Bayesian Spot Rate Analysis

Taylor Bosier

ABSTRACT This paper's focus is to highlight the efficacy of Hierarchical Bayesian Logistic Regression at providing insight into the likelihood of Shippers, in the Freight Spot Market, accepting the spot rate that a synthetic Carrier offers. This analysis allows the carrier to understand the behavior of the customer, the shipping lane that they're operating on, and the customer-to-shipping lane relationship.

Contents

1 Introduction

In the transportation industry, there is one such way in which goods are solicited to be delivered and how the pay is broken out, in what is called the Freight Spot Market, or simply the spot rate. This rate is a one-time payment from shipper to carrier, where the shipper is the party who owns the goods and the carrier is the party who delivers said goods. This market is driven by various forces, such as the Load-to-Truck ratio, which measures how many loads there are per available trucks in a given market. It is also driven by external market forces such as diesel fuel prices, the weather, and other various macroeconomic trends. Each carrier has their own internal pricing tools to ascertain a given rate to the shipper. Various costs such as insurance, driver pay, type of load and others influence these tools. If a carrier were to analyze where the shippers are at in their rates and rate acceptances, they would be able to provide optimal rates that balance acceptance probability and proftiability. This paper is focused in analyzing the log-odds or probability of a customer on a given lane accepting the rates that we offer. Hierarchical Bayesian Logistic Regression provides insight not only into these relationships, but how the overall Customer is accepting. This method is useful in finding hierarchical interactions on rates; each lane has their own micro-economy and the customer themselves represent their own micro-economy. Lanes are impacted by federal, state, and local level governments and general macroeconomic trends such as fuel rates, and customers themselves have their own internal costs. Walmart, for example, impacts the supply chain far more than what local businesses do, due to their economies-of-scale. This method hopes to highlight how these different customers interact with our business, how the lanes as a broader economy interact with our business, and then how the customer and lane interact with our business.

2 Data

The data for this paper is synthetic, however we provide test cases to assess the validity of the methodology. We generate data for 10 customers in 10 markets over a 1 month time period. Each day has 10 bids each, creating a 3000 sample set. We include the following test cases.

- A customer who always declines our rates in 2 randomly chosen markets. This simulates that some customers send rates to all carriers to ensure that if their first choice does not materialize, they have a backup rate.
- A customer who declines in every lane. This simulates the above scenario as well.
- A customer who accepts when the rate is 5% below the market rate.
- A customer who accepts when the rate is 10% below market, but with few acceptances here and there.
- A customer who always accepts, they utilize our service solely due to customer satisfaction.
- A customer who only operates in 1 lane. This is interesting to show that if they expanded their operations, how likely are they to accept our rate?
- Two customers who seemingly random accepts or declines our rates, whether we are above or below the market.
- A customer who will accept if we are within 2% margin to the market.
- A customer who will accept if we are providing market rates.

The following is the structure of the data and how it will be processed into a pandas DataFrame.

- Date
- Customer Name
- Lane - This represents the origin of pickup to the destination of the delivery. Typically in the format of DAL-HOU, which is Dallas to Houston
- Mileage of that lane - We establish a base line mileage for the overall lane and add perturbation to it. This helps simulate that origin pickup to destination drop offs don't all follow the same amount of mileage, customers are variable in their distances to travel.
- Cost per mile - We generate rates uniformly from 1 to 3 for each lane. This helps account for variable costs due to insurance, known weather patterns, wear and tear on the truck in that market, and other factors.
- Market Rate - The rate that the overall market is offering. This rate is typically gathered by scraping various load boards. This rate is determined through carrier costs and fuel. We blanket assume that all carriers in the market have similar cost per mile, and fuel does not change in the 1 month time period. We add a random effect of uniform distribution from 100 to 200.
- Offered Rate - The rate to which our internal mechanisms offered the shipper. From internal observation, a straight-forward reason as to why offered rates are variable. It ranges from market conditions to volume amounts to load to truck ratio to other effects that can be seen as random. Let's assume that our Account Managers have surmised that based on our operation size, we have the ability to provide better service through certain metrics, and we show that we tend to offer 4% above market. We utilize a triangular distribution here, with left parameter as .9, mode of 1.04, and 1.10 as right parameter. This indicates that we tend to offer 4% above market.
- Margin - How above or below is the carrier to the market.
- Acceptance - Did the shipper accept the rate.

3 Model

The Hierarchical Bayesian Logistic Regression model is formulated to predict the likelihood of rate acceptance by customers based on the offered margin, customer characteristics, lane characteristics, and their interactions. The model uses hierarchical priors to capture varying effects across different customers and lanes as well as interactions between them.

3.1 Model Specification

The probability of acceptance (y_i) for each transaction i is modeled using logistic regression, where the log-odds of acceptance are a linear function of intercepts and interaction effects:

$$\text{logit}(p_i) = \text{customer\_intercept}[i] + \text{lane\_intercept}[i] + \beta_{\text{margin}} \times \text{Margin}_i + \text{interaction\_effects}[i] \quad (1)$$

where:

- customer_intercept are random intercepts for customers,
- lane_intercept are random intercepts for lanes,
- β_{margin} is the coefficient for the margin,
- interaction_effects are random interaction effects between customers and lanes.

3.2 Hyperpriors

Hyperpriors are used to regulate the distribution of group-specific parameters, adding a level of hierarchy to the model:

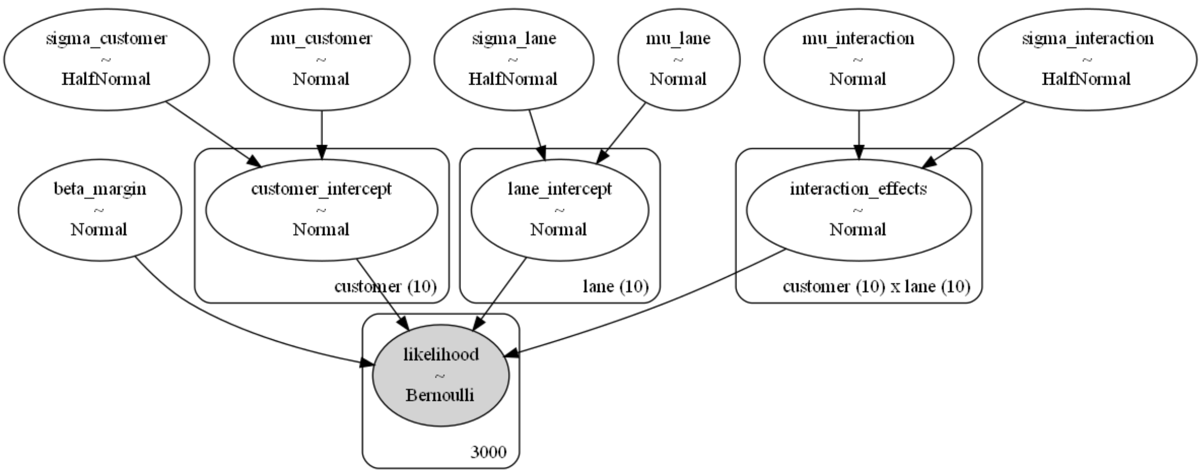
$$\mu_{\text{customer}}, \mu_{\text{lane}}, \mu_{\text{interaction}} \sim \mathcal{N}(0, 1)$$
$$\sigma_{\text{customer}}, \sigma_{\text{lane}}, \sigma_{\text{interaction}} \sim \text{HalfNormal}(1)$$

3.3 Priors

The priors for the intercepts and coefficients are drawn from normal distributions centered around the respective hyperparameters:

$$\text{customer\_intercept} \sim \mathcal{N}(\mu_{\text{customer}}, \sigma_{\text{customer}})$$
$$\text{lane\_intercept} \sim \mathcal{N}(\mu_{\text{lane}}, \sigma_{\text{lane}})$$
$$\beta_{\text{margin}} \sim \mathcal{N}(0, 1)$$
$$\text{interaction\_effects} \sim \mathcal{N}(\mu_{\text{interaction}}, \sigma_{\text{interaction}})$$

3.4 Graphical Representation of the Model



4 Results

4.1 Customer-level

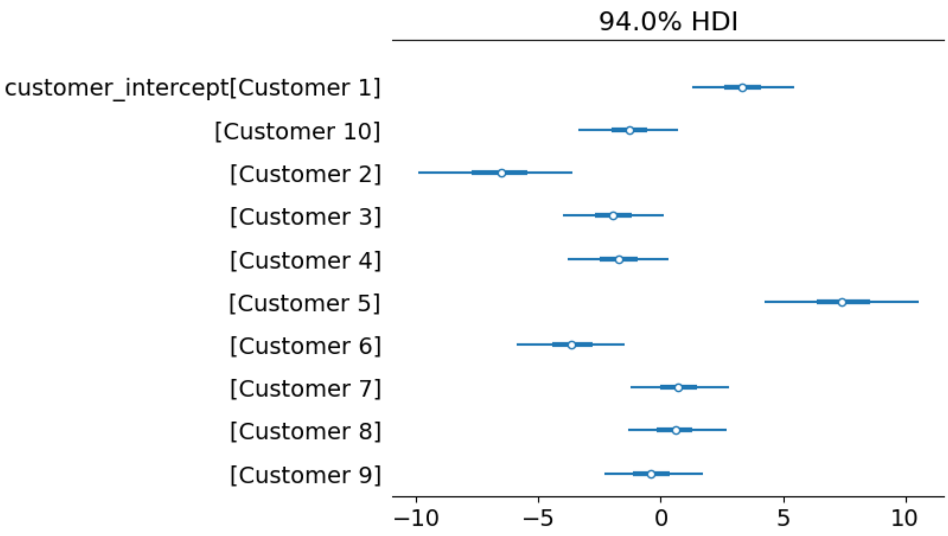


Figure 1: Customer-level analysis and forest plot.

Here we see how the customer(s) interact with our business at a high level. The model was able to show that the log-odds of accepting the margin to which we offer them is much higher than other customers. The model was able to also detect that Customer 2 has a low log-odds of accepting rates we offer them. This method would allow stakeholders to take action on an account level basis. If they need more bookings and are unsure of who to ask, this method would highlight customers that you are more likely to receive acceptances from. It is also an analysis that indicates which customer(s) might need further investigation. Why does Customer 2 always decline our rates?

4.2 Lane-level

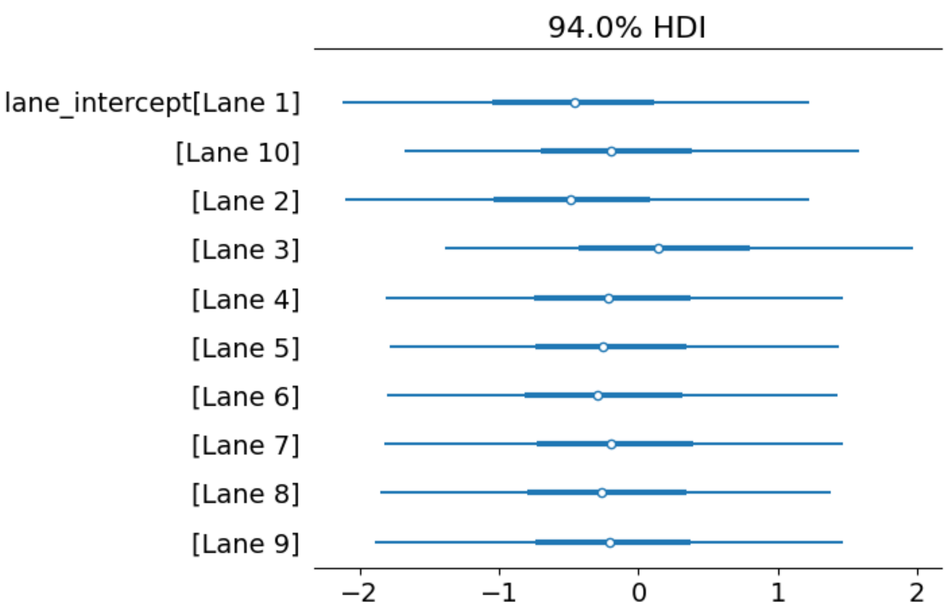


Figure 2: Lane-level analysis and forest plot.

The model indicates a challenging prospect of learning the lane behavior, due to the high variability in the 95% High Density Interval. However, this model does seem to indicate that Lane 3 is higher than other lanes. This method would give stakeholders further insight into their markets and where they should possibly target if they are in need of bookings soon. In this analysis, this might indicate that Lane 3 is a place to explore, while in Lane 2 it might be challenging to find a booking.

4.3 Customer-Lane-level

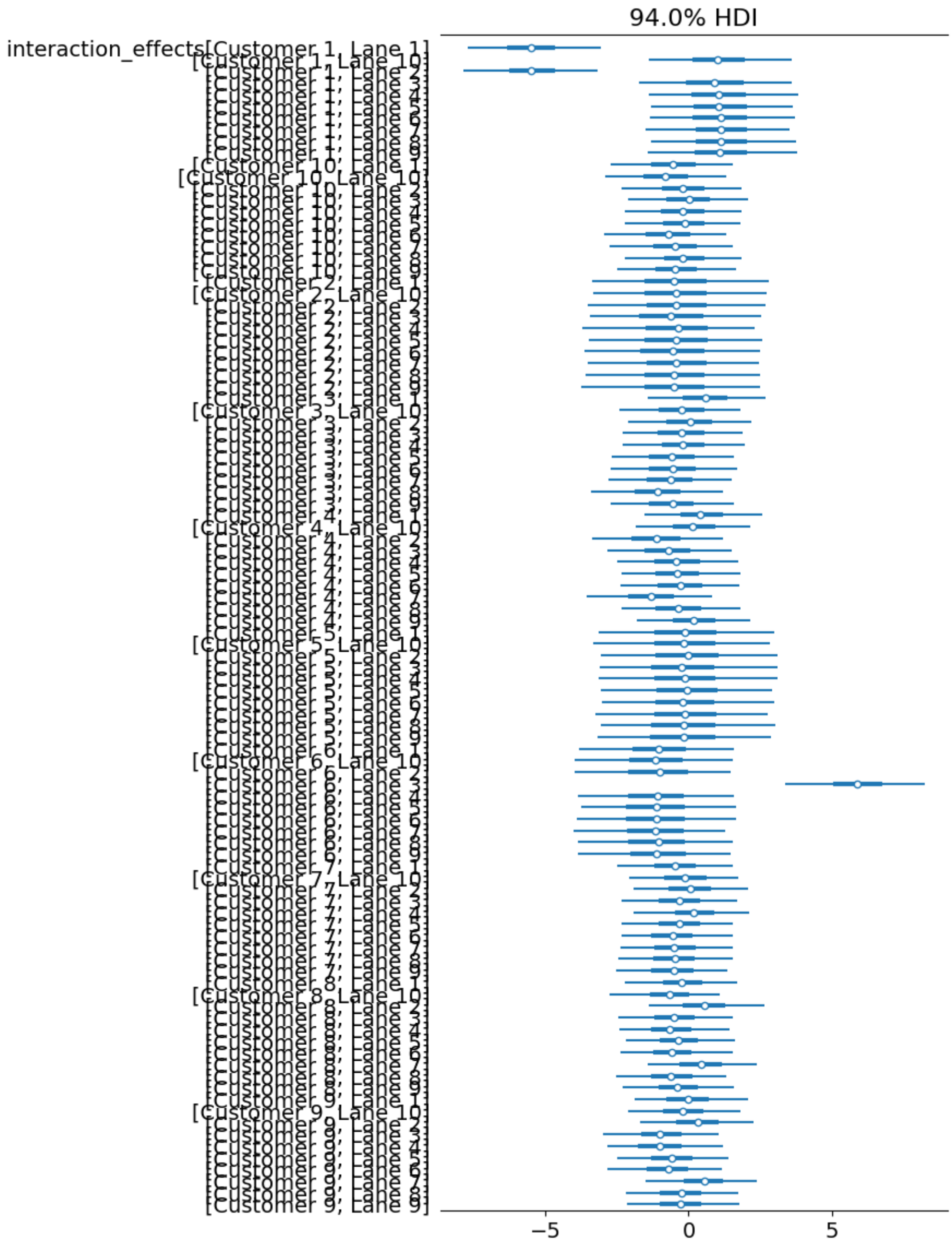


Figure 3: Customer Lane-level analysis and forest plot.

This output indicates the interaction-level that the model learned. The model learned our test-case of Customer 6 only accepting in Lane 3 while providing uncertainty around the log-odds of Customer 6 accepting in other lanes.

5 Frequentist Logistic Regression Comparison

In this section, we compare the results of the Hierarchical Bayesian Logistic Regression model to the implementation found in scikit-learn's packaging.

Table 1: Confusion Matrix

	Predicted Negative	**Predicted Positive**
Actual Negative	404	69
Actual Positive	91	186

Table 2: Classification Report

	Precision	**Recall**	**F1**	**Support**
0	0.82	0.85	0.83	473
1	0.73	0.67	0.70	277
accuracy	0	0	0.79	750
macro avg	0.77	0.76	0.77	750
weighted avg	0.78	0.79	0.78	750

These error metrics highlight the performance of the model, where the F1-score indicates relatively low ability to correctly identify whether or not a customer will accept. The model will return a point-estimate of the log-odds of a customer accepting our margin. This type of analysis does not address the typical concern found in industry, which is: so what? By that, we mean, we are able to return a point-estimate of the odds that a customer will accept, but what do we do with that information, and can we get further context? The Bayesian methods provide further context in the analysis of, we think that the customer in this lane are comfortable in this range, rather than a simple estimate. The Bayesian method provides further context and is able to equip the stakeholder with information that gives a holistic sense of the customer and market.

6 Critical Evaluation

There are assumptions to this model, and assumptions within the data itself. Is there truly a hierarchy within the data? If there is not, one can simply write a Bayesian Logistic Regression model to ascertain the interactions. The way in which one can detect if hierarchies exist is a comparative analysis utilizing hierarchical and non-hierarchical models and examine the metrics thereafter, typically utilizing the Watanabe-Aikake information critierion (WAIC). The higher WAIC score is typically chosen, but a consideration should be made regarding how the business would utilize the model. This method's output provides direct trace summary ouputs of how the hierarchies impact the rate to which the business will offer to the customer, whereas a non-hierarchical model will fail to model the hierarchy by definition.

Another point to consider is, why would a business use this over simply putting the data in a dashboard? An answer to this particular question is that a dashboard would fail to provide probabilistic estimates of the interactions and also fail at learning the hierarchies within the data.

7 Conclusions

In this paper, we analyzed the customer, lane, and customer-lane behavior on rates and the acceptance of rates. This method would allow stakeholders to take action in their respective accounts to ask for more freight volume or to ask their respective customer base for contractual agreements. Further implications would include the ability to provide a probabilistic estimate of the customer accepting future rates, granting stakeholders further insight into profitability of their section in the business.