McCormick
Northwestern Engineering
Northwestern University Transportation Center

Grain and Soybean Industry Dynamics and Rail Service

Econometric Analysis of Rail Transport Rates

Final Report

March 2016

Hani Mahmassani, Michael Hyland, Lama Bou Mjahed, and Breton Johnson
Northwestern University Transportation Center (NUTC)

Parr Rosson
Texas A&M University

McCormick
Northwestern Engineering
Northwestern University Transportation Center

Grain and Soybean Industry Dynamics and Rail Service

Econometric Analysis of Rail Transport Rates

BACKGROUND

The Staggers Act of 1980 gave railroads considerable freedom in terms of price discrimination. Despite expectations of the contrary, rail transport rates decreased for most shippers after the Staggers Act. The Surface Transportation Board (STB) is responsible for handling rate disputes between shippers and railroads in regions where competition is limited. The STB collects a 1% stratified sample of all waybills originated by major carriers in the United States. A version of the data is available to the public. The analysis presented in this report uses the STB's publicly available carload waybill sample (CWS) to analyze rail transport rates.. The analysis examines trends in rail transport rates for (a) all shipments, (b) specific commodity-types including grain, and (c) specific regions of the country including two regions that produce a large amount of grain, the Upper Midwest and the I-states. The analysis additionally attempts to examine (a) the shipment characteristics that impact rail revenue per ton-mile (RPTM) and revenue per carload-mile (RPCM), (b) how the impact of the characteristics differ as a function of commodity-type and region of origin, and (c) how this impact fluctuated between 2001 and 2013.

METHODOLOGY

The CWS is a very large data set; hence, it was necessary to conduct an in-depth exploratory analysis of the data. First and foremost, the exploratory analysis revealed errors in the dataset; therefore, the data was cleaned to remove nonsensical shipments and also filtered in order to remove extreme outliers. Second, data exploration uncovered correlations between variables in the dataset and also determined important temporal trends of specific variables. The data exploration results informed econometric models, which were developed to further analyze rail transport rates. Multivariate regression models were developed to analyze rail transport rates and answer the research questions posed in the Background section.

SUMMARY OF THE RESULTS

Data Exploration Results

- The correlation matrix for the CWS's fields shows that freight revenue per ton-mile (RPTM) is negatively correlated with distance, route density, weight, and carload number. These correlations are tested more systematically in the econometric models.
- Average RPTM increased, in real terms, between 2001 and 2013. Specifically, RPTM for grain and coal increased, as did the RPTM for export shipments.

Econometric Modeling Results

All Waybills

- Average RPCM for rail shipments increased between 2006 and 2012.
- The following shipment types were associated with lower RPCM: long distance shipments (see Figure 1), large shipments, and shipments bound for export.
- After accounting for shipment characteristics, the RPCM for bulk grain was lower than every other commodity-type examined, including: crude oil and natural gas, coal, food products, non-grain agricultural products, and chemicals.

•

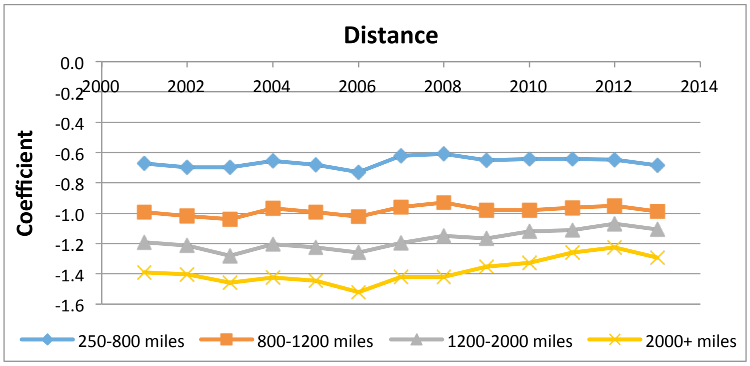


Figure 1: Coefficient values for the regression model's distance parameter. Points are relative to shipments traveling 20-250 miles.

Grain Waybills Only

- Average RPCM increased significantly between 2004 and 2011.
- Average RPCM for export grain shipments was lower than the RPCM for non-export shipments in the early 2000s; however in 2012 and 2013, export RPCM was higher than non-export.
- Interestingly, between 2010 and 2013 there was no noticeable difference between RPCM for 90+ carload shipments, 50-90 carload shipments, and 6-49 carload shipments.

Upper Midwest Waybills Only

- Average RPCM increased steadily between 2008 and 2013.
- RPCM for grain shipments was significantly lower than the RPCM of other commodities in the Upper Midwest between 2001 and 2013.
- Interestingly, between 2001 and 2003 the RPCM for shipments on very high density routes (100,000+ shipments annually) was significantly higher than the RPCMs for lower density routes. Conversely, between 2004 and 2013, the RPCM for very high density routes was lower than the RPCM of all lower density routes.

CONCLUSIONS

The econometric model results indicate that after controlling for shipment characteristics and other exogenous factors the average RPCM of rail shipments increased significantly, in real terms, between 2001 and 2013 with most of the increase occurring between 2004 and 2012. Further analysis shows that although RPCM increased for bulk grain shipments, the increase was consistent with a general increase in RPCM for all commodities during the period from 2001 to 2013. The regression model results show that the RPCM for bulk grain was consistently lower than the RPCM for many other commodity-types after holding other variables such as carload number, shipment distance, route density, and railcar ownership constant in the multivariate regression models. The methodology presented in the report provides a systematic means of determining trends in not only overall rail transport rates but also the shipment characteristics and exogenous factors that impact rates in different segments of the data.

McCormick

Northwestern Engineering

Northwestern University Transportation Center

Executive Summary

The production and transportation of grain and soybeans, especially in the North American western regions, are undergoing significant restructuring driven by the desire and need to achieve economies of scale and reach export markets. Within the framework of the Northwestern University Transportation Center (NUTC) study that is focused on gaining insight into the restructuring process via (1) modeling grain[1] and soybean transportation service and (2) analyzing the grain industry and its transportation needs, this third and final track of the study focuses on rail freight rates. The objective of the third track is to understand the *direction* and *magnitude* of rail transportation rates over time and the *predominant factors* impacting the rates in a statistical sense.

Understanding and capturing the dynamics behind freight rates, as this work seeks to do, is both important and complex. It is important for many reasons such as the implications for rail competitiveness and market share, the economic viability and global competitiveness of the United States grain sector, and the relationship between the grain industry and the transportation sector. As previous tracks of the analysis have shown, understanding and capturing the dynamics behind rates is complex because providing transportation service depends on a multitude of factors occurring in dynamic contexts. These factors are often hard to isolate and capture, and rates in general, in any given service and sector, are the result of a range of variables, market circumstances, and management discretion.

Over the past four decades, freight railroad rates have evolved considerably both in terms of their magnitudes and the freedoms/restrictions that federal regulators have placed on railroads in terms of rate discrimination. Often, in discussions of rail transport rates two regimes are identified for context: before and after the Staggers Act of 1980[2]. The Staggers Act gave

---

[1] In this report, the term 'grain' refers to bulk grains and bulk oilseeds (i.e. corn, wheat, and soybeans).

[2] Winston, Clifford. "The Success of the Staggers Rail Act of 1980." *AEI-Brookings Institute Center for*

[2] Winston, Clifford. "The Success of the Staggers Rail Act of 1980." *AEI-Brookings Institute Center for Regulatory Studies* (2005): <http://www.brookings.edu/~/media/research/files/papers/2005/10/railact-winston/10_railact_winston.pdf>.

railroads considerable freedom in terms of which shipments to provide service to and the rates charged to those shippers. Conversely, prior to Staggers Act, railroad companies were required by law to service certain shipments regardless of their relative profitability and without total control over the rates charged to shippers.

Instead of comparing those two regimes, this study focuses on the years between 2001 and 2013. The results uncovered in this analysis should be placed in the more recent context of the restructuring process of both production and grain logistics. The data used in the analysis were obtained through the publicly available carload waybill sample (CWS) that the Surface Transportation Board (STB) collects from major railroads.

In this report we perform an in-depth analysis of rail transportation rates using the CWS. The analysis examines all shipment types but emphasizes bulk grain shipments and shipments originating in the Upper Midwest (North Dakota, South Dakota, and western Minnesota). The analysis attempts to uncover trends in rail transport rates for (a) all shipments, (b) specific commodity-types including grain, and (c) specific regions of the country including two regions that produce a large amount of grain, the Upper Midwest and the I-states (Iowa, Indiana, Illinois, and Missouri). The analysis additionally attempts to examine (a) the shipment characteristics that impact rail revenue per ton-mile (RPTM) and revenue per carload-mile (RPCM), (b) how the impact of the characteristics differ as a function of commodity-type and region of origin, and (c) how the impact of the characteristics fluctuated between 2001 and 2013.

The key explanatory variables used in this study were chosen based on an extensive literature review of rail transportation rates. The literature review revealed that predominant factors affecting rail freight rates include: **distance traveled, shipment weight, carload number**, **commodity-type, export vs. domestic, and route density.** Those variables, along with freight rates, were examined in a two-step process. First, a *preliminary analysis* was conducted to understand correlations and trends in the CWS data. This paved the way to a more robust analysis of freight rates and their determinants through *multivariate regression models*.

The exploratory analysis included first a *correlation study* of major relationships in the data. The emanating results showed that **freight revenue per ton-mile (RPTM)** is *negatively* correlated with (1) **distance,** (2) **weight,** (3) **carload number,** (4) **route density,** (5) **grain shipments,** (6) **car capacity,** (7) **exports,** and (8) **number of interchanges.** This implies that *longer distance shipments, larger shipments (heavier shipments and more carload shipments), higher volume routes*, and shipments going through *more interchanges* were associated with *lower RPTM*. With the exception of the number of interchanges (which is not addressed in the literature), these relationships confirm literature findings. Correlation tests between the variables also showed that **weight** and **carload number** are highly correlated, suggesting that one of the two is sufficient to model RPTM and RPCM. Additionally, **higher route density** is associated with **longer distance shipments** and **fewer interchanges. Coal shipments were** strongly correlated with **carload number** and **shipment weight,** perhaps reflecting the fact that most coal is shipped using shuttle trains.

The exploratory analysis also examined trends in relevant variables. In general, shipment distances varied minimally over time; the average shipment distance in a given year ranged from 1,150 miles to 1,214 miles over 13 years. Moreover, a large percentage of the shipments in the data (around 30%) traveled between 250 and 800 miles. The average *shipment weight* in the data ranged between 87 tons and 100 tons. Shipment weights, like distances, were stable over time. Furthermore, most shipments in the CWS data were *one-carload* shipments (98%). Moreover, the majority of shipments in the data *did not go through any interchanges* (88% of shipments do not go through any interchanges). As for route density, the majority of the shipments in the data traveled on routes with *medium or heavy* density (i.e. between 1,000 and 100,000 M railcars) with 65% of shipments in this range. Finally, shippers' behavior changed in terms of choosing to use fewer railroad-owned railcars over time. The share of railroad-owned railcars in 2013 amounted to 17%, compared to 29% in 2001.

As for freight RPTM trends, preliminary results suggest that average RPTM for coal shipments was lower than the RPTM for grain shipments, and the RPTMs for coal and grain shipments were both lower than the RPTM for all shipments (including grain and coal). Nominal RPTM for coal increased from 1.7 cents per ton-mile in 2001 to 3.1 cents per ton-mile in 2013. Nominal RPTM for grain increased from 2.0 cents per ton-mile in 2001 to 3.6 cents per ton-mile in 2013. However, RPTM for **coal** had the highest compound annual growth rate (CAGR[3]) between 2001 and 2013 at 5%, followed by **grain** with a CAGR of 4.7%. The CAGRs for grain and coal between 2001 and 2013 were slightly higher than the CAGR for all shipments in the data (4.6%). Furthermore, the analysis showed that RPTM was lower for export shipments than all shipments. Within export shipments, RPTM was highest for coal shipments, followed by all-commodity exports, and finally grain exports. RPTM for coal exports increased at the highest pace (8% annually); the CAGR for both grain exports and all exports was 4.7%.

The regression models presented in the second phase of the analysis build upon the results presented in the exploratory phase. The regression models aim to determine trends in average RPCM for rail transport between 2001 and 2013 for (a) all shipments, (b) shipments of a specific commodity-type including bulk grain and coal, and (c) shipments originating from specific regions of the country including the Upper Midwest and the I-states. Regression modeling techniques allow one to determine the trends in RPCM while controlling for changes in shipment characteristics over time. The second purpose of the regression models is to examine the impact of shipment characteristics and other exogenous factors on RPCM for rail transport. Aside from commodity-type and origination region the analysis examines distance, shipment weight, carload number, route density, railcar ownership, number of railroad interchanges, and export vs. domestic. We segmented the data by commodity-type, origination region, and year in the analysis to determine how shipment characteristics impacted RPCM differently for various commodities and origination regions and how the impacts fluctuated over time.

---

[3] Compound annual growth rate (CAGR) is calculated as follows: $CAGR = \left(\frac{final\ value}{initial\ value}\right)^{\frac{1}{num\ years}} - 1$

Important and notable model results are highlighted below.

All Waybills

- Average RPCM[4] (after controlling for shipment characteristics and other exogenous factors):
    - was relatively constant between 2001 and 2006
    - noticeably increased between 2006 and 2012.
    - leveled off between 2012 and 2013
- The following shipment types were associated with lower RPCM:
    - Long distance shipments (exerts very large effect)
    - Shipments traveling between a high density origination-termination region pair
    - Large shipments (i.e. high carload number)
    - Export shipments
    - Railroad-owned railcars
- After controlling for shipment characteristics, the model results suggest that grain shipments were associated with lower RPCM than every other commodity-type examined, including: crude oil and natural gas, coal, coal and petroleum products, food products, non-grain agricultural products, chemicals, and pulp and paper.

Grain Waybills Only

- Average RPCM for grain shipments increased significantly between 2004 and 2011.
- The RPCM for export grain shipments was lower than the RPCM for non-export grain shipments in the early 2000s. However, in 2012 and 2013, the RPCM was higher for export grain shipments than non-export grain shipments.
- Route density did not have a consistent and significant effect on the RPCM for grain shipments
- RPCM was similar for 6-49 carload shipments, 50-90 carload shipments, and 90+ carload shipments between 2010 and 2013 for grain shipments. This result contradicts the model results for all waybills and economies of scale typically associated with rail transportation.

Upper Midwest Waybills Only

- Average RPCM for shipments originating in the Upper Midwest increased steadily between 2008 and 2013.
- The RPCM for grain shipments in the Upper Midwest was consistently lower than the RPCM for every other commodity-type examined between 2001 and 2013.
- Between 2001 and 2003 the RPCM for shipments on very high density routes (100,000+ shipments annually) was significantly higher than the RPCMs for all lower density route categories. Conversely, between 2004 and 2013, the RPCM for very high density routes was lower than the RPCMs of all lower density route categories.
- The RPCM for export shipments was consistently lower than the RPCM for non-export shipments in the Upper Midwest.

In summary, the econometric model results indicate that average RPCM increased significantly in real terms between 2001 and 2013 with most of the increase occurring between 2006 and 2012. Further analysis showed that although average RPCM increased in the Upper

---

[4] All revenues and RPCM were adjusted for inflation

Midwest and for grain shipments; the increase was consistent with a general increase in average RPCM for all shipments during the period from 2001 to 2013. The regression model results show that the RPCM for grain shippers was lower than the RPCM for many other commodity-types. The econometric regression models exhibited the aforementioned results while simultaneously taking into account and measuring the impact of shipment characteristics. Many of the shipment characteristics associated with lower RPCM in the econometric regression modeling analysis (long distance shipments, shipments on high-density routes, large shipments, and railroad-owned railcars) conform to previous findings in the literature[5]. In addition, the econometric analysis determined that the RPCM for export shipments was consistently lower than the RPCM for domestic shipments between 2001 and 2013 for all commodity-types. However, in 2012 and 2013 the RPCM for export grain shipments was higher than the RPCM for non-export grain shipments. The methodology presented in the report provides a systematic means of determining trends in not only overall rail transport rates but also the shipment characteristics and exogenous factors that impact rates in different segments of the data.

The analysis presented in this track of the NUTC study complemented the work completed in the previous two tracks; it combines both railroad operations and grain industry characteristics in an analysis of actual, observed rail freight rates. Through a two-step process of first understanding correlations and trends in the data and second studying the determinants of freight rates, this study uncovered major trends in rail freight rates and determinants of rail freight rates between 2001 and 2013. The results emanating from this analysis have important implications for both grain industry stakeholders as well as transportation providers.

Finally, the NUTC research group identified two potential areas for future research. The first, and most obvious path, requires access to the STB's confidential waybill sample. The confidential waybill data include a number of fields that would improve the regression model developed in the second part of the report, including: *fuel surcharge, estimated railroad variable cost, and better and finer geographical information.* The second opportunity to improve the regression model involves combining outside data sources with the CWS. For example, other researchers have examined how potential freight transport competition influences rail transport rates. They combined the CWS with information related to the nearest navigable inland waterway from the shipment's origin and termination points, and the number of competing railroads that could potentially serve the demand. Other data sources that could be integrated with the CWS include fuel prices, and aggregate rail indices.

---

[5] References to the existing literature are provided in the main report.

Contents

1 Introduction

Often in discussions of freight rail transport rates two regimes are identified for context: before and after the Staggers Act of 1980[6]. The Staggers Act of 1980 gave railroads considerable freedom in terms of which shipments to provide service to and the rates charged to shippers. Conversely, prior to the Staggers Act, railroad companies were required by law to service certain shipments regardless of their relative profitability and without total control over the rates charged to shippers.

The deregulation that emerged from the Staggers Act resulted in reduced rail transport rates to most shippers between 1980 and the end of the century. However, concerns still exist regarding rates charged to shippers, especially in regions of the country where shippers do not have alternative transportation options. Concerns have increased as the number of Class I railroads has decreased from thirty to seven over the past three decades due to rapid consolidation through mergers. Despite significant deregulation, rail rates are still subject to regulation and monitoring through the Surface Transportation Board (STB). The STB collects a 1% stratified sample of waybills from the Class I railroads each year in order to handle rate disputes and understand general trends in rail transportation rates. Fortunately, the STB releases a significant portion of the data (the released data includes all the waybills; however, a number of data fields are removed) to the public. The analysis presented in this report makes use of the publicly available carload waybill sample (CWS) provided to the public by the STB.

Understanding and capturing the dynamics behind freight rates is both important and complex. It is important for many reasons such as the implications for rail competitiveness and market share, the economic viability and global competitiveness of the United States grain sector, and the relationship between the grain industry and the transportation sector. As previous tracks of the analysis have shown, understanding and capturing the dynamics behind rates is complex because providing transportation service depends on a multitude of factors occuring in dynamic contexts. These factors are often hard to isolate and capture, and rates in general, in any given service and sector, are the result of a range of variables, market circumstances, and management discretion.

The work presented in this report addresses these difficulties by studying actual shipment records (waybills) over 13 years (from 2001 to 2013) to examine both the *direction and magnitude* of the changes in rates over time and the *predominant factors* determining them in a statistical sense. In particular, the analysis examines evidence for increases and decreases in

---

[6] Winston, Clifford. "The Success of the Staggers Rail Act of 1980." *AEI-Brookings Institute Center for Regulatory Studies* (2005): <http://www.brookings.edu/~/media/research/files/papers/2005/10/railact-winston/10_railact_winston.pdf>.

average revenue per carload-mile (RPCM) and average revenue per ton-mile (RPTM) for rail transport. By doing so, the report will seek to answer the following questions:

- *Have average RPCM and average RPTM for rail transport increased or decreased over the past 13 years?*
- *How has this increase or decrease been different for different commodities (i.e. grain versus coal, versus all others)?*
- *What shipment characteristics and other exogenous factors are associated with higher or lower average RPCM and RPTM?*

The report answers the previous set of questions through two main sections. Section 2 presents a **preliminary analysis of the data** that includes a description of key variables, correlations between the variables, and trends of individual variables over time. Section 3 details **the econometric models** used to uncover the predominant factors that can explain trends in RPCM and their relative importance.

2 Preliminary Analysis of Waybill Data

The first part of the report aims at understanding the data broadly with the objective of (1) providing **direction** for a deeper analysis of freight rail rates and (2) providing **context** for the analysis of the model results presented in the econometric modeling portion of the report. For that purpose, the data is described, filtered, and analyzed through (1) correlations between key variables and (2) individual trends.

2.1 Overview of Waybill Data

The data used to analyze rail rates in this study come exclusively from the CWS available on the STB's website[7]. The CWS is a 1% stratified sample of all rail movements originating or terminating in the U.S. by rail carriers terminating more than 4,500 revenue carloads annually.

The CWS dataset contains approximately 600,000 waybills per year from 1986 through 2013. Table 2-1 below displays the exact number of waybills corresponding to the years analyzed in this study (i.e. 2001-2013).

Each publicly available waybill contains 63 fields, the most interesting of which are discussed and analyzed throughout this section. The full list of fields is presented in Exhibit A: CWS Variables (as given by the STB) of the Appendix. The confidential waybill sample contains 130 additional fields that are removed from the publicly available sample. The most interesting variables excluded from the publicly available sample include: total variable cost (computed using the Uniform Railroad Costing System), the carrier, fuel surcharge, and variable cost and revenues broken down by carrier.

---

[7] STB Waybill's website: https://www.stb.dot.gov/stb/industry/econ_waybill.html.

Table 2-1: Total Waybills per Year in the STB's Carload Waybill Sample

Year	Waybills
2001	573,670
2002	589,826
2003	629,127
2004	632,482
2005	667,569
2006	688,170
2007	666,686
2008	622,318
2009	518,343
2010	580,717
2011	599,284
2012	622,884
2013	640,998

2.2 Data Filtering

The CWS, like any large dataset, contains errors and outliers that should be accounted for before the data are used in econometric models. Based on an extensive literature review as well as scientific judgment, the filters applied to the data are described below.

Filters Applied

The dataset was trimmed with respect to four factors: *origination region, weight, distance, and freight rates (i.e. RPTM)*.

Shipments Origination

The final analysis only includes shipments originating in the United States. This excludes waybills with Business Economic Area (BEA) origins falling in Canada or Mexico. The full list of origins removed from the data and their corresponding BEA codes, as defined by the STB, can be found in Exhibit B: Shipment Origins Removed from the Data of the Appendix. Between 3% and 4% of the data were removed after this filter was applied (6.6% removed for 2003).

Weight

It is common practice that shippers do not pay for empty shipments (i.e. **zero weight** shipments). Also, carload specifications dictate that no more than **130 tons/railcar** of laden weight can be shipped[8]. Therefore, shipments with reported weights of: less than or equal to zero, or greater than 130 tons were removed from the dataset. *Less than a half percent of the data were removed after this filter was applied.*

---

[8] ***Analysis of Freight Rail Rates for U.S. Shippers, Escalation Consultants Inc., March** 2014* suggests that Gross track weight limitations are 286,000 lbs. or 143 tons. The tare weight or empty weight of typical railcars range between 26 and 55 tons, adding 130 tons of laden to an empty car will exceed the 143-ton limit.

Distance

The literature suggests that shipments transported **less than 20 miles** are switching movements between railroads rather than actual shipments. Therefore, an analysis of those shipments would not reflect a true analysis of shipment rates. Hence, these shipments were removed from the dataset. Naturally, this filter accounts for the removal of zero-miles shipments. *Between 1% and 2% of the data were removed after this filter was applied.*

Freight Rates (RPTM)

Freight rates here are defined as the total revenue divided by shipment length and weight for each waybill, i.e. revenue per ton-mile (RPTM). Both shippers and carriers' perspectives were taken into account for this filtering. First, because carriers do not ship at a loss, **non-positive RPTMs** were removed from the dataset. Also, because shippers do not pay more than $**30,000/carload** for their shipments, waybills with revenue per carload exceeding this value were removed from the dataset. *Less than a half percent of the data was removed after these filters were applied.* Moreover, to avoid outliers in the data, and because it is suggested by the STB that between a quarter and half of 1% of the data contain errors, we removed the top and bottom 0.5% of the data after applying all the filters described above.

Filtered Data

The resulting dataset contains 94% of the publicly available CWS waybills for 2004 through 2013, 91% for 2003 and 93% for 2001 and 2002 (See Table 2-2).

Table 2-2: Pre- and Post-Filtered Waybill Data Summary (2001-2013)

	Total waybills in original CWS	Number of waybills after filtering	Percent of original CWS
Year	N_0	N	%
2001	573,670	535,781	93
2002	589,826	547,954	93
2003	629,127	569,943	91
2004	632,482	592,300	94
2005	667,569	627,023	94
2006	688,170	645,799	94
2007	666,686	626,963	94
2008	622,318	585,471	94
2009	518,343	489,548	94
2010	580,717	547,255	94
2011	599,284	566,164	94
2012	622,884	585,744	94
2013	640,998	602,283	94

The filtering process did not change the trend in the average RPTM. The average RPTM for different years, commodity-types, and regions of origin are described in detail in the following sections.

2.3 Data Description

The objective of this section is to explore the data via describing the key variables, their respective categories, and their evolution over time.

Variable Definitions

An extensive literature review revealed that key determinants of RPTM include distance traveled[9], shipment size[10], commodity-type[11], and shipment type[12]. Those variables are therefore important to the analysis and are described in Table 2-3 as found in the CWS. We also aimed to understand the effect of car ownership-type (private or railroad-owned) on RPTM and its evolution over time because railroads offer incentives to shippers using their own railcars. We contribute to other studies by additionally looking at the number of interchanges and the spatial distribution (See Table 2-4) of shipments. Finally, shipment weights are important because they are used to calculate RPTM for individual shipments.

More variables were of interest to this study than the basic variables given in the CWS. Using the basic variables in the CWS we created new variables; additionally, we created explicit categories for the variables displayed in Table 2-3. Table 2-4 describes the added variables used in the analysis. First, RPTM and other shipment characteristics vary substantially across specific commodities, therefore the analysis differentiates between commodity-types. Second, with the increased differentiation in performance and use of shuttle trains versus traditional trains, the carload number were segmented into different categories. Finally, route density is a factor that is increasingly discussed in the performance of railroads; therefore its implication on RPTM is examined in this analysis. The route density variable was created by summing the number of carloads moving between each origin-termination pair in the CWS for each year.

---

[9] **References:** Rail Tariff Rates for Grain by Shipment Size and D Distance Shipped, Marvin Prater Daniel O'Neil, Jr, AMS, USDA, June 2014; Review and Analysis of Corn Rail Rates, Informa Economics, June 2010; Rail Rate and Revenue Changes Since the Staggers Act, by Ken Casavant, Eric Jessup, Marvin E. Prater, Bruce Blanton, Pierre Bahizi, Daniel Nibarger, Johnny Hill, and Isaac Weingram; Study of Railroad Rates: 1985-2007, *Office of Economics, Environmental Analysis & Administration Section of Economics, STB, January 2009.*

[10] **Reference:** Rail Tariff Rates for Grain by Shipment Size and Distance Shipped, Marvin Prater Daniel O'Neil, Jr, AMS , USDA, June 2014

[11] **References:** Rail Tariff Rates for Grain by Shipment Size and Distance Shipped, Marvin Prater Daniel O'Neil, Jr, AMS , USDA, June 2014; Review and Analysis of Corn Rail Rates, Informa Economics, June 2010; Rail Rate and Revenue Changes Since the Staggers Act, by Ken Casavant, Eric Jessup, Marvin E. Prater, Bruce Blanton, Pierre Bahizi, Daniel Nibarger, Johnny Hill, and Isaac Weingram; Study of Railroad Rates: 1985-2007, *Office of Economics, Environmental Analysis & Administration Section of Economics, STB, January 2009.*

[12] **Reference:** Study of Railroad Rates: 1985-2007, Office of Economics, Environmental Analysis & Administration Section of Economics, STB, January 2009.

Table 2-3: Description of Freight Variables in the CWS Used in Econometric Models

Freight Variable	Description[13]
Freight Revenue	The total line-haul freight revenue, from origin to destination, shown in $
Billed Weight in Tons	The billed weight of lading, calculated in tons
Estimated Short Line Miles	The short line miles (shortest rail distance between origin and destination) rounded to the nearest ten miles
Carload Number	The total carload number on the sampled waybill
Commodity Code (STCC)	The standard Transportation Commodity Code (STCC) identifies the product designation for the transported commodity.
Type of Move [14]	Designation of export, import, or domestic shipment
Car Ownership Code	(P) Privately-owned car (R) Railroad-owned car (T) Trailer Train owned car
Car Capacity	Cubic foot capacity of car (for all car types except flat)
Number of Interchanges	The figure represents the total number of interchanges between railroads in the route
Miscellaneous Charges ($)	The total of all miscellaneous charges (excluding transit charges and freight revenue) shown in dollars
Origin and Termination BEA[15]	The Business Economic Area code for the reported waybill movement's origin or termination location. See "Department of Commerce – Bureau of Economic Analysis, Business Economic Area Codes" revised for 1997.

---

[13] The description is taken from STB's documents accompanying the CWS data

[14] Unfortunately a large portion of the shipments are labeled 'unknown' rather than import, export, or domestic. In the exploratory analysis and regression modeling sections we create a dummy variable for exports. All shipments labeled 'export' are given a value of one; whereas, shipments with any other label are given a value of zero, including, domestic, import, and unknown.

[15] The complete set of BEAs on a map can as given by STB data be found in Exhibit C: Spatial Distribution of BEA Origins in the CWS data (as given by the STB) of the Appendix.

Table 2-4: Description of Freight Flow Variables created for use in the Econometric Models

Freight Variable	Description
Commodity-Type[16]	Grain = Wheat, Corn, and Soybean; Crude oil and Natural gas; Coal; Chemicals; Farm non-grain; Pulp/Paper; Oil and Coal Products; Food; and Other.
Number of Interchanges	0; 1; 2; and 3+
Carload Number	1; 2-5; 6-49; 50-90; and 90+
Route Density (Continuous)	The carload number traveling between a given OD pair in a given year
Route Density (Categorical)	• Less than 100 carloads annually • Between 100 and 1,000 carloads • Between 1,000 and 10,000 carloads • Between 10,000 and 100,000 carloads • More than 100,000 carloads
Distance	• Short → 0-250 Estimated Short Line Miles • Low-Medium → 250-800 miles • High-Medium → 800-1200 miles • Long → 1200-2000 miles • Very Long → 2000+ miles
Spatial distribution [17]	• **Origin dummy variables:** *Northeast* (NE), *Southeast* (SE), *Penn_Ohio_Mich_Wisc, I-states, Upper Midwest* (UMW), Nebraska_Kansas_Colorado, *Texas_Oklahoma_NewMexico* and *Wyoming_Missouri_Idaho* • **Termination dummy variables:** *Texas, the Pacific Northwest (PNW), and Louisiana.*

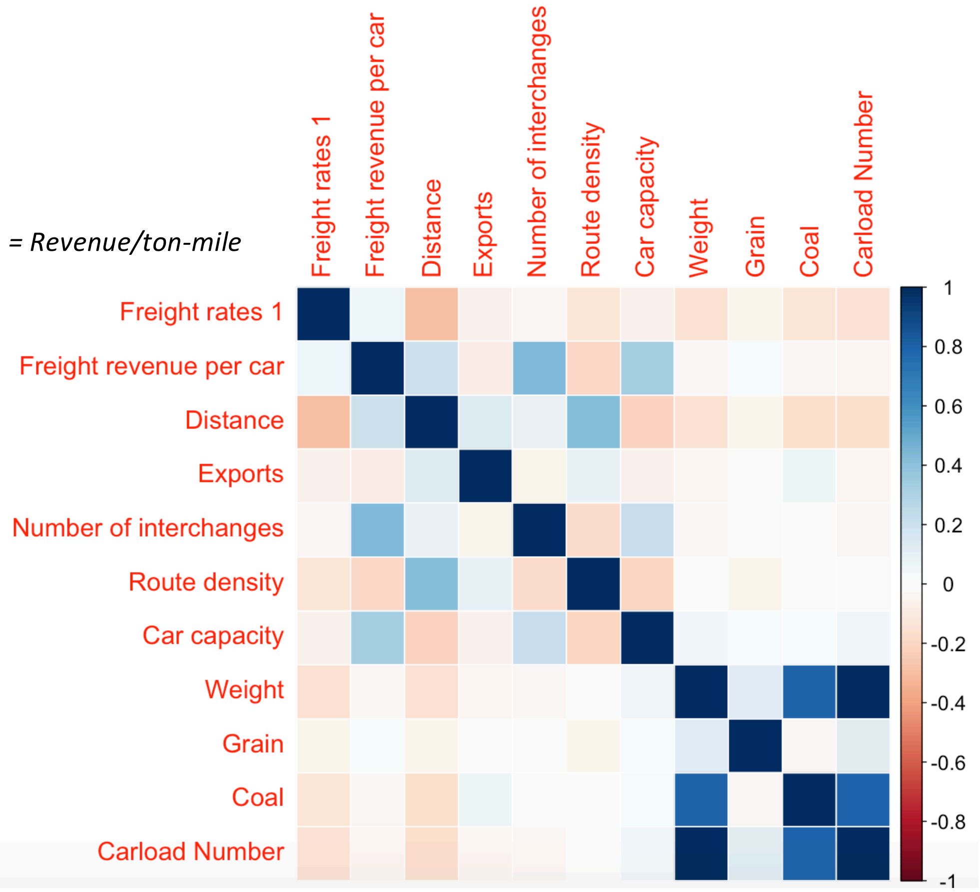
Correlations in the Data

This section examines correlations in the data to better understand prevalent relationships before delving deeper into individual trends. The matrix below examines the correlation between key variables discussed in the previous section. The matrix in Figure 2-1 corresponds to the 2001 CWS but is meant to describe relationships throughout the years, as preliminary analysis determined that the correlation between variables was consistent between 2001 and 2013.

---

[16] The corresponding STC codes can be found in Exhibit D: STCC Code for Commodities Examined in this Study of the Appendix.

[17] The complete list of regions can be found in Exhibit E: List of BEA Origins in the CWS data (as used in this study) of the Appendix.



**Freight rates 1 = Revenue/ton-mile*

Figure 2-1: Correlation Matrix of Key Variables in the Data

The correlation matrix displayed in Figure 2-1 is more helpful in describing the direction of correlations in the data than the strength of the correlations. Robust regression models presented in the next section serve this specific purpose.

Freight Rates (RPTM) Determinants

The correlation matrix shows that **RPTM** is *negatively* correlated with (1) **distance,** (2) **weight,** (3) **carload number,** (4) **route density,** (5) **grain,** (6) **car capacity,** (7) **exports,** and finally (8) **number of interchanges.**

First, this implies that *RPTM is higher for shorter distance shipments*. This conforms to the results in the literature which show that RPTM is lower for longer shipment distances than shorter shipment distances[18]. This is explained by the railroads' desire to increase asset utilization and revenue generation of their rolling stock.

Second, the results suggest that *RPTM is lower for larger shipments (more carload shipments and heavier shipments)*. This is consistent with previous studies including a recent

---

[18] **Reference:** Study of Railroad Rates: 1985-2007, Office of Economics, Environmental Analysis & Administration Section of Economics, STB, January 2009.

USDA report[19] suggesting that smaller shipments are associated with noticeably higher RPTM than large shipments. The report found that in 2011 RPTM for large shipments were 30 percent lower (equivalent to just over 1.3 cents/ton-mile less) than those for the smallest shipment size.

Third, the results also suggest that RPTM is lower for high volume routes. This also goes along the same direction as results found in the literature[20]. The matrix also shows that lower RPTM is associated with grain shipments and shipments going to export. This invites further investigation through the models developed in the second part of this report. Finally, the matrix indicates that lower RPTM is associated with a higher number of interchanges. The interpretation of this result is not very clear and is left for the second part of this report.

Correlations between the Variables in the Data

The correlation matrix shows that **carload number** and **shipment weight** are highly positively correlated. This suggests that using only one of the two variables in the regression model is sufficient for the analysis. In fact, using both variables might lead to multi-collinearity issues. In the following econometric analysis section, RPCM is used rather than RPTM. The correlation matrix suggests that results should be similar for both definitions. In fact, regression models confirmed that using RPCM and RPTM produced very similar results.

Second, the **number of interchanges** and **route density** are negatively correlated. This might reflect the fact that shippers and railroads both want to avoid the costs of switching rail operators. **Coal shipments** are strongly correlated with **carload number** and **shipment weight**. This likely reflects the fact that shuttle and unit trains are being used for coal shipments. Finally, a **higher route density** is associated with **longer distance shipments**. This could indicate that railroads are more efficient at transporting shipments long distances and that railroad pricing reflects these efficiencies.

Key Trends in the Data

After identifying key variable relationships in the data, this section turns to a more differentiated analysis of each variable. The variables discussed include: **(1)** distance, **(2)** weight, **(3)** railcar ownership, **(4)** carload number, **(5)** commodity-type, **(6)** exports **(7)** number of interchanges, **(8)** route density, **(9)** spatial distribution of shipments, and **(10)** *RPTM*.

Distance

In general, shipment distances varied very little over time ranging from 1,150 miles to 1,214 miles over 13 years. The majority of the shipments in the data (around 30%) are in the low-to-medium category (i.e. between 250 and 800 miles), followed by very long shipments (i.e.

---

[19] **Reference:** Rail Tariff Rates for Grain by Shipment Size and D Distance Shipped, Marvin Prater Daniel O'Neil, Jr, AMS, USDA, June 2014.

[20] **Reference:** Competition and Rail Rates for the Shipment of Corn, Soybeans, and Wheat, James M. MacDonald

more than 2000 miles) with a share of approximately 23% of the shipments, slightly higher than medium-to-high shipments (i.e. between 800 and 1200 miles) and long-distance shipments (i.e. between 1200 and 200 miles) with around 20% of the shipments each. The lowest share of 8% is for the short distances (i.e. between 20 and 250 miles).

Weight

The average weight in the data ranged between 87 tons (reached in 2013) and 100 tons (reached in 2009). Shipments weights, like distances were stable over time.

Railcar Ownership

Unlike distance and weight, railcar ownership changed over time, with shippers using fewer railroad-owned railcars over time. In 2013, 17% of shipments used railroad-owned railcars compared with 29% in 2001. This could be due to the premiums that railroads charge for using railroad-owned railcars[21].

Carload Number

Most shipments in the CWS data are one-carload shipments (98%). It is likely that one-carload shipments make up a huge majority of all rail waybills; however, it was brought to the NUTC's attention that sometimes each single carload on multi-car, unit, and shuttle trains are given their own waybill. This misreporting of carload number in the waybill sample may be the reason why the proportion of one-carload shipments is so high. The next highest share is for 2-5 carloads (1-1.5%), followed by 6-49 carloads (0.5% to 0.8%), and followed by 90+ carloads (0.3-0.5%) and finally 50-90 carloads (0.1-0.2%). The percentages were relatively stable over time.

Commodity-type

Table 2-5 below shows the percentages of different commodities in the data.

---

[21] It is worth noting here that shipments in the data can also be categorized as Trailer Train; however, the percentage of shipments designated 'Trailer Train' is very low. . Those were not taken into consideration in this analysis. Also, some shipments were unpopulated (neither private, nor railroad, nor Trailer Train). Those shipments as well were dropped from the analysis. The percentages here reflect the number out of both railroad-owned and private railcars.

Table 2-5: Commodity Percentages in the Data

Commodity	**Percentage of shipments**	**Percentage of carloads**
Chemical	8-9 %	6-8%
Bulk Grain (corn, wheat, soybean)	0.4%	3.1%
Coal	0.7%	22%
Oil and Gas	2 %	2%
Farm (non-grain)	1-2%	2-3%
Food	6-8%	5-6%
Pulp paper	3-4%	2-2.6%

The number of *coal shipments* decreased significantly between 2001 and 2013. *Grain shipments* fluctuated with peaks in 2007 and 2011. *Chemical shipments* increased over time with a drop between 2008 and 2009. Farm (non-grain) shipments significantly increased between 2007 and 2008 and continued rising until 2013.

Exports

Figure 2-2a and Figure 2-2b show that the percentage of all shipments and carloads labeled ‘export’ was relatively constant between 2001 and 2013, respectively. The percentage of shipments labeled export was between 4% and 7% every year from 2001 to 2013.

Figure 2-2a shows that the percentage of coal shipment labeled export decreased significantly between 2005 and 2006. Between 2001 and 2005 the percentage of coal shipments labeled export was between 35% and 60%; whereas, after 2005 the percentage ranged from 3% to 5%. However, Figure 2-2b shows that the percentage of coal carloads labeled export remained relatively stable between 2001 and 2013. The largest percentage of coal carloads labeled export was 3.5% and the lowest was 1.3% between 2001 and 2013.

Figure 2-2a and Figure 2-2b show that the percentage of grain shipments and carloads labeled export increased significantly and steadily between 2001 and 2011/2012, respectively. The percentage of grain shipments labeled export increased from 3-5% in the early 2000s to 20-30% between 2009 and 2013.

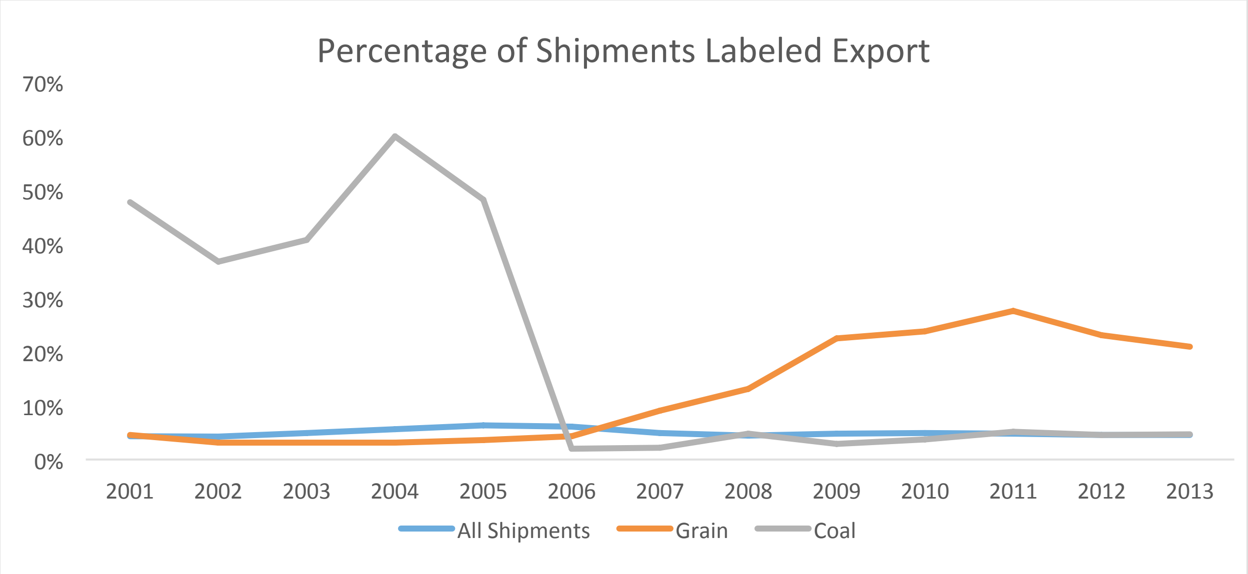


Figure 2-2a: The percentage of shipments labeled export by commodity-type

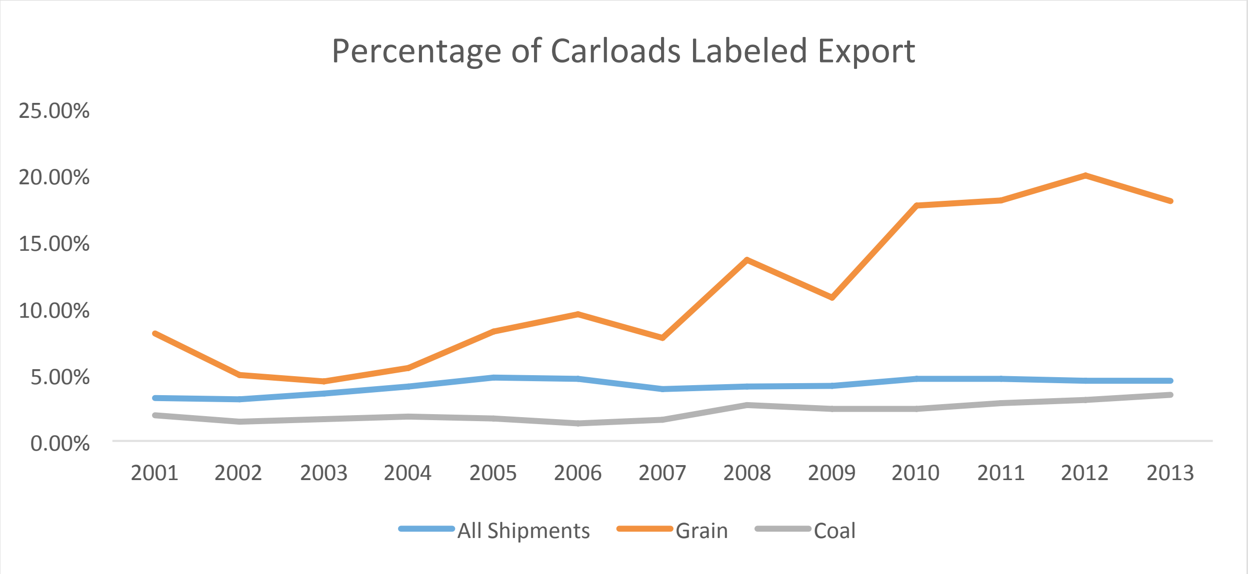


Figure 2-2b: The percentage of carloads labeled export by commodity-type

We use the phrase 'labeled export' because each year between 51% and 55% of all shipments are labeled 'unknown' in the move type data field which designates shipments import, export, or domestic. The results in Figure 2-2a and Figure 2-2b include the shipments labeled unknown; hence, the figures understate the percentage of shipments in each commodity category that are bound for export each year. Figure 2-3a and Figure 2-3b display the percentage of shipments bound for export after excluding the shipments labeled unknown. As expected the percentage going to export increases (i.e. it nearly doubles) for all shipments, grain, and coal. The percentage of grain shipments bound for export between 2010 and 2013 ranges from 45% to 60%; whereas, the percentage of grain carloads bound for export ranges from 35% to 43% over the same period. Despite the change in magnitude that occurs when the shipments labeled unknown are removed, the general trends displayed in Figure 2-2a and Figure 2-2b remain the same.

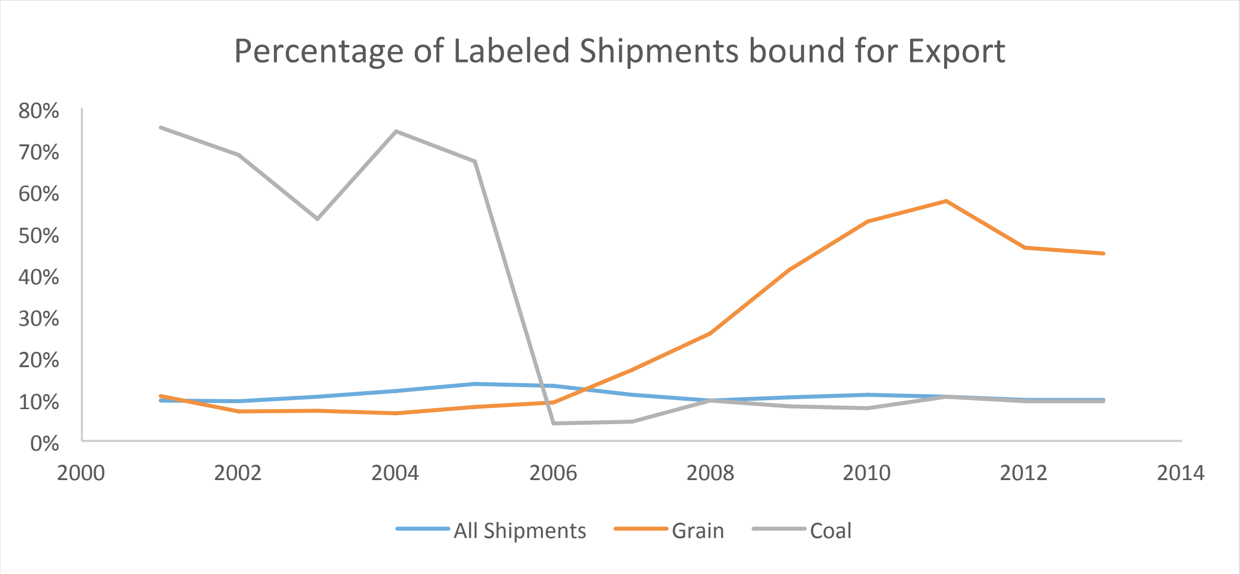


Figure 2-3a: Percent of shipments not labeled unknown that are bound for export by commodity-type

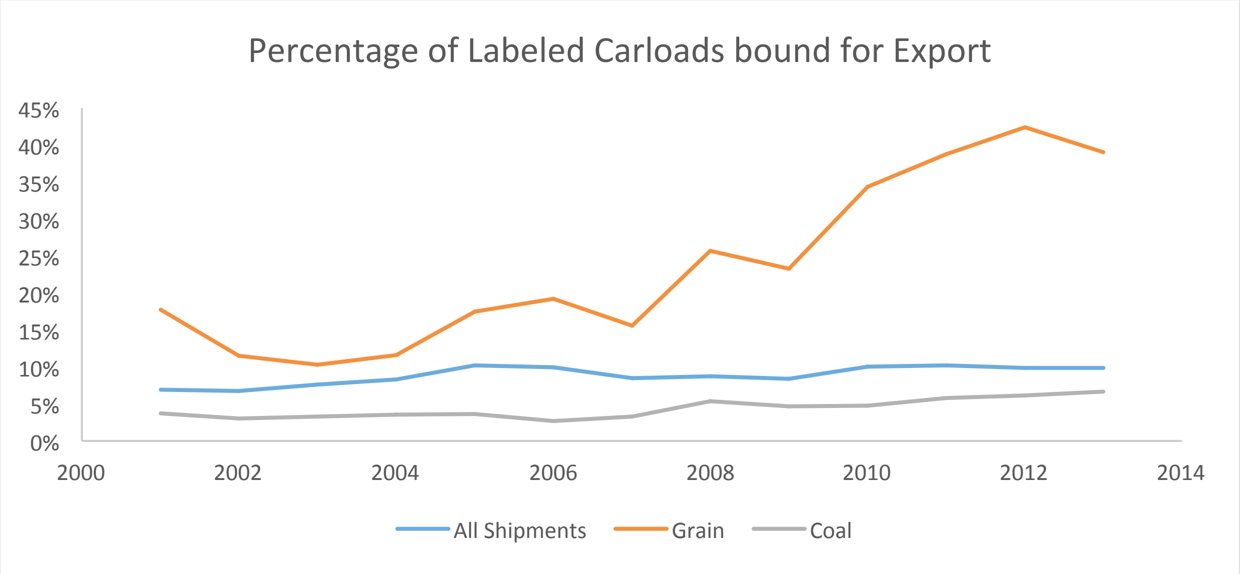


Figure 2-3b: Percent of carloads not labeled unknown that are bound for export by commodity-type

Number of Interchanges

The majority of shipments in the data do not go through any railroad interchange (88% yearly), 12% go through one interchange, 1% to 2% go through two interchanges and less than 0.5% go through three or more interchanges. The number of shipments and their percentages are relatively stable over time. The largest decrease occurred for shipments going through one interchange.

Route Density

Route densities for shipments in the CWS mostly fall under *medium* (between 1,000 and 10,000 carloads) and *heavy* (10,000 and 100,000 carloads) route density with 65% of shipments in these two categories. Next comes *really high* route density (i.e. more than 100,000 carloads) with a share of 23%. The next largest category is *low* route density (between 100 and 1,000 carloads) with a share of 7%, and finally *'no'* route density (less than 100 carloads) with a share of 4%. The percentages are relatively stable over time except for 2009 and 2010. In 2009 there were no shipments in the "really high" route density category. In 2010 there were no shipments in the small and medium congestion categories and all the majority of the shipments travelled on congestion-free routes.

Spatial Distribution of Shipments

Table 2-6 below shows the percentage of different shipment origins and terminations in the CWS dataset. The Upper Midwest origination point makes up 1.3% of all shipments and 4% of the carloads, indicating that the average carload number of shipments originating in the Upper Midwest is significantly higher than the national average.

Table 2-6: Percentage of Different Regions in the United States

Origin	**Percentage of shipments**	**Percentage of carloads**
Northeast	35%	28%
Pennsylvania- Ohio-Michigan-Wisconsin	23%	18%
Southeast	6.3%	10%
Texas-Oklahoma-New Mexico	6%	5%
Wyoming-Montana-Idaho	1.1%	12%
I-states (Indiana, Iowa, Illinois and Missouri)	3.8%	4%
Upper Midwest (North Dakota, South Dakota, western Minnesota)	1.3%	4%
Nebraska-Kansas-Colorado	1% and less	1.6%
Termination	**Percentage of shipments**	**Percentage of carloads**
Texas	3% to 4%	4%
Pacific Northwest (Northern California, Oregon, Washington)	2% to 5%	3%
Louisiana	1% and less	1% and less

2.4 Trends in Freight Rates (RPTM)

This last section of the first part of the report serves as a preliminary step in understanding trends in RPTM. Below we look at RPTM based on (1) **commodity-type:** all shipments versus grain shipments versus coal shipments and (2) **export vs. domestic:** all shipments versus export shipments and export shipments segmented into grain and coal.

Commodity-type

Figure 2-4a and Figure 2-4b show that the average RPTM for coal was less than the average RPTMs for bulk grain and all shipments. The only major difference between Figure 2-4a and Figure 2-4b is that adjusting for inflation flattens-out the increase in RPTM for all shipments, grain, and coal. Nominal RPTM for all shipments increased from 2.9 cents per ton-mile in 2001 to 5.3 cents per ton-mile in 2013. Nominal RPTM for grain (coal) shipments increased from 2.0 (1.7) cents per ton-mile in 2001 to 3.6 (3.1) cents per ton-mile in 2013. However, the average RPTM for coal had the highest compound annual growth rate (CAGR[22]) at 5.0%, followed by grain with a CAGR of 4.7%. The CAGR for average RPTM for all shipments was 4.6%.

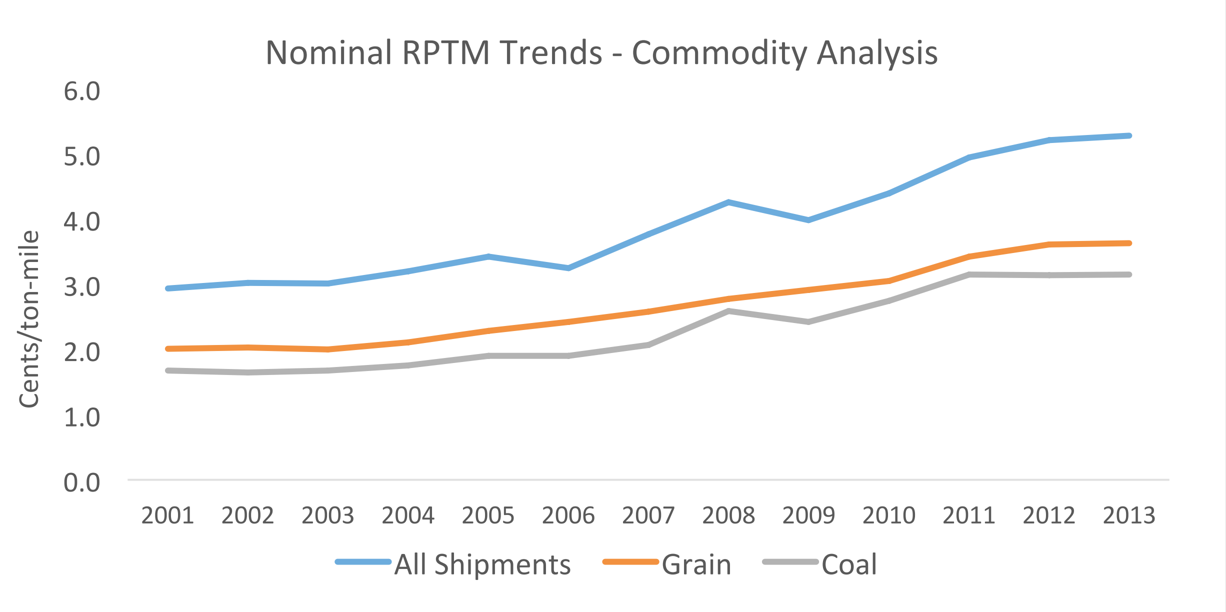


Figure 2-4a: Trends in average nominal RPTM by commodity-type

---

[22] Compound annual growth rate (CAGR) is calculated as follows: $CAGR = \left(\frac{final\ value}{initial\ value}\right)^{\frac{1}{num\ years}} - 1$

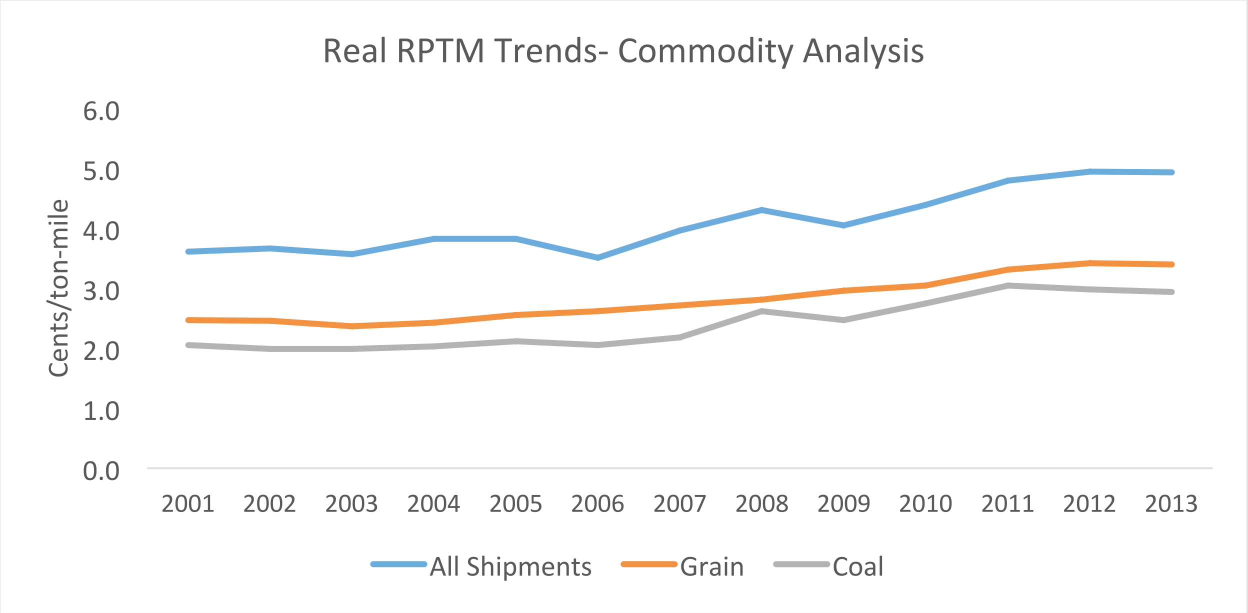


Figure 2-4b: Trends in average real (2010 dollars) RPTM by commodity-type

Exports

Figure 2-5 displays the average real RPTMs for grain export shipments and coal export shipments compared with the average real RPTMs for all exports and all shipments (export and domestic). The figure shows that the average real RPTM for grain shipments was consistently the lowest of the four categories; moreover, Figure 2-5 shows that the average real RPTM for all export shipments was consistently lower than the RTPM for all shipments. The result that the RPTM for export shipments is lower than the RPTM of all shipments is examined more formally in Section 3. Lastly, the figure shows that even in real terms, the average real RPTM for coal shipments significantly increased between 2001 and 2012. In 2001 the average RPTM (in 2010 dollars) for export coal shipments was 4.2 cents per ton-mile; whereas, in 2013 the RPTM (in 2010 dollars) for export coal shipments increased to 8.8 cents per ton-mile. Out of all exports, coal export RPTM increased at the highest rate (8% annually). Export grain increased annually at the same rate as all export shipments (4.8% annually).

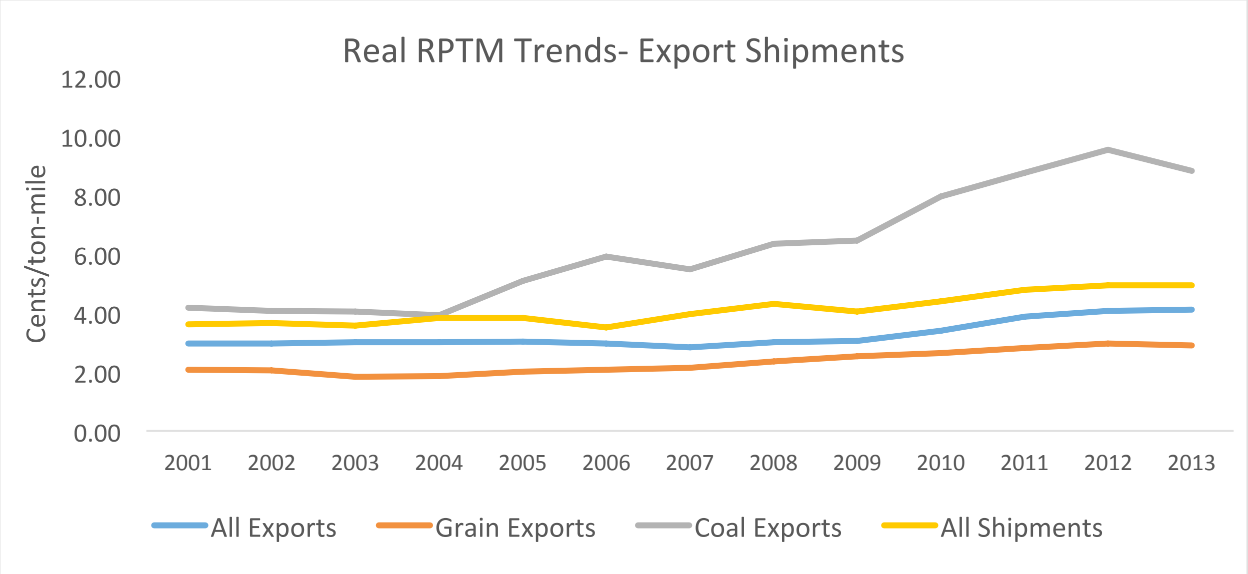


Figure 2-5: Trends in average RPTM for different export commodities in real terms

2.5 From Exploratory Analysis to Regression Modeling

The exploratory analysis undergone in the first part of the report serves as a preliminary step in understanding the data at hand through: **(1)** the identification of key variables, their statistics and trends, **(2)** examining correlations between variables, and **(3)** looking at trends in RPTM and other variables over time. While this is a crucial step to provide background, direction, and a sense of which variables are related and the degree of their relation, it does not tell the entire story. For example, weak correlations do not allow for strong conclusions; therefore a deeper statistical analysis is necessary. Furthermore, correlations and relationships between multiple variables need to be considered simultaneously to avoid spurious correlations and omitted variable bias. The next section of this reports builds on the results presented above to address the limitations of simple exploratory analysis and variable correlation through robust econometric models.

3 Multivariate Regression Models

The econometric regression models and the results presented in this section serve two main purposes. The first is to determine trends in average RPCM for rail transportation between 2001 and 2013 after controlling for shipment characteristics such as carload number, shipment distance, route density, railcar ownership, and export vs. domestic. We specifically determine whether or not trends in RPCM differed for grain shipments and shipments originating in the Upper Midwest (North Dakota, South Dakota, and western Minnesota) relative to overall trends in RPCM for all shipments. The second purpose of the regression models is to determine the shipment characteristics associated with higher and lower RPCM and examine how the associations varied across origination region and commodity-type. The regression models include the following shipment characteristics: shipment distance, carload number, number of railroad interchanges, route density, export vs. domestic, railroad or private railcar ownership, and shipment weight. The models also take into account the commodity-type, shipment origination and termination point as well as the year the shipment occurred.

Using the entire CWS, Section 3.1 presents a multivariate regression model to determine how specific shipment characteristics impacted RPCM. Additionally, the analysis determines whether average RPCM increased or decreased between 2001 and 2013 in real terms, controlling for the aforementioned shipment characteristics. The regression models in Section 3.1 are run on all the waybills that remained after the filtering process described in Section 2.2.

In Section 3.2 we segment the data by year (i.e. we run the same regression model on data from 2001, 2002, …, 2013 separately) in order to determine trends in the impacts of shipment characteristics on average RPCM. The combined model in Section 3.1 determines how, on average between 2001 and 2013, RPCM was impacted by different shipment characteristics, commodity-type, and region of origin. Section 3.2 examines whether and how the impact of the shipment characteristics, commodity-type and region of origin varied between 2001 and 2013.

Section 3.3 analyzes grain and coal rail shipments exclusively and Section 3.4 analyzes shipments originating in the Upper Midwest and in the I-states (Iowa, Indiana. Illinois, and Missouri) exclusively. The goal of these two sections is to analyze grain transport rates and transport rates in the Upper Midwest; however, we also analyze coal and the I-states in order to compare grain shipments and shipments originating in the Upper Midwest with comparable segments of the dataset. In both Section 3.3 and Section 3.4 we analyze the types of shipment characteristics associated with higher and lower average RPCM. Like Section 3.2, separate regression models are run for each year between 2001 and 2013. Section 3.5 focuses solely on grain shipments originating in the Upper Midwest. The analysis aims to determine the shipment characteristics associated with higher and lower average RPCM.

3.1 Pooled Model

In this section we examine the entire filtered CWS dataset, i.e. the data were not segmented by commodity, year, geographical region, or any other factor. The entire, filtered CWS dataset was the input to the multivariate regression model displayed in Equation 1.

$$
\begin{aligned}
\ln\left(\frac{revenue}{carloads - dist}\right) &= \beta_1 \frac{weight}{carloads} + \beta_2\, dist_{fact} + \beta_3\, RouteDen_{fact} + \beta_4\, Carloads_{fact} \\
&+ \beta_5\, Interchanges_{fact} + \beta_7\, Commod_{fact} + \beta_7\, Origin_{dummy} \\
&+ \beta_8\, Term_{dummy} + \beta_9\, Ownership_{dummy} + \beta_{10}\, Export_{dummy}
\end{aligned}
$$

Equation 1

Equation 1 is a log-linear model, a generalization of a linear model wherein the log of the dependent variable is used to reduce heteroscedasticity and model non-linear relationships. The term on the left-hand side of Equation 1 (i.e. the dependent variable) is the natural logarithm of RPCM. The first term on the right-hand side of Equation 1 is the ratio of a shipment's total weight divided by the number of carloads in the shipment. We expect the heavier the average carloads in a shipment the higher the RPCM, i.e., we expect β_1 to be positive. The next four explanatory variables $(dist_{fact}, RouteDen_{fact}, Carloads_{fact}, Interchanges_{fact})$ are ordinal variables. For example, the variable $dist_{fact}$ has five ordinal levels: 0-249 miles, 250-799 miles, 800-1199 miles, 1200-1999, and 2000 or more miles. The $Commod_{fact}$ variable defines the commodity-type that was transported. Each waybill is assigned a commodity-type by the carrier; in this analysis we narrow the number of commodity-types considerably. The $Commod_{fact}$ variable only includes chemicals, bulk grain, crude oil and natural gas, coal, food, non-grain agricultural products, and coal and petroleum products. All the remaining waybills with different commodity-types are grouped together and labeled 'other'. The variables with a subscript $'dummy'$ indicate binary variables. For example, $Ownership_{dummy}$ equals one if the railroad owns the railcars and zero if the railcars are privately-owned. The origin and termination variables actually represent a group of dummy variables. For example, we assigned dummy variables to shipments terminating in Texas, the Pacific Northwest, and Louisiana (three of the largest regions for export shipments). Presented below are the variable descriptions.

$revenue$:	price paid by shippers to railroad
$carloads$:	number of carloads on a waybill
$dist$:	distance the shipment traveled
$weight$:	total weight transported
$RouteDen$:	density (carloads/year) of the shipment's origin-termination pair
$Origin$:	a shipment's origination point
$Term$:	a shipment's termination point
$Commod$:	commodity-type
$Interchanges$:	number of railroad interchanges between origin and destination
$Ownership$:	1 if railroad owns railcars on waybill; 0 if railcars privately-owned
$Export$:	1 if the shipment is bound for export; otherwise, 0
β_i:	coefficient estimate of variable $i \in [0,10]$
$time$:	year the shipment was transported

Table 3-1 displays the results of the multivariate regression model, presented in Equation 1, run on the pooled dataset (i.e. all the filtered CWS's between 2001 and 2013). Each of the variables included in the model along with their relative factor levels, if relevant, are listed in the first column. The second column displays the coefficient value estimates for each of the variables in the model. The third column displays the p-value of each of those coefficient estimates. The p-values were calculated using heteroscedastic consistent standard errors[23].

---

[23] **Reference:** Long, J. Scott, and Laurie H. Ervin. "Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model." *The American Statistician* 54.3 (2000): 217. Web.

Table 3-1: Results of multivariate regression model run on entire CWS.

Parameter	Coefficient Estimate	P-value
(Intercept)	0.940	0.000
tons/ carloads	0.016	0.000
Distance (base = 0-199 miles)		
250-799 miles	-0.648	0.000
800-1199 miles	-0.977	0.000
1200-1999 miles	-1.165	0.000
2000+ miles	-1.386	0.000
Route Density (base = <100 carloads)		
100-999 carloads	-0.038	0.000
1,000-9,999 carloads	-0.087	0.000
10,000 -9,999 carloads	-0.234	0.000
100,000+ carloads	-0.438	0.000
Carloads (base = 1 Carload)		
2-5 Carloads	-0.103	0.000
6-49 Carloads	-0.272	0.000
50-90 Carloads	-0.336	0.000
90 + Carloads	-0.494	0.000
Interchanges (base = 0)		
1 Interchange	0.179	0.000
2 Interchanges	0.209	0.000
3+ Interchanges	0.154	0.000
Commodity (base = Other)		
Chemical	0.209	0.000
Coal	-0.041	0.000
Crude Oil and Natural Gas	0.304	0.000
Agriculture (excluding grain)	0.021	0.000
Food Products	-0.006	0.000
Grain	-0.144	0.000
Coal and Petroleum Products	0.258	0.010
Pulp and Paper	0.109	0.000
Regional Dummy Variables		
UMW Origin	-0.071	0.000
I-States Origin	-0.001	0.331
TxOkNm Origin	-0.142	0.000
Northeast Origin	-0.106	0.000
WyMoId Origin	-0.235	0.000
Texas Termination	0.171	0.000
PNW Termination	-0.120	0.000
Louis Termination	-0.105	0.000
Other Dummy Variables		
Railroad Owned	0.120	0.000
Export	-0.343	0.000
$R^2 = 0.7491$, Adjusted $R^2 = 0.7491$ Residual Standard Error: 31.03 on 4130347 degrees of freedom		

The interpretation of the parameter coefficient values in Table 3-1, due to the fact that we used a log-linear model, are as follows: a one unit increase in the parameter (e.g. export) produces an expected change in the log of the dependent variable (RPCM) of β_{10} (the coefficient in Equation 1 for the export dummy variable)[24]. More intuitively, the percent change in RPCM as a parameter increases by one unit is defined as follows:

$$let\ y = RPCM$$

$$\frac{y + \Delta y}{y} = \exp\left(\beta_{10} \times \Delta Export_{dummy}\right) = \exp(\beta_{10} \times 1)$$

$$\frac{\Delta y}{y} = \exp(\beta_{10}) - 1$$

As an example, the percent change in RPCM if a shipment is labeled export rather than 'non-export' is -29%.

$$\frac{\Delta y}{y} = \exp(-0.343) - 1 = 0.71 - 1$$

$$\frac{\Delta y}{y} = -0.29$$

The third column of Table 3-1 shows that all of the coefficient value estimates have p-values less than 0.001 except for the dummy variable designating shipments originating in the I-states.

An interesting result, displayed in Table 3-1, is that the RPCM was lower for export shipments than non-export shipments. This result takes into account carload number, shipment distance, route density, and commodity-type. This result is consistent with the exploratory analysis in Section 2. In contrast to the results of the exploratory analysis, which suggest that average RPCM decreased with the number of railroad interchanges, Table 3-1 shows that the RPCM is actually higher for one or two railroad interchanges than zero interchanges. It is unclear why the coefficient for three or more interchanges is smaller than the coefficient for one interchange and two interchanges. We can think of two possible explanations: (1) the number of shipments in the three or more category is relatively small suggesting that there is not enough data to make a strong conclusion about the magnitude of the parameter (however, the p-value for the parameter is 0.000 indicating that the standard-error is small relative to the coefficient estimate) and (2) given how rare it is for shipments to be transported by more than three railroads, it is possible that the number of interchanges is being misreported.

The coefficient values for the first five variables and the railroad ownership variable presented in Table 3-1 conform with railroad economic principles. The model results suggest: (1) The more weight a shipper put in a carload, the higher the average RPCM, likely due to the fact railroad costs per carload increase with heavier carloads. (2) Average RPCM was lower for

---

[24] **Reference:** Benoit, Kenneth. "Linear regression models with logarithmic transformations." *London School of Economics, London* (2011).

longer distance shipments, likely due to economies of distance. (3) Average RPCM was lower for shipments traveling between high density origin-termination pairs, likely due to Average fact that fewer inter-and intra-train switches are required for shipments on routes with high density. (4) Average RPCM was lower for larger shipments (i.e. shipments with more carloads), likely due to economies of scale. (5) Average RPCM was higher for shipments with more inter-railroad interchanges, likely due to the increased operational and accounting/financial burden of switching carloads from one carrier to another. (6) Lastly, average RPCM was lower for shippers using their own railcars than shippers using railroad-owned railcars.

The results in Table 3-1 also indicate that the RPCM for grain shipments, across the country, was lower on average than the RPCM for non-grain shipments. Specifically, the results indicate that the RPCM for grain was definitely lower on average than the RPCMs for chemicals, coal, crude oil and natural gas, food products, non-grain agricultural products, coal and petroleum products, and pulp and paper. Once again grain shipments only includes bulk grain: corn, wheat, and soybeans. Additionally, Table 3-1 indicates that the RPCM for shipments originating in the Upper Midwest was lower than the average RPCM of shipments that did not originate in the Upper Midwest. However, the RPCMs for shipments originating in the Wyoming-Montana-Idaho and Texas-Oklahoma-New Mexico regions were lower than the RPCM of shipments originating in the Upper Midwest.

The statements made in the previous paragraph regarding the lower RPCM of grain shipments and shipments originating in the Upper Midwest require the use of econometric methods. A naïve methodology, wherein, the analyst simply takes the average RPCM of grain shipments and compares it with the average RPCM of non-grain shipments, may or may not yield the same results as the econometric method. The naïve methodology does not account for external factors such as shipment distance, number of railroad interchanges, and carload number in its comparison of the RPCM across commodity-types or regions of origin. Not accounting for these factors can lead to erroneous results. For example, if the average grain shipment were shorter than the average non-grain shipment, and because the RPCM is lower for longer shipments (see Table 3-1), using the naïve methodology would erroneously suggest that average PRCM for grain is higher, in comparison with average RPCM for non-grain, than it truly are. Hence, it is necessary to account for these external factors when comparing RPCM across different commodities or shipment origination points. Multivariate regression methods explicitly separate the impacts of exogenous variables, such as, shipment distance and carload number from other explanatory variables such as commodity-type.

A second multivariate regression model was run on the entire CWS. The second model includes a variable that designates the year the shipment occurred ($time_{fact}$) in order to provide insights into how average RPCM varied between 2001 and 2013. Unfortunately, due to the enormous size of the data set and the software's random-access memory (RAM) storage procedure, the route density variable and most of the origination and termination dummy variables are not included in the second regression model. Adding the time variable to the regression model and removing the route density and origination and termination variables did not alter the coefficient estimates for the other parameters (displayed in Table 3-1 above) significantly. The coefficient estimates for the $time_{fact}$ variable are displayed in Figure 3-1. A point on the graph indicates the relative increase or decrease in average RPCM between the given year (x-axis) and the base year of 2001. Figure 3-1 shows that average RPCM was relatively constant between 2001 and 2006, steadily increased between 2006 and 2012, and then leveled off between 2012 and 2013. It is important to note that the revenue component of RPCM was adjusted for inflation; therefore Figure 3-1 shows the trend for real, as opposed to nominal, RPCM. The Consumer Pricing Index (CPI) was used in this analysis to adjust for inflation.

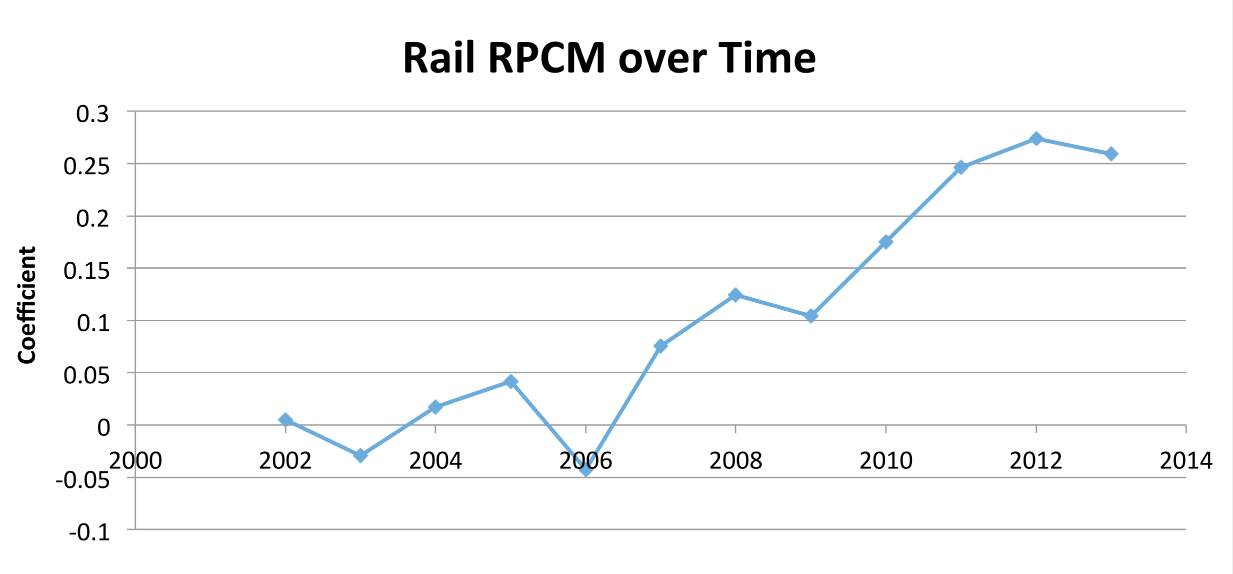


Figure 3-1: Coefficient values for the time variable in the regression model. The coefficient values are relative to the 2001 base level which was set to zero.

3.2 Segmentation by Year

Table 3-1 illustrates how a number of variables impact RPCM including carload number, shipment distance, commodity-type, and shipment origin. However, the results in Table 3-1 do not show whether the impact of these variables remained constant or changed between 2001 and 2013. There are two methods to determine how the impact of specific variables on RPCM changed between 2001 and 2013, taking into account the overall trend of higher RPCM displayed in Figure 3-1. The first method involves interacting each of the variables in Equation 1 with the year variable in the pooled data set. Due to the limitations of the software's RAM storage procedures, mentioned in Section 3.1, this method was infeasible. A second and equally effective method involves segmenting the dataset by year, running the same model on data from each year independently, and determining how the variable coefficients changed over time. The results of the second method are presented in Figure 3-2 through Figure 3-7.

Figure 3-2a and Figure 3-2b display the coefficient values for the origination and termination dummy variables in the multivariate regression model, respectively. The figures illustrate the trend in dummy variable coefficient values for a number of origination and termination points. It is imperative to note that these figures do not illustrate the overall trend in RPCM in the specific regions displayed, rather, they display the trend in the difference between average RPCM for specific regions and the average RPCM for all other regions. For example, Figure 3-2a illustrates that the RPCM for shipments originating in the Wyoming-Montana-Idaho region decreased steadily between 2001 and 2013 relative to shipments not originating in Wyoming-Montana-Idaho. It is possible, and likely, given the results presented in Figure 3-1, that average RPCM increased in Wyoming-Montana-Idaho between 2001 and 2013, but we definitively know, from Figure 3-2a, that average RPCM in Wyoming-Montana-Idaho decreased relative to the average RPCM of shipments not originating in Wyoming-Montana-Idaho between 2001 and 2013.

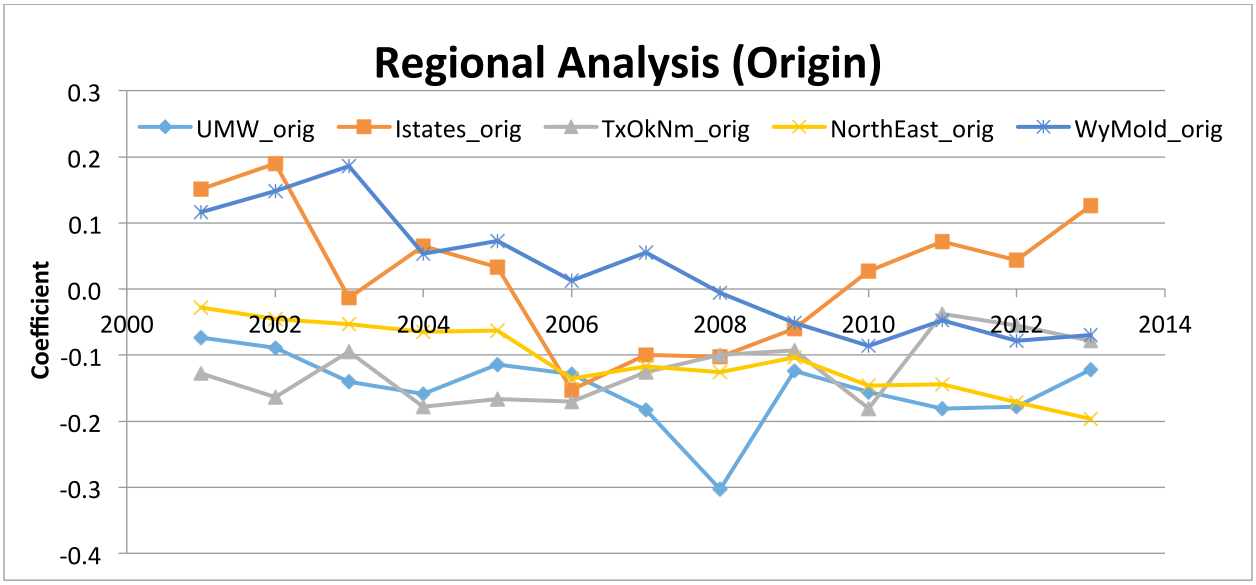


Figure 3-2a: Coefficient values for the origination region dummy variables. A separate regression model was run for every year between 2001 and 2013 in order to track trends.

The regression model coefficients displayed in Figure 3-2a show that the RPCM for shipments in the Upper Midwest was consistently lower than the RPCM for shipments not originating in the Upper Midwest (i.e. the light-blue 'UMW_orig' line is always below zero). Figure 3-2a also shows that the RPCM for shipments originating in the Texas-Oklahoma-New Mexico region was consistently lower than the RPCM for shipments not originating in this region. Additionally, the RPCM for shipments originating in both the Northeast and the Wyoming-Montana-Idaho region steadily decreased between 2001 and 2013 relative to shipments not originating in these two regions.

The analysis of three termination points, illustrated in Figure 3-2b, shows that the RPCM for shipments terminating in Texas was consistently higher than the RPCM for shipments not terminating in Texas; whereas, the RPCM for shipments terminating in the Pacific Northwest and Louisiana were lower than the average RPCM of other termination regions. However, the RPCM for the Pacific Northwest termination point increased slowly but steadily between 2001 and 2013 relative to shipments not terminating in the Pacific Northwest.

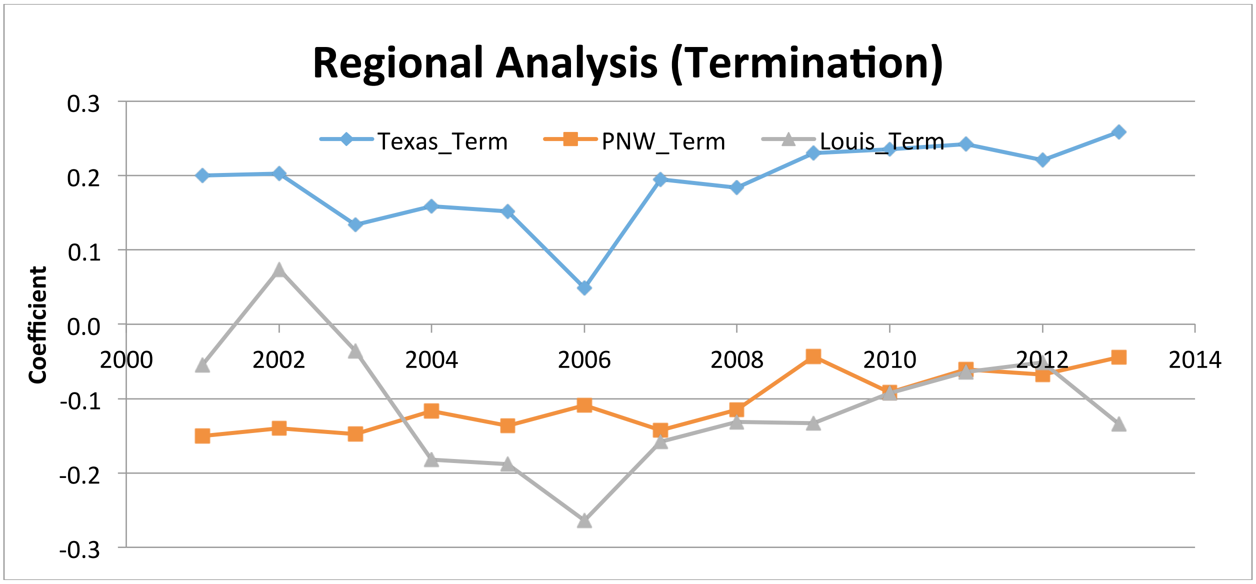


Figure 3-2b: Coefficient values for the termination region dummy variables.

Figure 3-3a and Figure 3-3b display the coefficient values for the commodity variable; each point in each of the figures represents the relative difference between the average RPCM of the specified commodity and the average RPCM of the 'all other' commodity category described in Section 3.1. Both figures compare the RPCM of grain shipments with the RPCMs of a few other commodities. Figure 3-3a illustrates that the RPCM for grain shipments was consistently lower than the RPCM for 'all other' commodities as well as the fossil fuel commodities. Figure 3-3a shows that the RPCM for crude oil and natural gas was highly variable. In 2006, 2007, and 2008 the RPCM for crude oil and natural gas shipments was significantly higher than other commodities; then in 2010 and 2011 the RPCM for crude oil and natural gas was lower than other commodities; then in 2012 and 2013 the RPCM for crude oil and natural gas was once again significantly higher than other commodities. The RPCM for coal shipments was similarly volatile between 2001 and 2013. The RPCM for coal and petroleum products steadily increased between 2001 and 2013 relative to other commodities. The RPCM for grain shipment was consistently lower than other commodities, although in 2006 and 2011 the difference was not as large as other years.

Figure 3-3b illustrates that the RPCM for bulk grain (corn, wheat, and soybeans) was also lower than the RPCMs for other, more similar, commodites including non-grain agricutural products, food products, as well as chemical and pulp-paper commoditites. Besides the average RPCM of pulp-paper shipments that steadily decreased relative to the other four commodities in Figure 3-3b between 2006 and 2013, the commodities in Figure 3-3b are nearly constant in their relative position compared with grain. That is, the RPCM of grain was always lower than the RPCM of non-grain agricultural, which, in turn, was always lower than the RPCM for food products, which was always lower than the RPCM for chemicals.

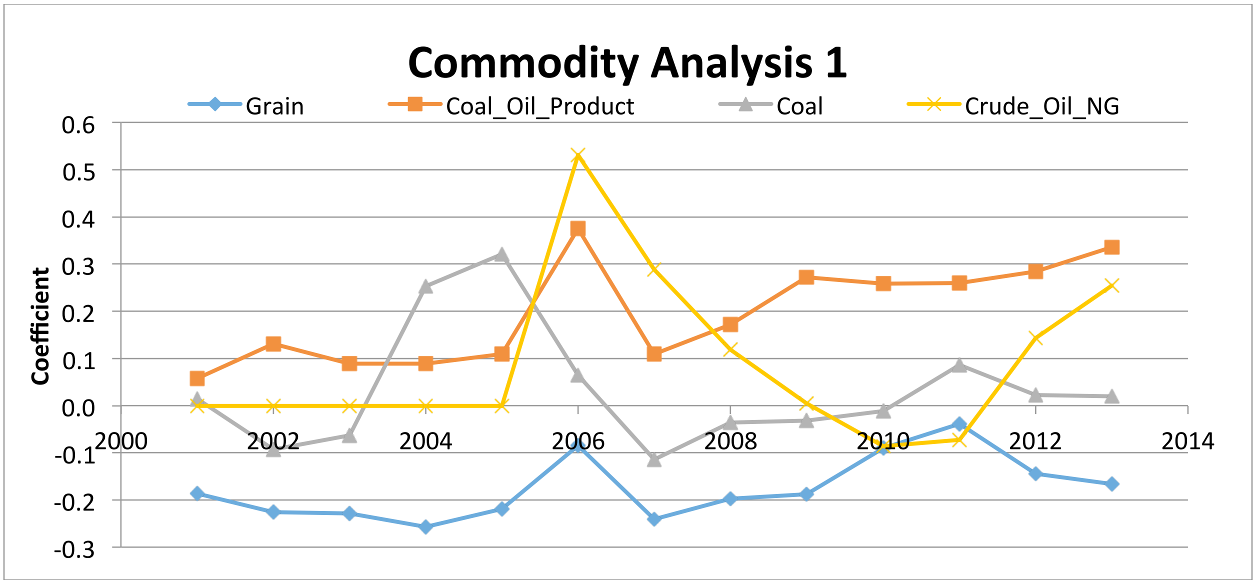


Figure 3-3a: Coefficient values the for commodity variables (values are relative to 'other' commodity which includes all the commodities not explicitly enumerated in the commodity analysis). Comparison of the average RPCM of grain with the average RPCM of fossil fuel.

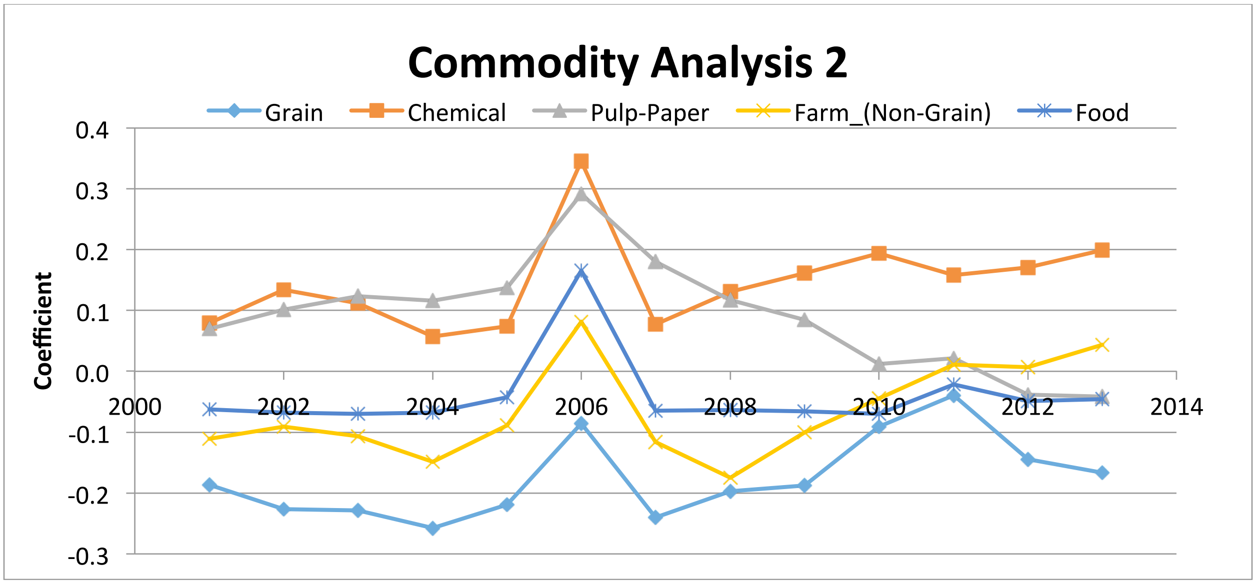


Figure 3-3b: More commodity variable coefficient values. Comparison of the average RPCM of grain with the average RPCM of related commodities.

The next four figures display the coefficient values for the the shipment characteristics included in the model in Equation 1. The figures display how the RPCMs of different shipment-types varied between 2001 and 2013. Figure 3-4 shows that the more carloads that are shipped together on a single waybill the lower the average RPCM of the shipment. This result is not entirely surprisingly but it is quite important for shippers to understand. After factoring in a number of other variables including shipment distance, and route density; the RPCM charged by railroads is systematically lower for larger shipments. Shippers should consider consolidating their products into larger shipments in order to receive lower per carload-mile rail transport costs. However, Figure 3-4 does show that the RPCM gap between all the non one-carload segments, and the one-carload segment, decreased between 2011 and 2013 and significantly between 2001 and 2013.

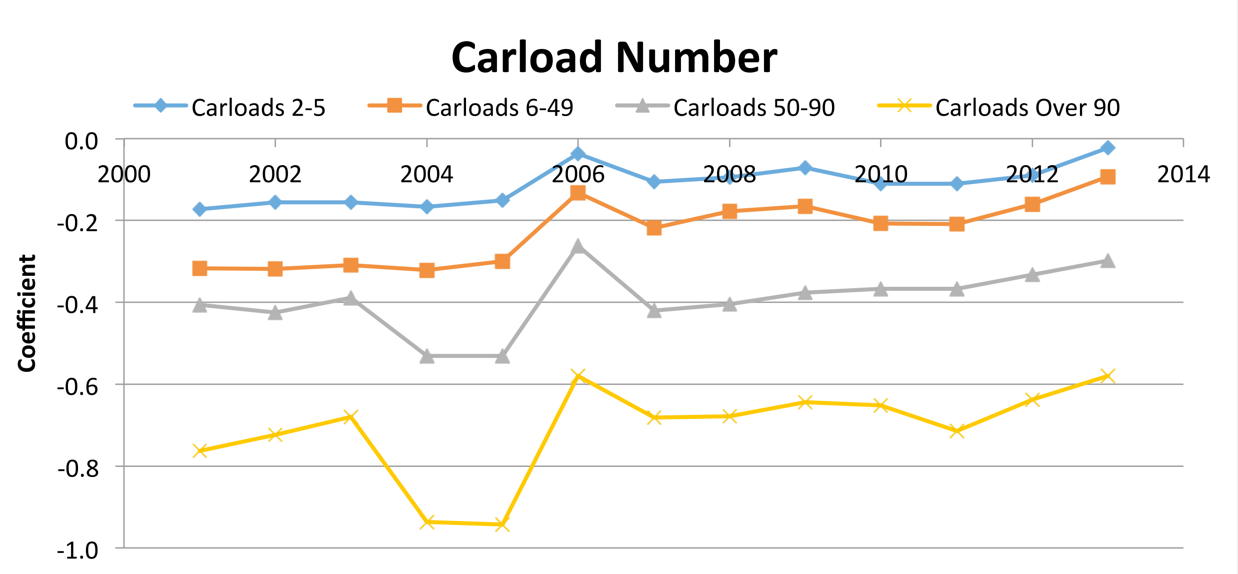


Figure 3-4: Coefficient values for carload number segments (values relative to one-carload shipments).

Figure 3-5 shows that RPCM is lower for longer distance shipments. The RPCM of shipments that traveled 250-800 miles was significantly lower than the RPCM of shipments that traveled less than 250 miles and shipments. Figure 3-5 shows that the decrease in RPCM for increasingly longer distance shipments continues all the way up to 2000+ miles. The coefficient values for these distance parameters are not only highly statistically significant, their magnitude in the real-world sense is also very high. The coefficient value for 2000+ mile shipments, relative to 20-250 mile shipments is approximately -1.4. This indicates that the RPCM for 2000+ mile shipments is 75% lower than the RPCM for 20-250 mile shipments.

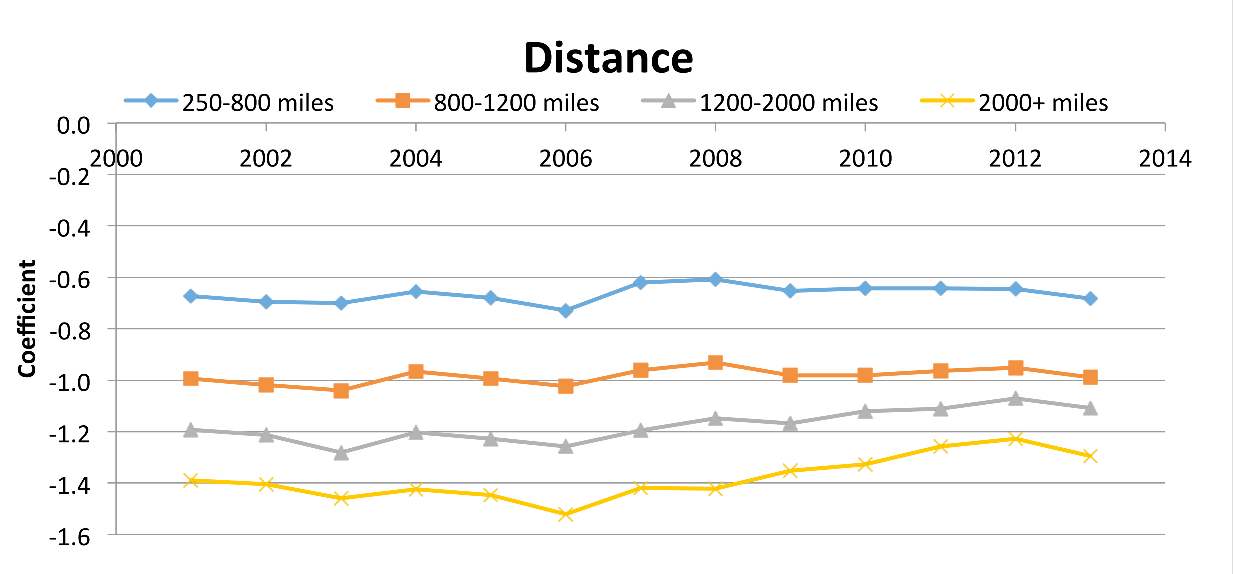


Figure 3-5: Coefficient values for shipment distance segments (values relative to 20-250 mile distance).

Figure 3-6 shows that the RPCM is lower for high density routes. The RPCM of very high density routes is significantly lower than the RPCM of lower density routes. The large discrepancy in RPCM across route density levels is likely due to the fact that carloads on these routes probably do not have to enter more than one classification yard. For example, a container shipment moving between Los Angeles and Chicago is going to pay a lower RPCM than a container shipment moving between Boston and Salt Lake City because, on any given day, there are huge volumes of carloads moving between Los Angeles and Chicago. Hence, a 100 or 120+ carload train can be formed in Los Angeles, and that train can go directly to Chicago without having to be disassembled and reassembled at classification yards. Whereas, there are significantly fewer containers moving from Boston to Salt Lake City; hence, a carload moving between these two cities will likely need to enter multiple classification yards, reducing efficiency and adding operation costs.

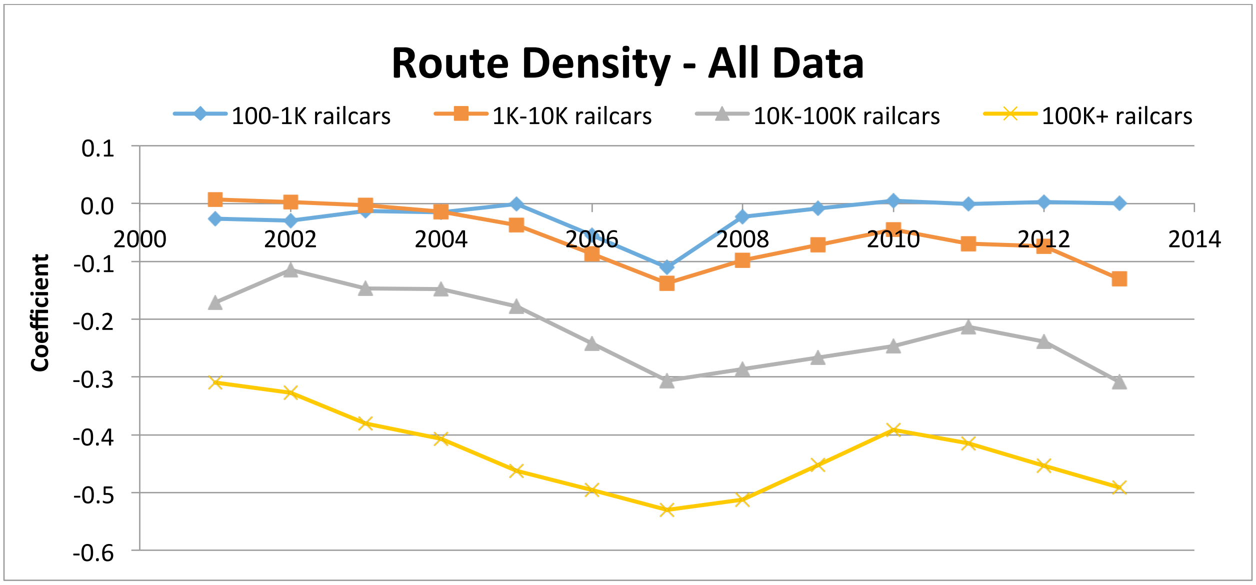


Figure 3-6: Coefficient values for shipment route density segments (values relative to 0 - 100 carloads).

Lastly, Figure 3-7 shows that, as expected, railroads charge higher a RPCM for railcars they own than privately-owned railcars. Figure 3-7 also shows that the RPCM of shipments bound for export was consistently lower than the RPCM of domestic shipments. The lower RPCM for export shipments is separate from the carload number, shipment distance, and route density effects described earlier in this section. It is interesting that the RPCM for export shipments is consistently lower than the RPCM for domestic shipments even after accounting for these other shipment characteristics.

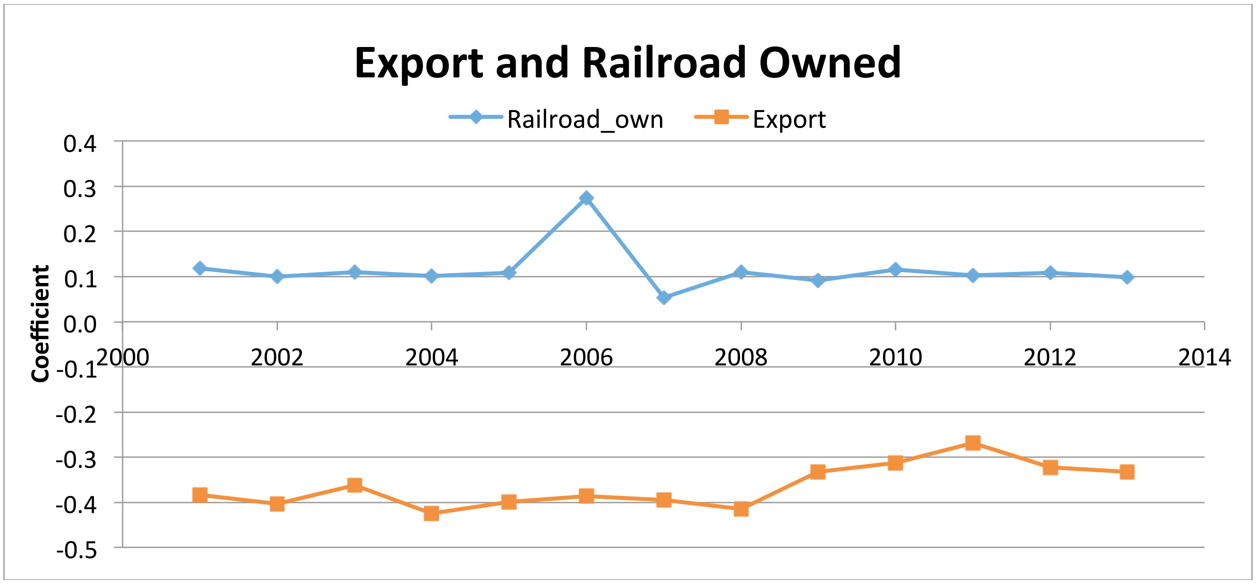


Figure 3-7: Coefficient values for the railroad ownership and export-bound dummy variables.

3.3 Segmentation by Commodity-type

In this section we examine grain and coal shipments exclusively. The results presented in this section mirror the results presented in Section 3.2 except rather than including all commodities, Section 3.3 only includes grain and coal shipments. Figure 3-8 illustrates that the RPCMs for grain and coal commodities increased between 2002 and 2013. It is important to note the fact that the coal line is higher than the grain line indicates that the RPCM for coal was higher in year X relative to the RPCM for coal in year 2001 than the RPCM for grain in year X relative to the RPCM for grain in 2001. Hence, Figure 3-8 shows that the RPCM for coal increased at a greater rate than the RPCM for grain between 2001 and 2013. The RPCM increases for grain and coal displayed in Figure 3-8 are very similar to the RPCM increases displayed in Figure 3-1 for all shipments. The RPCM for grain transport, like the RPCM for all commodities together, were relatively stable between 2001 and 2006. Additionally, the RPCMs for both grain and coal level-off between 2011 and 2013, similar to the results for all commodities displayed in Figure 3-1.

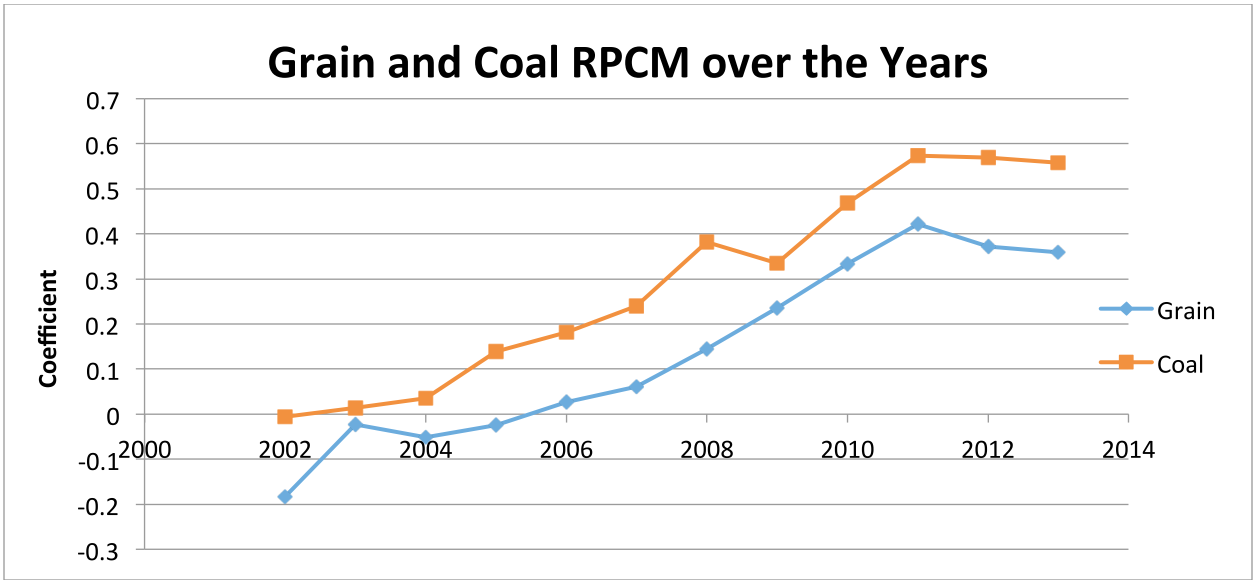


Figure 3-8: Coefficient values for the year variable (relative to the year 2001) for exclusively grain waybills and, separately, exclusively coal waybills.

The next four figures display the coefficient values for the origination and termination variables in the coal and grain segment models. Figure 3-9a and Figure 3-9b illustrate the effect of a shipment's origination location on the RPCMs of grain and coal, respectively. Figure 3-9c and Figure 3-9d illustrate the effect of a shipment's termination location on the RPCMs of grain and coal, respectively. Figure 3-9a shows that the RPCM for grain originating in the Upper Midwest was lower than the RPCM for grain not originating in the Upper Midwest, and specifically grain originating in the I-states. Figure 3-9a shows that Upper Midwest shippers never paid a premium for grain transport compared with non-Upper Midwest shippers between 2001 and 2013. Figure 3-9b shows that the RPCM of coal shipments originating in the Wyoming-Montana-Idaho region was consistently lower than the RPCM of coal shipment not originating in the Wyoming-Montana-Idaho region. This result is not surprising given that the Powder River Basin is located in Wyoming and 40% of U.S. coal comes from the Powder River Basin. Figure 3-9b also shows that relative to other regions, the RPCM for coal transport in the I-states and the Northeast region of the country steadily declined between 2001 and 2013.

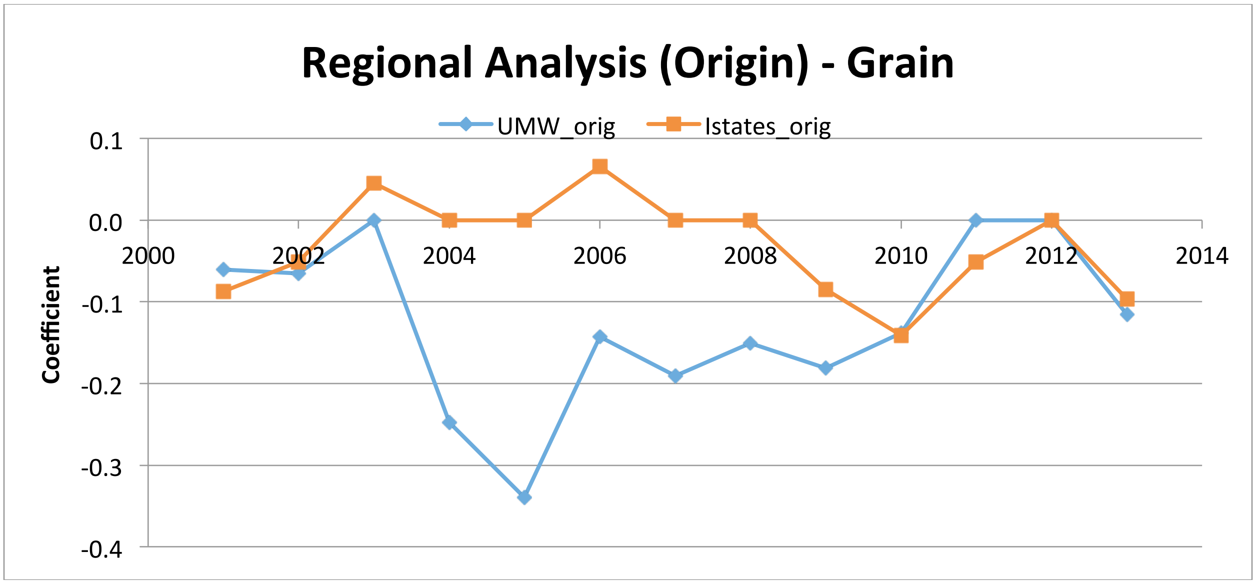


Figure 3-9a: Coefficient values for the origination region dummy variables – grain waybills only.

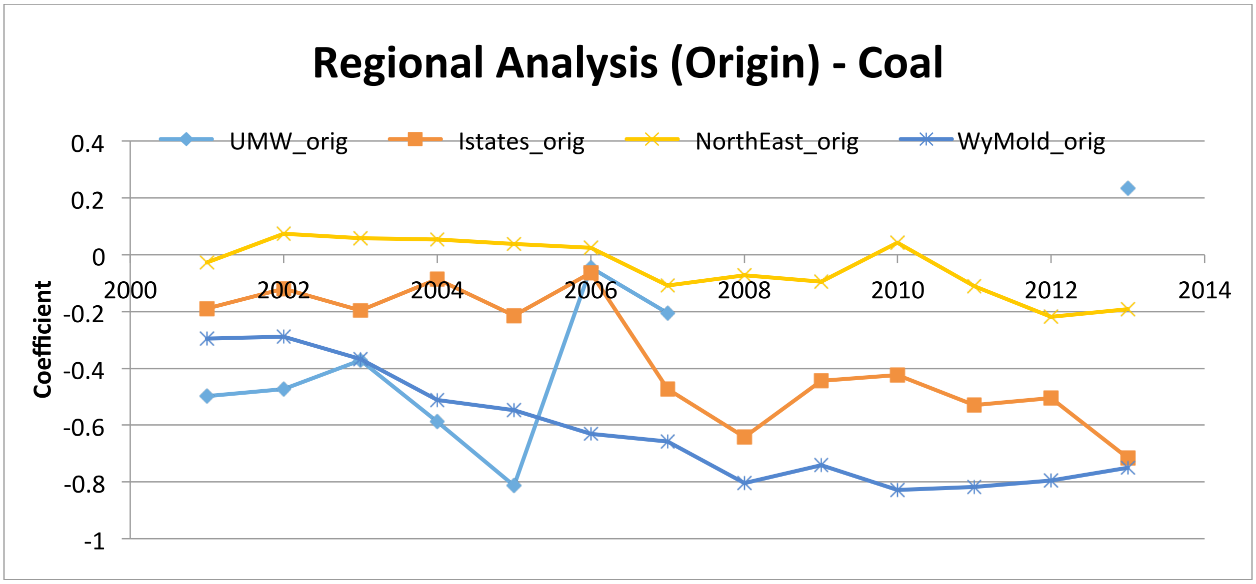


Figure 3-9b: Coefficient values for the origination region dummy variables – coal waybills only.

The results presented in Figure 3-9c, relating to the termination point of grain, are especially interesting. Figure 3-9c shows that in the early 2000s the RPCM of grain shipments terminating in Louisiana was lower than the RPCM for grain shipments not terminating in Louisiana; in contrast, in the early 2000s the RPCM for grain shipments terminating in Texas was significantly higher than the RPCM for grain shipments not terminating in Texas. However, between 2001 and 2010 the RPCM for Louisiana termination points began to increase steadily, whereas the RPCM for Texas termination points began to decrease steadily, and by 2010 and through 2013 the RPCM for grain shipments terminating in Louisiana and Texas were very similar. During this period between 2001 and 2013, the RPCM for grain terminating in the Pacific Northwest fluctuated considerably; however between 2010 and 2013 the RPCM for grain terminating in the Pacific Northwest was comparable to the RPCM for grain terminating in Texas and Louisiana.

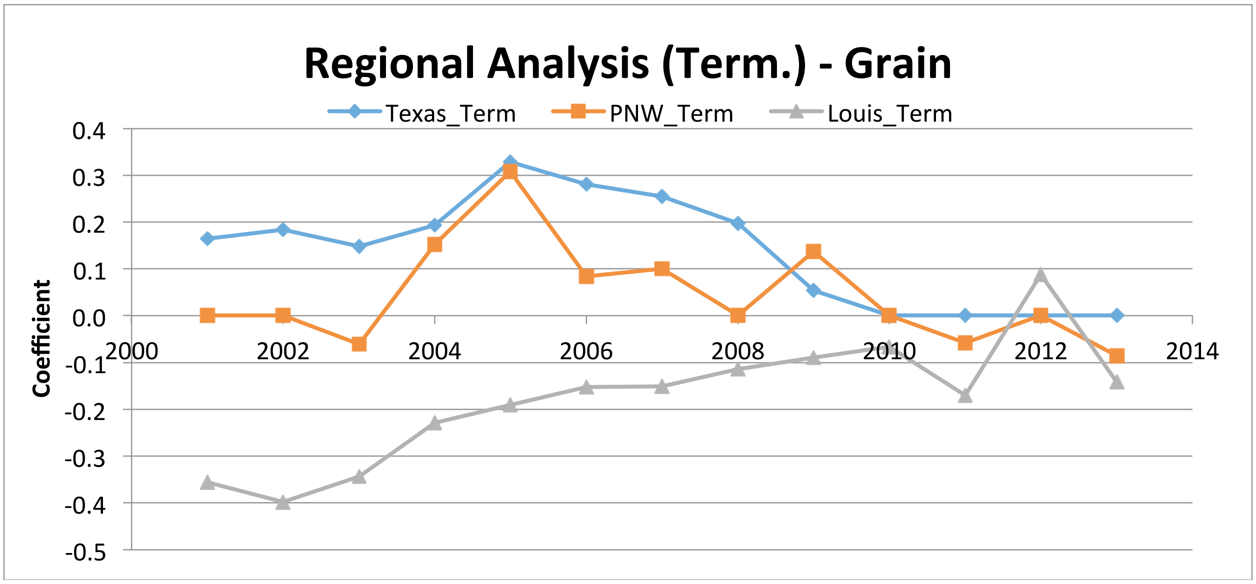


Figure 3-9c: Coefficient values for the termination region dummy variables – grain waybills only.

Figure 3-9d illustrates that the RPCM of coal terminating in Texas went from being significantly lower than the RPCM of coal not terminating in Texas in the early 2000s, to being significantly higher in the early 2010s. Comparing the Texas termination coefficient values in Figure 3-9c and Figure 3-9d, we see that in the early 2000s the coefficient for grain was positive and the coefficient for coal was negative; however, by the end of the 2000s and between 2011 and 2013, the Texas termination coefficient for grain is negative and the coal coefficient in positive.

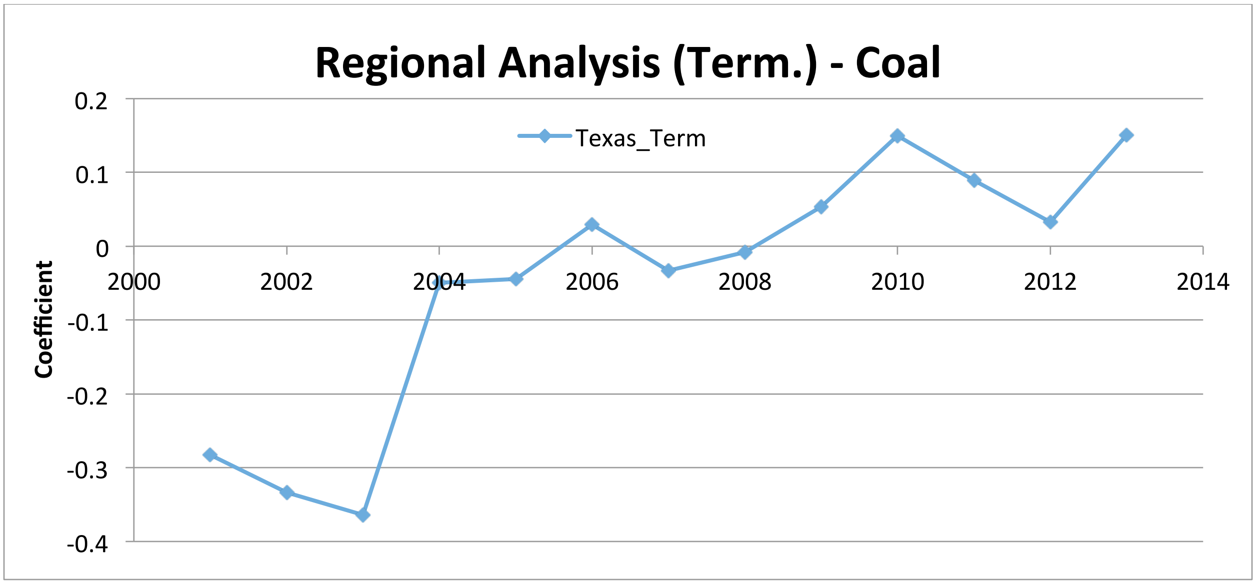


Figure 3-9d: Coefficient values for the termination region dummy variables – coal waybills only.

Figure 3-10a and Figure 3-10b display the coefficient values for the carload number variable for the grain and coal models, respectively. Figure 3-10a shows the same general results as Figure 3-4, which displays the carload number coefficient values for all commodities together. Interestingly, the coefficient for larger grain shipment sizes decreased steadily (indicating a lower RPCM relative to one-carload shipments) between 2006 and 2011 before quickly receeding back to pre-2006 levels in 2012 and 2013. Additionally, in Figure 3-10a, between 2009 and 2012 the RPCMs of 6-49 carloads and 50-90 carloads were comparable to the RPCM of shipments with greater than 90 carloads. These results for grain shipments in Figure 3-10a contradict Figure 3-4, which showed that adding more carloads to a shipment, up to and including 90+ carloads, resulted in a lower RPCM.

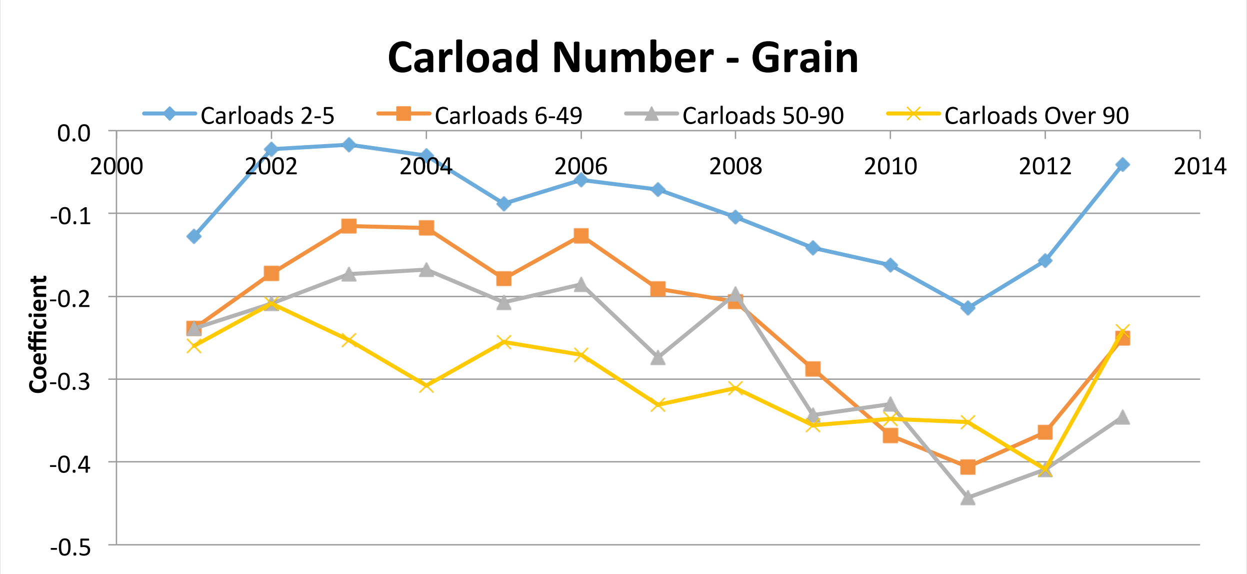


Figure 3-10a: Coefficient values for carload number segments (values relative to one-carload shipments) – grain waybills only.

Figure 3-10b heavily contradicts the principles of economies of scale and Figure 3-4 and Figure 3-10a. Figure 3-10b shows that the RPCM for coal shipments with 6-49 carloads was higher than the RPCM for one-carload shipments at least three years between 2001 and 2012. However, the RPCMs for 90+ railcar shipments and 50-90 railcar shipments were lower than RPCM for one-carload shipments. One possible reason for the contradictory results presented in Figure 3-10b is that railroads misreported the number of carloads traveling in a shipment. In fact, it was brought to the NUTC research team’s attention that some railroads create separate waybills for every single carload, even if those carloads traveled together on a unit or shuttle train from origin to termination point. Misreporting data can cause serious modeling issues.

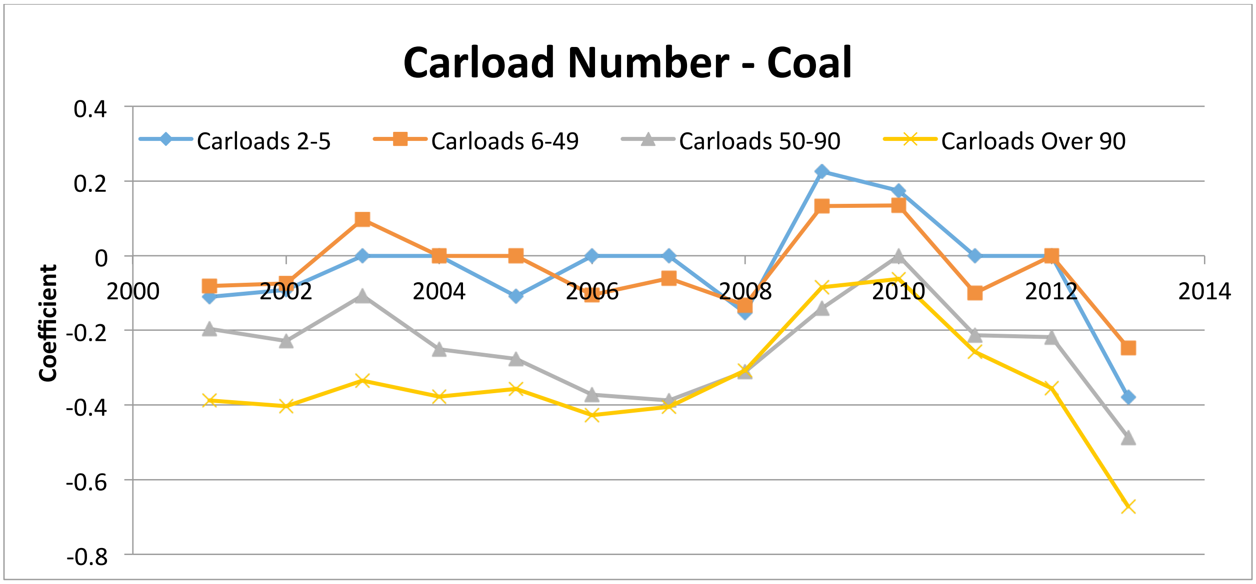


Figure 3-10b: Coefficient values for carload number segments (values relative to one-carload shipments) – coal waybills only.

The distance variable coefficient values for grain and coal shipments were consistent with the distance coefficient values for all commodity-types displayed in Figure 3-5. Hence, the graphs for coal and grain shipments are not displayed in this section; the charts are available in Exhibit G: Shipment Distance Coefficient Value Results of the Appendix. Similarly, the coefficient values for the railroad ownership dummy variable for the grain and coal models were also consistently significant and positive between 2001 and 2013 mirroring the results in Figure 3-7 for all commodity-types; hence, the results for grain and coal are in Exhibit H: Railcar Ownership and Export/Domestic Coefficient Value Results of the Appendix.

Interestingly, the route density results for coal and grain shipments for most years were not significantly different from zero. The reason why route density had a significant effect on all commodities but had no significant effect for grain and coal is unknown.

Figure 3-11 displays the coefficient values for the export dummy variable for coal and grain shipments between 2001 and 2013. In the analysis of all commodity-types (see Figure 3-7), the RPCM of export shipments was consistently lower than the RPCM for domestic shipments between 2001 and 2013. In contrast, the RPCMs for coal and grain shipments bound for export were significantly higher than the RPCMs for domestic shipments in 2012 and 2013. The RPCM for export grain shipments went from being significantly lower than the RPCM for domestic grain shipments in 2001 to being significantly higher in 2013. The coefficient value for the dummy export variable steadily increased for grain shipments between 2001 and 2013. This result is important for grain shippers searching for potential markets to sell their grain; it appears that after at least a decade of lower RPCM for export shipments, grain shippers might now have to pay a higher cost per carload-mile for export shipments.

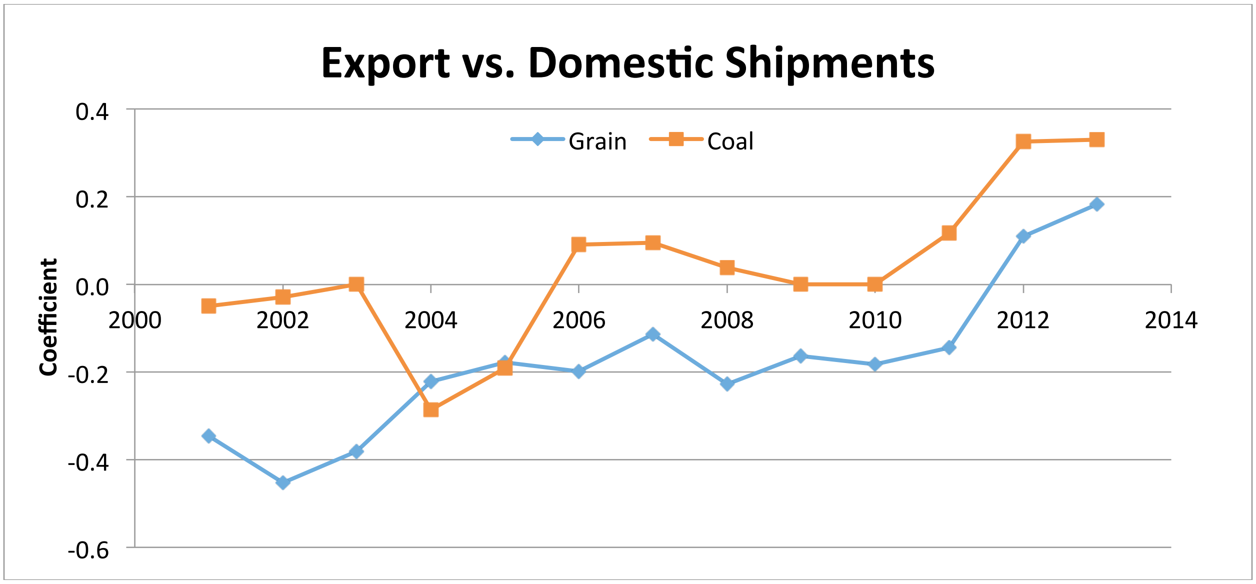


Figure 3-11: Coefficient values for the export dummy variable for exclusively grain waybills and, separately, exclusively coal waybills.

3.4 Segmentation by Origination Region

This section examines RPCM for shipments originating in the Upper Midwest and the I-states. The focus of the analysis is on shipments originating in the Upper Midwest but the I-states provide a useful region for comparison. Figure 3-12 illustrates that similar to the analysis of all shipments and grain and coal commodities, RPCM increased between 2001 and 2013 for shipments originating in the Upper Midwest and the I-states. However, the increase in average RPCM for the Upper Midwest mainly occurred between 2008 and 2013; whereas, average RPCM increased mainly between 2006 and 2012 for all shipments (see Section 3.1) as well as grain and coal shipments (see Section 3.3).

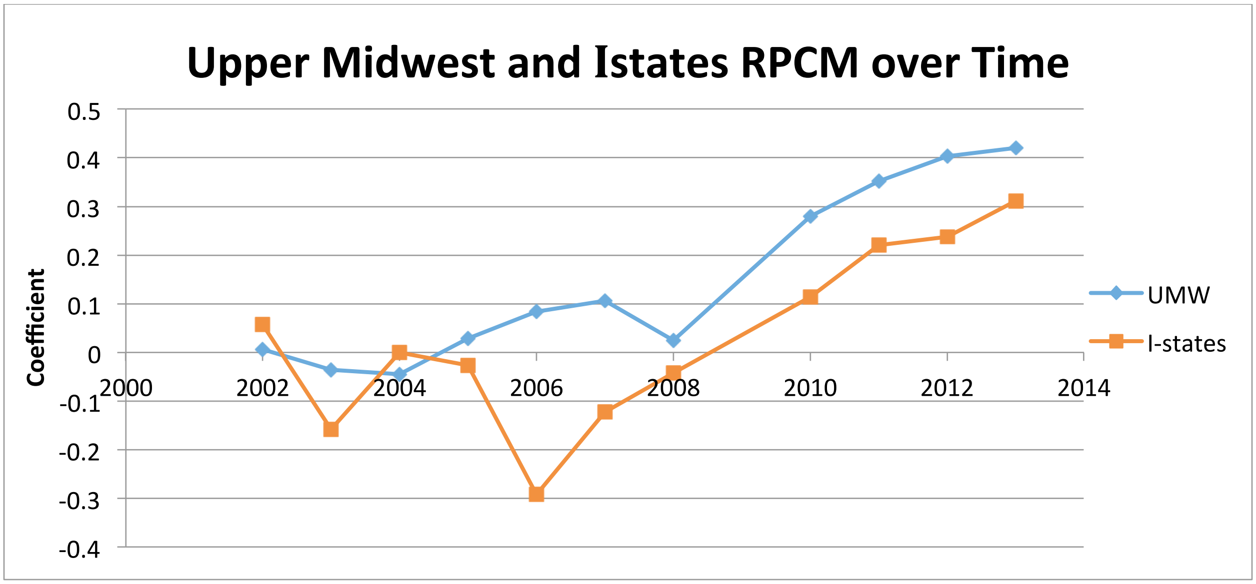


Figure 3-12: Coefficient values for the year variable (relative to the year 2001) for exclusively waybills originating in the Upper Midwest and, separately, exclusively waybills originating in the I-states.

Figure 3-13a and Figure 3-13b display the coefficient values for the termination region variables in the regression model for shipments originating in the Upper Midwest and the I-states, respectively. Figure 3-13a displays high variability in the coefficient values for each of the termination points. One useful result is that after 2003 the coefficient values for shipments originating in the Upper Midwest and terminating in either the Pacific Northwest or Louisiana were never greater than zero. In most years, aside from 2004 and 2012, the Texas termination coefficient was significantly greater than zero. The coefficient values for the I-state dataset, presented in Figure 3-13b, are considerably more stable between 2001 and 2013 than those in Figure 3-13a. The coefficient values for shipments originating in the I-states and terminating in Texas were positive and statistically significant every year between 2001 and 2013. In every year between 2001 and 2012 shipments originating in the I-states and terminating in Louisiana or the Pacific Northwest received a small but statistically significantly lower RPCM. However, in 2013 the RPCM for shipments terminating in the Pacific Northwest was greater than the RPCM for shipments not terminating in the Pacific Northwest. Despite the variability in Figure 3-13a, the results are quite similar to those in Figure 3-13b. The similarity in results is not surprising given that the Upper Midwest and the I-states are geographically close.

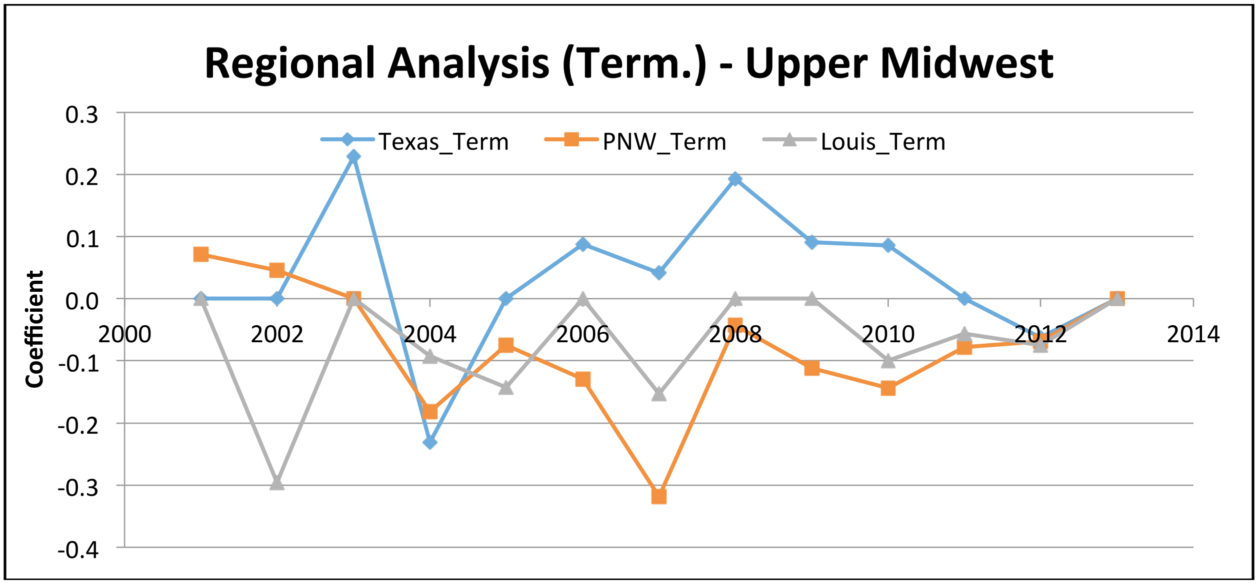


Figure 3-13a: Coefficient values for the termination region dummy variables – Upper Midwest waybills only.

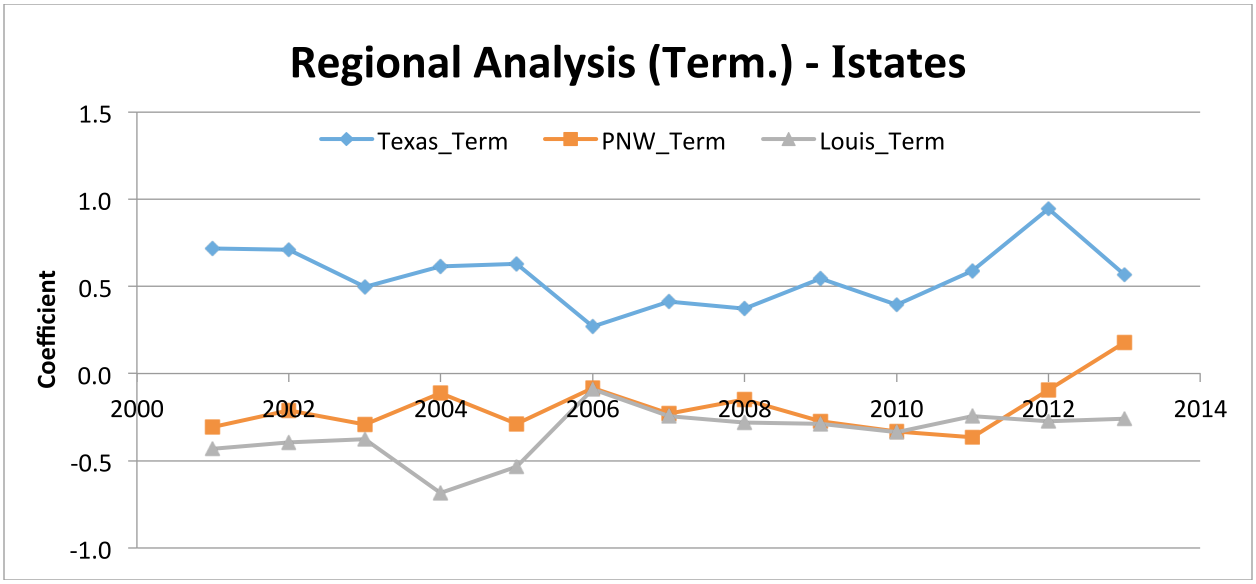


Figure 3-13b: Coefficient values for the termination region dummy variables – I-state waybills only.

Figure 3-14a through Figure 3-14d display the coefficient values for the commodity variables for shipments originating in both the Upper Midwest and the I-states. Figure 3-14a illustrates that the coefficient for grain in the Upper Midwest segment of the data was negative in 2004 through 2006, 2009, 2012 and 2013; positive in 2001 and 2011; and not significant in 2002, 2003, 2007, 2008 and 2010. An interesting result is that prior to 2008 no data for crude oil and natural gas in the Upper Midwest exists and after 2007 no data for coal shipments exists in the Upper Midwest. The RPCMs for coal and petroleum shipments originating in the Upper Midwest were significantly greater than the RPCM of all other commodities between 2001 and 2013. In the I-states, the grain and coal coefficients were negative and statistically significant every year between 2001 and 2013; conversely the coefficient for coal and petroleum products was most often positive between 2001 and 2013.

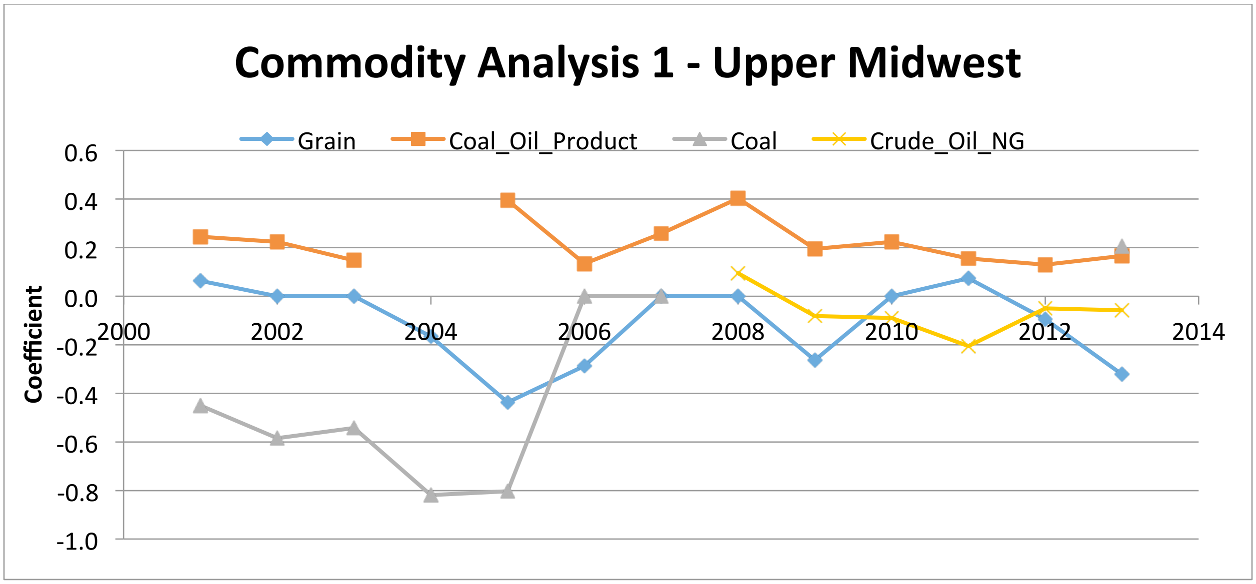


Figure 3-14a: Coefficient values for commodity variables (values are relative to 'other' commodity which includes all the commodities not explicitly enumerated in the commodity analysis) – Upper Midwest waybills only. Comparison of grain RPCM with fossil fuel RPCM.

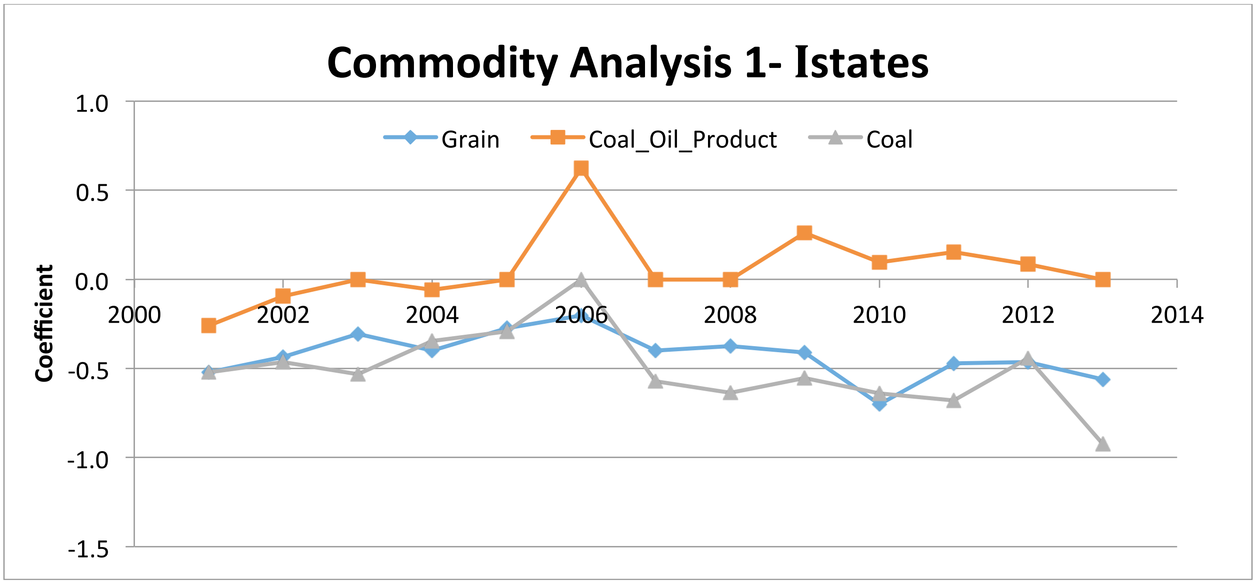


Figure 3-14b: Coefficient values for commodity variables – I-state waybills only. Comparison of grain RPCM with fossil fuel RPCM.

Figure 3-14c and Figure 3-14d illustrate the same results for grain commodities as Figure 3-14a and Figure 3-14b but rather than displaying the results next to fossil fuel commodities, the results are displayed next to non-grain agricultural products, food products, chemicals, and pulp-paper. Figure 3-14c shows that the RPCMs for non-grain agricultural and food product shipments were higher than the RPCM for all shipments between 2001 and 2004; however, after 2004, the RPCMs for non-grain agricultural and food products continued to decrease until by 2010 the RPCMs for non-grain agricultural and food product shipments were significantly lower than the RPCM for all other commodity-types. Every one of the commodity RPCMs in Figure 3-14c decreased between 2001 and 2013 relative to 'all other' commodities. Figure 3-14d shows considerably less variability; the RPCM for chemicals was consistently higher than food and non-grain agricultural products, which are both consistently higher than the RPCM for grain shipments.

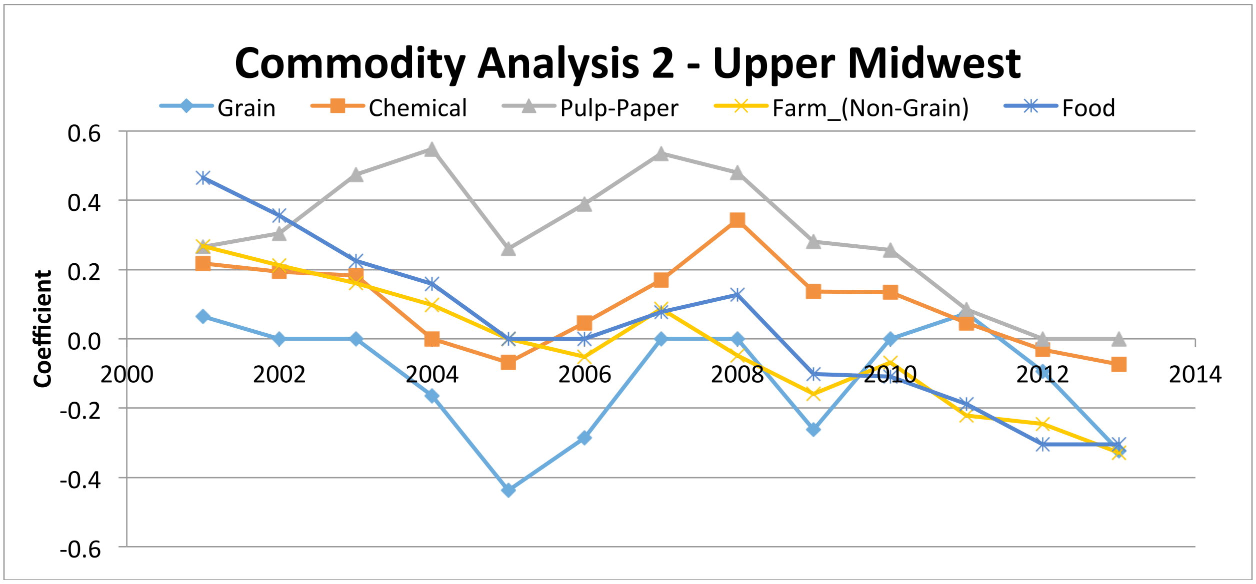


Figure 3-14c: Coefficient values for commodity variables – Upper Midwest waybills only. Comparison of grain RPCM with similar commodities.

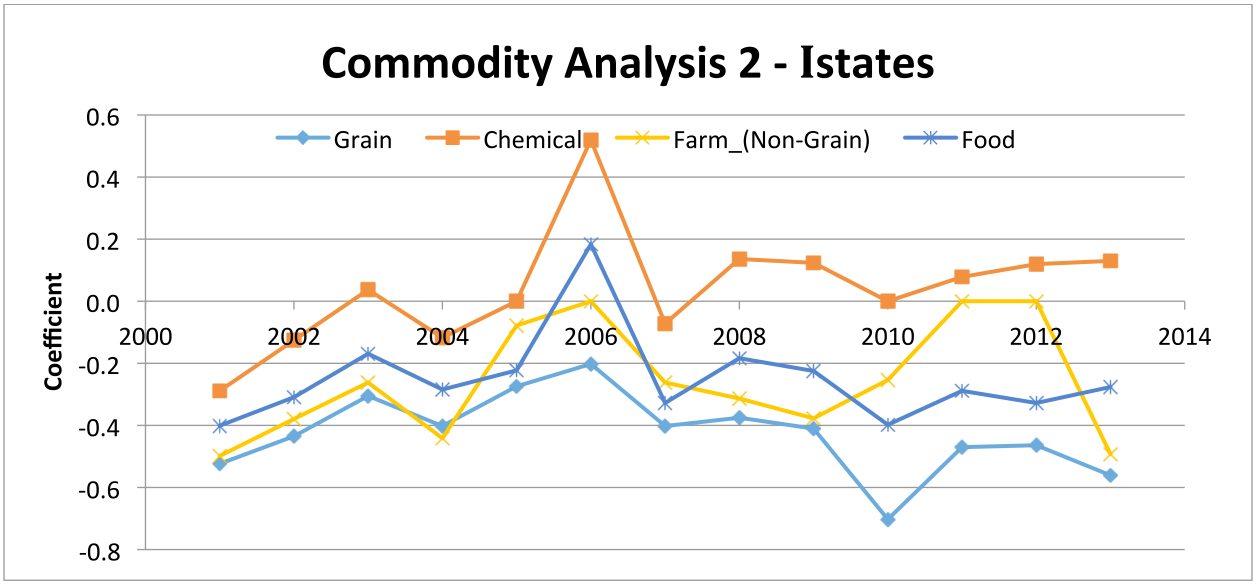


Figure 3-14d: Coefficient values for commodity variables – I-state waybills only. Comparison of grain RPCM with similar commodities.

Figure 3-15a and Figure 3-15b display the coefficient values for the carload number variable for shipments originating in the Upper Midwest and the I-states respectively. In both graphs, the RPCM for 90+ carload shipments was significantly lower than one-carload shipments; however, in 2007 and 2009, the RPCM for 90+ carload shipments in the Upper Midwest was noticeably closer to the RPCM of one-carload shipments than all other years between 2001 and 2013. Additionally, both graphs show that the RPCMs for 2-5 and 6-49 carload shipments were significantly lower than the RPCM for one-carload shipments between 2001 and 2003; however, between 2004 and 2013 the RPCM of one-carload shipments is not significantly different from the RPCMs of 2-5 carload and 6-49 carload shipments.

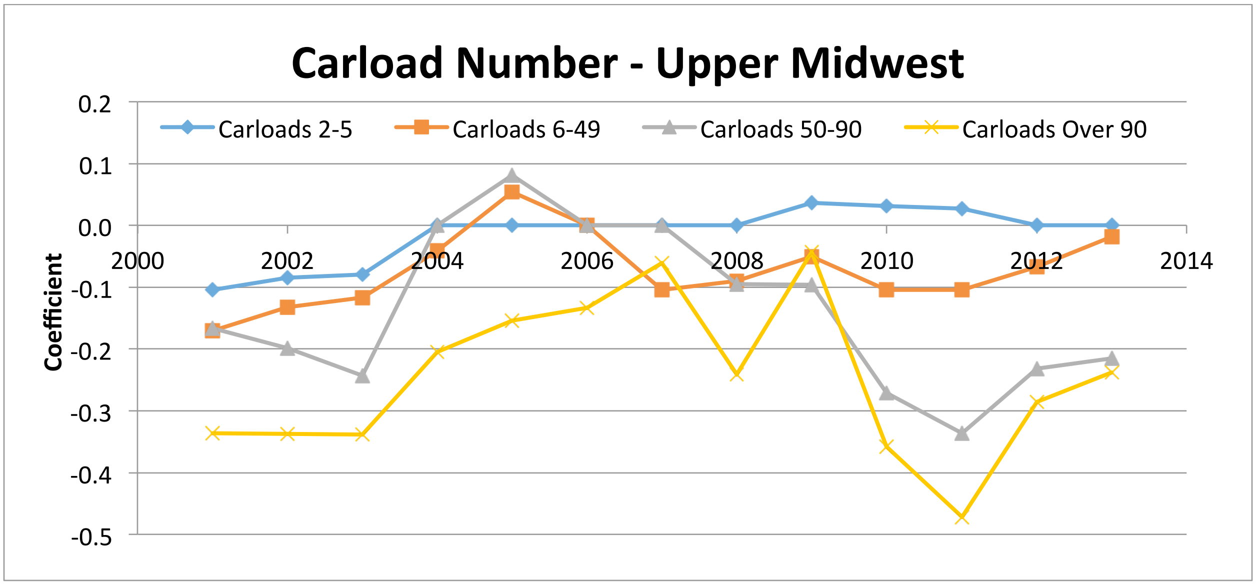


Figure 3-15a: Coefficient values for carload number segments (values relative to one-carload shipments) – Upper Midwest waybills only.

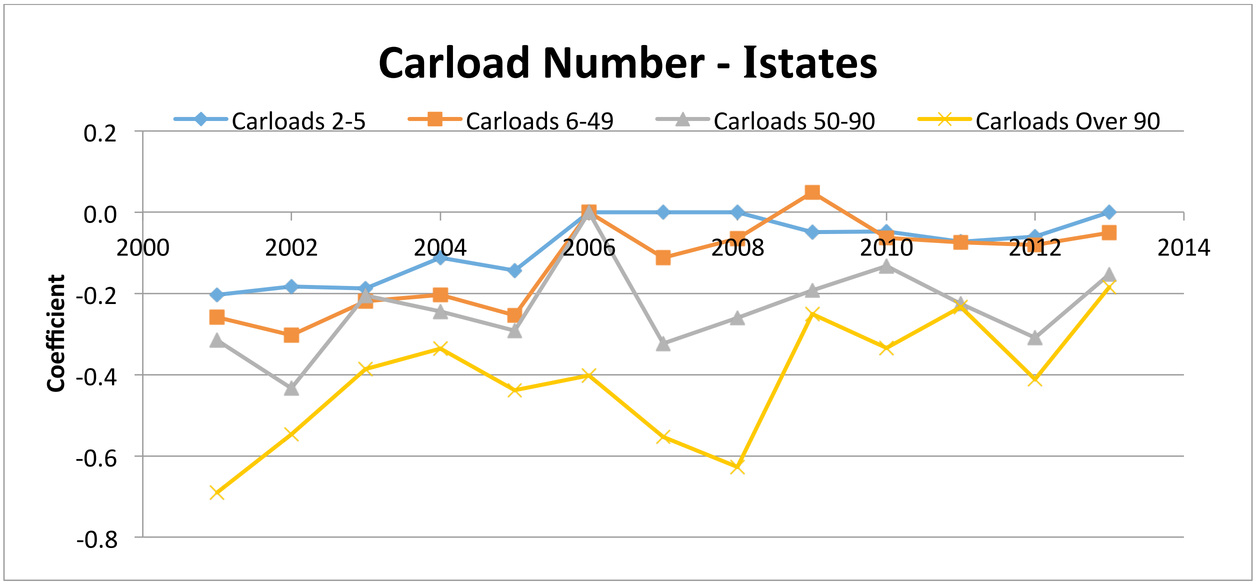


Figure 3-15b: Coefficient values for carload number segments (values relative one-carload shipments) – I-state waybills only.

Once again, the graphs displaying the coefficient values for the distance variable for shipments originating in the Upper Midwest and the I-states are in Exhibit G: Shipment Distance Coefficient Value Results of the Appendix because the results are very similar to the coefficient values for all shipments (see Figure 3-5). In addition, the graphs of the coefficient values for the export variable and the railroad ownership variable are in Exhibit H: Railcar Ownership and Export/Domestic Coefficient Value Results of the Appendix because the results were similar to the results for all shipments (see Figure 3-7); i.e. the coefficients for export were negative and the coefficients for railroad-owned were positive in both the Upper Midwest and the I-states consistently between 2001 and 2013.

Figure 3-16 illustrates the coefficient values for the route density variable for shipments originating in the Upper Midwest. None of the route density levels were statistically different from the low route density level for shipments originating in the I-states. Figure 3-16 shows that the RPCM for Upper Midwest shipments traveling on very high density routes as significantly higher than low density routes between 2001 and 2003. Conversely the figure shows that the RPCM for the high density route category was significantly lower than the RPCMs for the low route density route categories between 2004 and 2013.

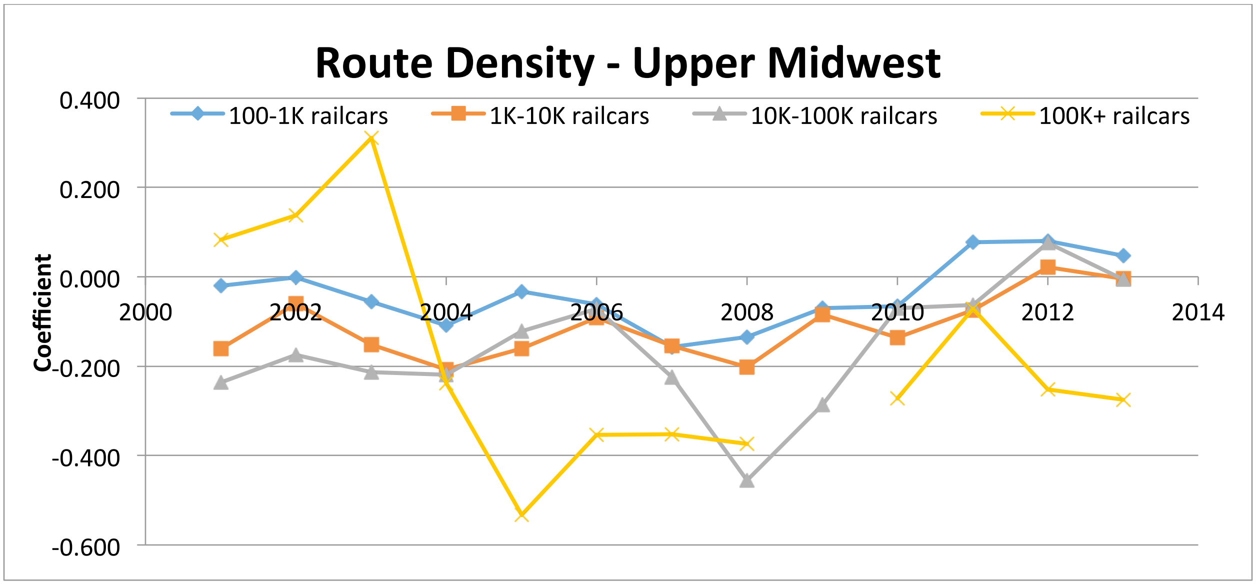


Figure 3-16: Coefficient values for shipment route density segments (values relative to 0-100 carload) – Upper Midwest waybills only.

3.5 Analysis of Grain RPCM in the Upper Midwest

In this final multivariate regression modeling section, we exclusively examine grain shipments originating in the Upper Midwest. Interestingly, Figure 3-17 shows that not until 2007 did the grain RPCM in the Upper Midwest exceed the grain RPCM in 2001. Comparing Figure 3-17 with Figure 3-1, Figure 3-8, and Figure 3-12 (graphs depicting increasing RPCM for different regional and commodity segments), although it will be little consolation to grain shippers in the Upper Midwest, the increase in RPCM they experienced between 2006 and 2013 was not unique, almost all regions of the country and shippers of all commodity-types experienced RPCM increases during this time period.

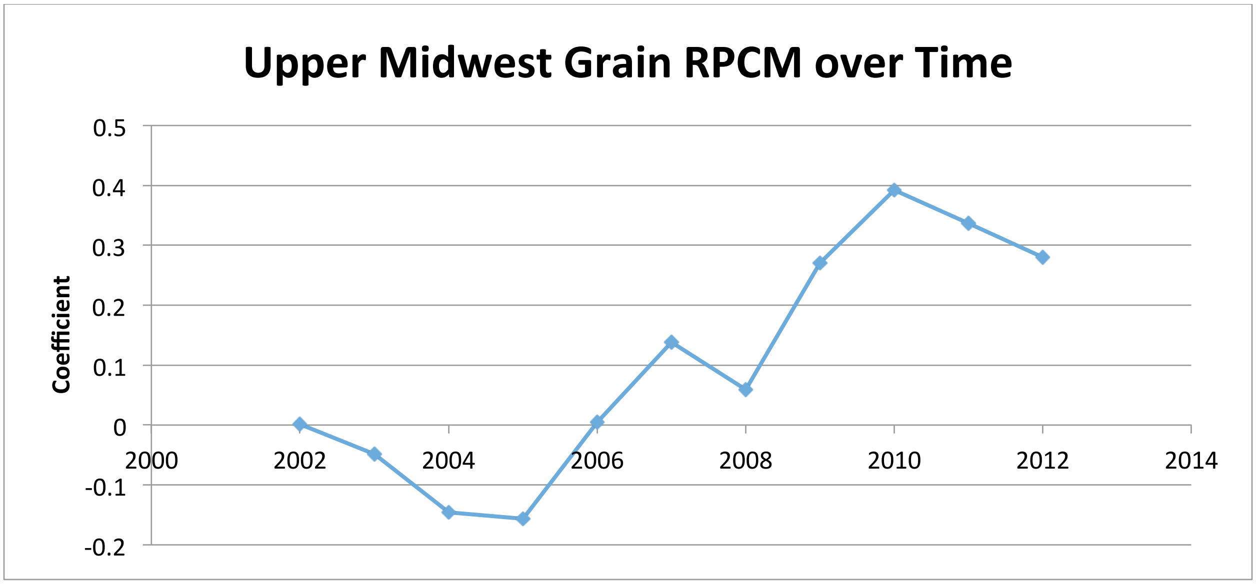


Figure 3-17: Coefficient values for the year variable (relative to the year 2001) for exclusively grain waybills originating in the Upper Midwest.

Figure 3-18 presents the coefficient values for the termination point variables in the Upper Midwest grain regression model. The figure shows that between 2001 and 2012 the RPCM of shipments terminating in the Pacific Northwest decreased steadily. However, in 2013 the RPCM for shipments terminating in the Pacific Northwest was not statistically significant. Figure 3-18 also shows that the RPCMs for shipments terminating in either Texas or Louisiana were either the same or higher than the RPCM for all termination points between 2001 and 2013.

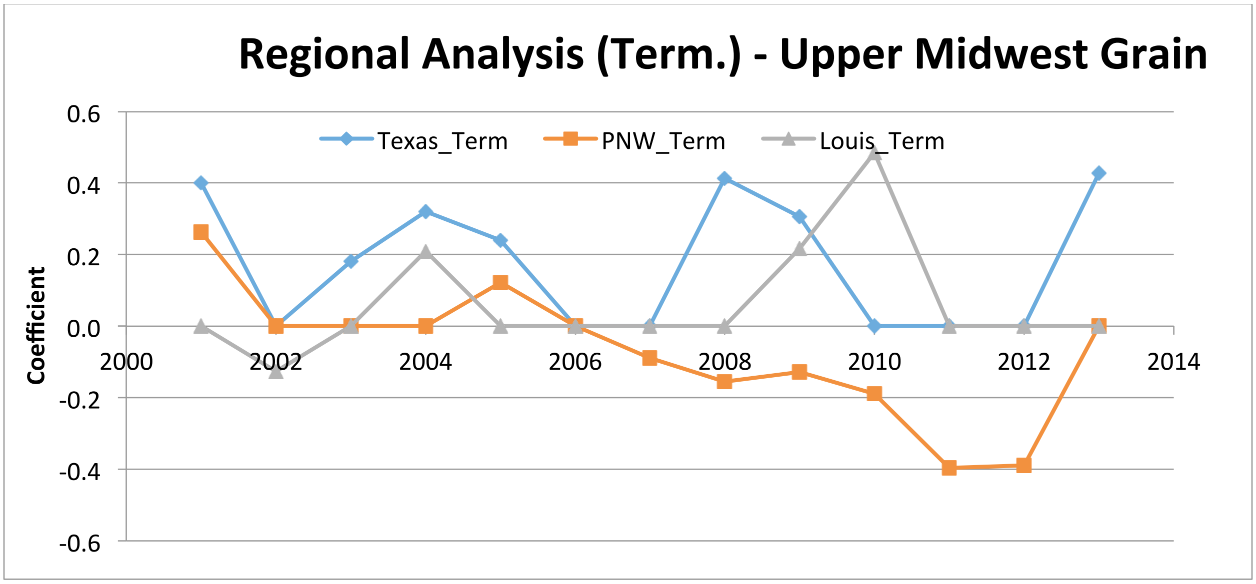


Figure 3-18: Coefficient values for the termination region dummy variables – Upper Midwest grain waybills only.

Figure 3-19 displays the carload number coefficient values for grain shipments originating in the Upper Midwest. Between 2006 and 2008, none of the coefficients were significantly different from one-carload shipments. In the other years, the coefficient values for the different carload segments do not form a clear trend. The results in Figure 3-19 strongly contradict previous results relating to carload number's relationship with RPCM as well as economies of scale in transportation.

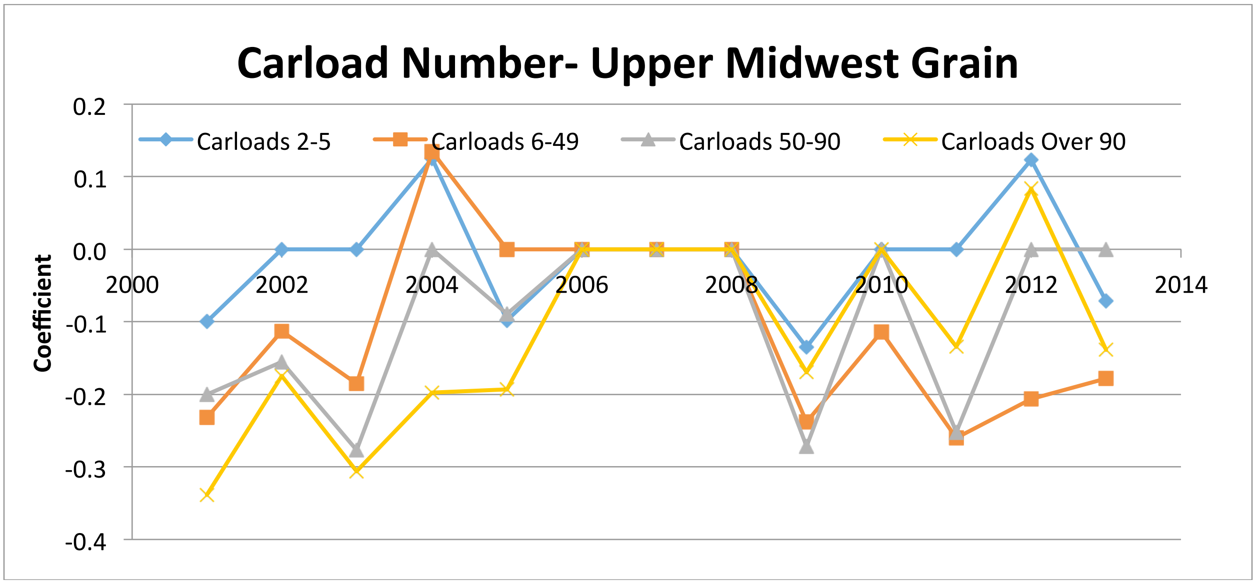


Figure 3-19: Coefficient values for carload number segments (values relative to one-carload shipments) – Upper Midwest grain waybills only.

Figure 3-20 shows the coefficient values for the shipment distance variable. The results in Figure 3-20 are significantly different from those in the model with all shipments (Figure 3-5) and the other segments. The coefficient values for the distance segments varied considerably between 2001 and 2013. Additionally, it does not appear that the RPCM of the longest shipments (2000+ miles) was significantly lower than the RPCM of shipments that traveled 1200-2000 miles or 800-1200 miles. This result is surprising given the operational advantages of shipping products over 2000 miles compared with only 800-1200 miles. It is possible that with more data the results in Figure 3-20 would be more consistent with the other distance coefficient results obtained in the model with all shipments.

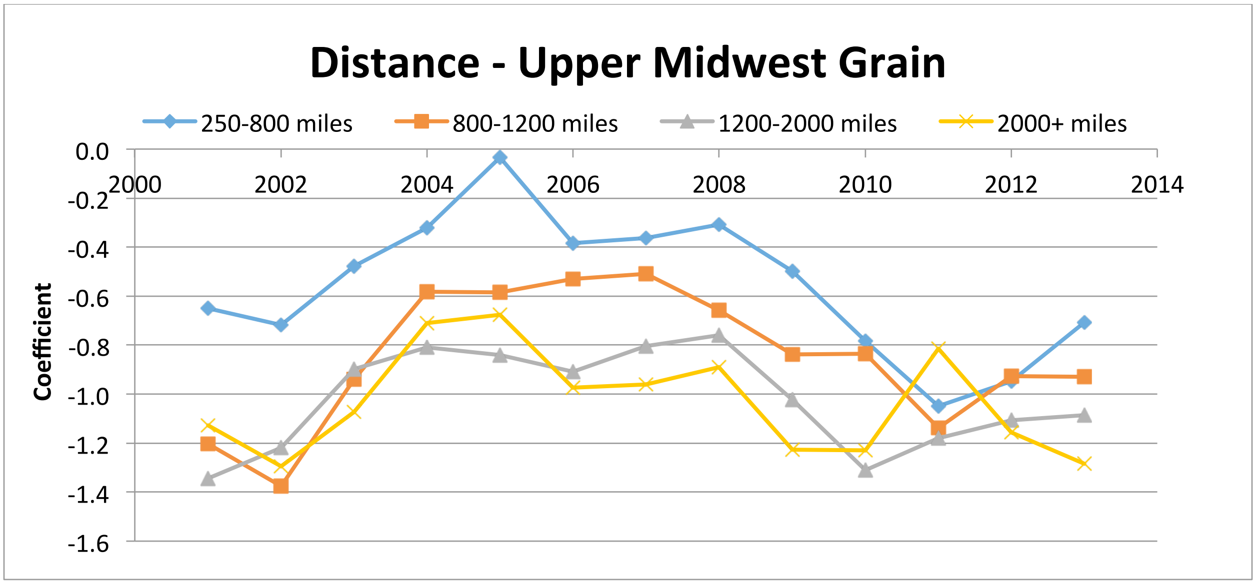


Figure 3-20: Coefficient values for shipment distance segments (values relative to 20-250 mile shipments) – Upper Midwest grain waybills only.

Figure 3-21 displays the coefficient values for the railroad ownership and export dummy variables. The results indicate that once again the RPCM for privately-owned railcars was lower than the RPCM for shipments transported with railroad-owned railcars. Figure 3-21 also shows that the RPCM for grain shipments originating in the Upper Midwest bound for export was either lower than the RPCM for domestic shipments or the RPCMs for the two segments were not statistically different. This result is especially interesting given the fact that we account for not only the distance and carload number, but we also have dummy variables for three of the most common grain export destinations: Texas, Louisiana and especially the Pacific Northwest.

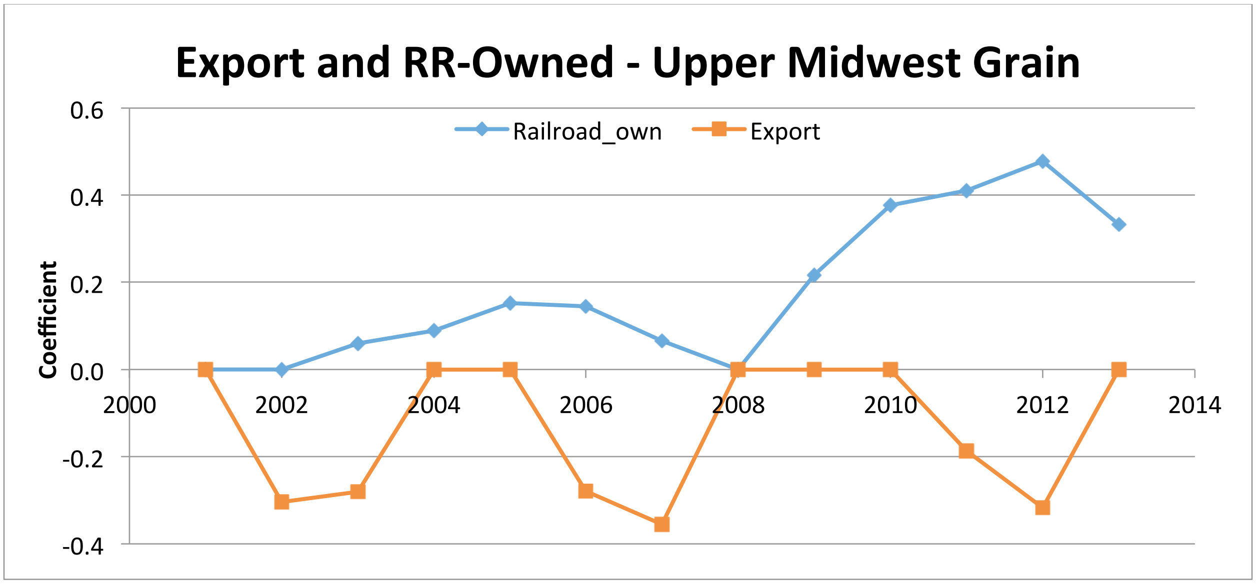


Figure 3-21: Coefficient values for the railroad ownership and export-bound dummy variables – Upper Midwest grain waybills only.

4 Conclusions and Future Research

This report set out to examine trends in rail transport rates and the shipment characteristics that impact rail transport rates in a robust statistical sense. We used the STB's CWS, which is a 1% stratified sample of all rail waybills originated by major carriers in the United States, to conduct the rate analysis. We conducted an in-depth exploratory analysis of the waybill data to determine correlations between different variables and analyze general trends in rail transport rates and other important variables, in order to inform our econometric models. A series of multivariate regression models were developed to answer the following questions:

- *Has average RPCM increased or decreased over the past 13 years?*
- *How has this increase or decrease varied by commodity-type (i.e. grain versus coal, versus all other commodities) and origination region?*
- *What shipment characteristics and other exogenous factors are associated with higher or lower RPCM?*
- *Do the shipment characteristics and other exogenous factors associated with higher or lower RPCM vary across time, commodity-type, and/or origination region?*

The results presented in the exploratory analysis section, suggest that RPTM is negatively correlated with distance, shipment weight, carload number, route density, car capacity, export shipments, and the number of interchanges. Additionally, higher route density was correlated with longer distance shipments and fewer railroad interchanges. Coal shipments were strongly correlated with carload number and shipment weight indicating that most coal is shipped in unit or shuttle trains. An analysis of average RPTM for grain and coal found that RPTM was lowest for coal but RPTM was also quite low for grain compared to all shipments. Interestingly, average RPTM was lower for all export shipments than all shipments. Lastly, the average RPTM for export coal shipments was significantly higher than the RPTM for export grain shipments.

The econometric model results indicate that average RPCM increased significantly in real terms between 2001 and 2013 with most of the increase occurring between 2006 and 2012. Further analysis showed that although average RPCM increased in the Upper Midwest and for grain shipments; the increase was consistent with a general increase in average RPCM for all shipments during the period from 2001 to 2013. In fact, the regression model results show that the RPCM for grain shippers was lower than the RPCM for many other commodity-types. The econometric regression models determined the aforementioned results while simultaneously taking into account and measuring the impact of shipment characteristics. Many of the shipment characteristics associated with lower RPCM in the econometric regression modeling analysis (long distance shipments, shipments on high-density routes, large shipments, and railroad-owned railcars) conform to previous findings in the literature. In addition, the econometric analysis determined that the RPCM for export shipments was consistently lower than the RPCM for non-export shipments between 2001 and 2013 for all commodity-types. However, in 2012 and 2013 the RPCM for export grain shipments was higher than the RPCM for non-export grain shipments. Another interesting result obtained in the analysis is that, in the Upper Midwest, between 2001 and 2003 the RPCM for shipments traveling on very high density routes was higher than the RPCM of shipments in lower route density categories; conversely, between 2004 and 2013, the RPCM for the very high route density category was significantly less than the RPCMs of the other route density categories. The methodology presented in the report provides an excellent means of determining trends in not only overall rail transport rates but also the shipment characteristics and exogenous factors that impact rates in different segments of the data.

The NUTC research group identified two potential areas for future research. The first, and most obvious path, requires the STB's confidential waybill sample. The confidential waybill data includes a number of fields that would improve the regression model developed in Section 2, including, fuel surcharge, railroad variable cost, and better and finer geographical information. The second opportunity to improve the regression model involves combining outside data sources with the CWS. For example other researchers examined how potential freight transport competition influences rail transport rates[25,26,27]. They combined the CWS with information related to the nearest navigable inland waterway from the shipment's origin and termination points and the number of competing railroads that could potentially serve the demand. Other data sources that could be integrated with the CWS include fuel prices, and aggregate rail indices.

---

[25] **Reference:** Laurits, R. "An update to the study of competition in the U.S. freight railroad industry: Final report. Christensen Associates, Inc. A report prepared for The Surface Transportation Board (STB)." (2010).

[26] **Reference:** Mac Donald, James M. "Railroad deregulation, innovation, and competition: Effects of the Staggers Act on grain transportation." *Journal of Law and Economics* (1989): 63-95.

[27] **Reference:** MacDonald, James M. "Competition and rail rates for the shipment of corn, soybeans, and wheat." *The RAND Journal of Economics* (1987): 151-163.

Appendix

Exhibit A: CWS Variables (as given by the STB)

FIELD	DATA DESCRIPTION
1	Waybill Date (mm/dd/yy)
2	Accounting Period (mm/yy)
3	Number of Carloads
4	Car Ownership Code
5	AAR Car Type
6	AAR Mechanical Designation
7	ICC Car Type
8	TOFC/COFC Plan Code
9	Number of TOFC/COFC Units
10	TOFC/COFC Unit Ownership Code
11	TOFC/COFC Unit Type Code
12	Hazardous/Bulk Material in Boxcar
13	Commodity Code (STCC)
14	Billed Weight in Tons
15	Actual Weight in Tons
16	Freight Revenue ($)
17	Transit Charges ($)
18	Miscellaneous Charges ($)
19	Inter/Intra State Code
20	Type of Move
21	All Rail/Intermodal Code
22	Type of Move via Water
23	Transit Code
24	Substituted Truck for Rail Service
25	Rebill Code
26	Estimated Short Line Miles
27	Stratum Identification
28	Subsample Code
29	Exact Expansion Factor
30	Theoretical Expansion Factor
31	Number of Interchanges
32	Origin BEA Area
33	Origin Freight Rate Territory
34	Interchange State #1
35	Interchange State #2
36	Interchange State #3
37	Interchange State #4
38	Interchange State #5
39	Interchange State #6
40	Interchange State #7
41	Interchange State #8
42	Interchange State #9
43	Termination BEA Area
44	Termination Freight Rate Territory
45	Waybill Reporting Period Length
46	Car Capacity
47	Nominal Car Capacity
48	Tare Weight of Car
49	Outside Length
50	Outside Width
51	Outside Height
52	Extreme Outside Height
53	Type of Wheel Bearings and Brakes
54	Number of Axles
55	Draft Gear
56	Number of Articulated Units
57	AAR Error Codes
57-A	Blank
58	Routing Error Flag
59	Expanded Carloads
60	Expanded Tons
61	Expanded Freight Revenue
62	Expanded Trailer/Container Count

Exhibit B: Shipment Origins Removed from the Data

Origin BEA Code	Area
173	Newfoundland
174	Nova Scotia
175	Prince Edward Island
176	New Brunswick
177	Quebec
178	Ontario
179	Manitoba
180	Saskatchewan
181	Alberta
182	British Columbia
183	Yukon/Northwest Territories
184	Puerto Rico
185	Mexico

Exhibit C: Spatial Distribution of BEA Origins in the CWS data (as given by the STB)

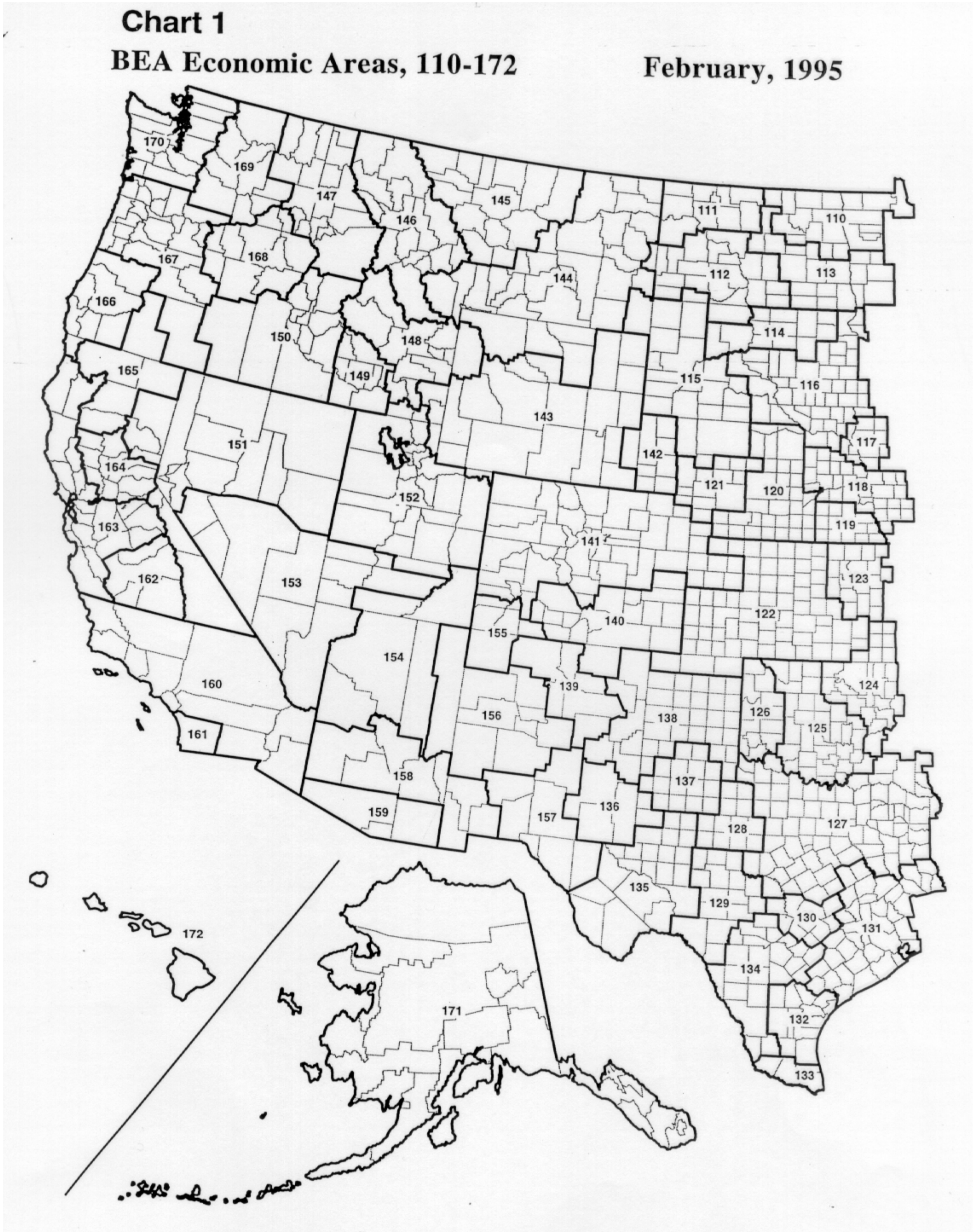


Chart 1

BEA Economic Areas, 110-172

February, 1995

Chart 2

BEA Economic Areas, 001-109 **February, 1995**

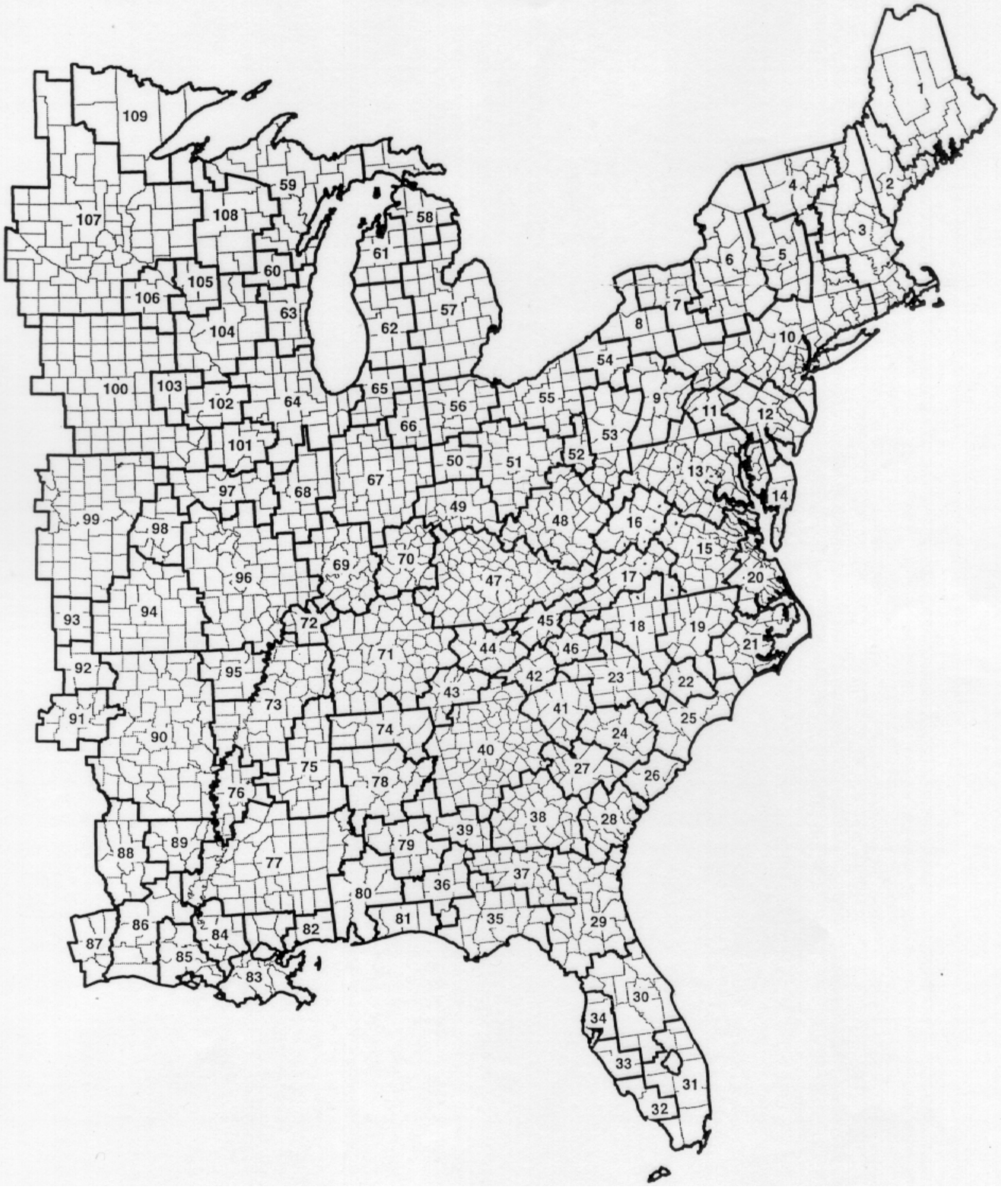


Exhibit D: STCC Code for Commodities Examined in this Study

Commodity Name	STCC Code in CWS
Grain: Wheat Corn and Soybean	1136, 1132, 1144
Crude Oil and Natural Gas	13000's and greater than 10,000
Coal	11000's and greater than 10,000
Chemical	28
Farm non-Grain	Less than 2,000
Pulp/Paper	26
Oil and Coal Products	29
Food	20

Exhibit E: List of BEA Origins in the CWS data (as used in this study)

Region	Origin BEA Code	**Area**
North East (NE)	001	Bangor, ME
	002	Portland, ME
	003	Boston-Worcester-Lawrence-Lowell-Brockton, MA-NH-RI-VT
	004	Burlington, VT-NY
	005	Albany-Schenectady-Troy, NY
	006	Syracuse, NY-PA
	007	Rochester, NY-PA
	008	Buffalo-Niagara Falls, NY-PA
	009	State College, PA
	010	New York-No. New Jer.-Long Island, NY-NJ-CT-PA-MA-VT
	011	Harrisburg-Lebanon-Carlisle, PA
	012	Philadelphia-Wilmington-Atl. City, PA-NJ-DE-MD
South East (SE)	013	Washington-Baltimore, DC-MD-VA-WV-PA
	014	Salisbury, MD-DE-VA
	015	Richmond-Petersburg, VA
	016	Staunton, VA-WV
	017	Roanoke, VA-NC-WV
	018	Greensboro-Winston-Salem-High Point, NC-VA
	019	Raleigh-Durham-Chapel Hill, NC
	020	Norfolk-Virginia Beach-Newport News, VA-NC
	021	Greenville, NC
	022	Fayetteville, NC
	023	Charlotte-Gastonia-Rock Hill, NC-SC
	024	Columbia, SC
	025	Wilmington, NC-SC
	026	Charleston-North Charleston, SC
	027	Augusta-Aiken, GA-SC
	028	Savannah, GA-SC
	029	Jacksonville, FL-GA
	030	Orlando, FL
	031	Miami-Fort Lauderdale, FL
	032	Fort Myers-Cape Coral, FL
	033	Sarasota-Bradenton, FL
	034	Tampa-St. Petersburg-Clearwater, FL
	035	Tallahassee, FL-GA
	036	Dothan, AL-FL-GA
	037	Albany, GA
	038	Macon, GA
	039	Columbus, GA-AL
	040	Atlanta, GA-AL-NC
	041	Greenville-Spartanburg-Anderson, SC-NC
	042	Asheville, NC

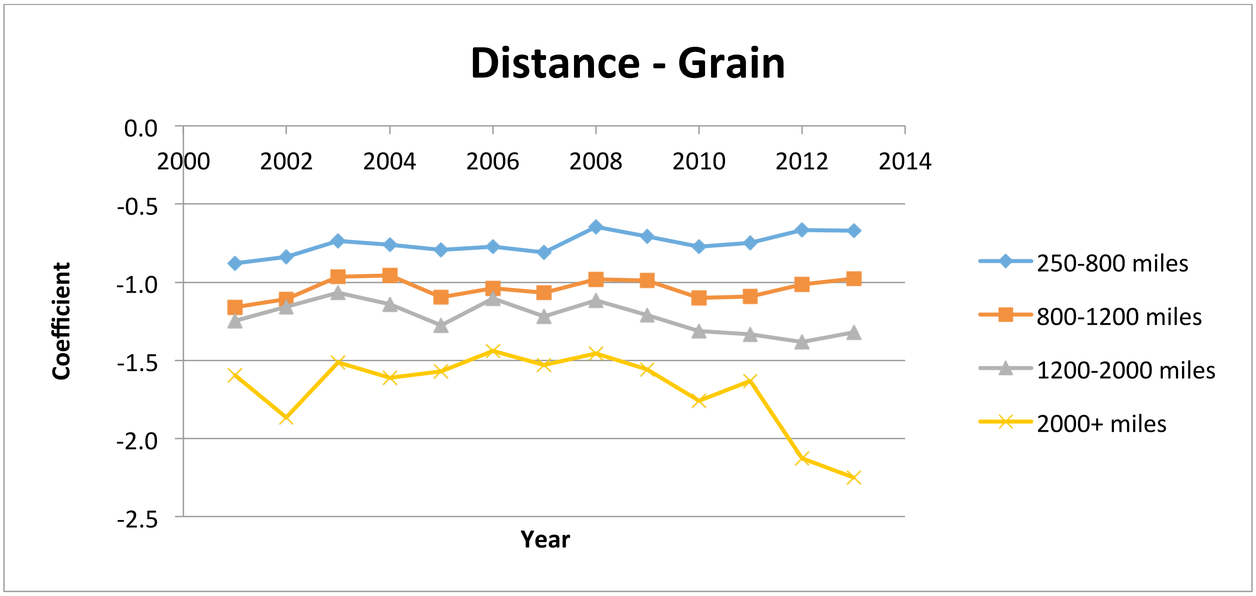
	043	Chattanooga, TN-GA
	044	Knoxville, TN
	045	Johnson City-Kingsport-Bristol, TN-VA
	046	Hickory-Morganton, NC-TN
	047	Lexington, KY-TN-VA-WV
	048	Charleston, WV-KY-OH
Penn/Ohio/Mich/Wisc	049	Cincinnati-Hamilton, OH-KY-IN
	050	Dayton-Springfield, OH
	051	Columbus, OH
	052	Wheeling, WV-OH
	053	Pittsburgh, PA-WV
	054	Erie, PA
	055	Cleveland-Akron, OH-PA
	056	Toledo, OH
	057	Detroit-Ann Arbor-Flint, MI
	058	Northern Michigan, MI
	059	Green Bay, WI-MI
	060	Appleton-Oshkosh-Neenah, WI
	061	Traverse City, MI
	062	Grand Rapids-Muskegon-Holland, MI
	063	Milwaukee-Racine, WI
	064	Chicago-Gary-Kenosha, IL-IN-WI
	065	Elkhart-Goshen, IN-MI
I_states	093	Joplin, MO-KS-OK
	094	Springfield, MO
	095	Jonesboro, AR-MO
	096	St. Louis, MO-IL
	097	Springfield, IL-MO
	098	Columbia, MO
	099	Kansas City, MO-KS
	100	Des Moines, IA-IL-MO
	101	Peoria-Pekin, IL
	102	Davenport-Moline-Rock Island, IA-IL
	103	Cedar Rapids, IA
Upper Midwest	104	Madison, WI-IA-IL
	105	La Crosse, WI-MN
	106	Rochester, MN-IA-WI
	107	Minneapolis-St. Paul, MN-WI-IA
	104	Madison, WI-IA-IL
	105	La Crosse, WI-MN
	106	Rochester, MN-IA-WI
	107	Minneapolis-St. Paul, MN-WI-IA
	108	Wausau, WI
	109	Duluth-Superior, MN-WI
	110	Grand Forks, ND-MN
	111	Minot, ND
	112	Bismarck, ND-MT-SD
	113	Fargo-Moorhead, ND-MN
	114	Aberdeen, SD

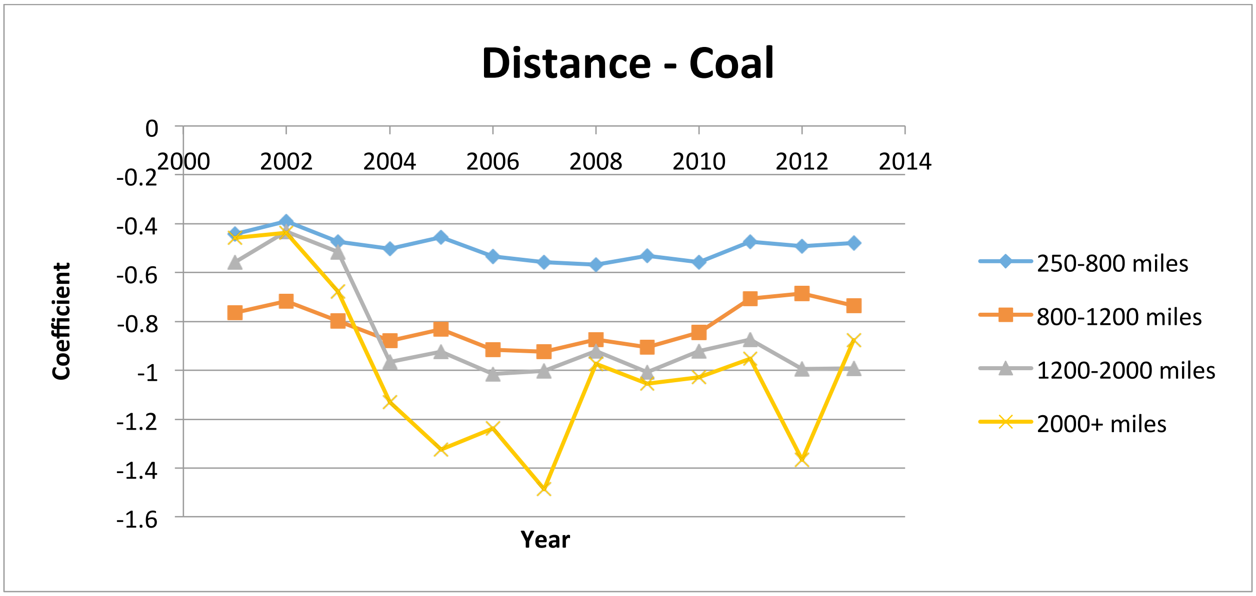
	115	Rapid City, SD-MT-ND-NE
	116	Sioux Falls, SD-IA-MN-NE
	117	Sioux City, IA-NE-SD
NE_KS_CO	118	Omaha, NE-IA-MO
	119	Lincoln, NE
	120	Grand Island, NE
	121	North Platte, NE-CO
	122	Wichita, KS-OK
	123	Topeka, KS
	124	Tulsa, OK-KS
	118	Omaha, NE-IA-MO
	119	Lincoln, NE
	120	Grand Island, NE
	121	North Platte, NE-CO
	122	Wichita, KS-OK
	123	Topeka, KS
	124	Tulsa, OK-KS
	125	Oklahoma City, OK
	121	North Platte, NE-CO
	122	Wichita, KS-OK
TX_OK_NM		
	126	Western Oklahoma, OK
	127	Dallas-Fort Worth, TX-AR-OK
	128	Abilene, TX
	129	San Angelo, TX
	130	Austin-San Marcos, TX
	131	Houston-Galveston-Brazoria, TX
	132	Corpus Christi, TX
	133	McAllen-Edinburg-Mission, TX
	134	San Antonio, TX
	135	Odessa-Midland, TX
	136	Hobbs, NM-TX
	137	Lubbock, TX
	138	Amarillo, TX-NM
	139	Santa Fe, NM
	140	Pueblo, CO-NM
WY_MO_ID	143	Casper, WY-ID-UT
	144	Billings, MT-WY
	145	Great Falls, MT
	146	Missoula, MT
	147	Spokane, WA-ID
	148	Idaho Falls, ID-WY
	149	Twin Falls, ID
	150	Boise City, ID-OR
	151	Reno, NV-CA

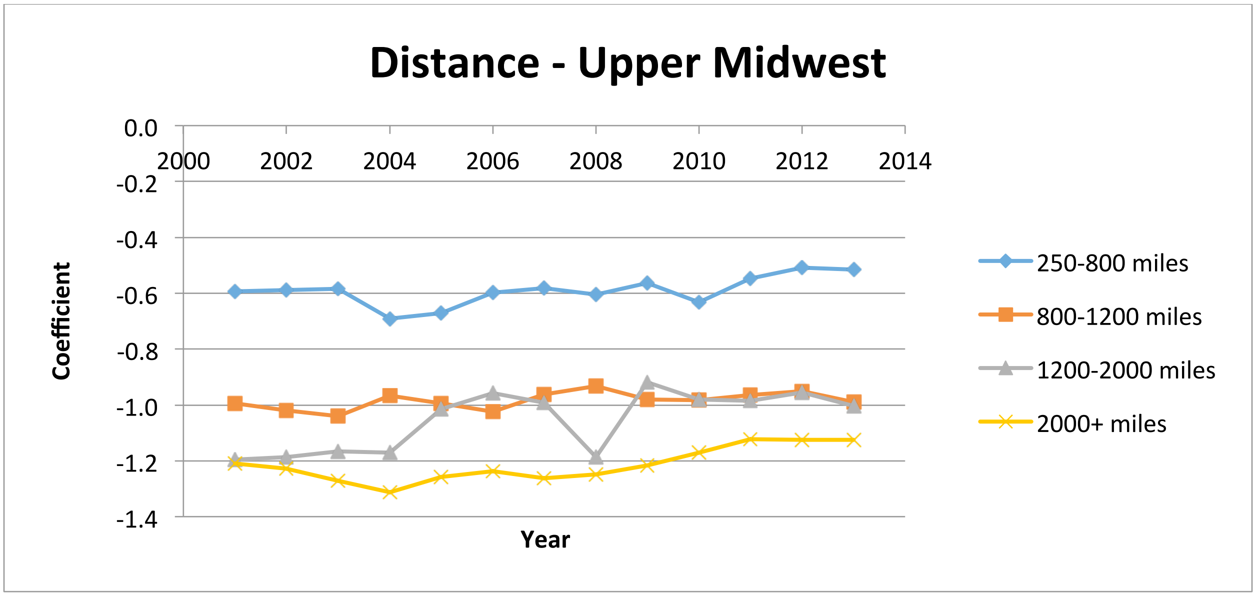
Exhibit F: List of BEA Termination Regions in the CWS data (as used in this study)

Louisiana	83	New Orleans, LA-MS
	84	Baton Rouge, LA-MS
	85	Lafayette, LA
	86	Lake Charles, LA
	87	Beaumont-Port Arthur, TX
	88	Shreveport-Bossier City, LA-AR
	89	Monroe, LA
TX	128	Abilene, TX
	129	San Angelo, TX
	130	Austin-San Marcos, TX
	131	Houston-Galveston-Brazoria, TX
	132	Corpus Christi, TX
	133	McAllen-Edinburg-Mission, TX
	134	San Antonio, TX
	135	Odessa-Midland, TX
	136	Hobbs, NM-TX
	137	Lubbock, TX
	138	Amarillo, TX-NM
Pacific Northwest	165	Redding, CA-OR
	166	Eugene-Springfield, OR-CA
	167	Portland-Salem, OR-WA
	168	Pendleton, OR-WA
	169	Richland-Kennewick-Pasco, WA
	170	Seattle-Tacoma-Bremerton, WA

Exhibit G: Shipment Distance Coefficient Value Results






Distance - Upper Midwest
Coefficient
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.2
-1.4
2000
2002
2004
2006
2008
2010
2012
2014
Year
250-800 miles
800-1200 miles
1200-2000 miles
2000+ miles

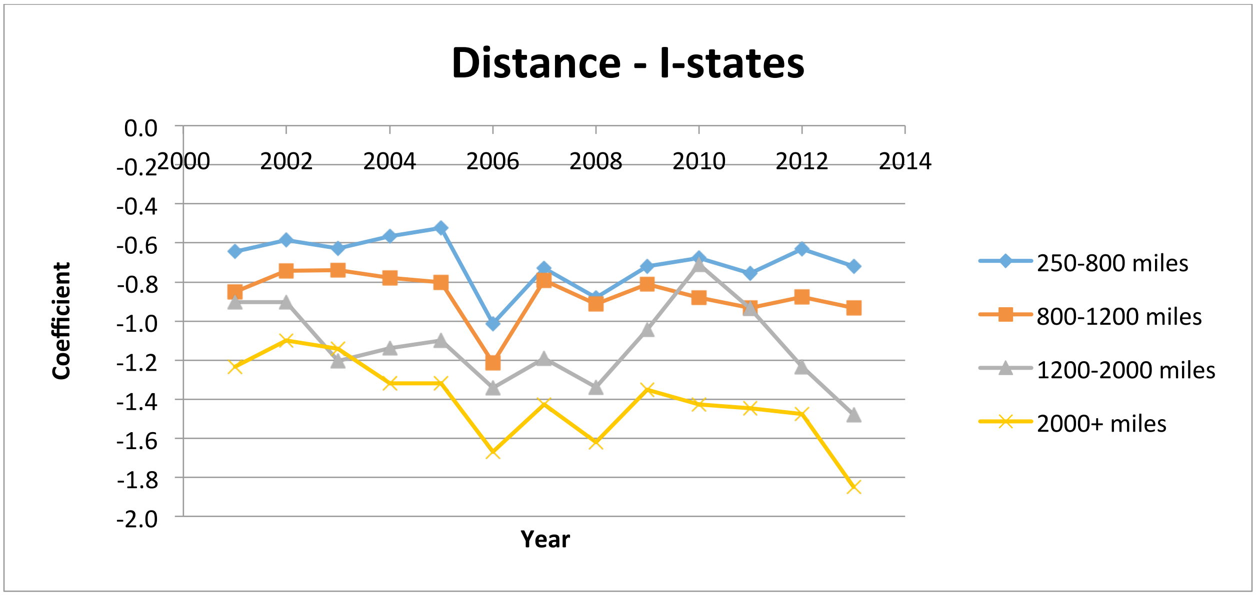
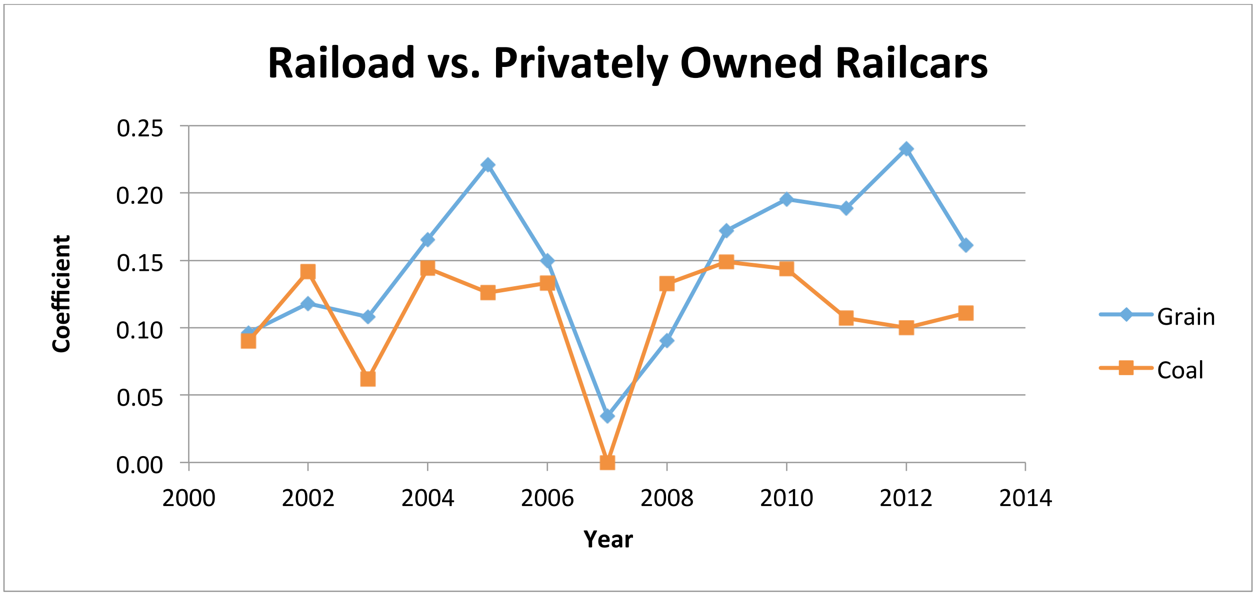
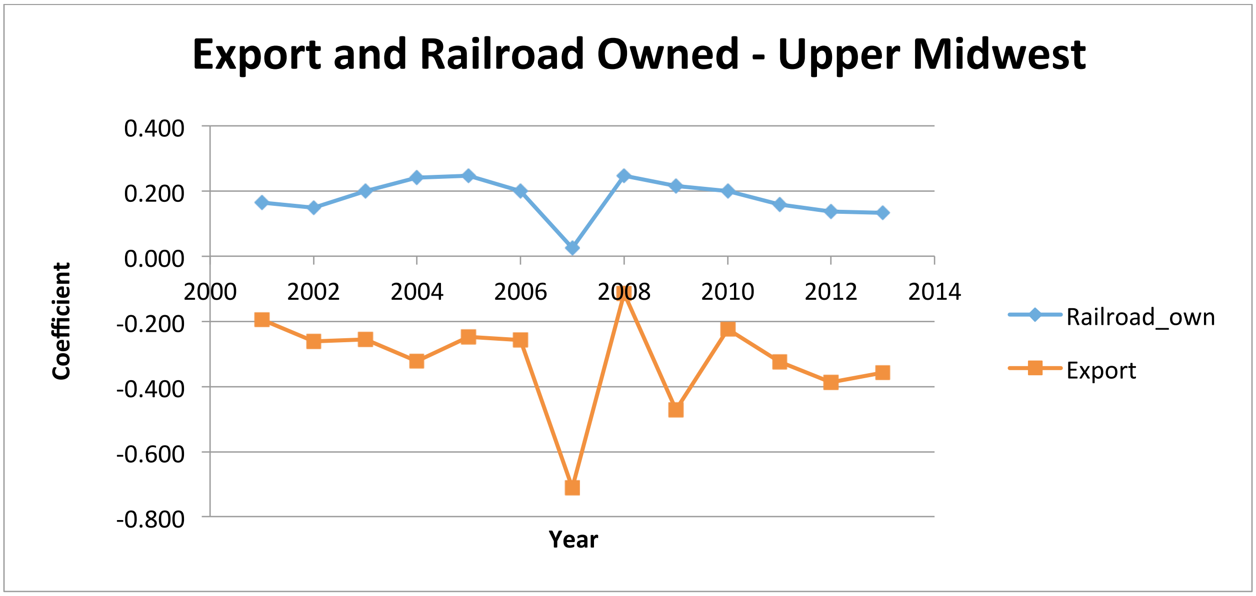
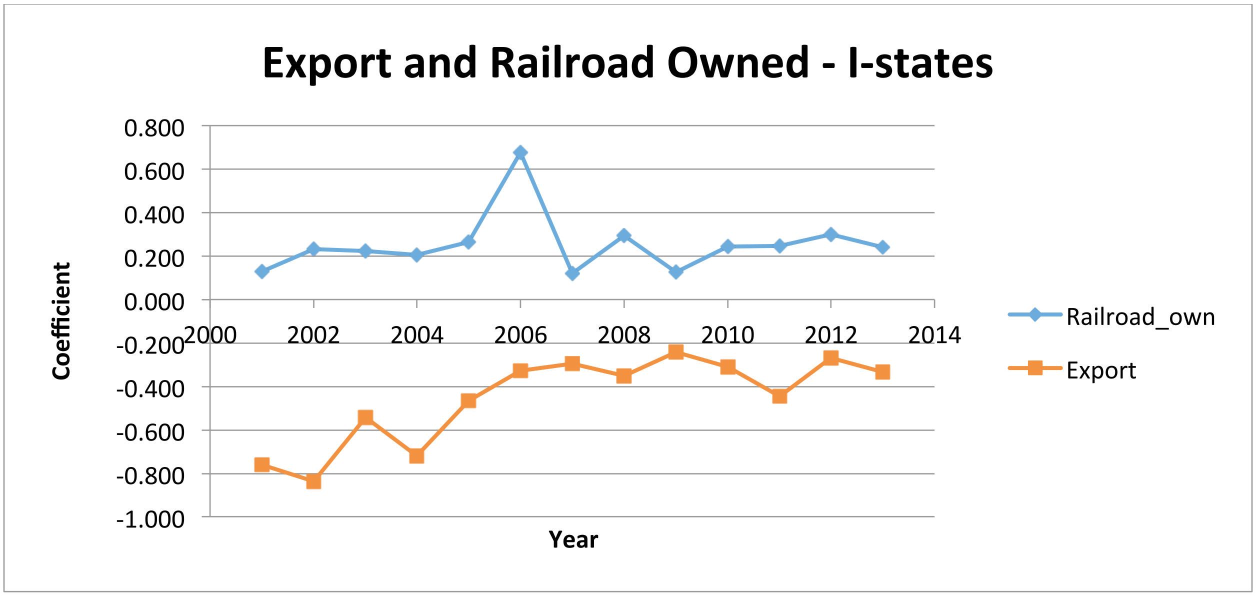

Distance - I-states
Coefficient
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.2
-1.4
-1.6
-1.8
-2.0
2000
2002
2004
2006
2008
2010
2012
2014
Year
250-800 miles
800-1200 miles
1200-2000 miles
2000+ miles

Exhibit H: Railcar Ownership and Export/Domestic Coefficient Value Results






Export and Railroad Owned - I-states
0.800
0.600
0.400
0.200
0.000
-0.200
-0.400
-0.600
-0.800
-1.000
Coefficient
2000
2002
2004
2006
2008
2010
2012
2014
Year
Railroad_own
Export