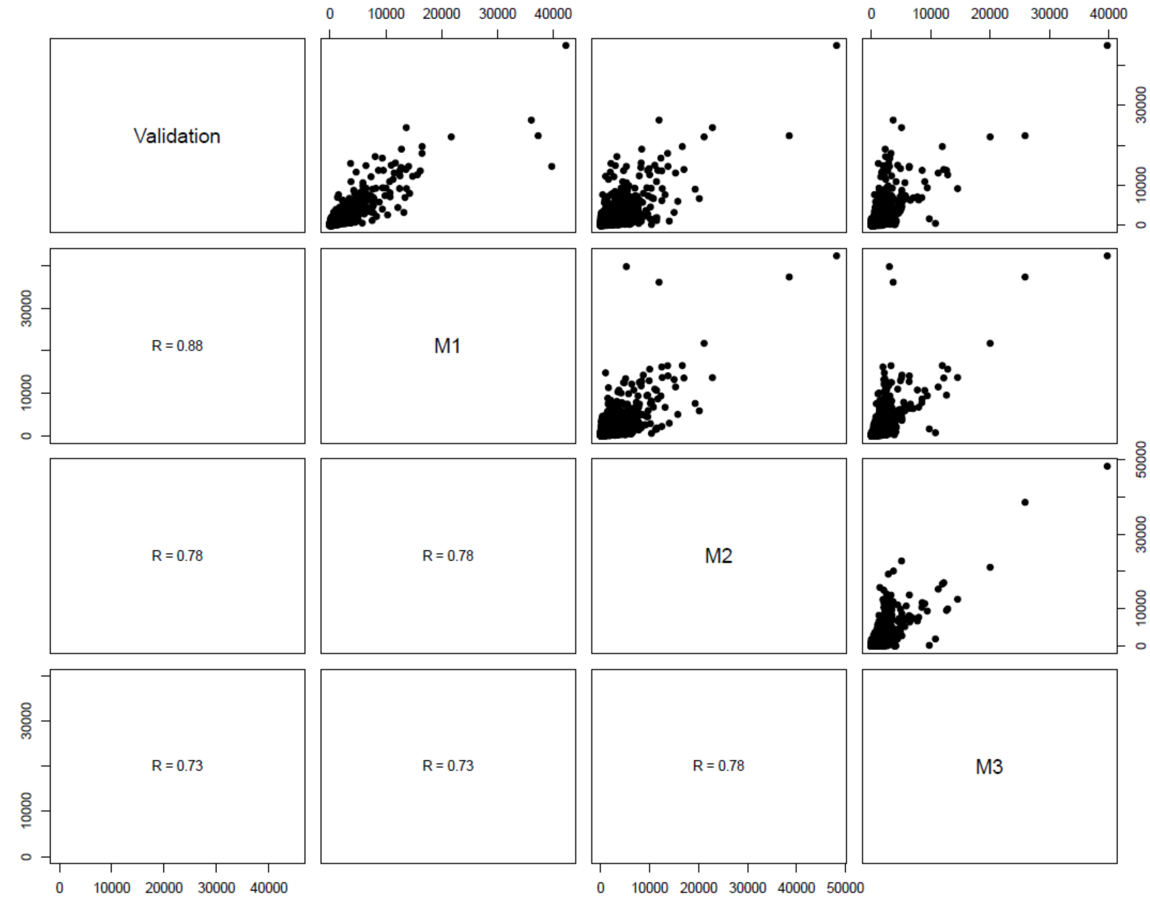
Source: BTS.

Figure 12. Pairwise Matrix Scatterplot Between Validation Data, M1, M2, and M3 With Pairwise Correlations: SCTG 34–99 Truck Origin Data



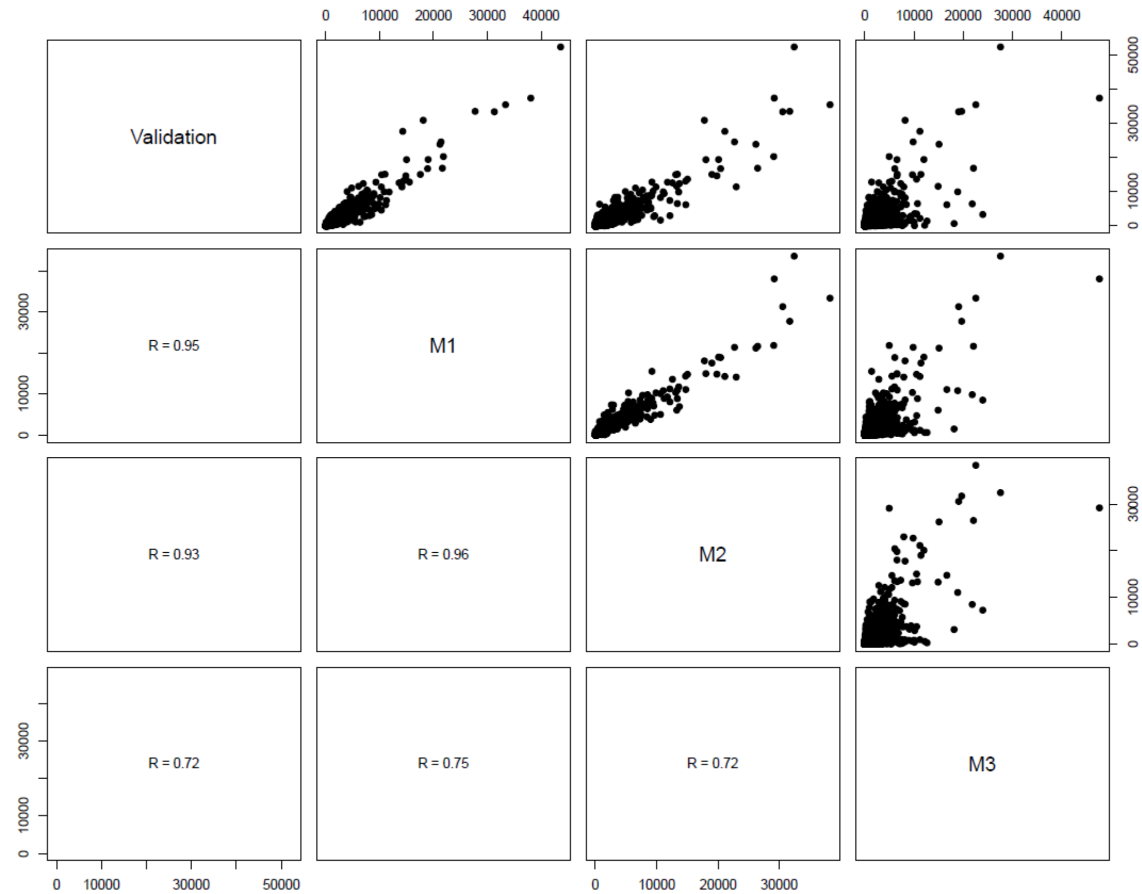
Source: BTS.

Figure 13. Pairwise Matrix Scatterplot Between Validation Data, M1, M2, and M3 With Pairwise Correlations: SCTG 01–09 Truck Destination Data



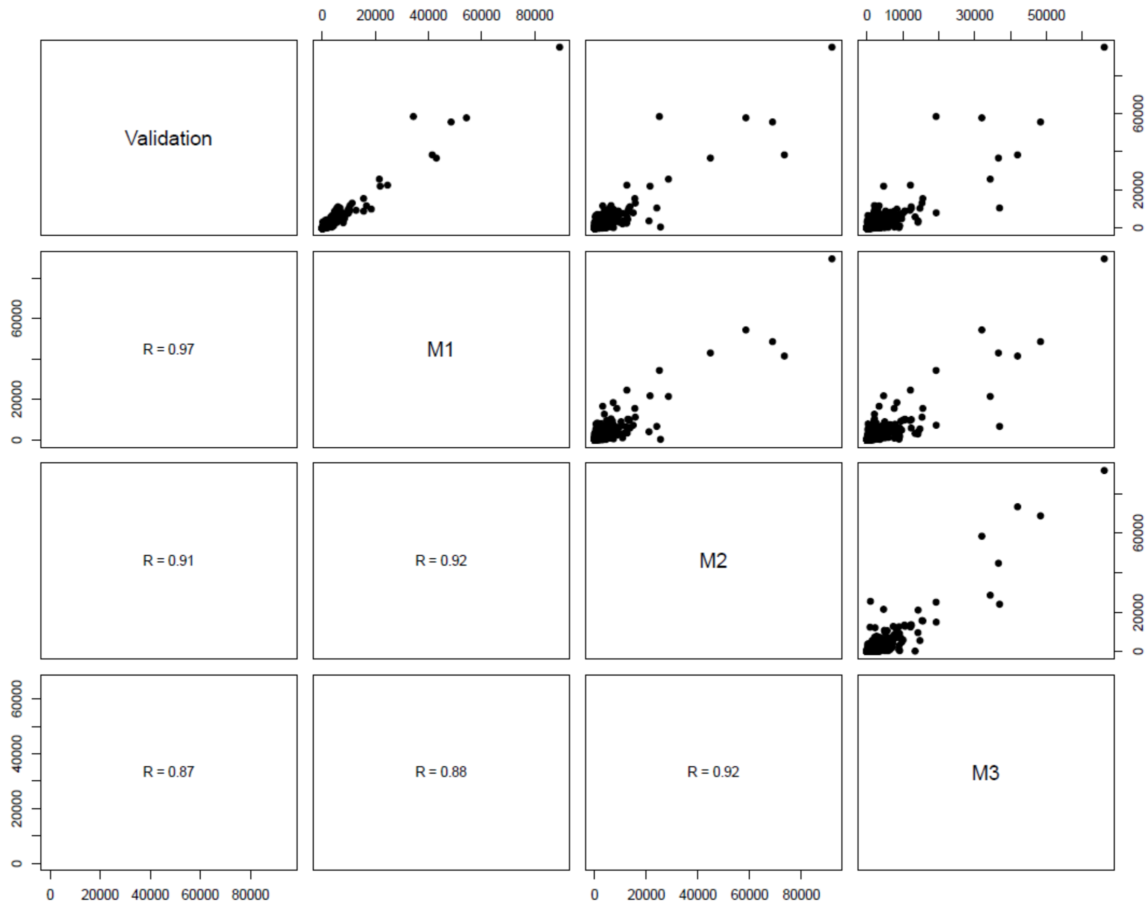
Source: BTS.

Figure 14. Pairwise Matrix Scatterplot Between Validation Data, M1, M2, and M3 With Pairwise Correlations: SCTG 10–14 Truck Destination Data



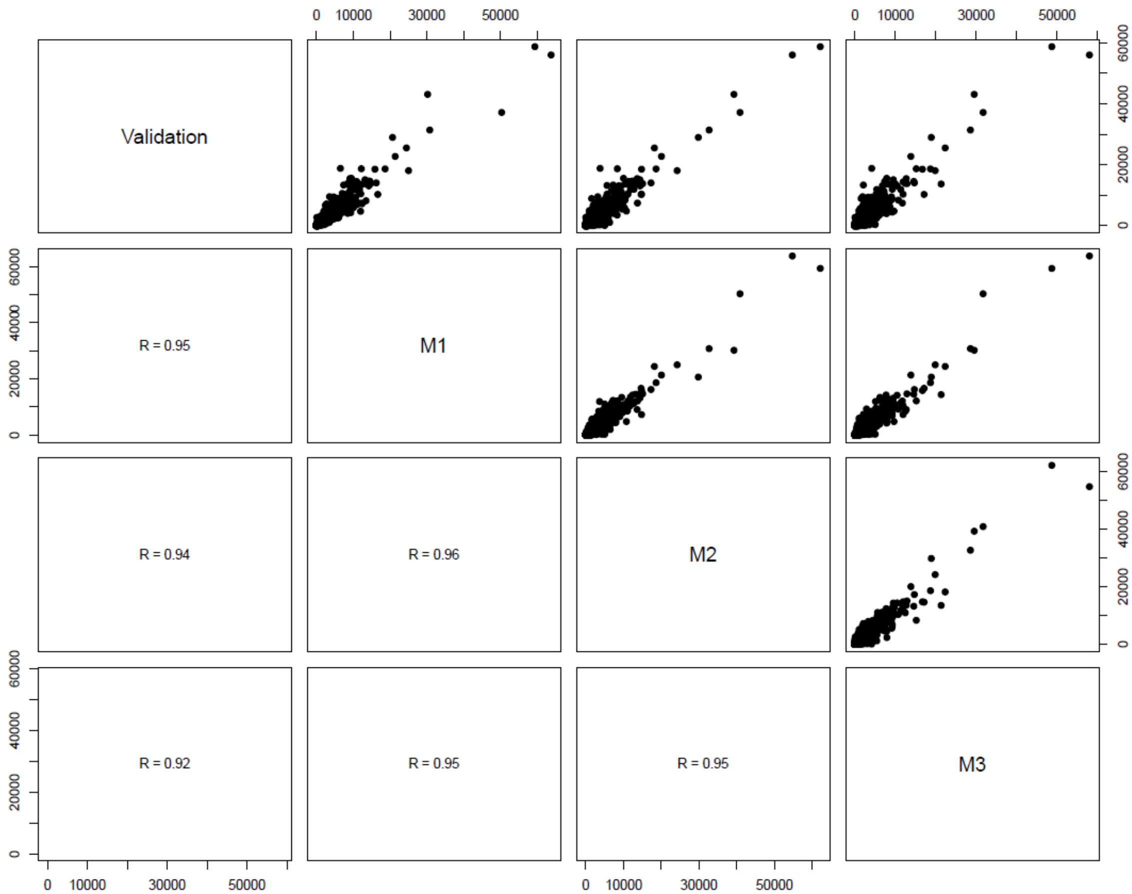
Source: BTS.

Figure 15. Pairwise Matrix Scatterplot Between Validation Data, M1, M2, and M3 With Pairwise Correlations: SCTG 15–19 Truck Destination Data



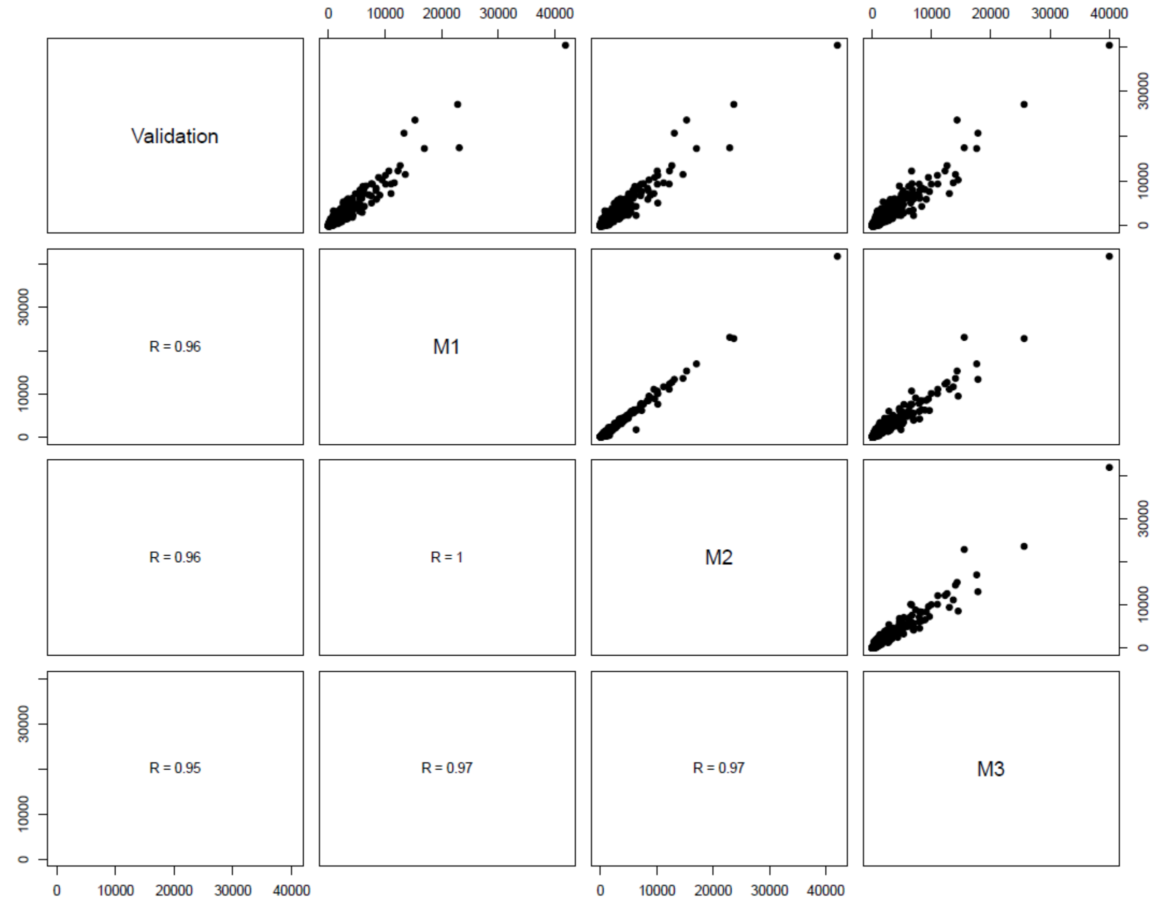
Source: BTS.

Figure 16. Pairwise Matrix Scatterplot Between Validation Data, M1, M2, and M3 With Pairwise Correlations: SCTG 20–33 Truck Destination Data



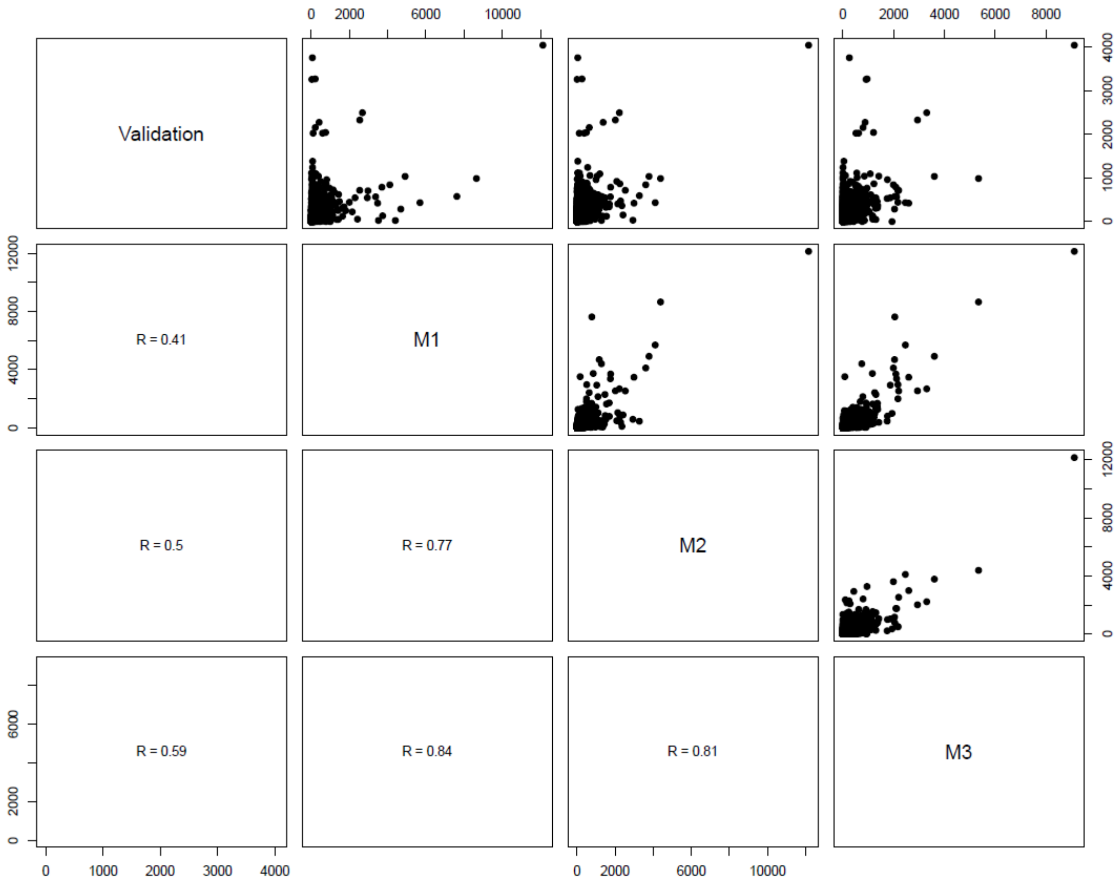
Source: BTS.

Figure 17. Pairwise Matrix Scatterplot Between Validation Data, M1, M2, and M3 With Pairwise Correlations: SCTG 34–99 Truck Destination Data



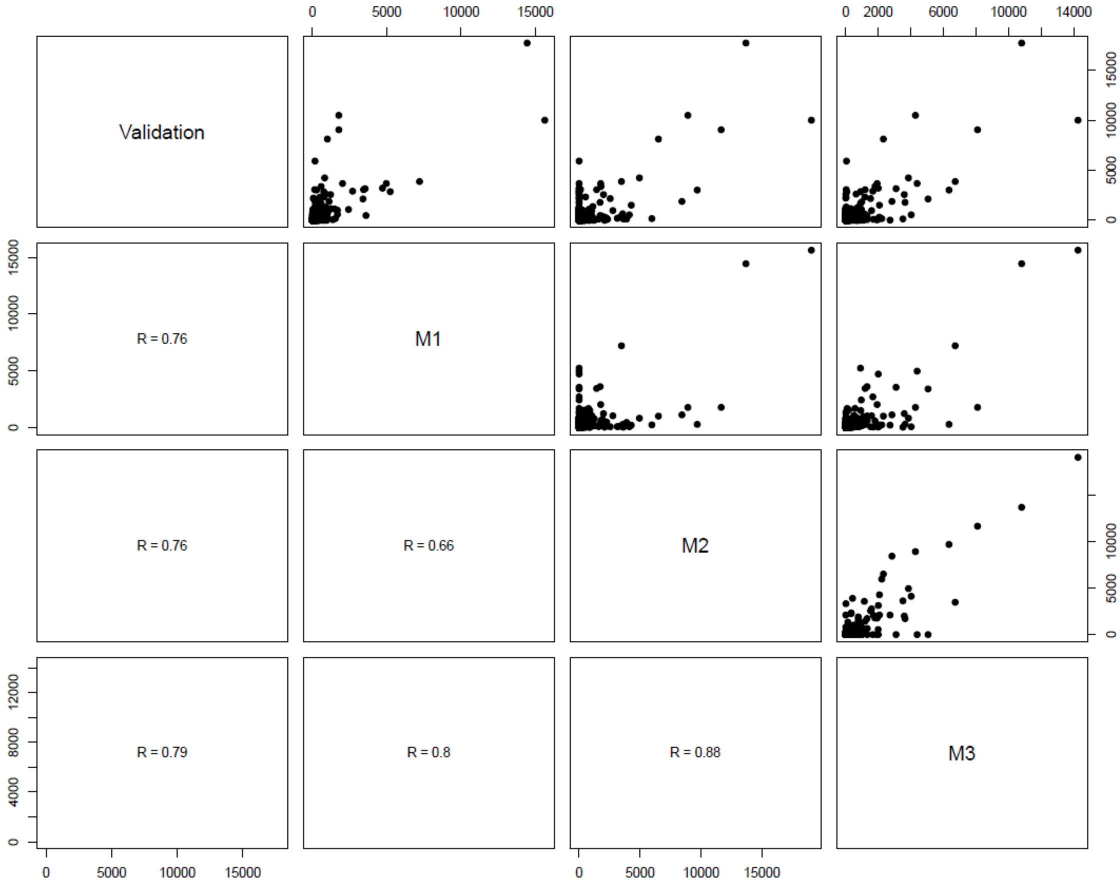
Source: BTS.

Figure 18. Pairwise Matrix Scatterplot Between Validation Data, M1, M2, and M3 With Pairwise Correlations: SCTG 01–09 Rail Origin Data



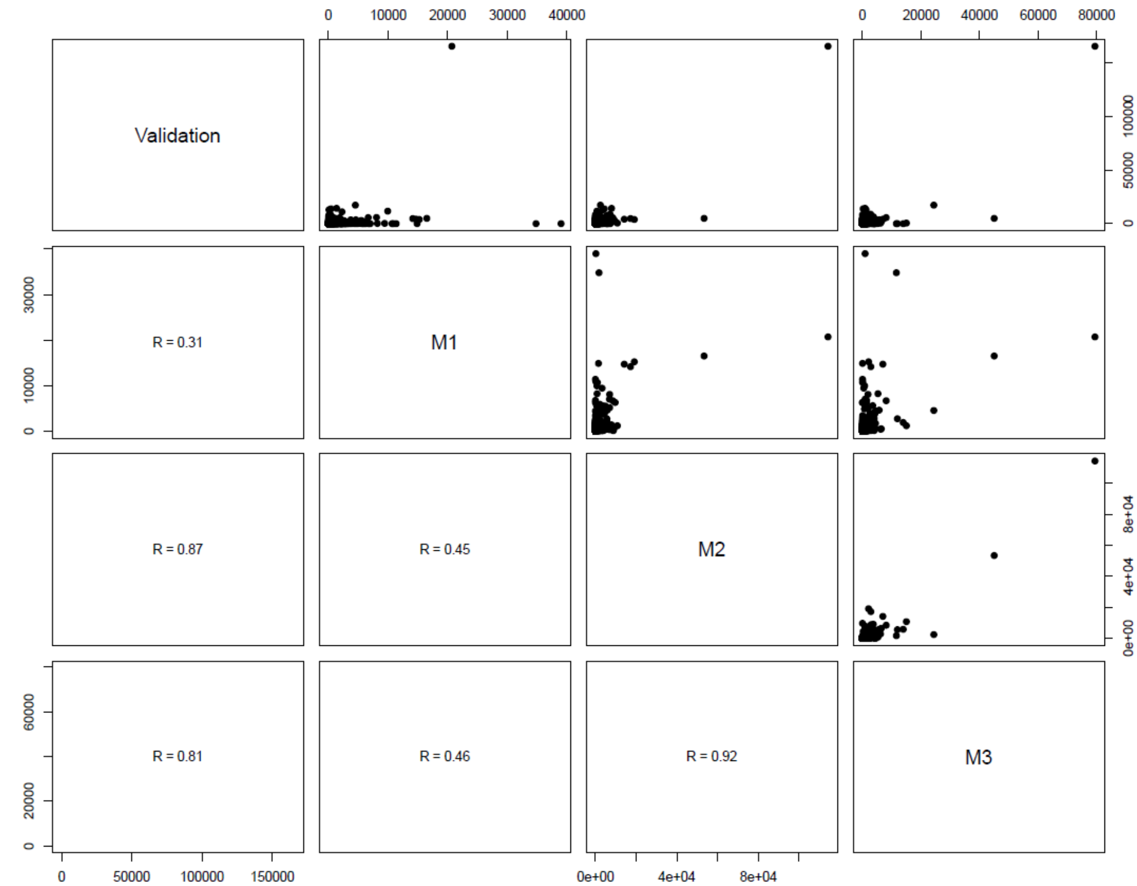
Source: BTS.

Figure 19. Pairwise Matrix Scatterplot Between Validation Data, M1, M2, and M3 With Pairwise Correlations: SCTG 10–14 Rail Origin Data



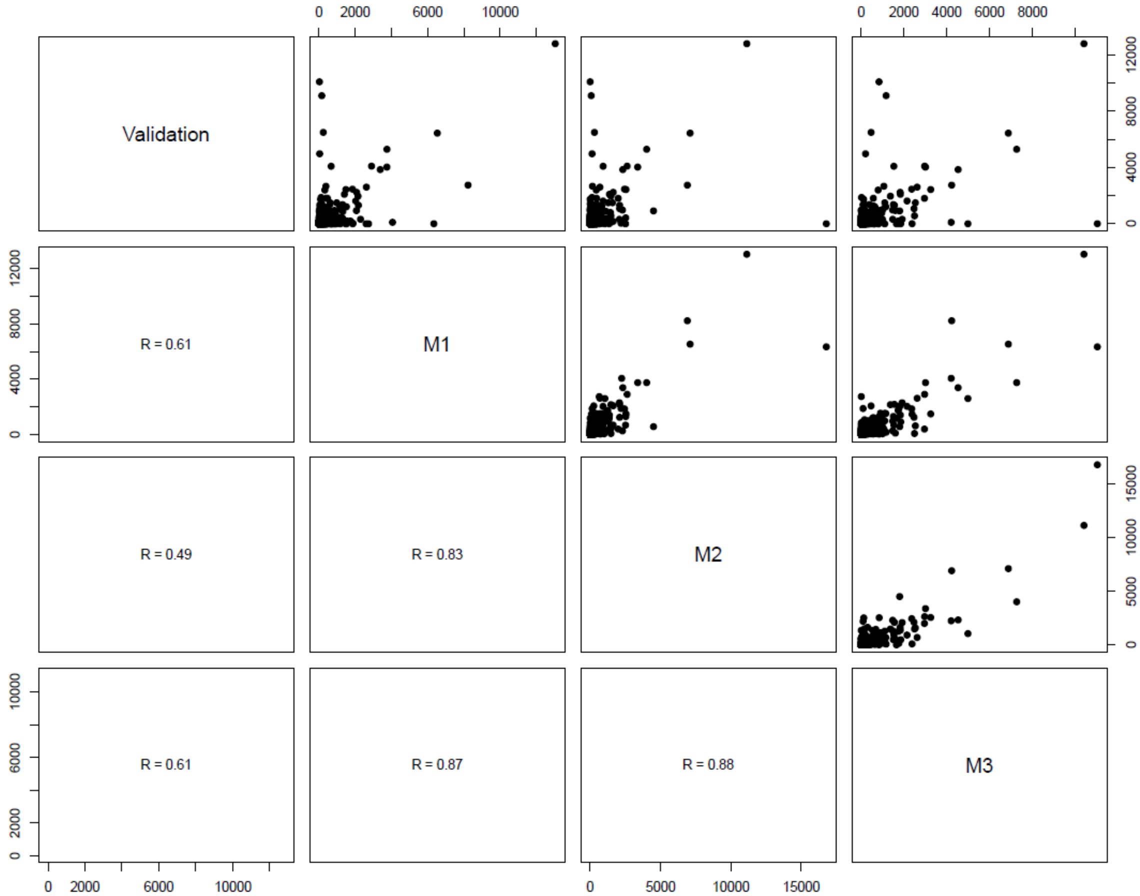
Source: BTS.

Figure 20. Pairwise Matrix Scatterplot Between Validation Data, M1, M2, and M3 With Pairwise Correlations: SCTG 15–19 Rail Origin Data



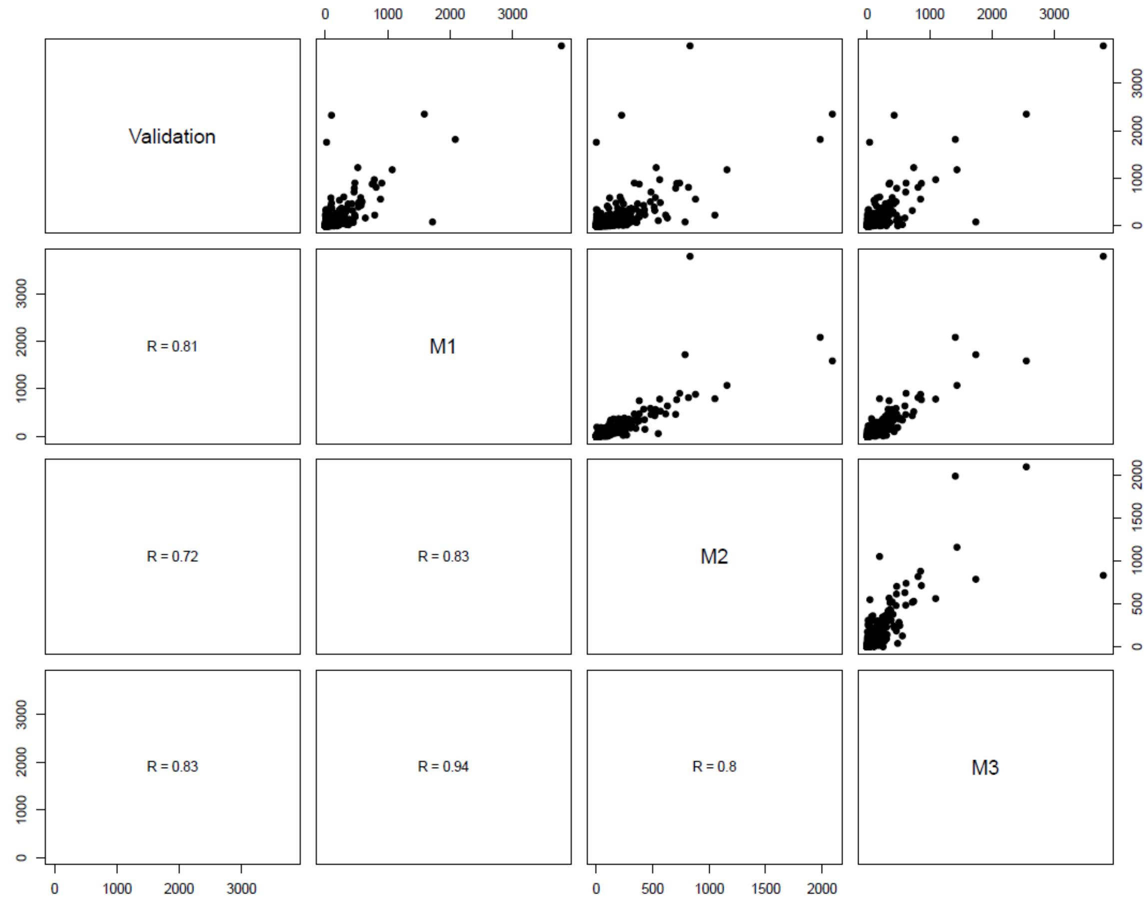
Source: BTS.

Figure 21. Pairwise Matrix Scatterplot Between Validation Data, M1, M2, and M3 With Pairwise Correlations: SCTG 20–33 Rail Origin Data



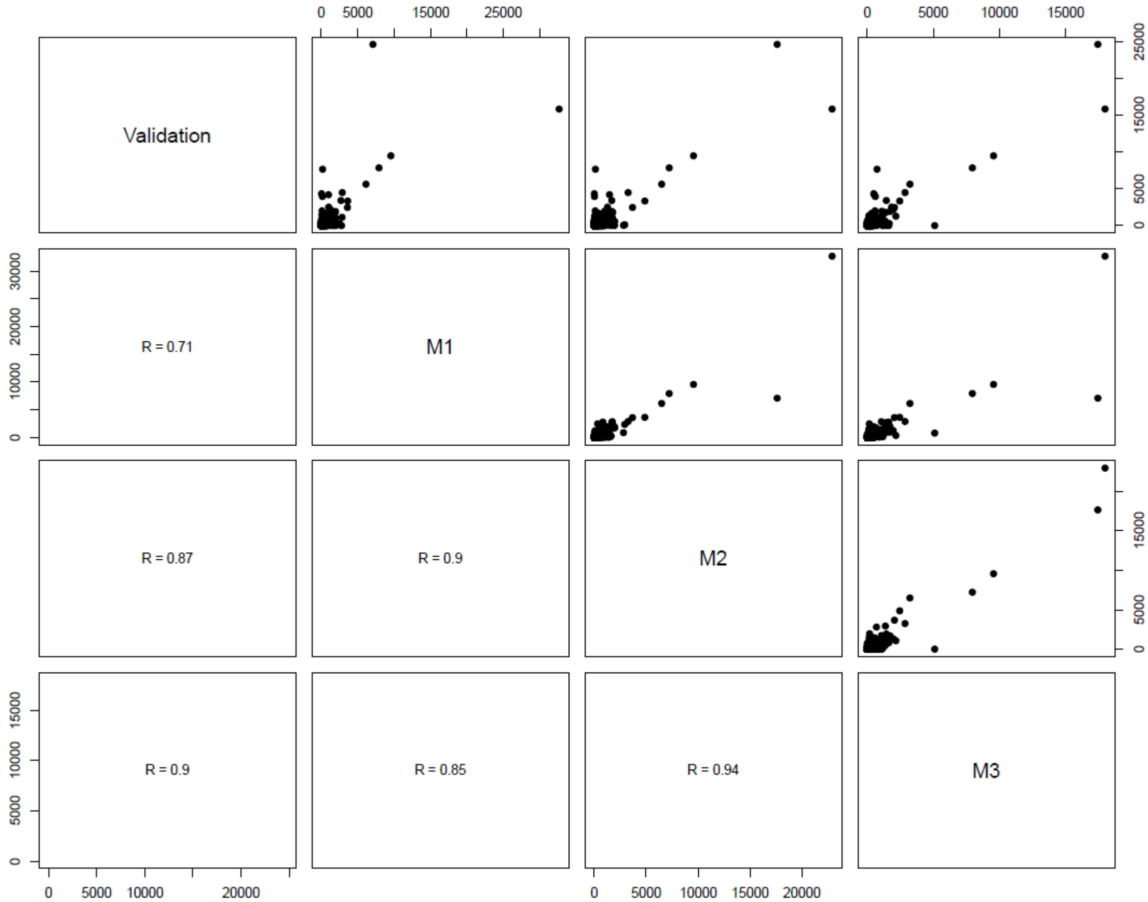
Source: BTS.

Figure 22. Pairwise Matrix Scatterplot Between Validation Data, M1, M2, and M3 With Pairwise Correlations: SCTG 34–99 Rail Origin Data



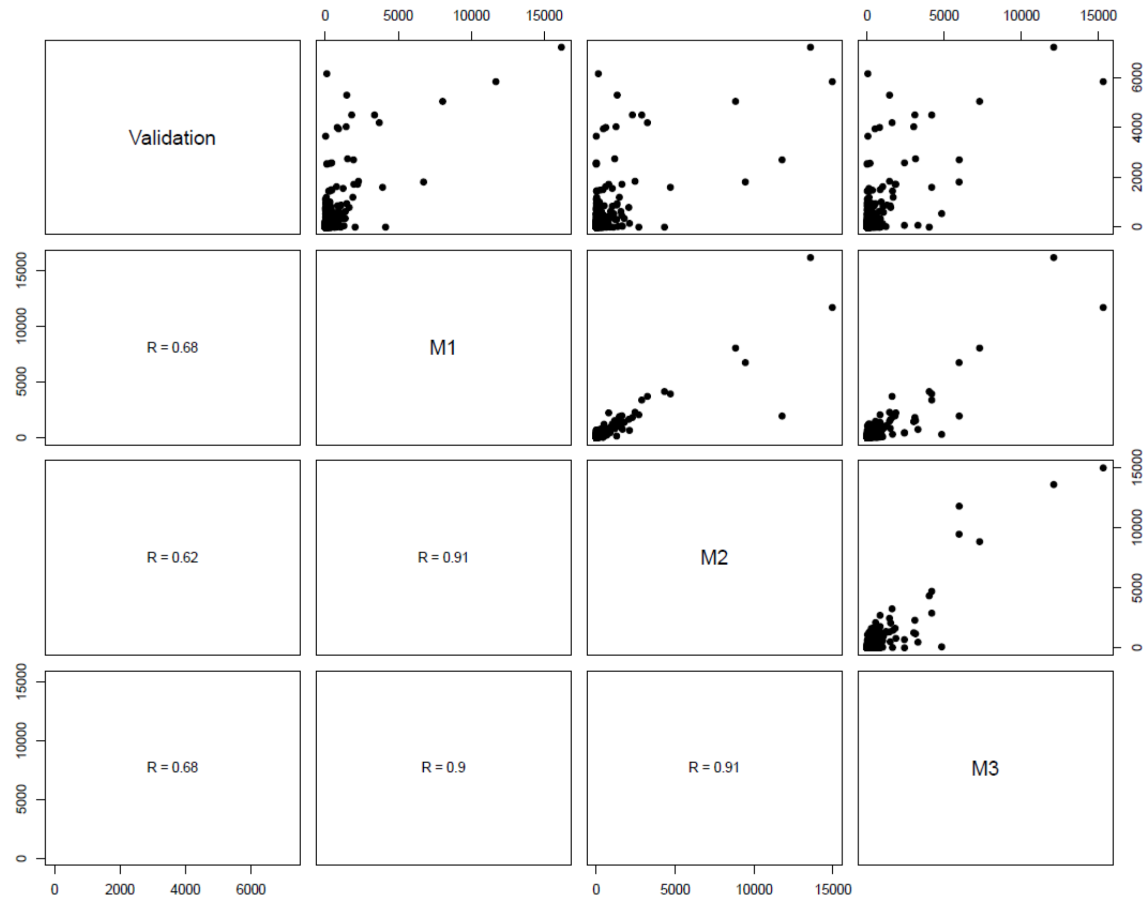
Source:BTS.

Figure 23. Pairwise Matrix Scatterplot Between Validation Data, M1, M2, and M3 With Pairwise Correlations: SCTG 01–09 Rail Destination Data



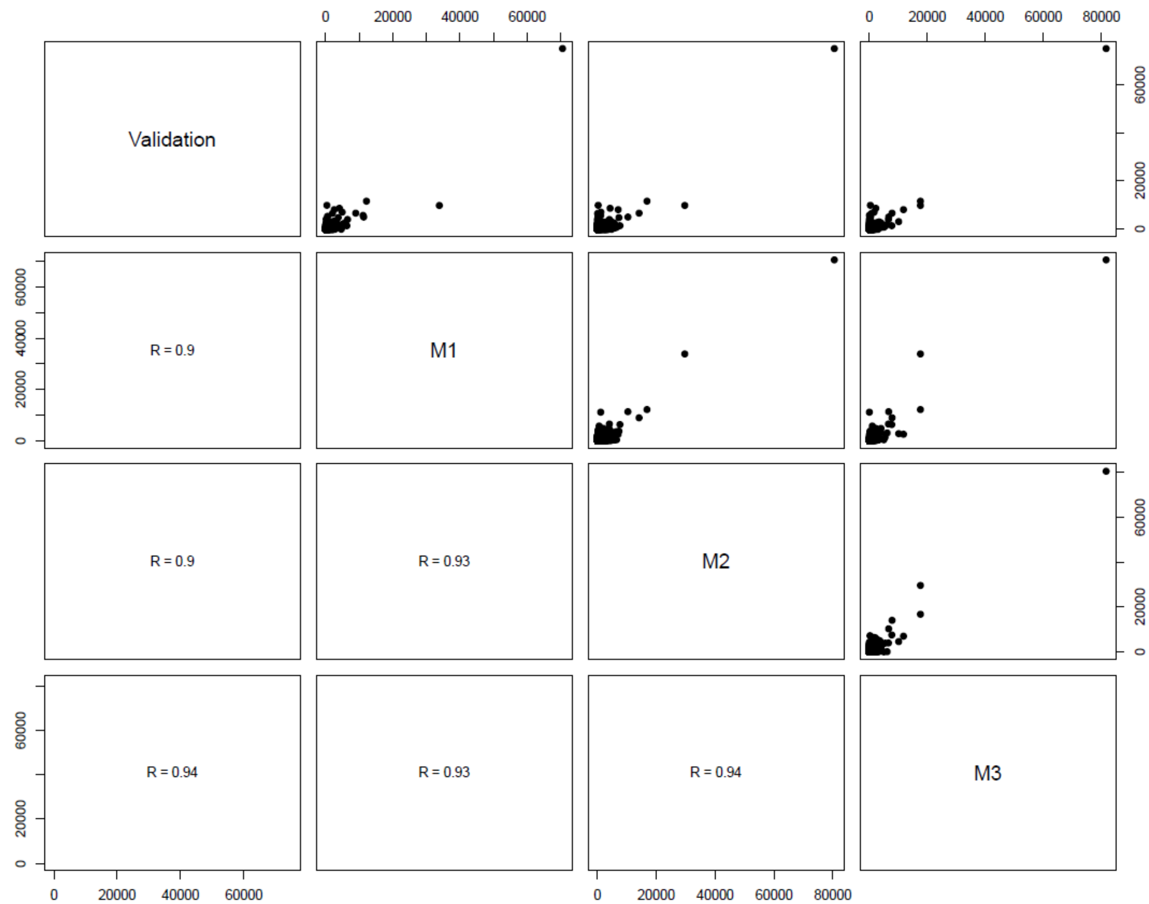
Source: BTS.

Figure 24. Pairwise Matrix Scatterplot Between Validation Data, M1, M2, and M3 With Pairwise Correlations: SCTG 10–14 Rail Destination Data



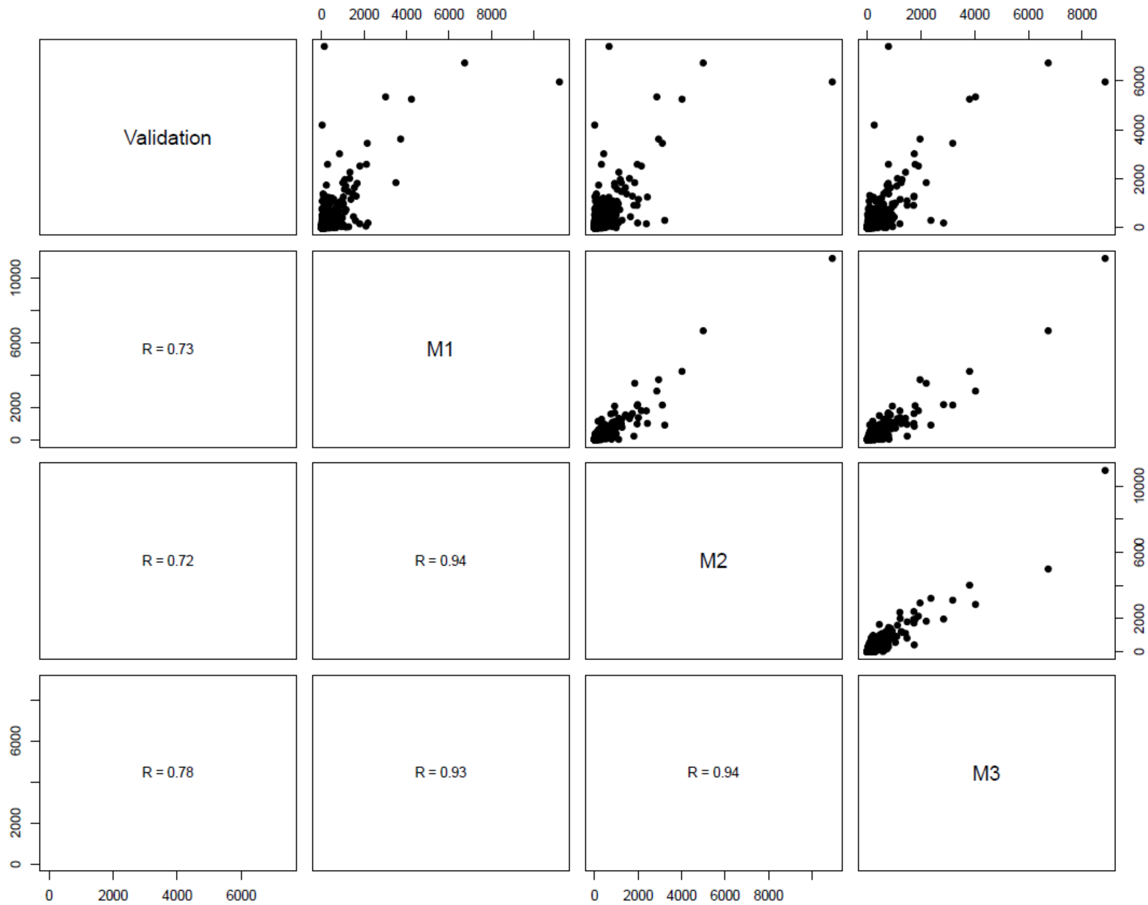
Source: BTS.

Figure 25. Pairwise Matrix Scatterplot Between Validation Data, M1, M2, and M3 With Pairwise Correlations: SCTG 15–19 Rail Destination Data



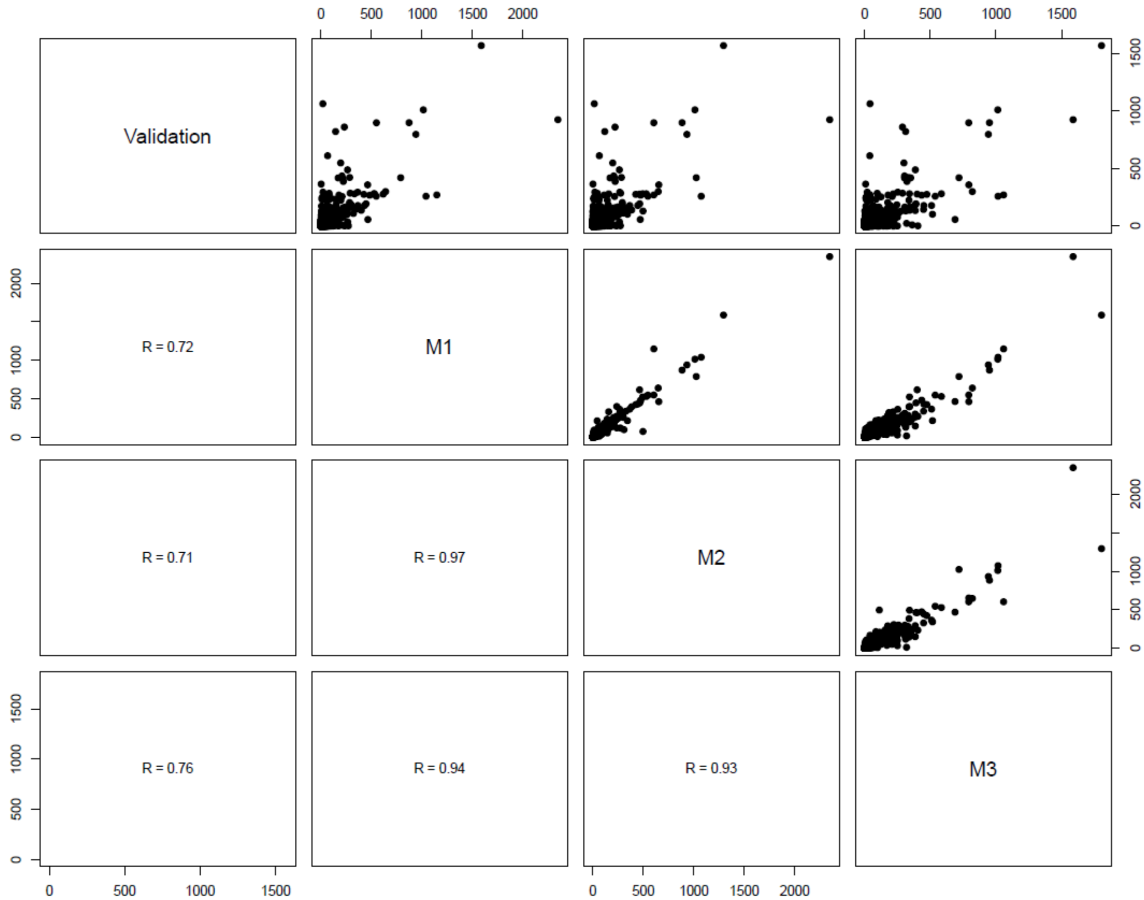
Source: BTS.

Figure 26. Pairwise Matrix Scatterplot Between Validation Data, M1, M2, and M3 With Pairwise Correlations: SCTG 20–33 Rail Destination Data



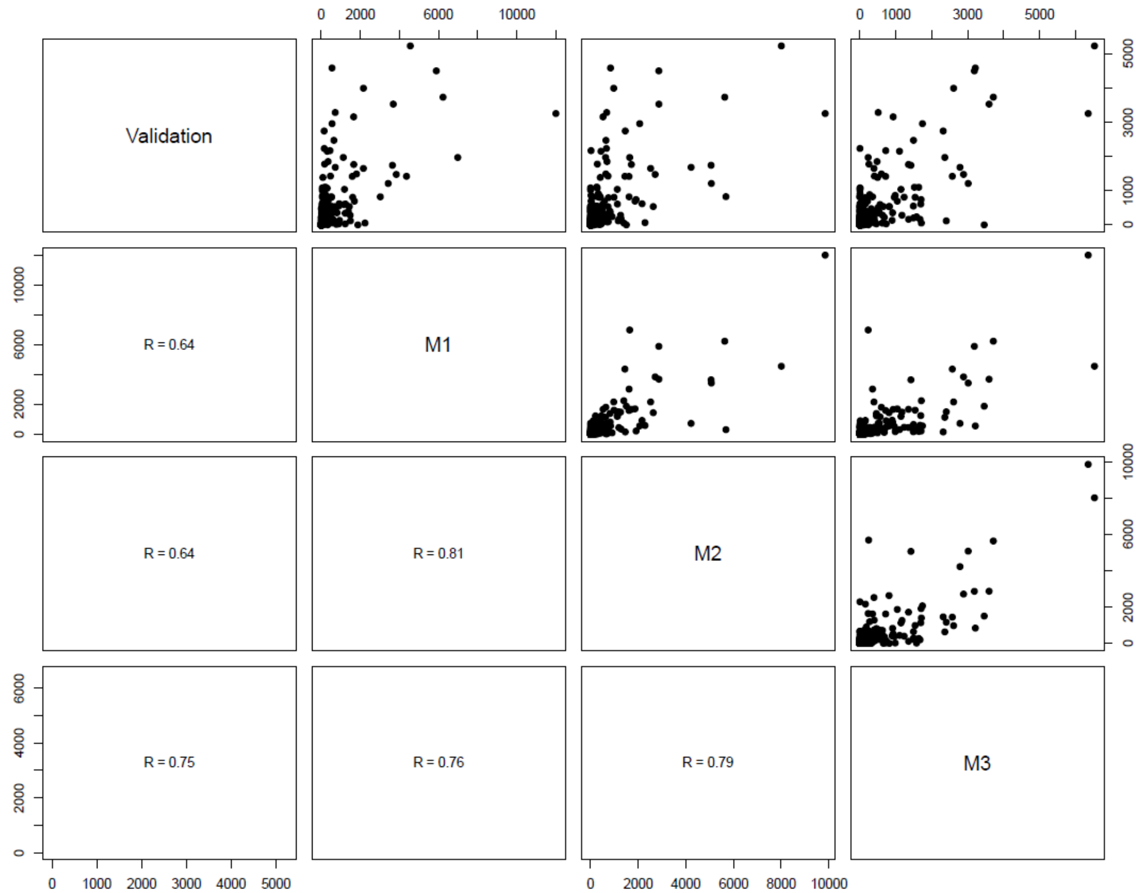
Source: BTS.

Figure 27. Pairwise Matrix Scatterplot Between Validation Data, M1, M2, and M3 With Pairwise Correlations: SCTG 34–99 Rail Destination Data



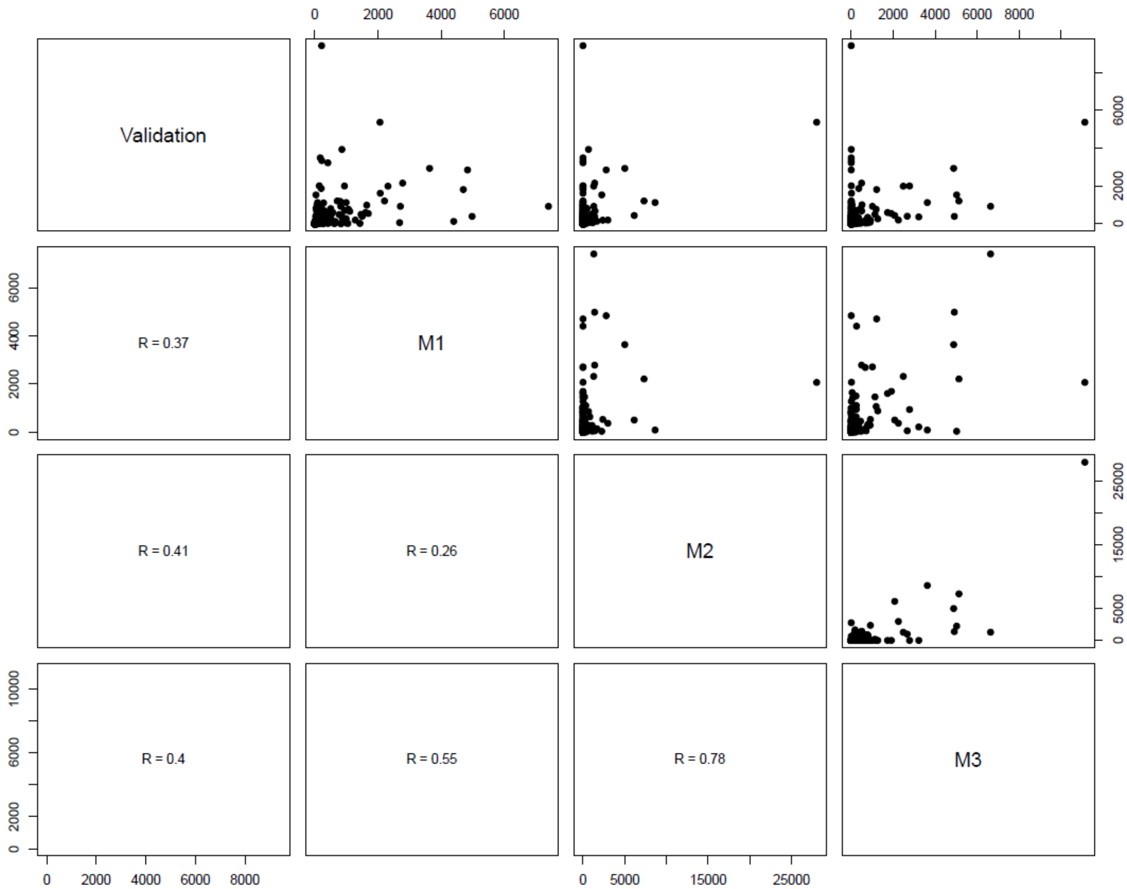
Source: BTS.

Figure 28. Pairwise Matrix Scatterplot Between Validation Data, M1, M2, and M3 With Pairwise Correlations: SCTG 01–09 Water Origin Data



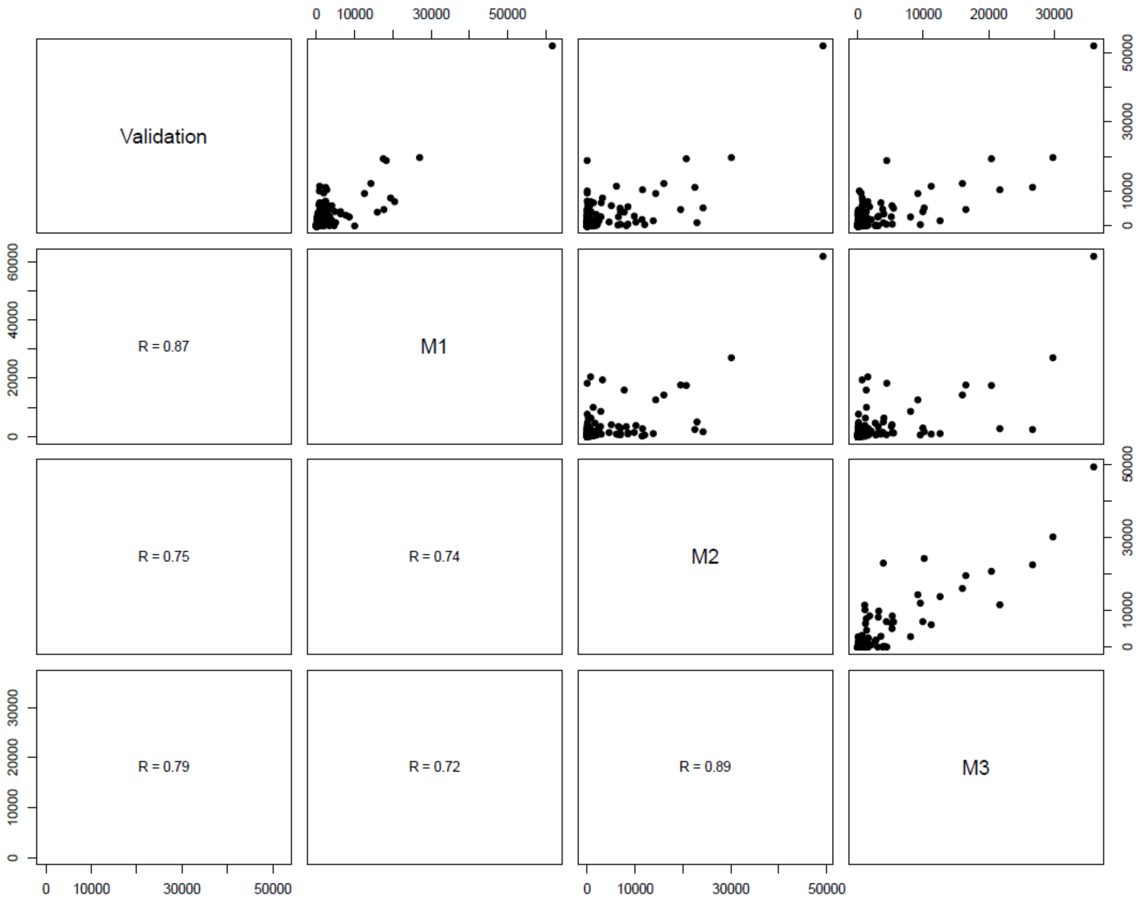
Source: BTS.

Figure 29. Pairwise Matrix Scatterplot Between Validation Data, M1, M2, and M3 With Pairwise Correlations: SCTG 10–14 Water Origin Data



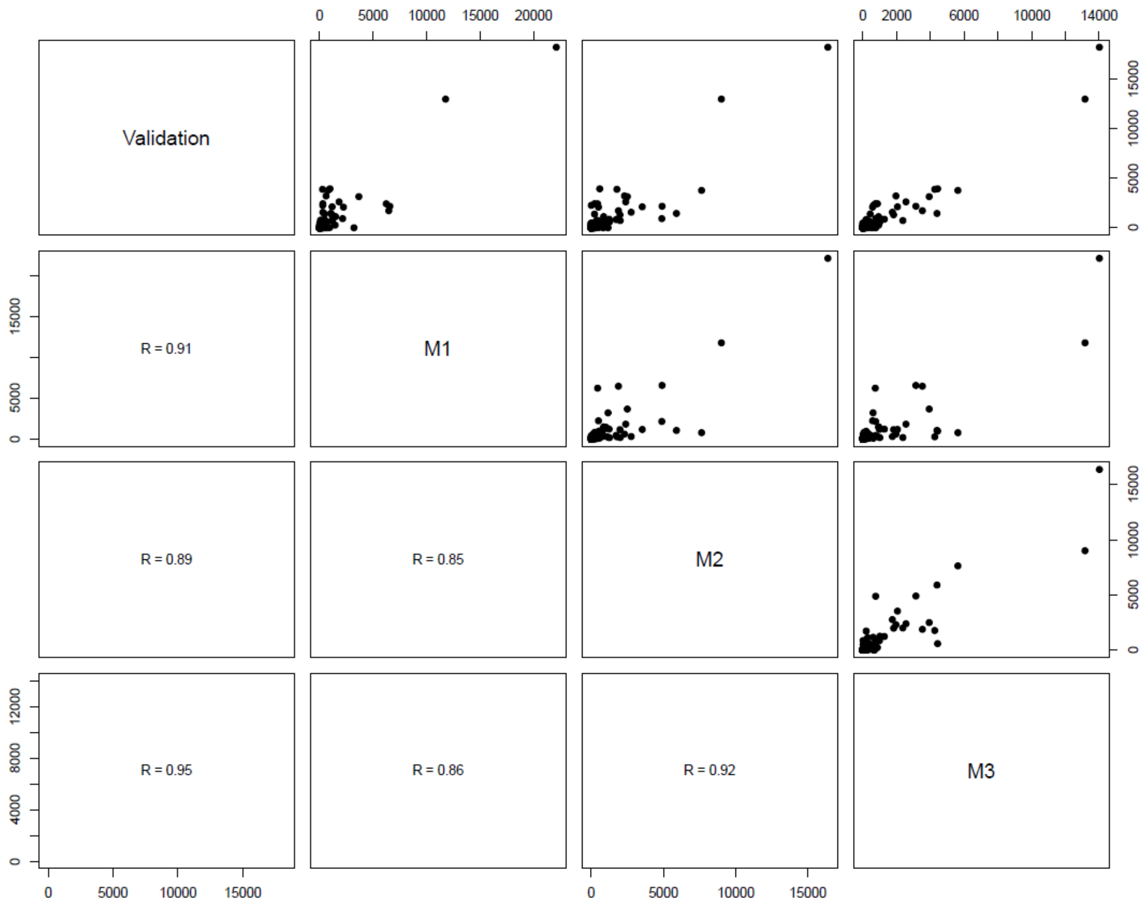
Source: BTS.

Figure 30. Pairwise Matrix Scatterplot Between Validation Data, M1, M2, and M3 With Pairwise Correlations: SCTG 15–19 Water Origin Data



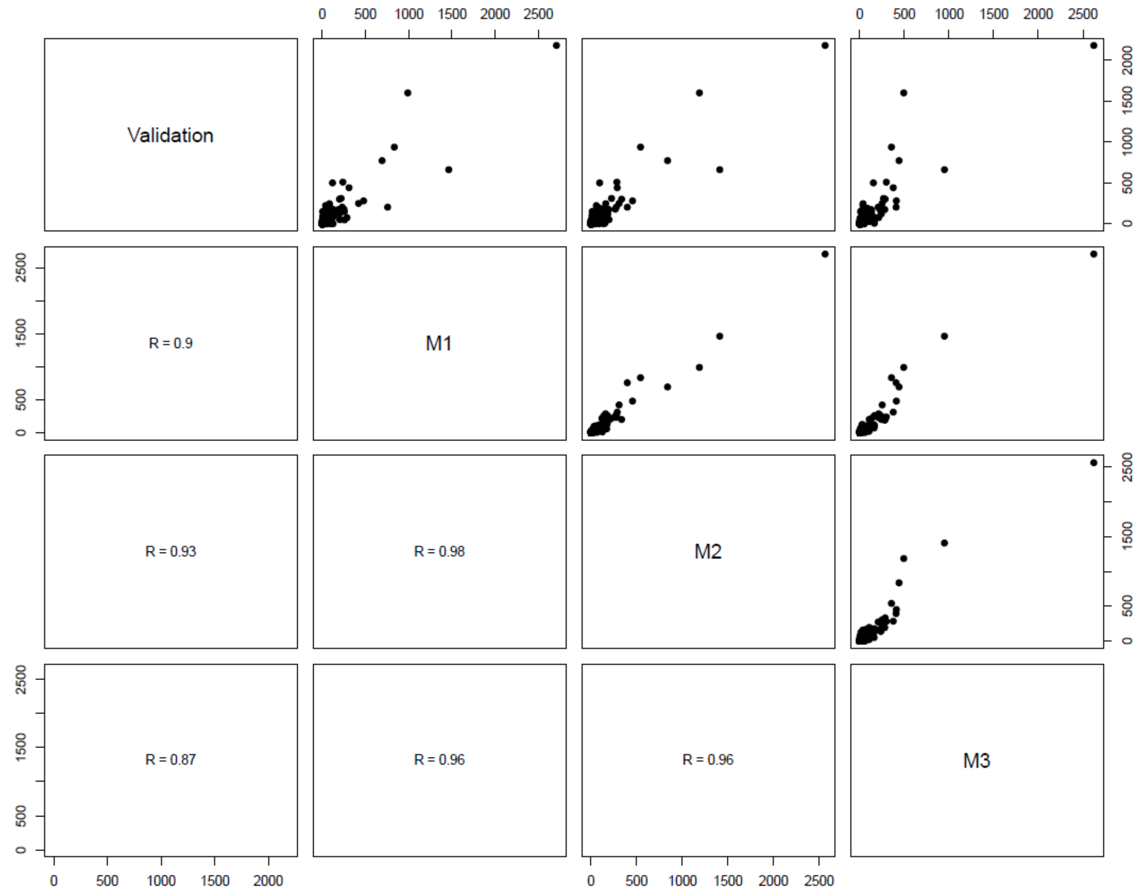
Source: BTS.

Figure 31. Pairwise Matrix Scatterplot Between Validation Data, M1, M2, and M3 With Pairwise Correlations: SCTG 20–33 Water Origin Data



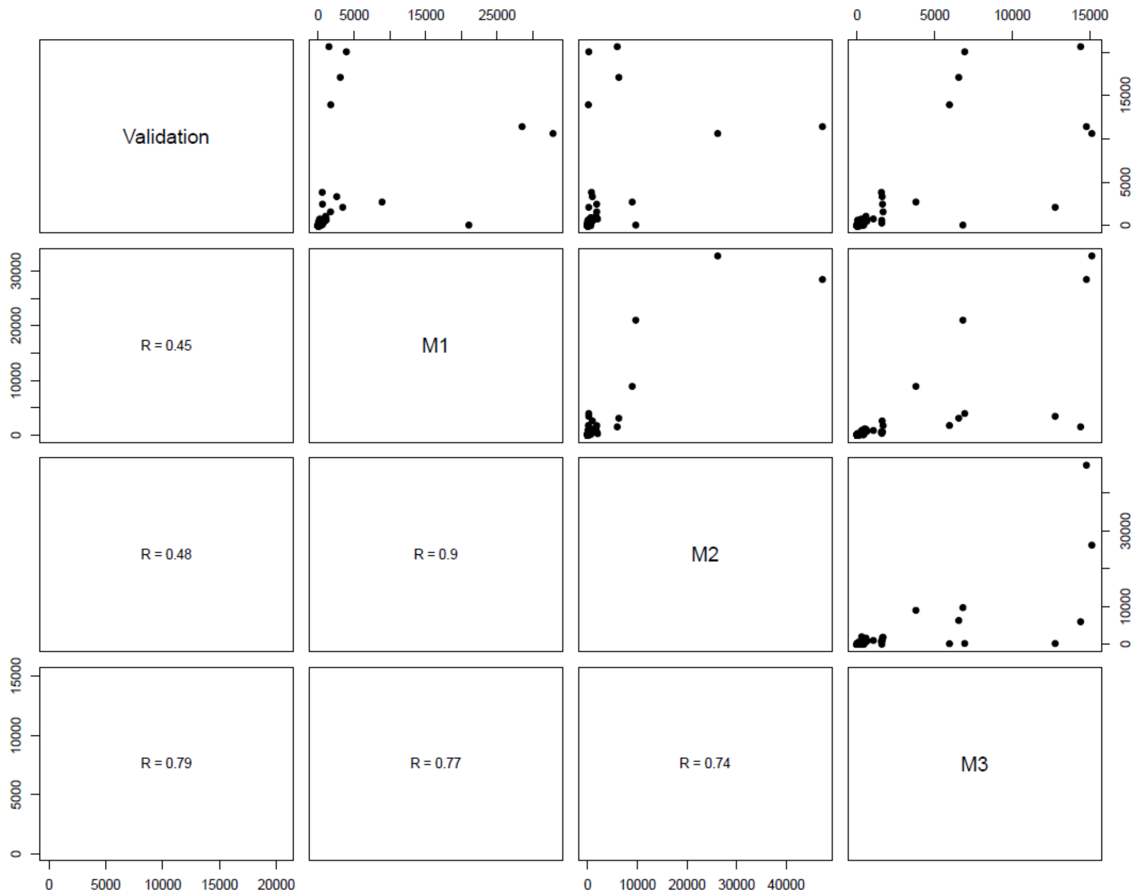
Source: BTS.

Figure 32. Pairwise Matrix Scatterplot Between Validation Data, M1, M2, and M3 With Pairwise Correlations: SCTG 34–99 Water Origin Data



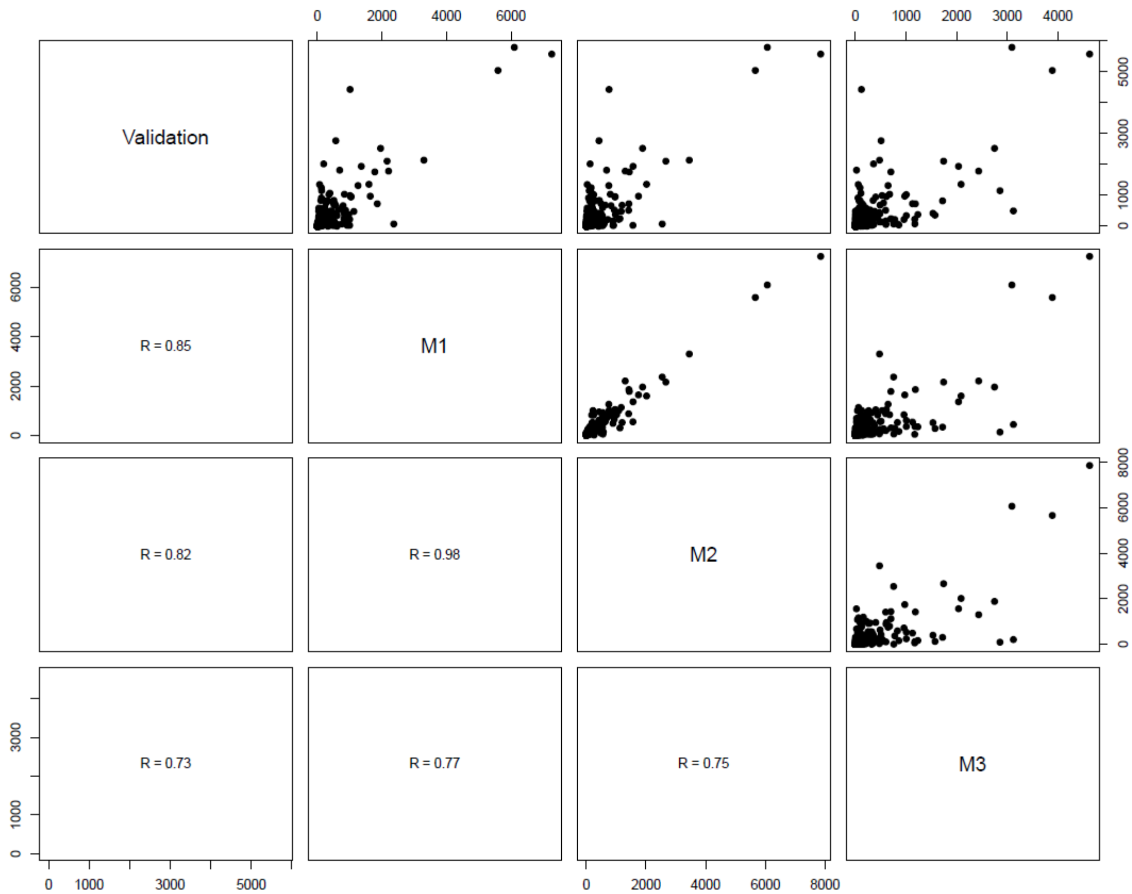
Source: BTS.

Figure 33. Pairwise Matrix Scatterplot Between Validation Data, M1, M2, and M3 With Pairwise Correlations: SCTG 01–09 Water Destination Data



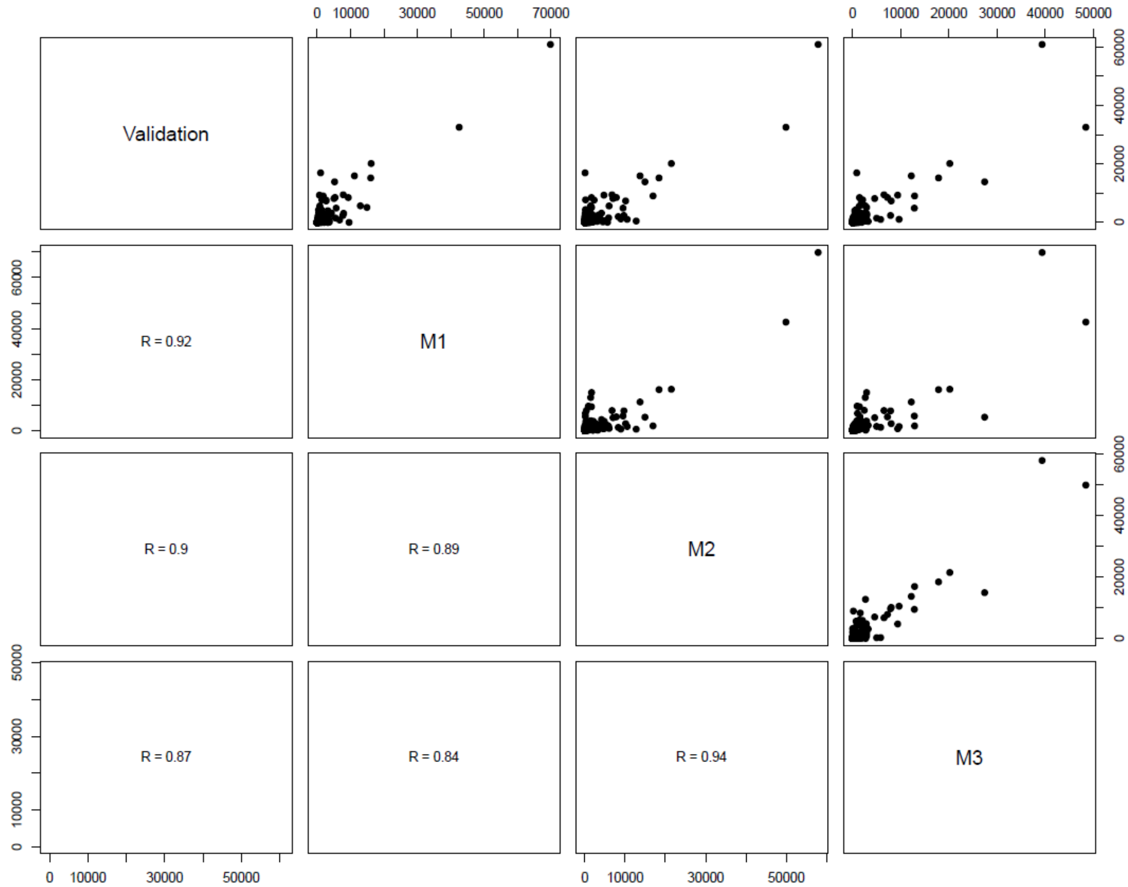
Source: BTS.

Figure 34. Pairwise Matrix Scatterplot Between Validation Data, M1, M2, and M3 With Pairwise Correlations: SCTG 10–14 Water Destination Data



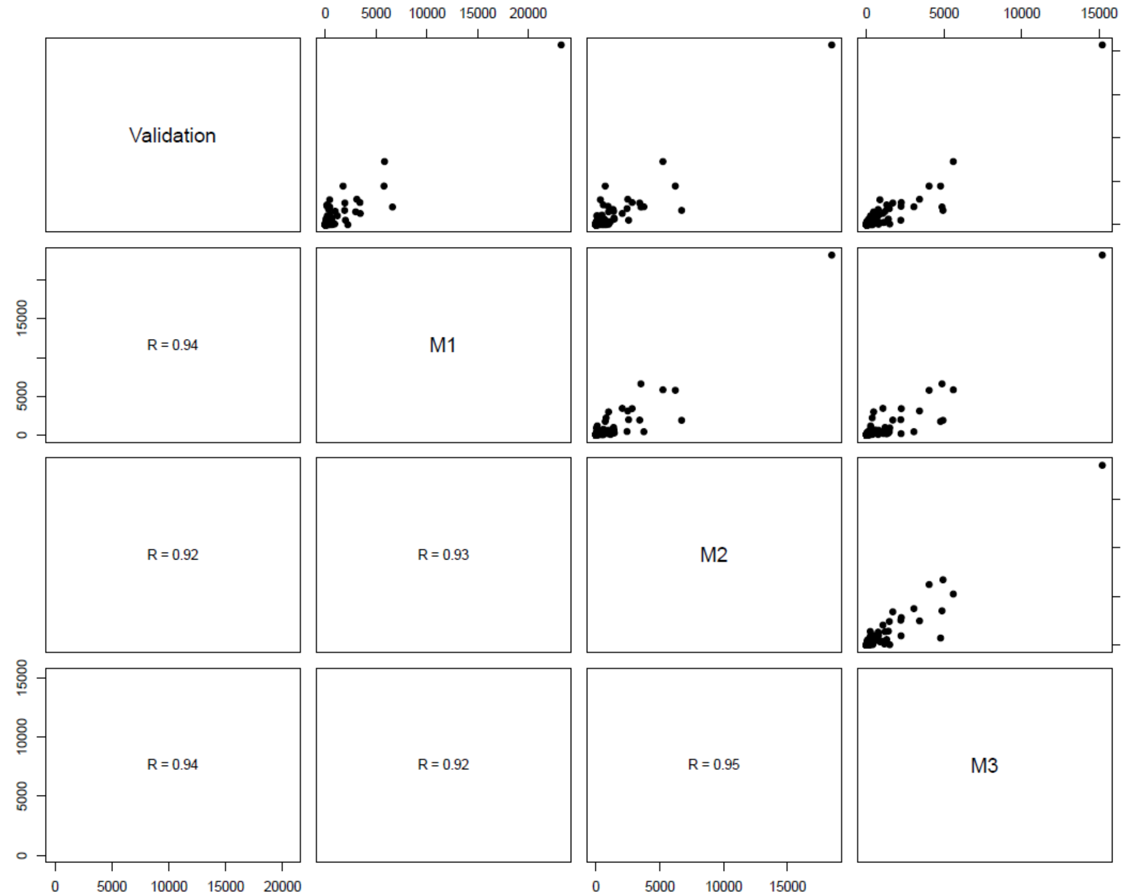
Source: BTS.

Figure 35. Pairwise Matrix Scatterplot Between Validation Data, M1, M2, and M3 With Pairwise Correlations: SCTG 15–19 Water Destination Data



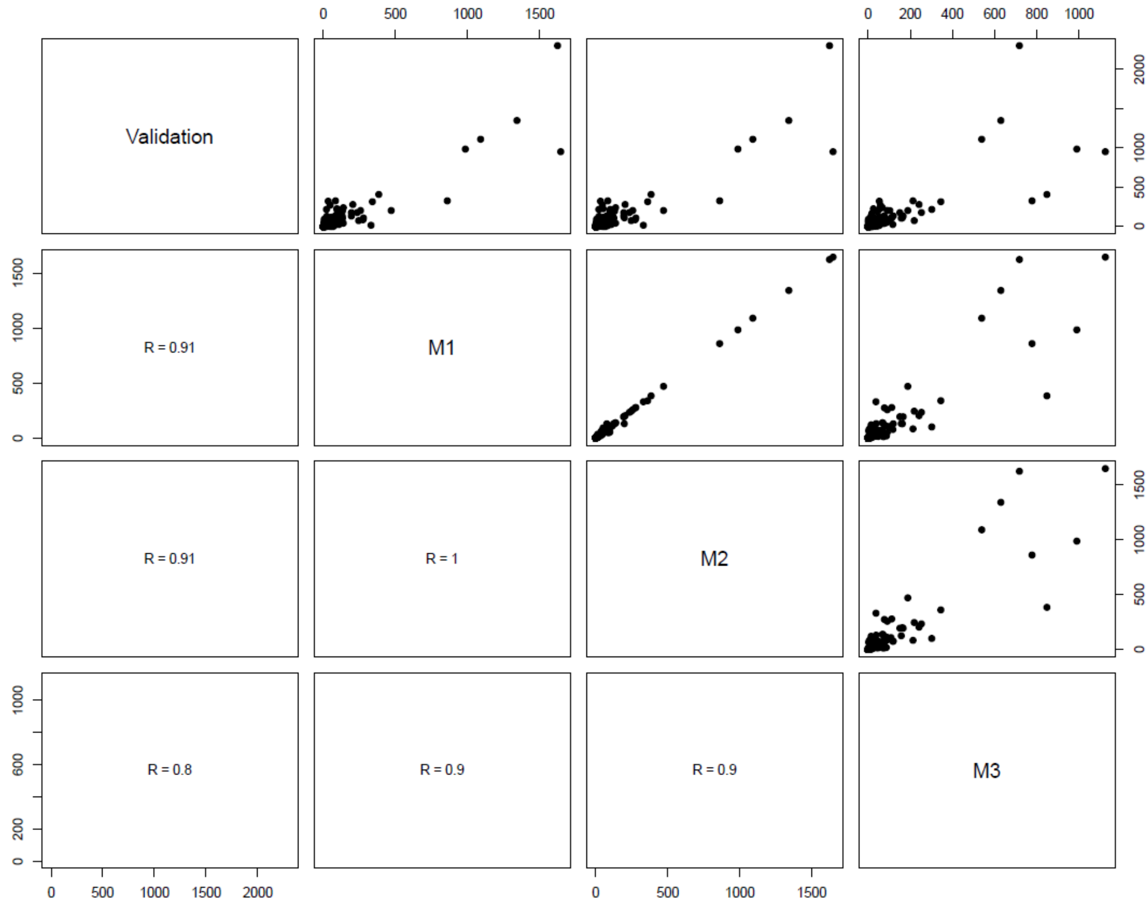
Source: BTS.

Figure 36. Pairwise Matrix Scatterplot Between Validation Data, M1, M2, and M3 With Pairwise Correlations: SCTG 20–33 Water Destination Data



Source: BTS.

Figure 37. Pairwise Matrix Scatterplot Between Validation Data, M1, M2, and M3 With Pairwise Correlations: SCTG 34–99 Water Destination Data



Source: BTS.

Appendix D. Regression Models for Composite Estimates

This appendix contains the detailed regression outputs for the composite estimation. Results are shown for truck, rail, and water.

Truck origin estimates for SCTG 01–09 are shown and discussed in the main text. Remaining results are presented in Table 17.

Table 17. Remaining Composite Estimate Regression Results for Truck Origins

SCTG	Model	Parameter	Estimate	Std. error	*t* statistic	*p*-value	Adj R-sq
10–14	OLS	*Intercept*	60.60	22.09	2.74	0.0061	0.7794
		M1	0.75	0.01	71.11	0.0000	
		M2	0.03	0.01	2.30	0.0215	
		M3	0.15	0.02	8.44	0.0000	
	Log-Log	*Intercept*	2.19	0.05	44.22	0.0000	0.6664
		log1p(*M*1)	0.60	0.01	48.95	0.0000	
		log1p(*M*2)	0.04	0.01	6.05	0.0000	
		log1p(*M*3)	0.03	0.01	2.98	0.0029	
15–19	OLS	*Intercept*	69.89	23.80	2.94	0.0034	0.8080
		M1	0.78	0.02	44.30	0.0000	
		M2	-0.01	0.01	-1.57	0.1172	
		M3	0.12	0.02	6.40	0.0000	
	Log-Log	*Intercept*	1.94	0.05	42.22	0.0000	0.6887
		log1p(*M*1)	0.66	0.01	50.21	0.0000	
		log1p(*M*2)	0.08	0.01	11.18	0.0000	
		log1p(*M*3)	-0.03	0.01	-2.78	0.0054	
20–33	OLS	*Intercept*	72.40	19.25	3.76	0.0002	0.8848
		M1	0.55	0.02	36.21	0.0000	
		M2	0.13	0.02	6.47	0.0000	
		M3	0.25	0.02	10.28	0.0000	
	Log-Log	*Intercept*	2.56	0.05	47.74	0.0000	0.6602
		log1p(*M*1)	0.60	0.02	38.75	0.0000	
		log1p(*M*2)	0.06	0.01	7.55	0.0000	
		log1p(*M*3)	-0.05	0.01	-3.79	0.0002	
34–99	OLS	*Intercept*	14.72	10.91	1.35	0.1771	0.8922
		M1	0.18	0.03	6.65	0.0000	
		M2	0.40	0.03	14.19	0.0000	
		M3	0.40	0.03	15.70	0.0000	
	Log-Log	*Intercept*	1.58	0.05	30.18	0.0000	0.6280
		log1p(*M*1)	0.72	0.03	28.05	0.0000	
		log1p(*M*2)	0.02	0.02	1.02	0.3091	
		log1p(*M*3)	-0.04	0.02	-2.31	0.0209	

Note: log1p() denotes log(value + 1).

Truck destination estimates for SCTG 01–09 are shown and discussed in the main text. Remaining results are presented in Table 18.

Table 18. Remaining Composite Estimate Regression Results for Truck Destinations

SCTG	Model	Parameter	Estimate	Std. error	*t* statistic	*p*-value	Adj R-sq
10–14	OLS	*Intercept*	0.90	0.02	40.46	0.0000	0.9196
		M1	0.12	0.02	6.18	0.0000	
		M2	0.01	0.01	0.89	0.3747	
		M3	1.02	0.06	16.96	0.0000	
	Log-Log	*Intercept*	0.93	0.02	39.68	0.0000	0.7262
		log1p(*M1*)	-0.11	0.02	-6.82	0.0000	
		log1p(*M2*)	-0.01	0.01	-1.25	0.2112	
		log1p(*M3*)	0.90	0.02	40.46	0.0000	
15–19	OLS	*Intercept*	0.94	0.01	86.74	0.0000	0.9519
		M1	0.07	0.01	5.82	0.0000	
		M2	0.03	0.01	2.84	0.0045	
		M3	1.18	0.05	22.20	0.0000	
	Log-Log	*Intercept*	0.79	0.02	40.35	0.0000	0.7391
		log1p(*M1*)	0.06	0.01	4.54	0.0000	
		log1p(*M2*)	-0.06	0.01	-6.27	0.0000	
		log1p(*M3*)	0.94	0.01	86.74	0.0000	
20–33	OLS	*Intercept*	0.58	0.02	29.84	0.0000	0.9296
		M1	0.32	0.02	15.37	0.0000	
		M2	0.10	0.02	4.92	0.0000	
		M3	1.56	0.05	29.82	0.0000	
	Log-Log	*Intercept*	0.69	0.02	37.06	0.0000	0.7169
		log1p(*M1*)	0.15	0.02	10.24	0.0000	
		log1p(*M2*)	-0.10	0.01	-9.96	0.0000	
		log1p(*M3*)	0.58	0.02	29.84	0.0000	
34–99	OLS	*Intercept*	0.59	0.05	11.75	0.0000	0.9408
		M1	0.07	0.05	1.46	0.1452	
		M2	0.36	0.02	17.14	0.0000	
		M3	1.03	0.05	21.55	0.0000	
	Log-Log	*Intercept*	0.93	0.06	14.71	0.0000	0.7313
		log1p(*M1*)	-0.09	0.06	-1.45	0.1473	
		log1p(*M2*)	-0.05	0.01	-3.44	0.0006	
		log1p(*M3*)	0.59	0.05	11.75	0.0000	

Note: log1p() denotes log(value + 1).

Table 19 shows composite estimates for rail origin data.

Table 19. Composite Estimate Linear Regression Without Transformation (OLS) and With Log Transformation (Log-Log) by Commodity Group: Rail Origin Data

SCTG	Model	Parameter	Estimate	Std. error	*t* statistic	*p*-value	Adj R-sq
01–09	OLS	*Intercept*	77.41	4.90	15.79	0.0000	0.3797
		M1	-0.19	0.02	-10.89	0.0000	
		M2	0.11	0.02	5.76	0.0000	
		M3	0.54	0.03	20.98	0.0000	
	Log-Log	*Intercept*	1.36	0.05	25.16	0.0000	0.5303
		log1p(*M*1)	0.43	0.03	14.90	0.0000	
		log1p(*M*2)	0.24	0.02	10.86	0.0000	
		log1p(*M*3)	0.07	0.02	2.98	0.0029	
10–14	OLS	*Intercept*	29.25	10.51	2.78	0.0055	0.6930
		M1	0.47	0.03	17.27	0.0000	
		M2	0.31	0.03	12.48	0.0000	
		M3	0.17	0.04	4.18	0.0000	
	Log-Log	*Intercept*	0.26	0.06	4.45	0.0000	0.5415
		log1p(*M*1)	0.64	0.02	26.00	0.0000	
		log1p(*M*2)	0.01	0.02	0.56	0.5723	
		log1p(*M*3)	0.18	0.03	6.17	0.0000	
15–19	OLS	*Intercept*	37.85	45.62	0.83	0.4069	0.7655
		M1	-0.24	0.03	-8.40	0.0000	
		M2	1.04	0.04	27.93	0.0000	
		M3	0.19	0.05	3.82	0.0001	
	Log-Log	*Intercept*	1.00	0.08	12.21	0.0000	0.3440
		log1p(*M*1)	0.21	0.03	7.14	0.0000	
		log1p(*M*2)	0.06	0.02	2.59	0.0097	
		log1p(*M*3)	0.38	0.03	12.31	0.0000	
20–33	OLS	*Intercept*	32.81	8.43	3.89	0.0001	0.4216
		M1	0.48	0.04	12.51	0.0000	
		M2	-0.34	0.03	-9.97	0.0000	
		M3	0.59	0.04	14.07	0.0000	
	Log-Log	*Intercept*	1.17	0.06	18.96	0.0000	0.3464
		log1p(*M*1)	0.19	0.03	6.07	0.0000	
		log1p(*M*2)	0.09	0.02	3.68	0.0002	
		log1p(*M*3)	0.34	0.03	11.32	0.0000	
34–99	OLS	*Intercept*	1.52	1.69	0.90	0.3680	0.6965
		M1	0.20	0.04	4.76	0.0000	
		M2	0.15	0.03	4.93	0.0000	
		M3	0.63	0.04	16.51	0.0000	
	Log-Log	*Intercept*	0.12	0.03	3.52	0.0004	0.6096
		log1p(*M*1)	0.39	0.04	9.78	0.0000	
		log1p(*M*2)	0.12	0.03	3.89	0.0001	
		log1p(*M*3)	0.34	0.03	10.20	0.0000	

Note: log1p() denotes log(value + 1).

Table 20 shows composite estimates for rail destinations.

Table 20. Composite Estimate Linear Regression Without Transformation (OLS) and With Log Transformation (Log-Log) by Commodity Group: Rail Destination

SCTG	Model	Parameter	Estimate	Std. error	*t* statistic	*p*-value	Adj R-sq
01–09	OLS	*Intercept*	-0.37	0.02	-20.71	0.0000	0.8500
		M1	0.61	0.03	19.40	0.0000	
		M2	0.82	0.03	27.83	0.0000	
		M3	0.27	0.08	3.36	0.0008	
	Log-Log	*Intercept*	0.48	0.04	11.91	0.0000	0.4066
		log1p(*M*1)	0.11	0.03	4.06	0.0000	
		log1p(*M*2)	0.22	0.03	6.63	0.0000	
		log1p(*M*3)	-0.37	0.02	-20.71	0.0000	
10–14	OLS	*Intercept*	44.82	6.16	7.28	0.0000	0.4871
		M1	0.37	0.03	11.68	0.0000	
		M2	-0.13	0.03	-4.74	0.0000	
		M3	0.33	0.03	11.05	0.0000	
	Log-Log	*Intercept*	0.68	0.05	12.72	0.0000	0.4588
		log1p(*M*1)	0.57	0.04	16.02	0.0000	
		log1p(*M*2)	0.01	0.03	0.42	0.6763	
		log1p(*M*3)	0.16	0.03	6.17	0.0000	
15–19	OLS	*Intercept*	48.45	12.88	3.76	0.0002	0.8855
		M1	0.19	0.02	9.17	0.0000	
		M2	0.10	0.02	5.01	0.0000	
		M3	0.58	0.02	27.66	0.0000	
	Log-Log	*Intercept*	0.36	0.12	2.91	0.0037	0.2929
		log1p(*M*1)	0.63	0.04	15.33	0.0000	
		log1p(*M*2)	0.07	0.03	2.57	0.0103	
		log1p(*M*3)	0.09	0.04	2.62	0.0089	
20–33	OLS	*Intercept*	9.55	4.92	1.94	0.0523	0.6134
		M1	0.08	0.04	1.83	0.0677	
		M2	-0.16	0.05	-3.57	0.0004	
		M3	1.02	0.05	21.47	0.0000	
	Log-Log	*Intercept*	0.06	0.08	0.78	0.4371	0.4659
		log1p(*M*1)	0.23	0.04	5.45	0.0000	
		log1p(*M*2)	0.34	0.03	9.99	0.0000	
		log1p(*M*3)	0.30	0.04	8.62	0.0000	
34–99	OLS	*Intercept*	9.25	1.04	8.93	0.0000	0.5761
		M1	0.16	0.05	3.00	0.0027	
		M2	-0.12	0.05	-2.32	0.0205	
		M3	0.61	0.03	17.76	0.0000	
	Log-Log	*Intercept*	0.27	0.04	6.65	0.0000	0.4851
		log1p(*M*1)	0.64	0.08	8.13	0.0000	
		log1p(*M*2)	-0.36	0.07	-4.90	0.0000	
		log1p(*M*3)	0.52	0.04	12.57	0.0000	

Note: log1p() denotes log(value + 1).

Table 21 shows composite estimates for water origin data.

Table 21. Composite Estimate Linear Regression Without Transformation (OLS) and With Log Transformation (Log-Log) by Commodity Group: Water Origins

SCTG	Model	Parameter	Estimate	Std. error	*t* statistic	*p*-value	Adj R-sq
01–09	OLS	*Intercept*	102.68	23.61	4.35	0.0000	0.5694
		M1	0.10	0.04	2.46	0.0143	
		M2	0.04	0.05	0.94	0.3471	
		M3	0.57	0.05	10.86	0.0000	
	Log-Log	*Intercept*	0.57	0.12	4.98	0.0000	0.7013
		log1p(*M*1)	0.77	0.05	14.62	0.0000	
		log1p(*M*2)	0.13	0.04	2.90	0.0040	
		log1p(*M*3)	-0.02	0.04	-0.52	0.6013	
10–14	OLS	*Intercept*	130.22	31.77	4.10	0.0000	0.2403
		M1	0.31	0.05	5.72	0.0000	
		M2	0.18	0.03	5.35	0.0000	
		M3	-0.05	0.07	-0.76	0.4479	
	Log-Log	*Intercept*	0.44	0.09	5.01	0.0000	0.7591
		log1p(*M*1)	0.84	0.03	25.36	0.0000	
		log1p(*M*2)	0.03	0.03	0.84	0.4004	
		log1p(*M*3)	0.03	0.04	0.81	0.4173	
15–19	OLS	*Intercept*	160.25	70.12	2.29	0.0228	0.8200
		M1	0.54	0.03	21.13	0.0000	
		M2	-0.08	0.04	-2.02	0.0444	
		M3	0.42	0.05	9.20	0.0000	
	Log-Log	*Intercept*	0.32	0.11	3.04	0.0025	0.8092
		log1p(*M*1)	0.68	0.04	16.79	0.0000	
		log1p(*M*2)	-0.09	0.02	-3.77	0.0002	
		log1p(*M*3)	0.34	0.04	7.63	0.0000	
20–33	OLS	*Intercept*	2.65	13.62	0.19	0.8461	0.9366
		M1	0.30	0.02	14.10	0.0000	
		M2	-0.05	0.03	-1.36	0.1748	
		M3	0.76	0.03	21.97	0.0000	
	Log-Log	*Intercept*	0.55	0.08	6.81	0.0000	0.7173
		log1p(*M*1)	0.59	0.06	10.22	0.0000	
		log1p(*M*2)	-0.15	0.05	-3.21	0.0014	
		log1p(*M*3)	0.40	0.06	6.87	0.0000	
34–99	OLS	*Intercept*	4.51	2.60	1.73	0.0841	0.8597
		M1	-0.06	0.09	-0.64	0.5201	
		M2	1.04	0.09	12.16	0.0000	
		M3	-0.14	0.07	-1.99	0.0476	
	Log-Log	*Intercept*	0.28	0.05	6.17	0.0000	0.7997
		log1p(*M*1)	0.14	0.08	1.74	0.0822	
		log1p(*M*2)	0.37	0.06	5.80	0.0000	
		log1p(*M*3)	0.39	0.06	6.83	0.0000	

Note: log1p() denotes log(value + 1).

Table 22 shows composite estimates for water destinations.

Table 22. Composite Estimate Linear Regression Without Transformation (OLS) and With Log Transformation (Log-Log) by Commodity Group: Water Destinations

SCTG	Model	Parameter	Estimate	Std. error	*t* statistic	*p*-value	Adj R-sq
01–09	OLS	*Intercept*	44.02	55.58	0.79	0.4289	0.6887
		M1	-0.41	0.06	-7.40	0.0000	
		M2	0.10	0.05	2.11	0.0358	
		M3	1.36	0.06	24.49	0.0000	
	Log-Log	*Intercept*	0.09	0.07	1.16	0.2459	0.7724
		log1p(*M*1)	0.51	0.08	6.44	0.0000	
		log1p(*M*2)	0.04	0.07	0.67	0.5006	
		log1p(*M*3)	0.35	0.06	5.62	0.0000	
10–14	OLS	*Intercept*	26.34	15.08	1.75	0.0814	0.7277
		M1	0.82	0.11	7.52	0.0000	
		M2	-0.13	0.10	-1.33	0.1836	
		M3	0.22	0.05	4.75	0.0000	
	Log-Log	*Intercept*	0.30	0.10	2.94	0.0035	0.7445
		log1p(*M*1)	0.80	0.06	12.35	0.0000	
		log1p(*M*2)	-0.09	0.05	-1.80	0.0721	
		log1p(*M*3)	0.19	0.04	4.60	0.0000	
15–19	OLS	*Intercept*	93.40	61.62	1.52	0.1303	0.8763
		M1	0.51	0.03	15.87	0.0000	
		M2	0.20	0.05	4.09	0.0001	
		M3	0.20	0.05	4.25	0.0000	
	Log-Log	*Intercept*	0.38	0.10	3.73	0.0002	0.7983
		log1p(*M*1)	0.62	0.05	12.16	0.0000	
		log1p(*M*2)	-0.01	0.04	-0.18	0.8561	
		log1p(*M*3)	0.29	0.04	6.74	0.0000	
20–33	OLS	*Intercept*	0.93	14.29	0.06	0.9483	0.9219
		M1	0.43	0.03	12.66	0.0000	
		M2	0.03	0.05	0.60	0.5490	
		M3	0.58	0.05	11.03	0.0000	
	Log-Log	*Intercept*	0.34	0.10	3.53	0.0005	0.7050
		log1p(*M*1)	0.43	0.07	6.29	0.0000	
		log1p(*M*2)	0.08	0.06	1.36	0.1751	
		log1p(*M*3)	0.39	0.07	5.78	0.0000	
34–99	OLS	*Intercept*	3.46	2.98	1.16	0.2463	0.8356
		M1	1.57	0.49	3.19	0.0015	
		M2	-0.53	0.49	-1.08	0.2814	
		M3	-0.18	0.06	-2.81	0.0051	
	Log-Log	*Intercept*	0.11	0.05	2.29	0.0223	0.8096
		log1p(*M*1)	0.81	0.17	4.70	0.0000	
		log1p(*M*2)	-0.29	0.17	-1.70	0.0906	
		log1p(*M*3)	0.44	0.07	6.76	0.0000	

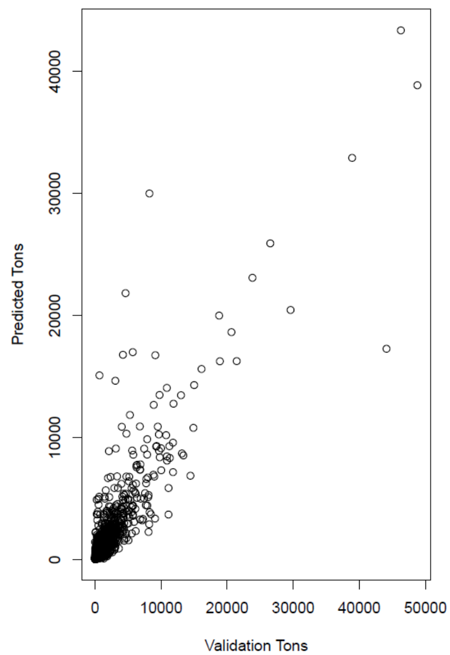
Note: log1p() denotes log(value + 1).

Appendix E. Predicted Versus Observed Values: Scatterplots

Figure 38–Figure 51 are scatterplots illustrating predicted versus validated tons by SCTG groups and modes. Truck origin scatter plots for SCTG 01–09 are shown and discussed in the main text.

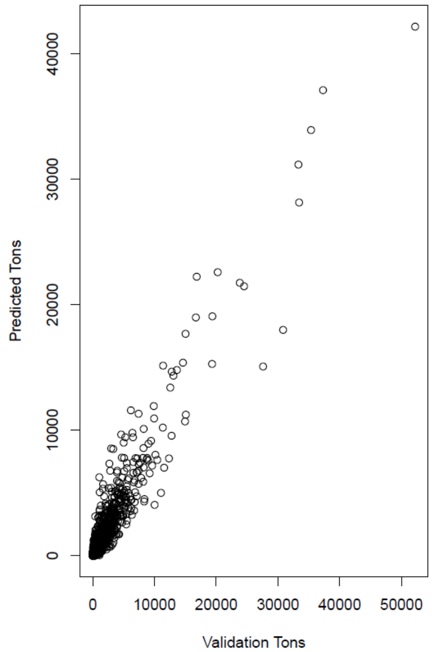
Figure 38. Predicted Versus Validated Tons for SCTG 10–14 Truck Origins and Destinations

A. Origins



Source: BTS.

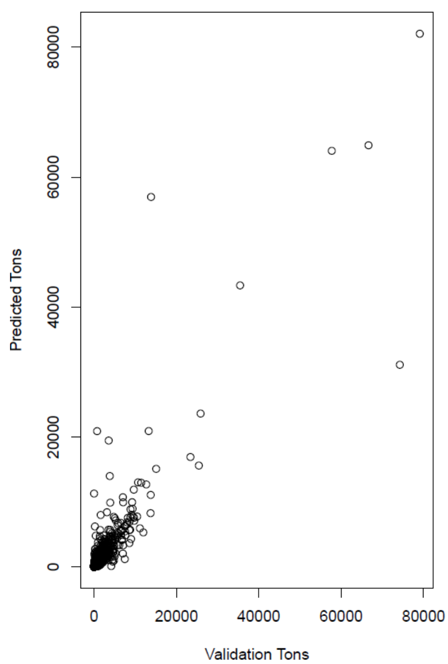
B. Destinations



Source: BTS.

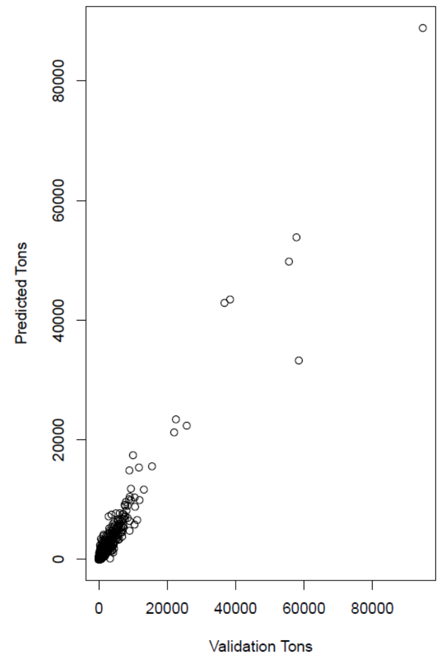
Figure 39. Predicted Versus Validated Tons for SCTG 15–19 Truck Origins and Destinations

A. Origins



Source: BTS.

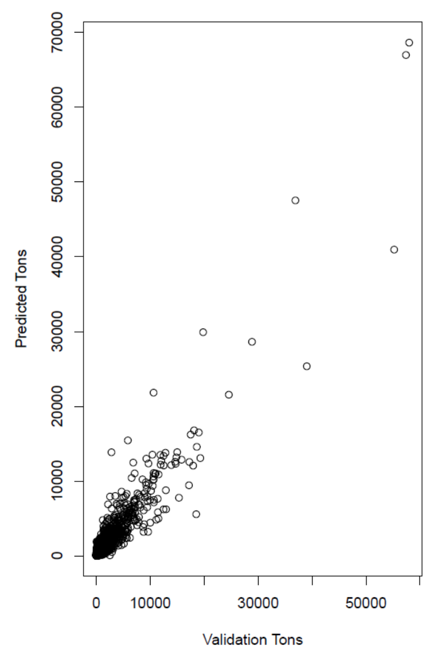
B. Destinations



Source: BTS.

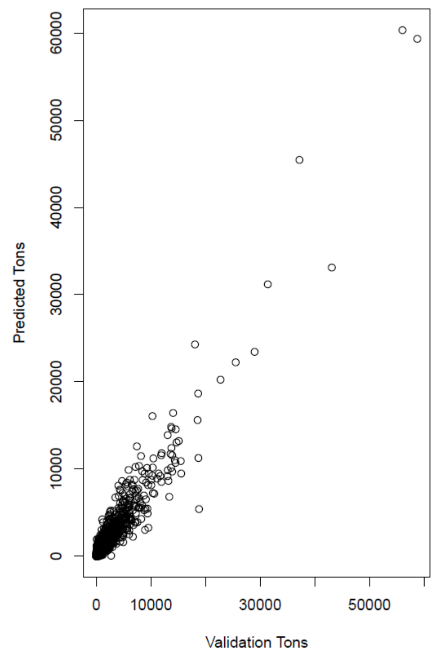
Figure 40. Predicted Versus Validated Tons for SCTG 20–33 Truck Origins and Destinations

A. Origins



Source: BTS.

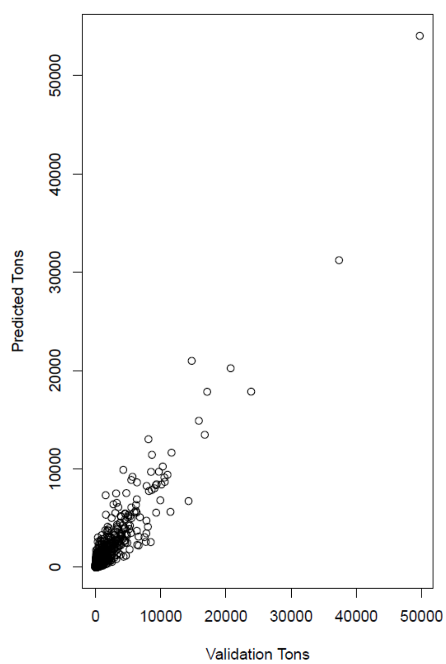
B. Destinations



Source: BTS.

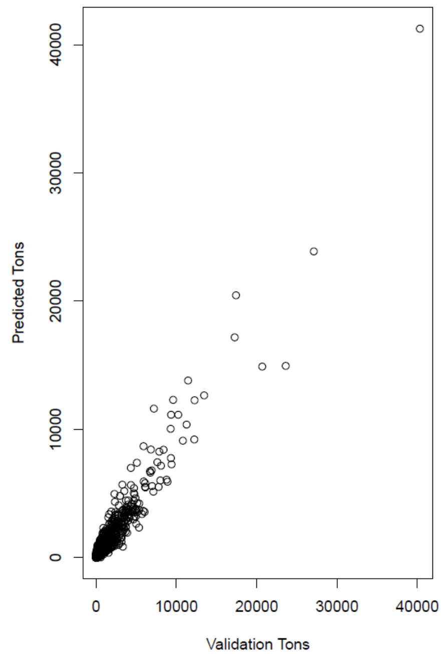
Figure 41. Predicted Versus Validated Tons for SCTG 34–99 Truck Origins and Destinations

A. Origins



Source: BTS.

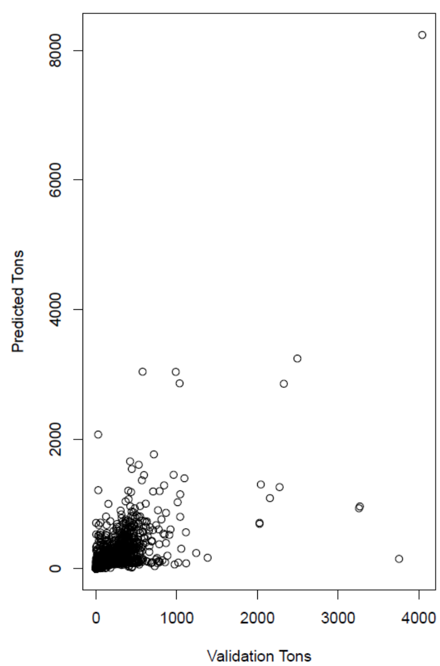
B. Destinations



Source: BTS.

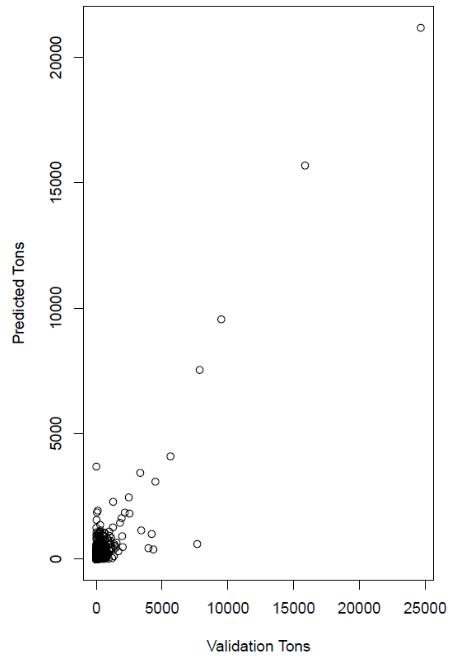
Figure 42. Predicted Versus Validated Tons for SCTG 01–09 Rail Origins and Destinations

A. Origins



Source: BTS.

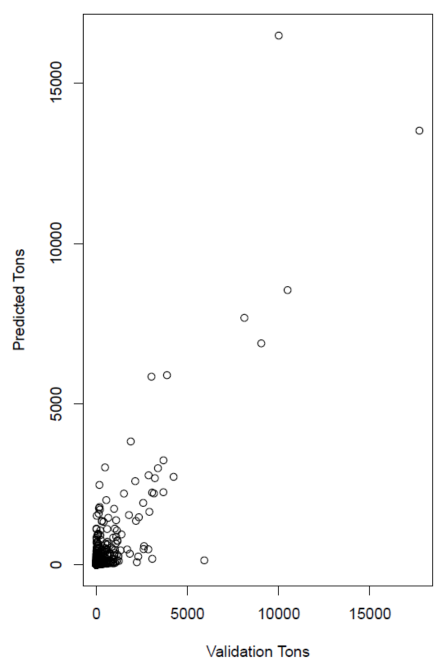
B. Destinations



Source: BTS.

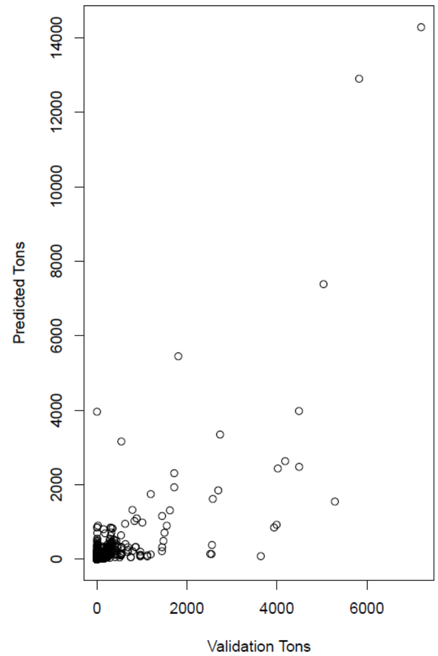
Figure 43. Predicted Versus Validated Tons for SCTG 10–14 Rail Origins and Destinations

A. Origins



Source: BTS.

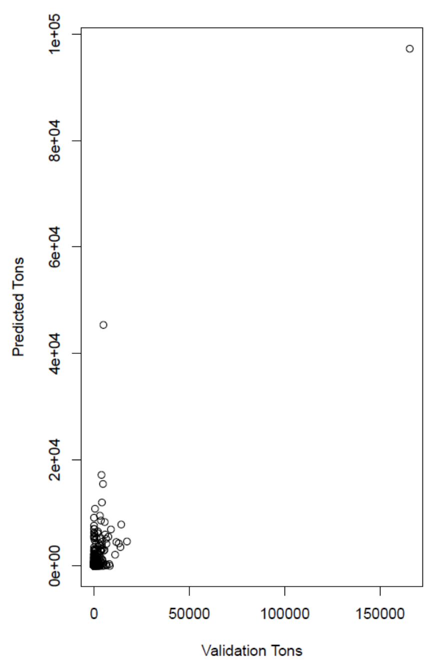
B. Destinations



Source: BTS.

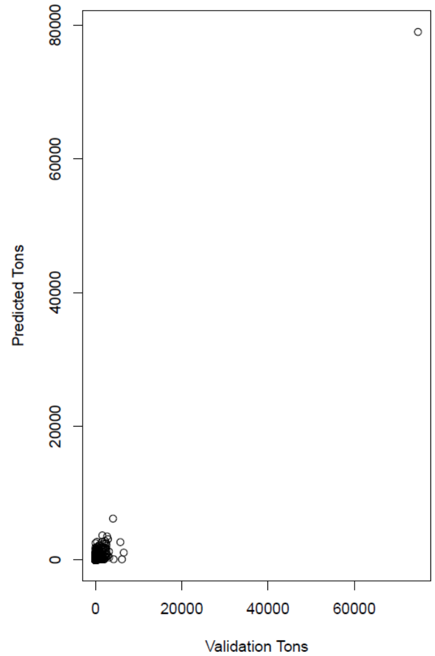
Figure 44. Predicted Versus Validated Tons for SCTG 15–19 Rail Origins and Destinations

A. Origins



Source: BTS.

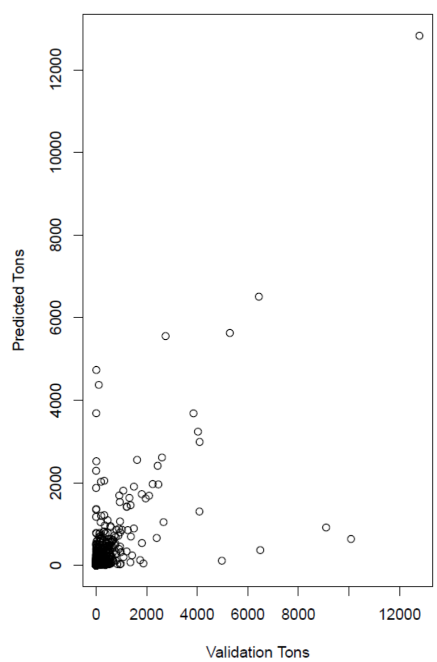
B. Destinations



Source: BTS.

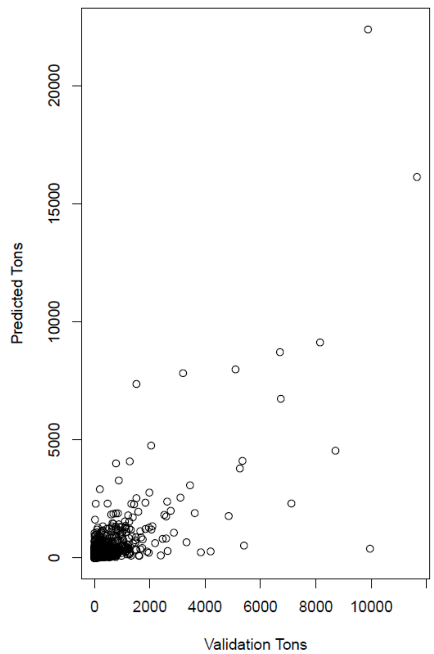
Figure 45. Predicted Versus Validated Tons for SCTG 20–33 Rail Origins and Destinations

A. Origins



Source: BTS.

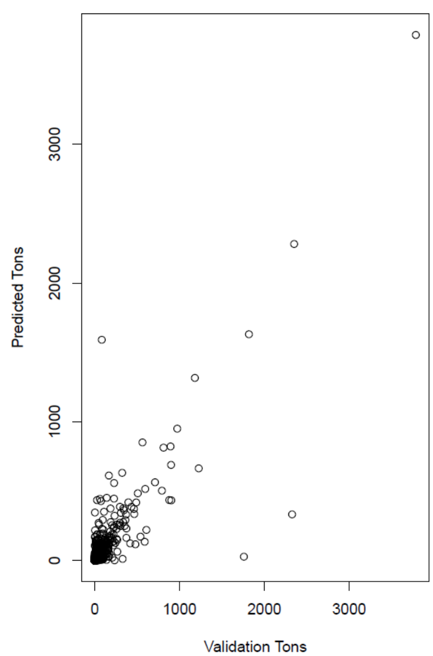
B. Destinations



Source: BTS.

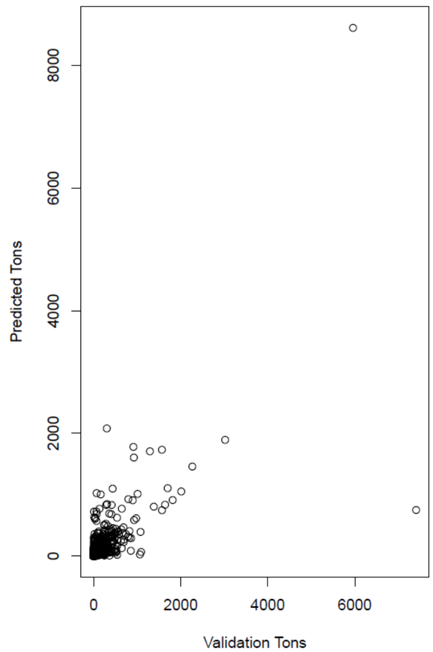
Figure 46. Predicted Versus Validated Tons for SCTG 34–99 Rail Origins and Destinations

A. Origins



Source: BTS.

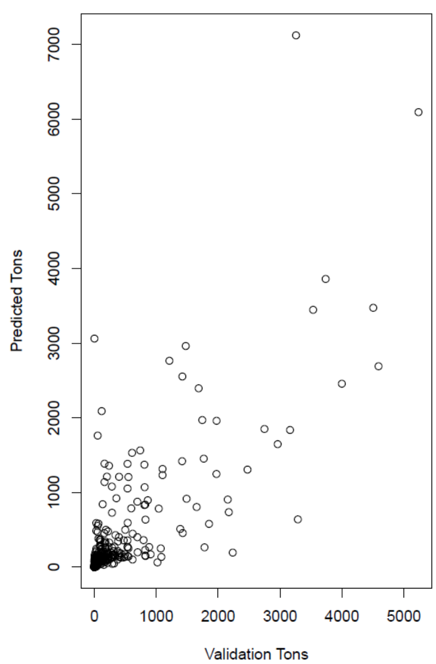
B. Destinations



Source: BTS.

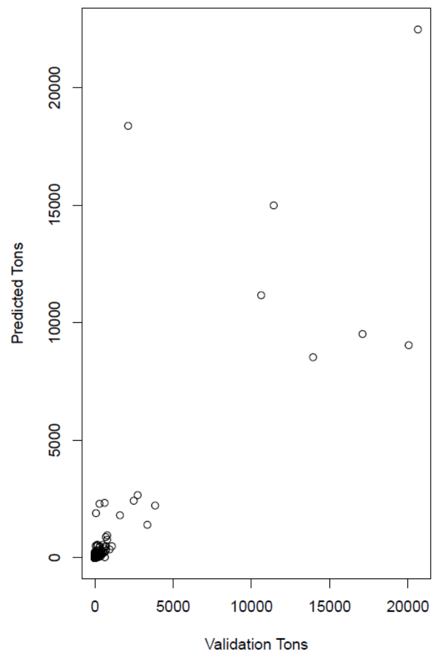
Figure 47. Predicted Versus Validated Tons for SCTG 01–09 Water Origins and Destinations

A. Origins



Source: BTS.

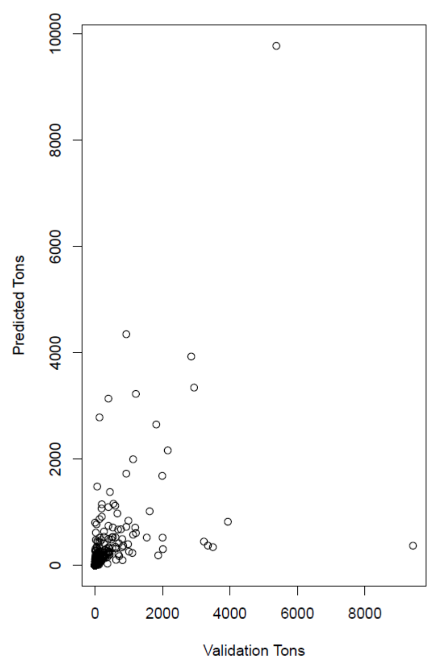
B. Destinations



Source: BTS.

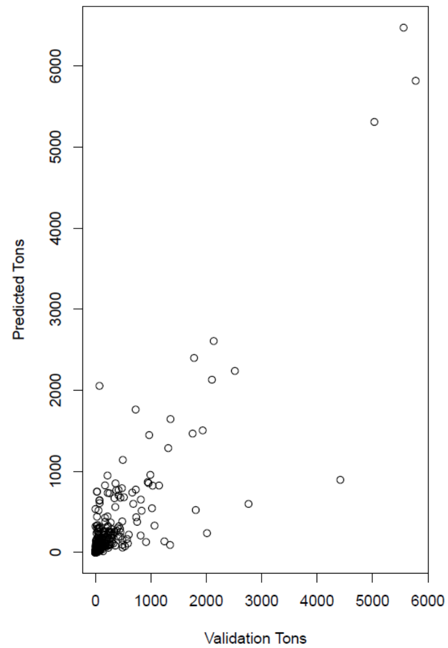
Figure 48. Predicted Versus Validated Tons for SCTG 10–14 Water Origins and Destinations

A. Origins



Source: BTS.

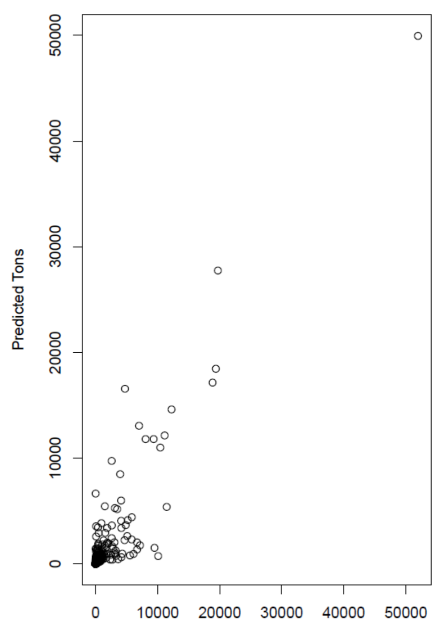
B. Destinations



Source: BTS.

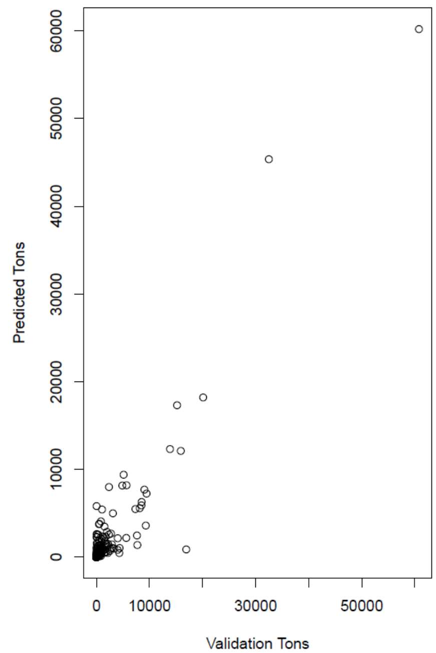
Figure 49. Predicted Versus Validated Tons for SCTG 15–19 Water Origins and Destinations

A. Origins



Source: BTS.

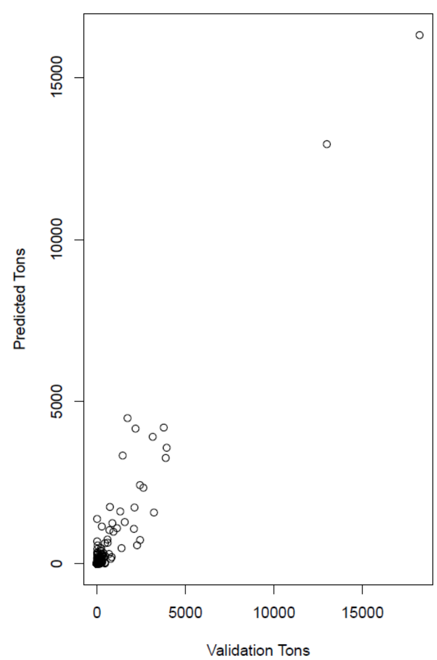
B. Destinations



Source: BTS.

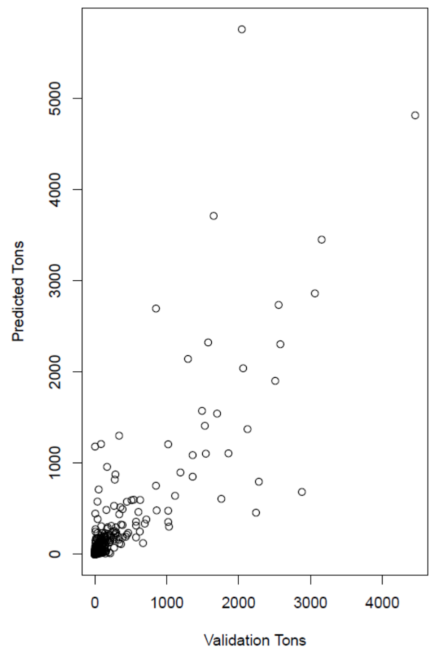
Figure 50. Predicted Versus Validated Tons for SCTG 20–33 Water Origins and Destinations

A. Origins



Source: BTS.

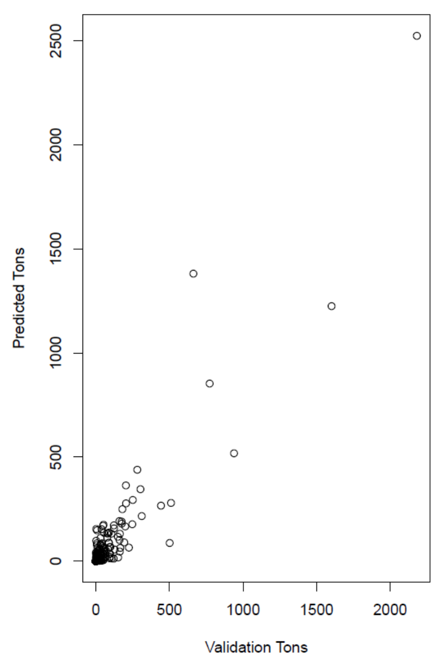
B. Destinations



Source: BTS.

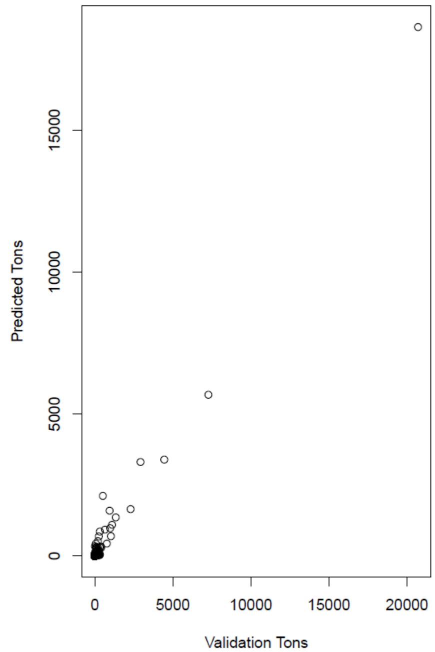
Figure 51. Predicted Versus Validated Tons for SCTG 34–99 Water Origins and Destinations

A. Origins



Source: BTS.

B. Destinations



Source: BTS.

Appendix F. Statistical Summary of Results

Table 23–Table 50 summarize results by SCTG group and mode. Truck origin and destination summaries for SCTG 01–09 are shown and discussed in the main text.

Table 23. Summary Statistics of Validation Data, M1, M2, M3, and Rescaled OLS and Log-Log Composite Estimates: SCTG 10–14 Truck Origins

Variable	Min	Median	Mean	Max	Sum
Validation	0.01	217.37	845.54	48,788.71	2,658,388
M1	0.00	176.14	845.54	47,348.32	2,658,388
M2	0.00	0.00	836.02	63,480.65	2,628,450
M3	0.00	30.83	845.49	41,982.84	2,658,224
OLS	33.35	212.73	845.54	43,359.11	2,658,388
Log-Log	6.99	284.23	845.54	36,620.23	2,658,388

Table 24. Summary Statistics of Validation Data, M1, M2, M3, and Rescaled OLS and Log-Log Composite Estimates: SCTG 15–19 Truck Origins

Variable	Min	Median	Mean	Max	Sum
Validation	0.01	138.48	645.99	79,077.51	2,030,982
M1	0.00	106.03	645.99	85,068.88	2,030,982
M2	0.00	0.00	580.19	91,875.09	1,824,107
M3	0.00	119.15	645.94	67,309.29	2,030,823
OLS	0.00	144.00	645.99	82,059.72	2,030,982
Log-Log	0.52	158.65	645.99	64,710.77	2,030,982

Table 25. Summary Statistics of Validation Data, M1, M2, M3, and Rescaled OLS and Log-Log Composite Estimates: SCTG 20–33 Truck Origins

Variable	Min	Median	Mean	Max	Sum
Validation	0.01	297.34	1079.37	57,995.52	3,393,525
M1	0.00	252.09	1079.37	72,723.68	3,393,525
M2	0.00	270.20	1077.14	65,057.79	3,386,519
M3	0.00	347.59	1079.28	68,411.32	3,393,253
OLS	48.16	335.32	1079.37	68,642.62	3,393,525
Log-Log	9.85	450.53	1079.37	52,709.39	3,393,525

Table 26. Summary Statistics of Validation Data, M1, M2, M3, and Rescaled OLS and Log-Log Composite Estimates: SCTG 34–99 Truck Origins

Variable	Min	Median	Mean	Max	Sum
Validation	0.00	103.94	516.35	49,710.27	1,623,406
M1	0.00	94.54	516.35	56,108.86	1,623,406
M2	0.00	88.90	514.11	52,712.44	1,616,367
M3	0.00	114.27	516.33	54,528.10	1,623,343
OLS	11.60	111.63	516.35	54,054.43	1,623,406
Log-Log	4.79	153.58	516.35	44,854.84	1,623,406

Table 27. Summary Statistics of Validation Data, M1, M2, M3, and Rescaled OLS and Log-Log Composite Estimates: SCTG 10–14 Truck Destinations

Variable	Min	Median	Mean	Max	Sum
Validation	0.01	213.79	845.54	52,217.95	2,658,388
M1	0.35	235.86	845.54	43,627.76	2,658,388
M2	0.00	166.47	836.02	38,275.58	2,628,450
M3	0.00	175.50	845.49	47,681.02	2,658,224
OLS	0.32	233.29	845.54	42,160.92	2,658,388
Log-Log	1.74	298.18	845.54	39,348.87	2,658,388

Table 28. Summary Statistics of Validation Data, M1, M2, M3, and Rescaled OLS and Log-Log Composite Estimates: SCTG 15–19 Truck Destinations

Variable	Min	Median	Mean	Max	Sum
Validation	0.01	138.13	645.99	94,743.21	2,030,982
M1	1.08	145.91	645.99	89,452.74	2,030,982
M2	0.00	51.84	580.19	91,808.06	1,824,107
M3	0.00	196.26	645.94	66,166.80	2,030,823
OLS	1.12	147.66	645.99	88,862.31	2,030,982
Log-Log	4.82	175.24	645.99	72,337.67	2,030,982

Table 29. Summary Statistics of Validation Data, M1, M2, M3, and Rescaled OLS and Log-Log Composite Estimates: SCTG 20–33 Truck Destinations

Variable	Min	Median	Mean	Max	Sum
Validation	0.01	297.56	1079.37	58,756.96	3,393,525
M1	0.23	343.29	1079.37	63,812.65	3,393,525
M2	0.01	316.58	1077.14	62,162.18	3,386,519
M3	0.00	523.91	1079.28	57,960.56	3,393,271
OLS	0.13	384.57	1079.37	60,334.01	3,393,525
Log-Log	4.54	445.69	1079.37	52,282.56	3,393,525

Table 30. Summary Statistics of Validation Data, M1, M2, M3, and Rescaled OLS and Log-Log Composite Estimates: SCTG 34–99 Truck Destinations

Variable	Min	Median	Mean	Max	Sum
Validation	0.00	112.84	516.35	40,321.60	1,623,406
M1	0.17	136.66	516.35	41,877.03	1,623,406
M2	0.17	121.98	514.11	42,048.33	1,616,367
M3	0.00	169.42	516.33	40,026.92	1,623,343
OLS	0.11	149.08	516.35	41,252.67	1,623,406
Log-Log	2.89	167.95	516.35	36,019.13	1,623,406

Table 31. Summary Statistics of Validation Data, M1, M2, M3, and Rescaled OLS and Log-Log Composite Estimates: SCTG 01–09 Rail Origins

Variable	Min	Median	Mean	Max	Sum
Validation	0.00	42.58	141.12	4,040.61	298,609
M1	0.00	20.97	134.87	12,126.53	285,380
M2	0.00	19.18	137.68	12,161.39	291,333
M3	0.00	12.54	135.36	9,115.99	286,427
OLS	0.02	64.00	143.83	8,236.59	304,336
Log-Log	0.02	37.46	143.83	9,362.71	304,336

Table 32. Summary Statistics of Validation Data, M1, M2, M3, and Rescaled OLS and Log-Log Composite Estimates: SCTG 10–14 Rail Origins

Variable	Min	Median	Mean	Max	Sum
Validation	0.00	4.91	139.71	17,715.70	228,559
M1	0.00	8.58	110.37	15,651.01	180,557
M2	0.00	0.00	123.01	19,073.11	201,236
M3	0.00	2.46	116.10	14,274.23	189,939
OLS	0.01	21.71	143.55	16,469.35	234,845
Log-Log	0.01	16.67	143.55	14,236.84	234,845

Table 33. Summary Statistics of Validation Data, M1, M2, M3, and Rescaled OLS and Log-Log Composite Estimates: SCTG 15–19 Rail Origins

Variable	Min	Median	Mean	Max	Sum
Validation	0.00	14.65	351.51	165,628.94	655,923
M1	0.00	27.36	321.58	39,000.99	600,073
M2	0.00	0.00	317.78	114,253.54	592,981
M3	0.00	14.87	322.95	79,395.87	602,634
OLS	0.02	29.82	361.10	97,297.18	673,820
Log-Log	0.02	51.86	361.10	51,124.47	673,820

Table 34. Summary Statistics of Validation Data, M1, M2, M3, and Rescaled OLS and Log-Log Composite Estimates: SCTG 20–33 Rail Origins

Variable	Min	Median	Mean	Max	Sum
Validation	0.00	16.03	110.04	12,786.38	261,015
M1	0.01	15.20	105.29	13,049.56	249,739
M2	0.00	10.33	104.02	16,784.95	246,729
M3	0.00	11.52	104.82	11,040.00	248,628
OLS	0.08	30.53	110.99	12,826.81	263,258
Log-Log	0.21	27.63	110.99	8,596.47	263,258

Table 35. Summary Statistics of Validation Data, M1, M2, M3, and Rescaled OLS and Log-Log Composite Estimates: SCTG 34–99 Rail Origins

Variable	Min	Median	Mean	Max	Sum
Validation	0.00	2.68	28.79	3,786.73	66,437
M1	0.02	3.64	28.36	3,786.81	65,448
M2	0.00	3.00	27.12	2,091.22	62,596
M3	0.00	3.91	28.07	3,786.81	64,797
OLS	0.48	4.97	29.05	3,786.81	67,041
Log-Log	0.36	5.34	29.05	3,786.81	67,041

Table 36. Summary Statistics of Validation Data, M1, M2, M3, and Rescaled OLS and Log-Log Composite Estimates: SCTG 01–09 Rail Destinations

Variable	Min	Median	Mean	Max	Sum
Validation	0.00	20.28	134.45	24,677.32	298,609
M1	0.00	29.57	126.89	32,682.35	281,813
M2	0.00	20.77	129.37	22,912.05	287,320
M3	0.00	53.97	125.06	18,000.43	277,764
OLS	0.00	50.79	135.70	21,182.97	301,379
Log-Log	0.01	50.54	135.70	26,647.56	301,379

Table 37. Summary Statistics of Validation Data, M1, M2, M3, and Rescaled OLS and Log-Log Composite Estimates: SCTG 10–14 Rail Destinations

Variable	Min	Median	Mean	Max	Sum
Validation	0.00	13.93	97.22	7,206.25	228,559
M1	0.00	13.18	92.71	16,184.25	217,951
M2	0.00	6.63	90.78	14,954.88	213,429
M3	0.00	7.32	92.33	15,322.55	217,074
OLS	0.01	22.56	103.52	14,284.13	243,382
Log-Log	0.01	19.88	103.52	12,497.55	243,382

Table 38. Summary Statistics of Validation Data, M1, M2, M3, and Rescaled OLS and Log-Log Composite Estimates: SCTG 15–19 Rail Destinations

Variable	Min	Median	Mean	Max	Sum
Validation	0.00	40.33	301.44	74,760.51	655,923
M1	0.13	65.23	292.96	70,510.97	637,470
M2	0.00	8.20	281.89	80,546.31	613,400
M3	0.00	38.48	289.45	81,621.68	629,839
OLS	0.27	82.73	309.41	78,979.10	673,284
Log-Log	0.27	92.04	309.41	73,854.44	673,284

Table 39. Summary Statistics of Validation Data, M1, M2, M3, and Rescaled OLS and Log-Log Composite Estimates: SCTG 20–33 Rail Destinations

Variable	Min	Median	Mean	Max	Sum
Validation	0.00	21.70	105.38	7,410.59	261,015
M1	0.21	30.12	104.28	11,208.03	258,292
M2	0.04	22.32	103.09	10,933.25	255,342
M3	0.00	41.83	102.99	8,860.32	255,098
OLS	3.58	47.85	106.27	8,608.88	263,230
Log-Log	1.40	38.70	106.27	9,527.26	263,230

Table 40. Summary Statistics of Validation Data, M1, M2, M3, and Rescaled OLS and Log-Log Composite Estimates: SCTG 34–99 Rail Destinations

Variable	Min	Median	Mean	Max	Sum
Validation	0.00	3.81	26.10	1,569.75	66,437
M1	0.02	5.05	25.92	2,346.39	65,976
M2	0.01	4.29	24.86	2,346.22	63,279
M3	0.00	5.12	25.71	1,804.16	65,425
OLS	0.36	8.66	26.35	1,733.03	67,051
Log-Log	0.29	7.05	26.35	1,506.43	67,051

Table 41. Summary Statistics of Validation Data, M1, M2, M3, and Rescaled OLS and Log-Log Composite Estimates: SCTG 01–09 Water Origins

Variable	Min	Median	Mean	Max	Sum
Validation	0.00	44.08	303.82	5,235.48	130,641
M1	0.00	33.09	306.78	12,009.71	131,916
M2	0.00	21.41	306.50	9,847.84	131,794
M3	0.00	11.42	272.94	6,526.95	117,364
OLS	0.00	94.60	306.89	7,122.94	131,965
Log-Log	0.00	40.85	306.89	11,194.65	131,965

Table 42. Summary Statistics of Validation Data, M1, M2, M3, and Rescaled OLS and Log-Log Composite Estimates: SCTG 10–14 Water Origins

Variable	Min	Median	Mean	Max	Sum
Validation	0.00	14.17	228.53	9,427.58	97,582
M1	0.00	11.03	228.55	7,401.10	97,589
M2	0.00	0.00	213.38	28,017.75	91,115
M3	0.00	1.05	201.60	11,124.72	86,083
OLS	0.00	33.47	228.55	9,775.66	97,589
Log-Log	0.00	13.86	228.55	7,370.51	97,589

Table 43. Summary Statistics of Validation Data, M1, M2, M3, and Rescaled OLS and Log-Log Composite Estimates: SCTG 15–19 Water Origins

Variable	Min	Median	Mean	Max	Sum
Validation	0.00	67.77	967.95	51,955.43	430,739
M1	0.00	50.28	972.47	61,744.93	432,749
M2	0.00	0.00	953.99	49,222.82	424,525
M3	0.00	38.11	854.88	36,045.10	380,421
OLS	0.00	123.69	972.49	49,942.50	432,756
Log-Log	0.00	78.27	972.49	55,056.56	432,756

Table 44. Summary Statistics of Validation Data, M1, M2, M3, and Rescaled OLS and Log-Log Composite Estimates: SCTG 20–33 Water Origins

Variable	Min	Median	Mean	Max	Sum
Validation	0.00	7.13	208.97	18,236.96	96,960
M1	0.00	6.24	208.97	22,130.76	96,962
M2	0.00	1.12	208.89	16,386.81	96,923
M3	0.00	2.30	202.54	14,021.44	93,976
OLS	0.00	5.74	209.14	16,302.35	97,043
Log-Log	0.00	7.17	209.14	18,229.94	97,043

Table 45. Summary Statistics of Validation Data, M1, M2, M3, and Rescaled OLS and Log-Log Composite Estimates: SCTG 34–99 Water Origins

Variable	Min	Median	Mean	Max	Sum
Validation	0.00	1.35	34.54	2,181.80	16,512
M1	0.00	0.98	34.58	2,710.70	16,530
M2	0.00	0.58	34.59	2,564.45	16,534
M3	0.00	0.35	29.61	2,621.32	14,153
OLS	0.00	1.97	34.60	2,523.79	16,538
Log-Log	0.00	1.20	34.60	2,530.27	16,538

Table 46. Summary Statistics of Validation Data, M1, M2, M3, and Rescaled OLS and Log-Log Composite Estimates: SCTG 01–09 Water Destinations

Variable	Min	Median	Mean	Max	Sum
Validation	0.00	2.63	322.57	20,642.38	130,640
M1	0.00	4.03	322.78	32,804.61	130,727
M2	0.00	2.30	322.78	47,309.72	130,727
M3	0.00	3.94	279.56	15,149.95	113,221
OLS	0.00	9.09	322.78	22,474.70	130,727
Log-Log	0.00	5.52	322.78	27,416.98	130,727

Table 47. Summary Statistics of Validation Data, M1, M2, M3, and Rescaled OLS and Log-Log Composite Estimates: SCTG 10–14 Water Destinations

Variable	Min	Median	Mean	Max	Sum
Validation	0.00	28.53	206.74	5,779.32	97,582
M1	0.00	37.30	206.74	7,258.56	97,579
M2	0.00	15.67	193.04	7,858.16	91,114
M3	0.00	22.95	170.67	4,626.03	80,556
OLS	0.00	50.68	206.76	6,473.82	97,589
Log-Log	0.00	41.64	206.76	6,329.45	97,589

Table 48. Summary Statistics of Validation Data, M1, M2, M3, and Rescaled OLS and Log-Log Composite Estimates: SCTG 15–19 Water Destinations

Variable	Min	Median	Mean	Max	Sum
Validation	0.00	57.04	891.83	60,781.96	430,753
M1	0.00	70.80	895.38	69,851.61	432,469
M2	0.00	16.98	878.30	57,938.86	424,218
M3	0.00	41.44	831.86	48,434.87	401,791
OLS	0.00	119.83	896.99	60,207.75	433,244
Log-Log	0.00	92.75	896.99	59,334.80	433,244

Table 49. Summary Statistics of Validation Data, M1, M2, M3, and Rescaled OLS and Log-Log Composite Estimates: SCTG 20–33 Water Destinations

Variable	Min	Median	Mean	Max	Sum
Validation	0.00	18.91	201.58	20,711.52	96,960
M1	0.00	21.94	201.53	23,248.27	96,934
M2	0.00	14.26	201.49	18,466.92	96,918
M3	0.00	19.71	186.85	15,203.44	89,874
OLS	0.00	26.60	201.58	18,640.60	96,960
Log-Log	0.00	26.91	201.58	18,750.36	96,960

Table 50. Summary Statistics of Validation Data, M1, M2, M3, and Rescaled OLS and Log-Log Composite Estimates: SCTG 34–99 Water Destinations

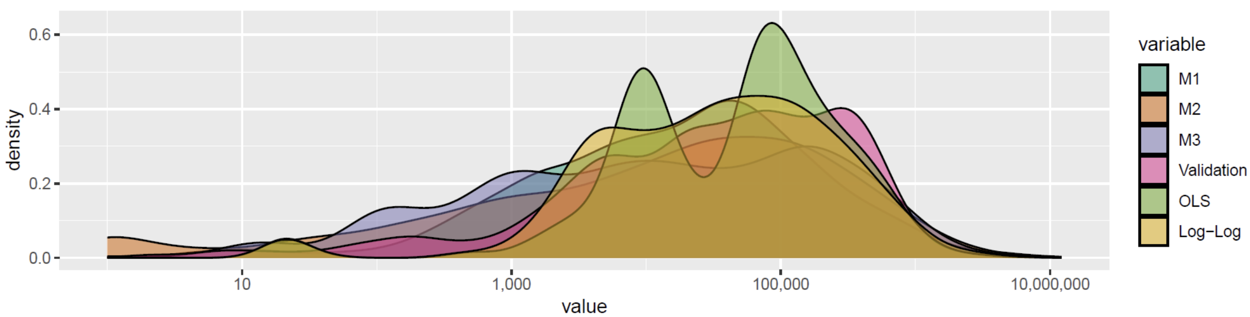
Variable	Min	Median	Mean	Max	Sum
Validation	0.00	1.24	35.36	2,297.37	16,512
M1	0.00	1.52	35.18	1,646.16	16,428
M2	0.00	1.19	35.17	1,648.25	16,426
M3	0.00	1.31	25.69	1,125.68	11,997
OLS	0.00	2.09	35.36	1,650.67	16,512
Log-Log	0.00	1.68	35.36	1,844.79	16,512

Appendix G. Overlaid Density Curves

Figure 52–Figure 55 show overlaid density curves by SCTG group and mode. Truck origin and destination density curves are shown and discussed in the main text.

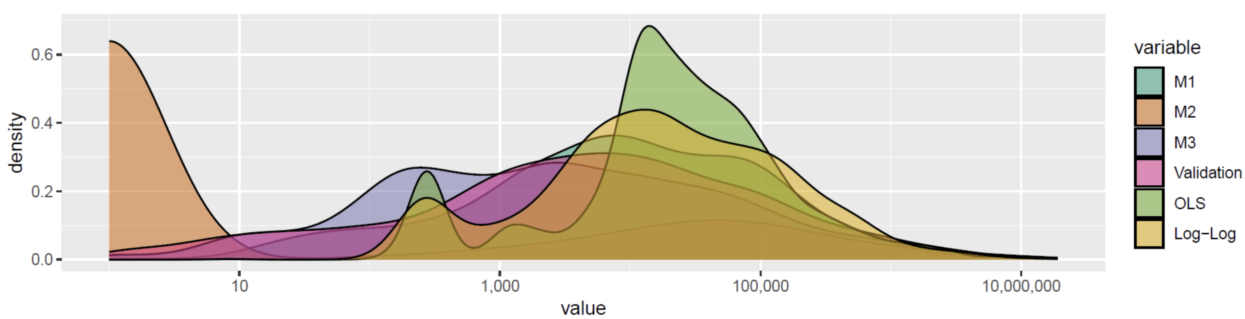
Figure 52. The Overlaid Density Curves for M1, M2, M3, Validation Data, Rescaled Predicted Values of OLS, and Log-Log: Rail Origins

A. SCTG 01–09



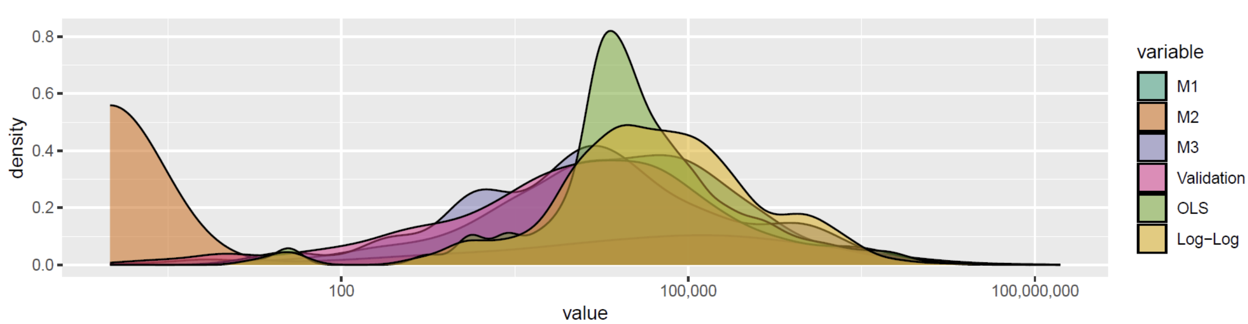
Source: BTS.

B. SCTG 10–14



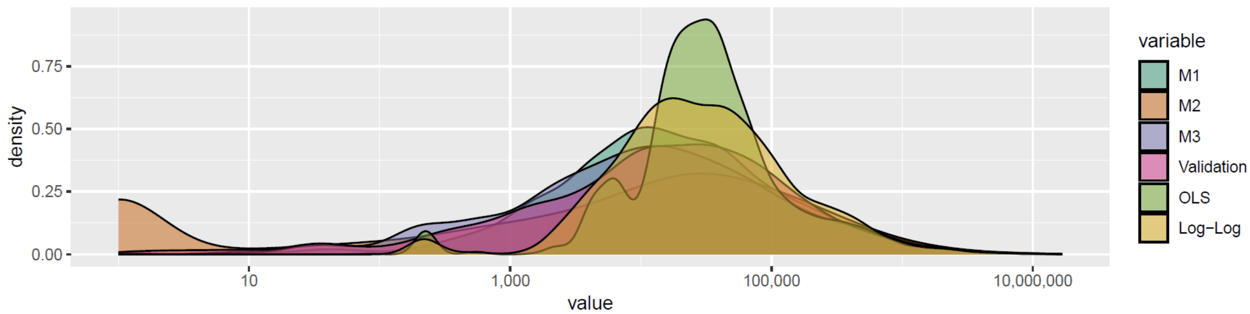
Source: BTS.

C. SCTG 15–19



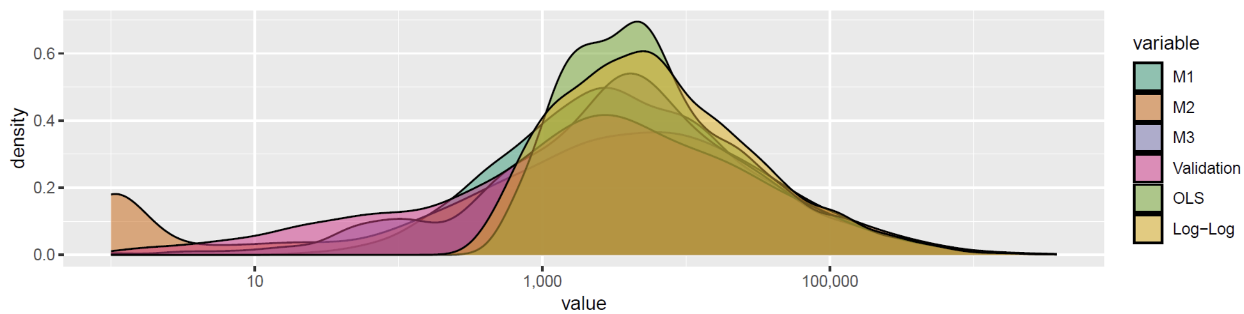
Souce: BTS.

D. SCTG 20–33



Source: BTS.

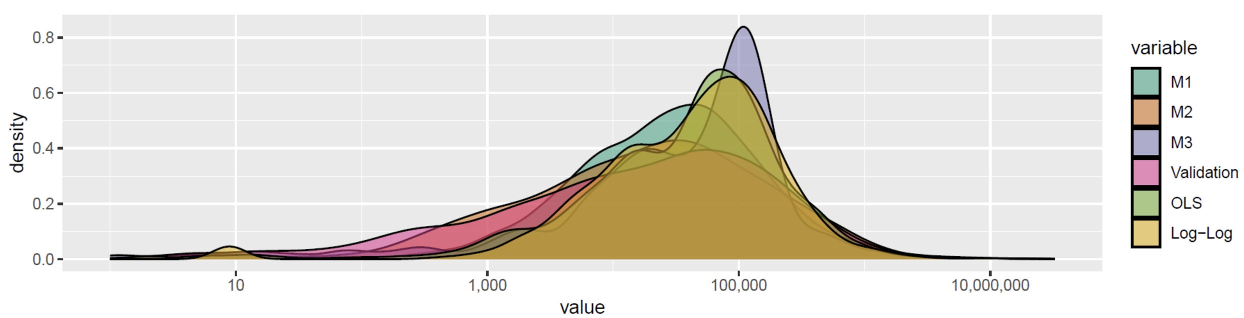
E. SCTG 34–99



Source: BTS.

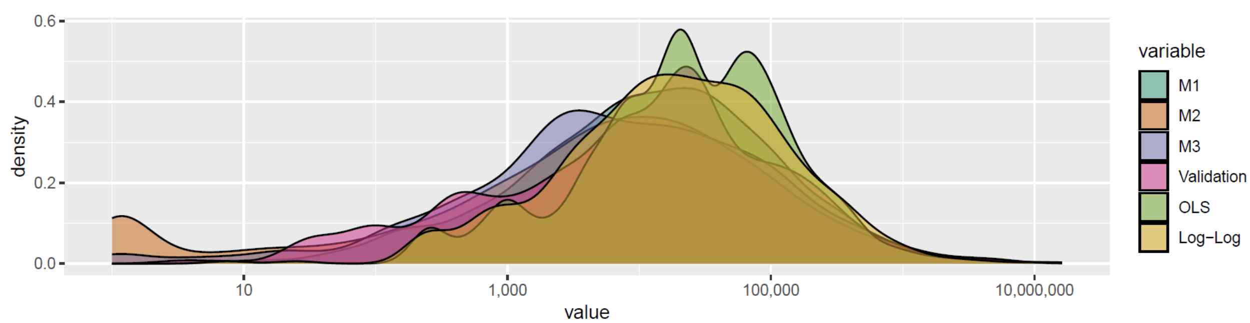
Figure 53. The Overlaid Density Curves for M1, M2, M3, Validation Data, Rescaled Predicted Values of OLS, and Log-Log: Rail Destinations

A. SCTG 01–09



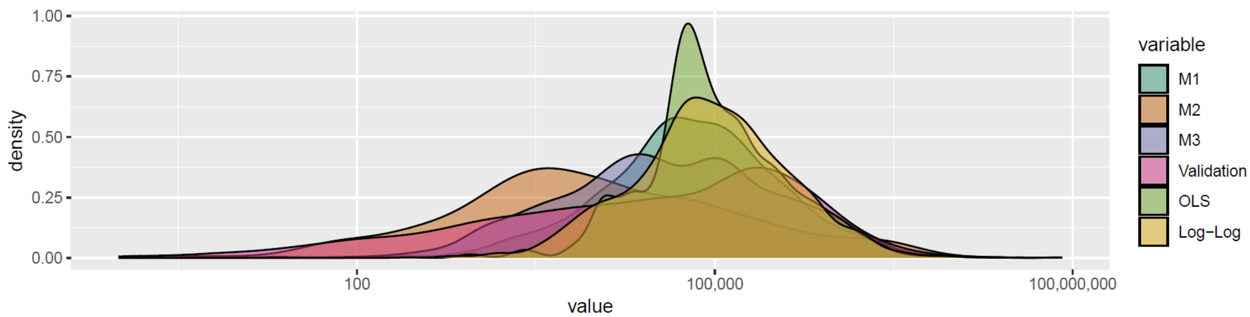
Source: BTS.

B. SCTG 10–14



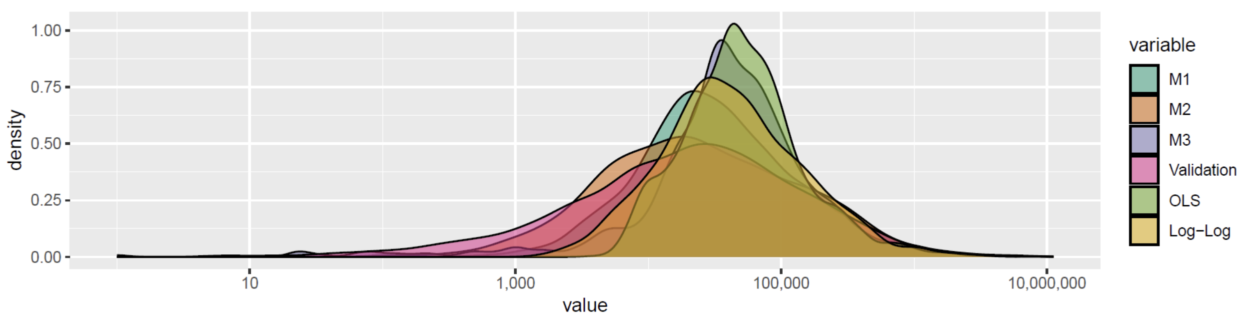
Source: BTS.

C. SCTG 15–19



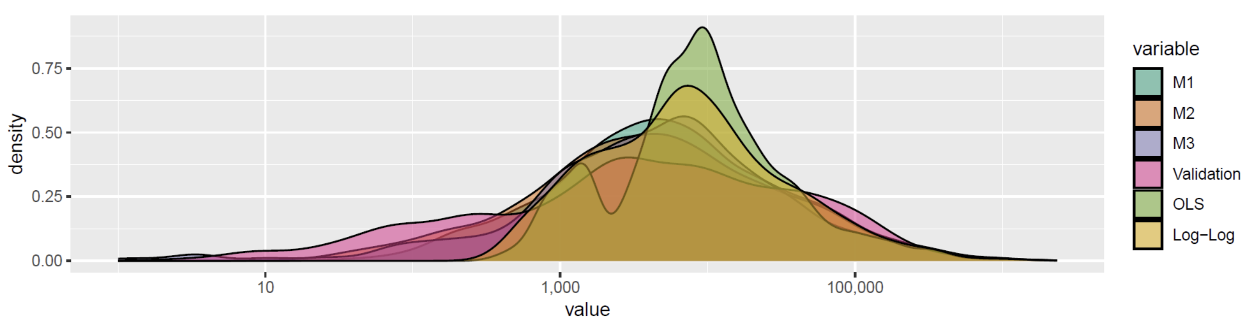
Souce: BTS.

D. SCTG 20–33



Source: BTS.

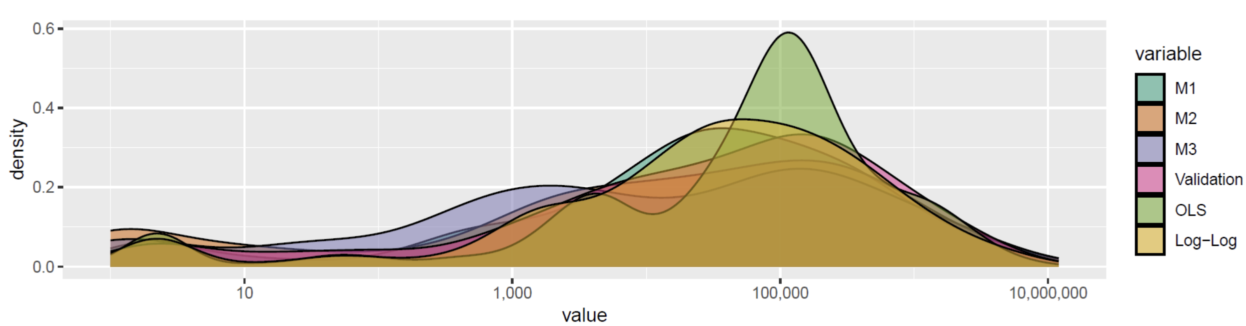
E. SCTG 34–99



Source: BTS.

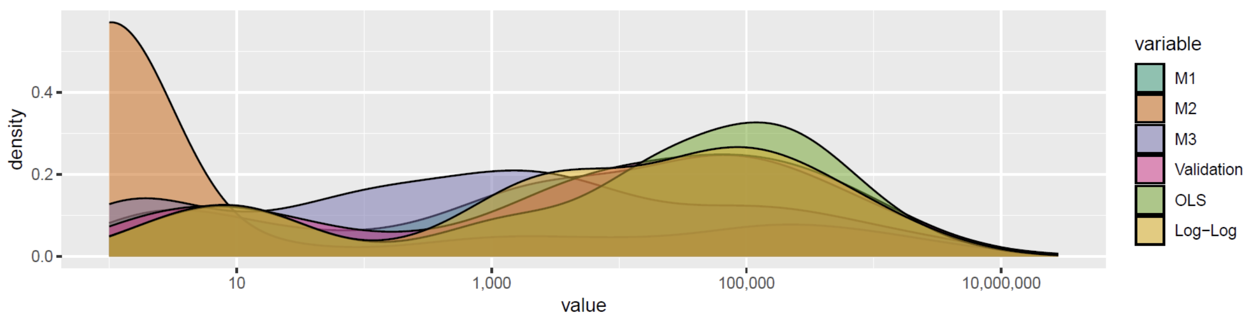
Figure 54. The Overlaid Density Curves for M1, M2, M3, Validation Data, Rescaled Predicted Values of OLS, and Log-Log: Water Origins

A. SCTG 01–09



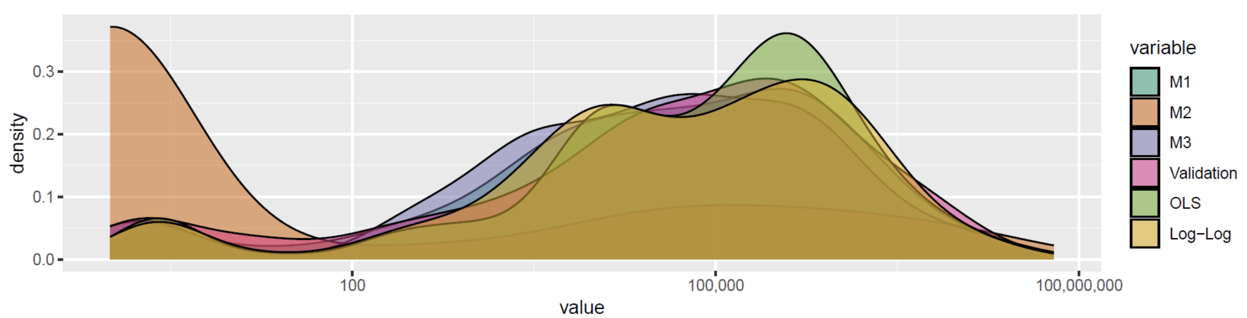
Source: BTS.

B. SCTG 10–14



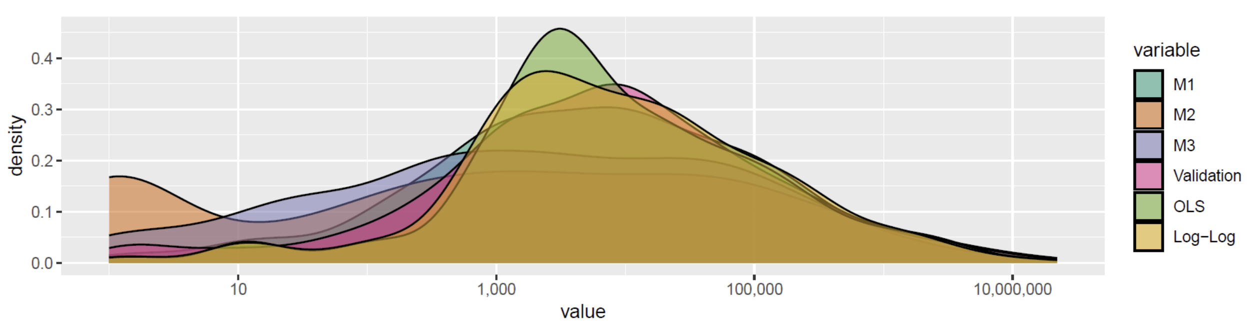
Source: BTS.

C. SCTG 15–19



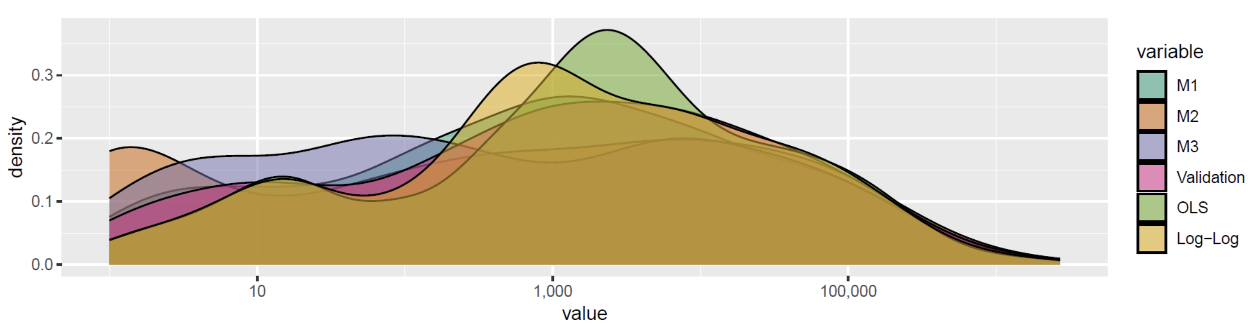
Souce: BTS.

D. SCTG 20–33



Source: BTS.

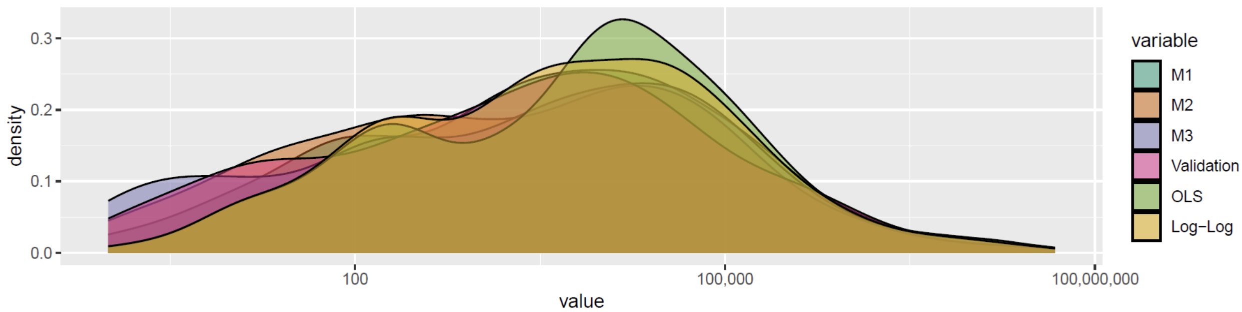
E. SCTG 34–99



Source: BTS.

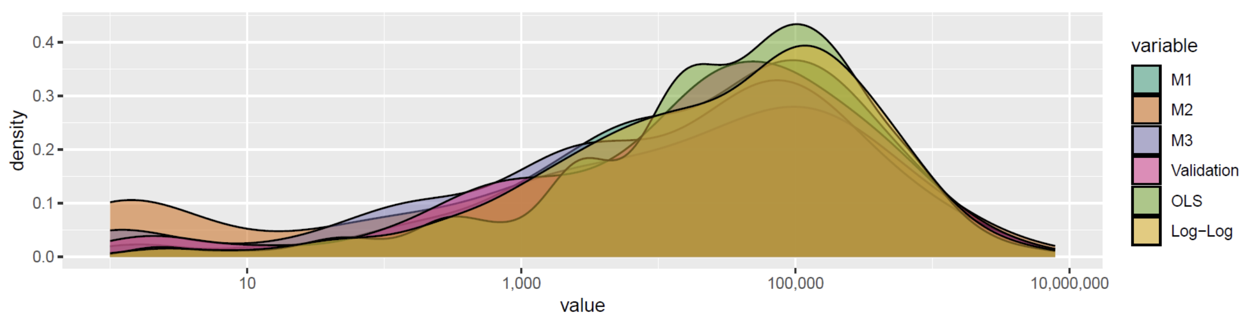
Figure 55. The Overlaid Density Curves for M1, M2, M3, Validation Data, Rescaled Predicted Values of OLS, and Log-Log: Water Destinations

A. SCTG 01–09



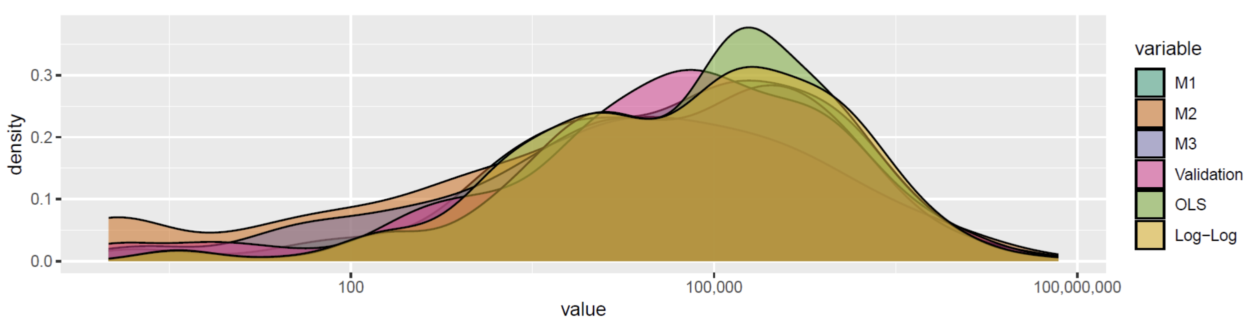
Source: BTS.

B. SCTG 10–14



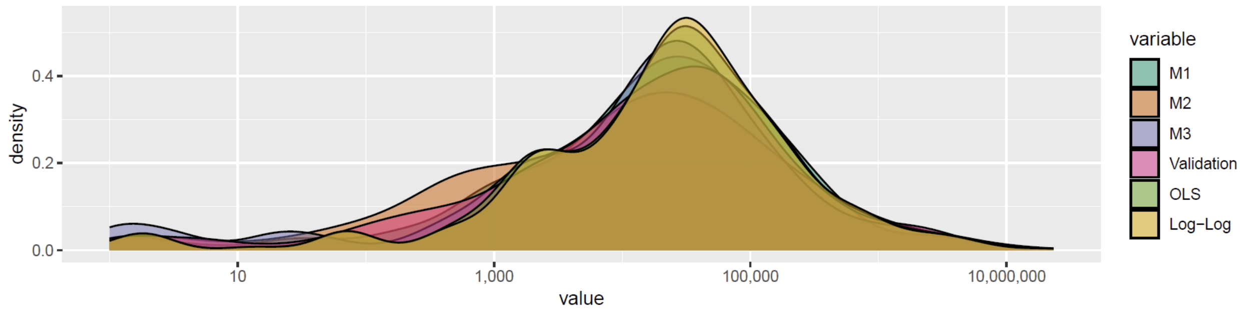
Source: BTS.

C. SCTG 15–19



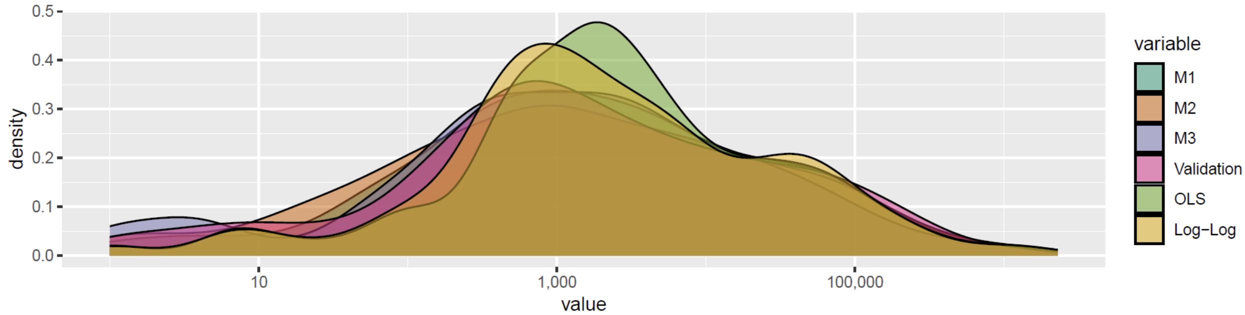
Souce: BTS.

D. SCTG 20–33



Source: BTS.

E. SCTG 34–99



Source: BTS.