MODELING CARRIER TRUCKLOAD FREIGHT RATES IN SPOT MARKETS

Christopher Lindsey
Andreas Frei
Hamed Alibabai
Hani S. Mahmassani*

Transportation Center, Northwestern University, 600 Foster Street, Evanston, IL 60208

Young-Woong Park
Diego Klabjan

Department of Industrial Engineering and Management Sciences, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208

Michael Reed
Greg Langheim
Todd Keating

Echo Global Logistics, 600 West Chicago Avenue, Ste. 725, Chicago, IL 60654

Word Count: 5,465 words + 3 figures + 3 tables = 6,965 words

*Corresponding Author: Hani S. Mahmassani

Phone: 847-491-7287

email: masmah@northwestern.edu

Submitted for presentation at the 92nd Annual Meeting of the Transportation Research Board

January 2013, Washington D.C.
MODELING CARRIER TRUCKLOAD FREIGHT RATES IN SPOT MARKETS

ABSTRACT

Most transportation research has focused on the cost determinants of long-term motor carrier contracts for specific lanes. However, with the emergence of third-party logistics (3PL) providers in the U.S. following deregulation in the 1980s, a significant amount of capacity for shipments is secured via spot market transactions as opposed to contracts. Carrier rates for shipments with even the same origin and destination can vary widely from transaction to transaction in this scenario. This research investigates the factors behind this occurrence and identifies the major determinants of carrier costs in spot market transactions at both an individual shipment level and at a more aggregate lane level. Additionally, it also explores a tactical planning scenario in which a 3PL provider addresses chronic fiscal underperformance on certain lanes. The research has found that factors such as distance, characteristics of the shipping lane and the required truck type are among the most important determinants of motor carrier rates at both the individual shipment and the lane level. Also, seasonality and overall market conditions play a major role in determining rates for truckload shipments. The study then goes on to show that the results of the cost determinant analysis may be used to set better baseline prices on underperforming lanes.
INTRODUCTION

Since the deregulation of the U.S. freight and logistics industry in the 1980s, the freight market has experienced the entry of third parties into the logistics process. The Council of Supply Chain Management defines third-party logistics (3PL) providers as “a firm which provides multiple logistics services for use by customers…These firms facilitate the movement of parts and materials from suppliers to manufacturers, and finished products from manufacturers to distributors and retailers (11).” Deregulation and the subsequent emergence of 3PL providers allowed the freight and logistics industry to realize the efficiencies gained through specialization and outsourcing that were not previously possible in the regulated environment. By outsourcing the management of the supply chain companies could concentrate on their core business activities. However, deregulation also gave rise to an extremely competitive market in which shippers, carriers and third parties all try to leverage information and technology to their advantage.

Primarily, 3PL providers manage portions, or the entirety of the supply chain on behalf of a shipper; in addition, they may supply capacity over the long-term or in spot markets. 3PL providers supply these services using either their own physical assets or those of others which leads to a distinction among 3PL providers. Besides the breadth of offered services and areas of specialization, 3PL providers are often further distinguished by their ownership, or lack thereof, of transportation and logistics assets (22). Asset-based 3PL providers are those that own or control rolling stock, warehouse space, or any other physical assets critical to the movement of goods. Non asset-based 3PL providers are those that do not own any freight assets. Instead, these companies treat their industry knowledge and specialized skills as the primary assets available to shippers. Because non asset-based 3PL providers rely solely on knowledge and skill for survival, the use of information and decision-making tools is especially important in such a competitive and economically significant industry. According to Armstrong and Associates, Inc. (33), in 2010 the 3PL industry accounted for over a $127 billion in revenue. Of that, non asset-based providers generated the second largest share at $36.8 billion.

This research demonstrates a potential use of historical shipment data by a non asset-based 3PL provider. The first objective of the study is to empirically investigate the determinants of carrier linehaul costs in shipping lanes over time, and also in transactional spot markets. The spatial nature of freight movement gives us a cross-sectional unit, lanes, over which to examine carrier costs over time. We seek to determine which factors most affect these costs.

A common task of 3PL brokers is to source capacity for recently attained loads. These loads are not part of a pre-determined shipping schedule in which a contract is negotiated. Instead, they are the result of a “cold call” from a shipper to the 3PL provider.
in which the 3PL provider agrees to find capacity for a negotiated fee: a spot market. Knowing how characteristics of a shipment affect the expected carrier costs would prove useful to the 3PL broker.

The second objective of this study is to demonstrate how this information may be used in a tactical planning situation, specifically the avoidance of habitual non-profitable transactions. The assumption is that while some non-profitable transactions are the result of market conditions or human fallacy, others share common characteristics that indicate the 3PL provider is habitually “missing” some critical aspect of carrier costs. If identified and addressed, the result would be fewer non-profitable transactions.

In addition to its practicality, this research fills a gap in the freight and logistics literature. There appear to be no studies regarding the factors affecting motor carrier rates for shipments that occur in spot markets. This is extremely useful to 3PL providers who often operate in this type of environment.

**BACKGROUND AND LITERATURE REVIEW**

**Spot Markets in the Procurement of Transportation Services**

The Internet and information and communication technologies (ICT) together have played a major role in enabling the direct exchange between shippers and carriers in the procurement of transportation services. In these marketplaces, shippers receive quotes from carriers for the provision of transportation services ranging from one-time shipments to long-term contract services. Online marketplaces such as NTE (www.nte.com), BestTransport (www.BestTransport.com) and LeanLogistics (www.LeanLogistics.com) are examples. Lin et al. (44) found many shippers to be aware of and willing to utilize the services provided by these marketplaces. Much of the scholarly work regarding the procurement of transportation services in spot markets assumed business models similar to the examples given. An exception is Huang et al. (55) who used a spot market facilitated by a 3PL provider as done here. However, they examined profitability and broker efficiency as opposed to carrier rates as this work does.

Largely, researchers have employed auctions as the methodological tool for examining the operation and improving the performance of transportation spot markets (66, 77, 8 8,99). Though all of these studies center on transportation spot markets, the basic business model is different from the one used in this paper. In this study, the 3PL provider acts as a broker in the spot market and facilitates deals between shippers and carriers. Also, none of these research efforts investigated the determinants of motor carrier rates, the basic goal of this paper.
Motor Carrier Rates

There has not been much research into the factors affecting carrier freight rates in spot markets. Much more attention has been devoted to the determinants of base motor carrier cost structures. The impetus for much of this research was the deregulation of the U.S. freight and logistics industry in the 1980s. Researchers were primarily concerned with identifying economies of scale, economies of specialization (e.g., truckload and less-than-truckload operators versus general carriers), and gaining insights into the possible effects of deregulation on the trucking industry (1010, 1111, 1212). Research in this vein of the literature largely confirmed that service-quality, the number of shipments, weight of shipments, and the distance shipped are the predominant factors determining carrier rates to shippers. Though these factors are obviously important, they describe rates given for long-term contracts over specific lanes, not spot market transactions.

Other works in the literature focus solely on less-than-truckload operations (1314, 14, 151516, 1617, 17Error! Reference source not found.). Smith et al. (1515) modeled net freight rates for less-than-truckload (LTL) shipments using data from a nationwide motor carrier. They were primarily interested in ascertaining the determinants of freight rates offered to shippers and also in assessing carrier policy for offering discounts to published freight rates, i.e. net freight rates. Their work confirmed that results of previously conducted work regarding the major determinants of freight rates. They then used insights from their model to describe how the model could be used to reassess rate discount policies for certain shippers or at particular terminals.

Using a regression-based methodology, Ozkaya et al. (1616) estimated LTL market rates using data from a nationwide carrier. Their approach was to combine quantitative data with qualitative market knowledge to improve the predictive ability of their models. Ozkaya et al. found that distance and weight are among the most important determinants of LTL rates. Also, the incorporation of qualitative information into the analysis proved to be useful.

Our research is unique because it examines motor carrier rates within a spot market setting as opposed to long-term contracts as done by all others. Because these rates are determined in spot markers, there is an added dimension to the prediction of and explication of the factors affecting TL rates. Also, this research treats the spatial distribution of shipments differently than other works. Instead of relying on pre-determined carrier regions in ascertaining their effect on market rates, we allow the data to reveal these regions of influence. This research is useful to 3PL providers because it gives them greater insight into the factors affecting carrier rates beyond current market averages and rules of thumb.
DATA

The data for this study comes from a U.S.-based 3PL provider operating in North America. Data is from the year 2011 and contains information denoting the date of the shipment, origin and destination, carrier, equipment type, cost and the number of stops among other information. Over 4,000 carriers, both large firms and owner-operators, are included in the data set. The representativeness among carriers in the data set makes it an excellent source of insight into the determinants of carrier costs.

LANE-LEVEL ANALYSIS

Here, the aggregate effects over time of different variables on carrier prices at the lane level are explored. A panel data set is formed by aggregating individual shipment data at lane-level cross-sections \((i)\) and months longitudinally \((t)\). Observations are measured as their median values in a particular one-month span. With this treatment of the data, we are in a position to explore broad market dynamics.

Data Mining

In order to draw useful inferences from the data, we construct a data mining framework in which several tasks are done. The data mining process described here, and also the alternative process for shipment-level analyses in the following section of the paper, can serve as a framework for dealing with transactional shipment data. Since 3PL providers will likely continue to serve as excellent sources of data to researchers for years to come, this framework provides an excellent method for data treatment. The process, shown in FIGURE 1, starts with the clustering of data.
Those in the freight and logistics industry typically describe the spatial aspect of goods movement in terms of lanes: origin-destination pairs. In industry, lanes are usually defined at the level of cities, metropolitan areas, or states. This presents a problem when creating statistical models for spot market transactions. At even the state level, there are often too few observations to draw any statistically significant inference from the data. In addition to 3PL providers often having greater market share in some areas over others, a significant difference in activity between centers of goods movement (e.g., Los Angeles or Chicago) and secondary markets (e.g., St. Louis) can create severe imbalances in data representation at common geographic levels. Clustering is the solution to this problem.

We cluster the data using the k-means algorithm ($18\text{18}$). With k-means, each coordinate is first weighted proportionally to its frequency at both the origin and destination. Then, according to a predetermined amount, clusters are formed by minimizing the weighted distance between coordinates.

Once the data is clustered, lanes are created by forming all possible cluster pairs. In the framework, lanes are directional (i.e., shipments from cluster A to cluster B are distinct from shipments from cluster B to cluster A). Directionality is an important aspect of the lane-level data mining procedure.
characteristic of shipping lanes as the issue of empty backhauls is widely recognized as important within the freight and logistics industry. Carriers adamantly avoid empty backhauls since it amounts to lower utilization of equipment and personnel and ultimately results in less profit. Flow imbalances are better captured when there is a directional component to lanes.

Variable Creation and the Removal of Outliers

Variables included in the analysis may be classified into two types: shipment-specific variables, lane-specific variables, and a single market variable. Shipment-specific variables are those defined according to a particular aspect of the shipment. In the analysis these include the distance shipped (‘Miles’), the number of stops (‘NumOfStops’), and the type of equipment used for transport (i.e., Van, Flatbed, Refrigerated, or Hot Shot). Lane-specific variables are likewise defined. They include the lane's market price (i.e., the median price on that lane based on historical data in the predefined period), the ratio of shipments out of and into a lane's destination cluster (‘OutInRatio’), and a proxy for congestion on a lane (‘Volume-to-Capacity’). The ratio of volume to capacity is measured by dividing the number of shipments in a lane in a month by the number of carriers active in the lane over the same period. The ‘OutInRatio’ captures the potential for attaining a backhaul. This is believed to make a shipment more desirable to potential carriers because of the need to balance flows and more efficiently use equipment. The lone market variable is ‘Market Index,’ which is the Cass Truckload Linehaul Index (1919). It is an indicator of the per-mile price fluctuations in the truckload market that undoubtedly affect carrier rates in spot markets. It is all the more valuable since it is an exogenous source of information for our model.

The analysis is limited to only those shipments within the contiguous U.S. This is because shipments that cross international borders have costs and challenges associated with them that are not common to domestic loads. Prior to estimation, we removed observations in the data set believed to be outliers. These include those with costs that seem unreasonable. Also, only those observations with no accessorial costs are included. Accessorial costs are those that are charged for activities beyond the basic shipping service (e.g., loading/ unloading, etc.).

Twenty clusters are used resulting in 380 unique lanes in our analysis. Twenty clusters were chosen because of diminishing returns in accuracy with an increasing number of clusters. That is, the gains in terms of statistical fit did not continue to increase with the addition of more clusters. There are 38 months in the data set. However, because there is not an observation for every lane in every month, we are left with an unbalanced panel consisting of 10,980 observations.
Model Estimation

Linehaul carrier price per mile (‘PPM’) is the dependent variable. The model is estimated using both fixed and random effects regression. Random effects regression assumes that the cross-section entities, in our case lanes, represent a random sample from the population and as such any lane-specific effect on carrier prices is randomly distributed across lanes and uncorrelated with other regressors similar to a disturbance term. Fixed effects regression does not assume that lane-specific effects are randomly distributed. As such, lane-specific effects should be explicitly accounted for in the analysis via a lane-specific dummy variable, else the parameter estimates would be inconsistent due to omitted variables (2020). Results from both approaches are reported in the analysis.

Furthermore, all models were estimated using heteroskedasticity-robust asymptotic variance-estimators in order to account for possible within-group correlation (21, 2221). The commercial software Stata (2222) was used to conduct the analysis. The model specification is as follows:

\[ PPM_{it} = \beta_0 + \beta_1 \cdot \text{Miles}_{it} + \beta_2 \cdot \text{Miles}_{it}^2 + \beta_3 \cdot \text{Volume-to-Capacity}_{it} + \beta_4 \cdot \text{OutInRatio}_{it} + \beta_5 \cdot \text{OutInRatio}_{it}^2 + \beta_6 \cdot \text{Market Index}_t + \beta_7 \cdot \text{Equipment}_{it} + \beta_8 \cdot \text{Quarter}_{it} + \varepsilon. \]

The inclusion of the squared values of distance and the out-to-in ratio detect any nonlinear effects they may have on carrier cost. In the lane-level analysis, equipment is determined as the most frequently used truck type on a given lane in a given month. In the model, it is treated as a categorical variable indicating one of several equipment types. In this case, those include Van, Refrigerated, Flatbed or Hot Shots (a form of expedited service). Market index is indexed only by time \( t \) since it is the same for all lanes. ‘Quarter’ is a categorical variable that captures any possible seasonal effects on spot market costs. The results of the analysis are found in TABLE 1.
From TABLE 1, shipment distance and its squared value are associated with negative and positive influences on carrier price per mile, respectively. Since the carrier cost is already normalized by distance, this result is intuitive because it implies that increased shipment distances still have effect on the general carrier rate. A higher market index value is associated with higher carrier prices. This simply implies that the general rates given to the 3PL provider are consistent with market conditions. The results suggest that flatbed trucks and shipments that occur in the second calendar quarter are associated with lower carrier prices when compared to the baseline scenario (i.e., a shipment requiring a van and occurring in the 1st calendar quarter). An unexpected result of the analysis, however, is the suggestion that a positive out-to-in ratio is associated with higher carrier prices.

**SHIPMENT-LEVEL ANALYSIS**

In this portion of the analysis, we explore the disaggregate effects of different variables on carrier prices at the shipment level. By analyzing the data at the shipment level, we reveal the determinants of carrier costs in spot market transactions.
Data Mining

The data mining procedure for the shipment-level analysis follows the lane-level analysis. The procedures diverge upon completion of the clustering step. As opposed to aggregating data by lanes and by month in order to create variables, instead they are created on a rolling horizon basis. It is a rolling horizon in the sense that variables are measured according to market conditions at the time the transaction has been made. We accomplish this by selecting a time period over which to observe the market, e.g., a certain number of days prior to the transaction in question, and measure the value of the variable for that shipment using data in the delineated period. This is done for each shipment in the data set until all are cycled through. By taking this approach, a truer picture is painted regarding carrier pricing behavior because it reflects conditions at the time the pricing decision occurred.

The exact number of days selected for each rolling horizon period is determined by the number of observations (i.e., shipments) present in each possible horizon meeting a specific criterion. The criterion is that there must be at least 30 observations present in that horizon for it to be valid. For example, the first rolling horizon period is the previous 30 days. If there are not at least 30 observations in this period, then the data mining procedure examines the previous 45 days. If this period too does not have at least 30 observations, then the next 60 days is examined and so on.

The same variable types and outlier removal conditions as described in the Lane-level Analysis section is employed for the shipment-level analysis. FIGURE 2 outlines the entire procedure.
FIGURE 2 Shipment-level data mining procedure.

**Model Estimation**

Again, linehaul cost per mile (‘PPM’) is the dependent variable in the analysis. After an exhaustive search through numerous possibilities, the following model specification was chosen:

\[ PPM = \beta_0 + \beta_1 \times \text{Miles} + \beta_2 \times \text{NumOfStops} + \beta_3 \times \text{OutInRatio} + \beta_4 \times \text{OutInRatio}^2 + \beta_5 \times \text{MarketIndex} + \beta_6 \times \text{Equipment} + \beta_7 \times \text{Quarter} + \beta_8 \times \text{Lane} + \varepsilon. \]

Because the dependent variable is already normalized by distance, its inclusion as a covariate detects any nonlinear effects it has on carrier cost. Likewise, the squared value of the out-to-in ratio behaves similarly. ‘Equipment’, ‘Lane,’ and ‘Quarter’ are categorical variables while all others are continuous. Like the lane-level analysis, we use 20 clusters resulting in 380 unique lanes for this analysis. The model is estimated using ordinary least squares regression (2020). Results of the analysis are in TABLE 2.
TABLE 2 Shipment-level model estimation results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.152</td>
<td>-6.246</td>
</tr>
<tr>
<td>Miles</td>
<td>-0.0001135</td>
<td>-3.712</td>
</tr>
<tr>
<td>NumOfStops</td>
<td>0.07674</td>
<td>5.772</td>
</tr>
<tr>
<td>OutInRatio</td>
<td>-0.03893</td>
<td>-1.990</td>
</tr>
<tr>
<td>OutInRatio^2</td>
<td>0.007702</td>
<td>1.868</td>
</tr>
<tr>
<td>Index</td>
<td>0.03731</td>
<td>23.09</td>
</tr>
<tr>
<td>Equip. - Flatbed</td>
<td>0.08147</td>
<td>6.579</td>
</tr>
<tr>
<td>Equip. - HotShots</td>
<td>-0.01917</td>
<td>-1.291</td>
</tr>
<tr>
<td>Equip. - Reefer</td>
<td>0.1261</td>
<td>3.454</td>
</tr>
<tr>
<td>Quarter 2</td>
<td>0.2319</td>
<td>20.07</td>
</tr>
<tr>
<td>Quarter 3</td>
<td>0.02256</td>
<td>1.715</td>
</tr>
<tr>
<td>Quarter 4</td>
<td>0.4047</td>
<td>27.70</td>
</tr>
<tr>
<td>R-Squared</td>
<td>0.1838</td>
<td></td>
</tr>
<tr>
<td>Adj. R-Squared</td>
<td>0.1692</td>
<td></td>
</tr>
</tbody>
</table>

As shown in TABLE 2, the results reveal several insights about carrier rates in spot markets. The number of stops has a positive effect on carrier prices. That is, an additional stop triggers a corresponding increase in carrier price. The market index also exhibits a positive effect on carrier prices as expected. Carriers charge higher or lower rates conditional on higher or lower rates in the market.

The out-to-in ratio and the distance shipped, however, both negatively affect costs. Because the dependent variable is already normalized by distance, the results reveal that longer-distance trips command slightly smaller carrier rates. The out-to-in ratio captures the ability of potential backhauls to lower carrier prices. Its squared value reflects its diminishing influence on carrier rates. Relative to vans (the base scenario), use of a flatbed or refrigerated truck increases the carrier cost while ‘hot shots’ decreases cost.

Of the 379 lanes included as dummy variables in the analysis (a single lane is reserved as the base category), at the 10% significance level, 261 are statistically significantly different from the reference lane.

COMPARISON OF THE LANE- AND SHIPMENT-LEVEL ANALYSIS

RESULTS

To a degree, the shipment-level analysis reflects many of the insights gleaned from the lane-level analysis. Characteristics of the lane such as the out-to-in ratio of its
destination cluster and also its volume-to-capacity ratio (which was not included in the shipment-level analysis because it was insignificant) are important determinants of carrier price. Shipment characteristics such as distance and the required equipment type play major roles in carrier pricing, as well. Also, seasonal effects reflected in the calendar quarter in which the shipment occurs and the general price level affects carrier prices.

However, where the two analyses differ is in the direction of influence on carrier prices for certain variables. In both analyses, shipment distance and its squared value are associated with negative and positive impacts on carrier price per mile, respectively. The same is true of the effect of the market index. Regarding the out-to-in ratio, however, the panel data analysis suggests that positive values are associated with higher carrier prices while the shipment-level analysis suggests the opposite. Also, the panel data analysis suggests that costs related to flatbed and refrigerated equipment types are not significantly different from vans. In addition, the panel analysis suggested that shipments that occur in the second calendar quarter are associated with lower carrier prices though this is not the case with the shipment-level analysis.

A possible reason for these discrepancies is the aggregate nature of the panel data. The panel analysis yields a broad view of what is generally occurring across lanes, not pricing on a shipment-by-shipment basis as the pooled analysis does. Higher priced lanes may well be generally associated with destination clusters with higher out-to-in ratios, while a shipment destined for a cluster with a high out-to-in ratio commands a lower price in a spot market transaction. Also, equipment-specific effects may not be relevant at this level.

TACTICAL PLANNING SCENARIO: UNDERPERFORMING LANES

While most non-profitable transactions are sporadic, others are habitual and share common characteristics that suggest the 3PL provider repeatedly overlooks some critical aspect of carrier costs. Otherwise, a sufficient cost would have been charged to the carrier to ensure a profit. The obvious manner in which to investigate this occurrence is by lane. Lanes are the common denominator by which we can identify habitual non-profitable transactions and begin to address the problem.

Identifying Underperforming Lanes

In order to identify underperforming lanes, we examine what we have termed lane 'profit-to-loss ratios.' The profit-to-loss ratio (PLR) is the number of profitable transactions on a lane divided by the number of unprofitable transactions on the same lane over the study time frame.
Here, a profitable transaction is defined as one in which the difference between the linehaul cost charged to the shipper and the linehaul cost paid to the carrier is positive. Unprofitable transactions are those in which this difference is less than, or equal to, zero. FIGURE 3 shows the distribution of PLR values over the lanes included in the data set. The actual PLR value which would normally be reported on the vertical axis is omitted in order to respect the privacy of the 3PL provider.

FIGURE 3 Distribution of profit-to-loss ratios.

A relatively low PLR value for a lane indicates poor profit performance while higher PLR values imply good performance. By examining the distribution of PLR values and setting an appropriate base performance level for a lane (e.g., a PLR value equal to 3, or perhaps the 25th percentile for instance), a 3PL provider can identify underperforming lanes. Once those lanes are singled out, the firm can then devise strategies to improve their performance. Thus, the data mining procedure combined with the PLR present a methodology by which 3PL provider performance can be assessed on a lane-by-lane basis.
Application of the Shipment-Level Model Results to Underperforming Lanes

Once underperforming lanes have been identified according to the pre-defined baseline metric, the question that naturally follows is 'How can they be improved?' We address this issue by again estimating a regression model. However, the primary goal is now prediction instead of explication. A new variable, ‘Market Price,’ is introduced into the model specification. This variable is the combined median value of shipper and carrier rates on a particular lane and is the method by which market conditions are captured using the 3PL provider’s own data; this is greatly preferred to relying on third-party data. Unlike, the market index used previously, this market indicator is lane specific. The model specification is as follows:

 Carrier linehaul cost = $\beta_0 + \beta_1 \times \text{Miles} + \beta_2 \times \text{Market Price} + \beta_3 \times \text{NumOfStops} + \beta_4 \times \text{OutInRatio} + \beta_5 \times \text{OutInRatio}^2 + \beta_6 \times \text{Equipment} + \beta_7 \times \text{Lane} + \epsilon.$

The model is estimated using the primary set of data and while its predictive ability is gauged using a holdout sample (a portion of data over which the models were not calibrated). Results are in TABLE 3. The metric by which the model is assessed is the mean absolute error (MAE). Absolute error is the absolute difference between the predicted and true value; the MAE is the average value across all observations.

### TABLE 3 Predictive Model Estimation Results

<table>
<thead>
<tr>
<th></th>
<th>Traditional Estimate</th>
<th>t-statistic</th>
<th>Log-Transform Estimate</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-41.87</td>
<td>-0.7257</td>
<td>1.984</td>
<td>38.23</td>
</tr>
<tr>
<td>Miles</td>
<td>1.199</td>
<td>36.53</td>
<td>0.0004127</td>
<td>27.45</td>
</tr>
<tr>
<td>MarketPrice</td>
<td>0.4565</td>
<td>57.37</td>
<td>0.6722</td>
<td>98.00</td>
</tr>
<tr>
<td>NumOfStops</td>
<td>75.37</td>
<td>7.059</td>
<td>0.008811</td>
<td>1.790</td>
</tr>
<tr>
<td>OutInRatio</td>
<td>-67.24</td>
<td>-4.460</td>
<td>-0.03859</td>
<td>1.900</td>
</tr>
<tr>
<td>OutInRatio$^2$</td>
<td>21.61</td>
<td>7.062</td>
<td>0.007028</td>
<td>4.763</td>
</tr>
<tr>
<td>Equip. - Flatbed</td>
<td>27.23</td>
<td>2.755</td>
<td>0.02256</td>
<td>5.111</td>
</tr>
<tr>
<td>Equip. - HotShots</td>
<td>-30.85</td>
<td>-2.590</td>
<td>-0.01525</td>
<td>-2.871</td>
</tr>
<tr>
<td>Equip. - Reefer</td>
<td>179.4</td>
<td>6.220</td>
<td>0.05973</td>
<td>4.654</td>
</tr>
<tr>
<td>Adj. R-Squared</td>
<td>0.8565</td>
<td></td>
<td>0.8750</td>
<td></td>
</tr>
</tbody>
</table>

With adjusted R-squared values of 0.8530 and 0.8750 for the traditional and log-transform models, respectively, much of the variance in carrier costs is explained. The MAE for the traditional model specification was calculated to be 425.19. Dividing the absolute error by the true linehaul cost for each shipment and then taking the average
yields a value equal to 0.2515. This value implies that, on average, the predicted values are within approximately 25% of the true value.

If underperforming lanes are singled using a predefined criterion, for instance a PLR less than or equal to 3, it can be judge how well the model estimates carrier prices on those lanes. Using the lanes in the holdout data set with a PLR less than or equal to 3, the predicted values of the linehaul costs are within 21% of their true value. This shows that the model does a good job of predicting what the actual linehaul costs for a shipment may be. A 3PL broker with this information beforehand could use it to set a baseline price for shipper negotiations.

**CONCLUSIONS**

This research has investigated the determinants of carrier linehaul costs in spot markets; it has defined a methodology by which historical shipment data may be processed and mined; and lastly it has demonstrated some of the potential uses of historical shipment data by non asset-based 3PL providers. Regarding the determinants of carrier costs, the distance a shipment is transported along with the rate at which the destination generates outbound loads are among the most important factors.

Also, the study examined the transactional data as a panel data set by aggregating transactions at the lane level and in monthly intervals. In so doing, a broad view of the temporal and spatial dynamics of carrier pricing was taken yielding insights into overall market behavior. Many of the same characteristics identified in the shipment analysis proved to be consistent with the lane-level analysis. However, the direction of influence was sometimes opposite.

Lastly, the research developed a methodology by which underperforming lanes are identified using historical shipment data. Lanes that exhibit below par PLRs can be singled out for improvement by the 3PL provider and efforts can be undertaken to improve performance. As part of those efforts, the regression model used for prediction can suggest to the 3PL broker baseline prices to begin negotiations with a potential shipper.

In further work, measures other than the median values could be tested in the lane-level analysis. In addition, pooling data across several 3PL providers would provide a more complete perspective of the market.

**REFERENCES**


