Predicting the Likelihood of a Shipment Write-Off

by

Oscar Bonet Olano
B.S. in Computer Science and B.A. in Business Administration, University of Deusto, 2019

and

Joshua Weston
B.S. in Economics and B.S. in Supply Chain Management, Arizona State University, 2020

SUBMITTED TO THE PROGRAM IN SUPPLY CHAIN MANAGEMENT
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE IN SUPPLY CHAIN MANAGEMENT
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2024

Signature of Author: ____
Oscar Bonet Olano, Department of Supply Chain Management
May 10, 2024

Signature of Author: ____
Joshua Weston, Department of Supply Chain Management
May 10, 2024

Certified by: ____
Dr. Chris Caplice
Executive Director, Center for Transportation and Logistics
Director, MicroMasters Credential Program in SCM
Founder and Director, MIT FreightLab
Capstone Advisor

Certified by: ____
Dr. Devadrita Nair
Postdoctoral Associate, Center for Transportation and Logistics
Capstone Co-Advisor

Accepted by: ____
Prof. Yossi Sheffi
Director, Center for Transportation and Logistics
Elisha Gray II Professor of Engineering Systems
Professor, Civil and Environmental Engineering

Predicting the Likelihood of a Shipment Write-Off

by

Oscar Bonet Olano

and

Joshua Weston

Submitted to the Program in Supply Chain Management

on May 10, 2024 in Partial Fulfillment of the

Requirements for the Degree of Master of Applied Science in Supply Chain Management

ABSTRACT

In the U.S. trucking industry, freight brokerages act as vital intermediaries between shippers and carriers, but they face financial risks due to write-offs from unpaid services. Despite the recognized importance of mitigating these financial risks, the sponsoring company, a freight brokerage, does not currently have a predictive model to assess the likelihood and magnitude of write-offs, making it challenging to prevent financial losses before they occur. This study tackles this issue by analyzing shipment data and historical write-off incidents to identify key predictors of financial write-offs. Utilizing logistic and linear regression models, it quantifies the risk associated with each shipment, enabling the brokerage to prioritize transactions with lower risk profiles. The analysis revealed that specific shipment characteristics, such as mode of transportation, significantly influence the likelihood and magnitude of write-offs. Predictive models developed in this study were able to predict the probability of write-offs occurring, capturing 65-70% of all write-offs in the test set, offering a tool for more informed decision-making. The findings demonstrate the potential for predictive modeling to significantly reduce financial risks for freight brokerages by enabling preemptive identification of high-risk shipments. By applying this predictive approach, freight brokerages can enhance their financial stability and operational efficiency, contributing to the overall health of the trucking industry's economic ecosystem.

Capstone Advisor: Dr. Chris Caplice
Executive Director, Center for Transportation and Logistics
Director, MicroMasters Credential Program in SCM
Founder and Director, MIT FreightLab

Capstone Co-Advisor: Dr. Devadrita Nair
Postdoctoral Associate, Center for Transportation and Logistics

ACKNOWLEDGMENTS

We would like to express our gratitude to everyone who contributed to the success of this project. First, we extend our thanks to our sponsor company for providing us with the opportunity to work on this exciting and meaningful project that is critical to the business. We appreciate their time, feedback, and continuous support given throughout the process.

We would also like to thank both of our advisors, Dr. Chris Caplice and Dr. Devadrita Nair, for their guidance, patience, and wisdom throughout this project. We are grateful to Dr. Chris Caplice for providing focus and clarity throughout the process, ensuring we stayed on the right track and provided valuable insights. Dr. Devadrita Nair was instrumental in pushing us to think more broadly about our approach, challenging us to explore innovative solutions and expand our analytical perspectives. Their combined expertise has profoundly shaped this project, and their ongoing encouragement was a constant source of motivation.

We are grateful to Toby Gooley for her meticulous attention to formatting and editing our paper, carefully reviewing multiple drafts until the final version was perfected. Additionally, we extend a special thank you to our colleagues and the CTL staff for providing a wonderful experience throughout the program.

Finally, we would like to provide personal notes below.

First, I want to thank my family for their moral support throughout this project. Mom, Dad, and James, I will always remember your love and support throughout this year. Finally, I owe a big thank you to Oscar Bonet. I could not have asked for a better capstone partner for this project. It was a pleasure working with you this past year, and I am proud of the work and progress we were able to create together.

Joshua Weston

I want to thank my father for granting me this opportunity and for his continuous support and motivation. Thank you to my brother and grandmother for always believing in me and helping me to dream big. I am grateful to my girlfriend, Andrea, who has been a huge emotional support during this capstone and throughout this journey. I also want to thank my partner, Josh, for being an amazing teammate this past year and for offering great support during the ups and downs of the project. Finally, a special thank you to my mother and grandfather, who have been my constant inspiration in every challenge.

Oscar Bonet

Table of Contents

1 Introduction

The trucking industry is a massive market in the United States. The full truckload market in the United States is valued at $403.8 billion, while the dedicated or private market is valued at $395.8 billion. Furthermore, the less-than-truckload market represents a total transportation spend of $96.3 billion in the US (Kearney, 2023). In this large and dynamic market, freight brokerages have emerged as key intermediaries.

Freight brokers serve as the middlemen that match shippers, companies that ship goods to their customers, with carriers, logistics companies that offer for-hire transportation services for customer shipments. Brokers allow shippers to quickly secure transportation capacity for their shipments from a variety of carriers while also granting carriers access to a broader range of potential customers.

The trucking industry in the United States is fragmented and dynamic due to the low barriers to entry for carriers. As of 2024, there were approximately 2.04 million heavy and tractor-trailer truck drivers employed in the United States (US Bureau of Labor Statistics, 2023). Additionally, there are approximately 773,886 carriers that own or lease tractor trailers, and of these carriers, 95.8% operate 10 or fewer trucks (US Department of Transportation, 2023). Consequently, brokers provide value to the fragmented carrier market as carriers and shippers utilize broker services to provide smoother communication between demand and supply (DAT, 2023).

While freight brokerages serve as the intermediary between carriers and shippers, they are still exposed to financial risks in the form of payment defaults. When shippers request that goods be moved, they enter into a contractual agreement with a freight brokerage. Payment terms between brokerages and shippers tend to be net 30 or 60, meaning brokerages can expect payment in 30 or 60 days (Denim, NA). After the brokerage has matched a carrier to move the shipper's goods and the carrier has delivered those goods with a signed bill of lading, the carrier generally expects payment from the brokerage more quickly than a shipper will pay the brokerage for the service provided. A payment default for the freight brokerage occurs when a shipper does not pay all of the fees associated with services provided by the carrier, since the freight brokerage has already paid the carrier. While freight brokerages have processes in place to help collect on these unpaid receivables, if all or partial payment is considered uncollectible, the freight broker must write off this unreceived revenue as an expense, incurring a loss. This is a common problem for freight brokerages. For that reason, brokers budget an allowance for doubtful accounts.

1.1 Motivation

The project sponsor is a freight brokerage interested in reducing the financial impact associated with write-offs where shippers do not fully pay for services carriers provide. The sponsor has written off millions of dollars in past years. These write-offs include both total write-offs (where an entire payment for a service provided is forgone) and partial write-offs (where the company receives partial compensation from the shipper). The sponsor aims to assess the probability of a write-off before brokering each deal to reduce its financial risk and improve revenue collection rates.

The organization seeks to unravel the relationship between various shipment characteristics and the propensity for write-offs, enabling strategic decision-making to avoid high-risk shipments and prioritize revenue-collection efforts. As of October 2023, the organization has developed a Standard Operating Procedure (SOP) for its agents to follow when signing agreements. The process emphasizes the importance of communication with customers to align expectations on the level of service, base prices of the deal, and accessorial fees that could be added if there are changes to the shipment. This SOP is one of the actions the company has in scope to tackle write-off impact, along with the investigation taking place in this capstone.

1.2 Problem Statement

The company has quantified the financial impact and documented the reason codes for each of these write-offs; however, they have not successfully developed a predictive model to identify write-offs in advance or identify the root cause(s) for these write-offs. The sponsor has access to a large amount of data surrounding each shipment but seeks guidance on which characteristics of a shipment would be most relevant in predicting outcomes or in preventing these write-offs from occurring in the first place.

The questions this study looked to answer in the context of write-offs were the following:

1. What are the leading indicators that suggest a likely shipment write-off during the tendering process?
2. How does each leading indicator contribute to the potential write-off?
3. Can these indicators be utilized to identify future shipments at risk of being written off?
4. Can these indicators be utilized to identify the root cause of these write-offs?
5. If a shipment is at risk of being written off, what is the magnitude of that expected write-off?

1.3 Project Goals

The project provides the company with a series of models that identify the probability that a write-off will occur for each future shipment that the sponsor evaluates. The results from these models indicate the risk of each deal for two types of write-offs: partials (where less than 100% of the shipment is written off) and full write-offs (where 100% of the shipment is written off). A separate model provides an estimate of the monetary value of the write-off. Furthermore, the project makes recommendations that the sponsor can employ to potentially resolve the root cause for certain write-offs to reduce the occurrence and magnitude of these write-offs.

This study hypothesized that through the analysis of anonymized shipment data and the development of a predictive model, it is possible to discern the weight of different indicators contributing to the probability of a write-off occurring. Furthermore, the output of this hypothesized model can be used as an input into a different continuous regression model to identify the magnitude or dollar value of a predicted write-off.

The final deliverables for the sponsor company were as follows:

1. **Identified leading indicators:** Identified crucial variables and indicators that influence the likelihood of a shipment being written off.
2. **Predictive model for write-off probability:** Developed a predictive model to accurately assess the probability of write-offs, enabling strategic decision-making.
3. **Predictive model for write-off magnitude:** Constructed a subsequent model to estimate the anticipated value of the write-off.
4. **List of strategic recommendations:** Presented actionable strategies for the sponsor to consider in order to mitigate the occurrence of write-offs.

1.4 Outcomes

Initially, descriptive statistics were utilized to gain insights into the frequency and distribution of written-off shipments across different parameters. Subsequently, correlation analysis was conducted to identify potential relationships between the identified leading indicators and the occurrence of write-offs. Furthermore, predictive modeling techniques, such as logistic regression and continuous linear regression, were employed to develop a predictive model for assessing the likelihood of a shipment being written off and the value of a predicted write-off. Strategic recommendations were provided to the sponsor company, leveraging the coefficients from the results from these models.

The rest of this study is organized as follows: Chapter 2 presents a review of relevant literature; Chapter 3 presents an overview of the current process used by the sponsor company; Chapter 4 outlines data characteristics and methodology used; Chapter 5 describes the data analysis and modeling results; and Chapter 6 provides a summary and recommendations for future research.

2 State of the Practice

This state of the practice chapter explores the methodologies and frameworks relevant to modeling financial distress, or write-offs, in the freight brokerage industry. It reviews various predictive modeling techniques and their applications in predicting financial risks. By examining academic studies and industry practices, this review aims to establish a foundation for developing a predictive model tailored to the specific needs and challenges of the freight brokerage sector.

2.1 Approach to Modeling Financial Distress (Write-Offs)

In managing financial risks within the freight brokerage industry, a tailored approach to predictive modeling is key. Wu et al. (2021) provides a foundational methodology that aims to improve the accuracy of firm valuation. Their research, which focuses on predicting asset write-offs using Support Vector Machine (SVM) and Support Vector Regression (SVR), combines both financial and managerial incentive factors, offering insights into developing a two-stage predictive model for firms in order to provide a more accurate view of business performance (Wu et al., 2021). While not industry-specific, their approach can be adapted to the freight brokerage industry. Their findings showed that these SVM models slightly outperformed logistic regression in the binary prediction of a write-off. However, the difference was not found to be significant in comparison to logistic regression. In addition, SVR was more accurate than logistic regression when viewing the continuous prediction of the magnitude of a write-off. This accuracy improvement is likely attributable to a finding that these learning techniques become more accurate as the number of nominal variables increases proportionately to quantitative variables (Hans et al., 1996).

While machine-learning models such as neural networks offer high accuracy in their predictions, they lack the explainability that simpler models like logistic regression provide (Wanner et al., 1970). Logistic regression offers a clear mathematical framework where the influence of each independent variable on the outcome is quantified through coefficients. These coefficients directly represent the strength and direction of the relationship between each independent variable and the dependent variable.

3 Current Process

The current process chapter provides an overview of the existing processes used by the sponsoring company for managing accounts receivable and allowances for uncollectibles, or write-offs. By highlighting the strengths and limitations of the current system, it is possible to set the stage for how a predictive model can enhance the organization's ability to anticipate and manage financial risks associated with write-offs.

3.1 Understanding Allowance for Write-Offs

From a high level, the firm budgets for uncollectible accounts (write-offs) and reports variances to this budget quarterly. Currently, the freight brokerage creates a contra-account allowance for uncollectible revenues in anticipation of future delinquencies in payments from shippers. Using historical write-off data, the firm has created a set of percentage values of the total accounts receivable that will likely be written off. The firm creates different percentages based on how long the receivable is past payment. As accounts receivables age, the likelihood that it is uncollectible increases. This method applies these percentages to the receivable revenues in each age bucket to develop a "budget" for uncollectible revenues or write-offs. By using this process, the firm anticipates the financial impact of future write-offs.

3.2 Collection Process for Accounts Receivable

The accounts receivables team at the firm ensures that accounts receivables from shippers are received in full. When shippers pay in full and on time, no action is needed from the team. However, when payments are not received, this team has a set of operating procedures to collect these expected revenues. These procedures include communicating directly with shippers to understand their situation, creating new terms for receiving funds, such as breaking down payment installations and utilizing debt collector services. The team reserves legal action for significant collectibles or when fraudulent or illegal activities are suspected to have occurred.

3.3 Performing a Write-Off

Once all avenues for collecting revenues have been exhausted, the firm will write off the uncollectible revenue as an expense to the uncollectible revenues account. The firm currently has a process in place of writing off uncollectible accounts receivable once the receivable reaches the 120-days late mark. If the accounts receivable team deems a revenue as uncollectible before this timeframe, a reduction is made to the accounts receivable by a transfer to the bad debt account, classifying the debt as uncollectible.

The current system focuses on accurately forecasting the value of write-offs in a given future period. However, the firm does not have a system in place to predict whether a shipment will be written off. Therefore, the firm cannot take action in real time to prevent a write-off from occurring.

4 Methodology

In this chapter, the dataset provided by the sponsor company is explored, the methodology for correlation analysis and predictive modeling is explained, and alternative approaches considered during the study are discussed.

4.1 Data Preview

The foundation of this capstone is based on the sponsor company's existing processes and datasets, which span approximately five years (from July 2018 to December 2023). Information was provided from two sources, which was then treated and merged into a consolidated dataset. The main set of data is related to the shipment characteristics, including the shipment pickup date, terms of payment, mode of transportation, product category, geographical information of pickup and delivery, and the rate applied to the customer for the specific shipment; this dataset additionally shows details specific to the customer, like the customer's industry, first shipment with the sponsor, or customer segment. Detailed data on these shipment characteristics can be found in Table A1 in the Appendix. Shipment information for each write-off the finance department had performed was provided, including information specific to the write-off. This write-off dataset contains 56 different types of codes, describing the reason for write-off occurrence. Each shipment can have multiple write-offs. The data dictionary describing the characteristics of each provided feature in the dataset is included in Table A1 (see Appendix).

A notable characteristic of the dataset is the disproportionately low representation of shipments that have been written off relative to those without issues, signifying a marked imbalance within the data. This discrepancy necessitates modifications in the modeling approach to accommodate this imbalance adequately. Specifically, when the minority class—shipments written off—constitutes less than 1% of the dataset, the model encounters difficulties discerning patterns that predict write-offs due to insufficient examples. Therefore, an oversampling technique was applied to increase the representation of the underrepresented shipment write-offs. This method duplicates the minority class entries until there is an even representation of both shipments with write-offs and without, enhancing the model's ability to identify and learn the relevant patterns for predicting write-offs.

4.2 Summary Statistics

As an overview of the summary statistics of each feature, Table 1 presents the mean, minimum, 25% quartile, median (or 50% quartile), 75% quartile, and maximum values for each numerical feature, while Table 2 details the mode and the percentage occurrence for each categorical feature. In total, there were approximately 4.3 million observations in the provided dataset. These tables include only those features deemed statistically significant in the models. Further analysis of these features is discussed in Section 5.

It is important to note that the most frequently occurring datapoint (mode) for Terms of Payment in the dataset is 30 days, with 73% of instances holding this value. Therefore, the 25% and 75% quartiles show the same value. The minimum value for payment terms is 0 days with a small representation in the dataset, where customers are requested to pay upfront. Additionally, the 'days since the last shipment' has a mean of 3 and a median of 0, illustrating that the data is skewed to the right, with a few customers who have not contracted another shipment with the sponsor company in many days.

Table 1

Summary Statistics (2018-2023) for the significant numerical features

Feature	Mean	Min	Q25%	Median	Q75%	Max
Terms of Payment (days)	35	0	30	30	30	90
Miles (miles)	562	0	136	373	774	8,496
Days Since Last Shipment (days)	3	0	0	0	1	2,070
Credit Limit (US dollars)	1,511,683	0	18,000	150,000	890,000	25,000,000
Agent Tenure (days)	857	0	331	668	1,239	9,033
Days as customer (days)	1,219	0	447	1,025	1,792	4,416

Table 2

Summary Statistics (2018-2023) for the significant categorical features

Feature	Mode	Mode share %
Customer Segment	Small and Medium Sized Business (SMB)	46.69%
Mode	Full Truck Load (FTL)	66.16%

4.3 Selected Data and Targets

For modeling, data prior to 2022 was excluded in order to avoid trends that are unique to the COVID-19 pandemic period. Having excluded those years, the models were trained based on more than one million shipments, to be tested against approximately two hundred thousand shipments in the first quarter of 2023 to validate each model's performance. Furthermore, recent 2023 data was excluded because the write-off business process of the sponsor company takes a minimum of 120 days to complete. Therefore, the risk of testing these models on data that does not yet include the complete count of write-offs is avoided. Once the date range was determined, the same range was used for training and testing across all models to ensure a fair side-by-side comparison.

Different underlying correlations were identified for shipments that were fully written off compared to those that were partially written off. Consequently, this behavior was investigated within the predictive modeling framework, utilizing both types of write-offs as separate target variables to enhance the respective metrics of the models.

4.4 Modeling

This section explains the methodology used to predict both full and partial write-offs using binary logistic regression models. It focuses on deriving practical insights, preferring logistic regression for its clarity over more complex models.

4.4.1 Statistical Models

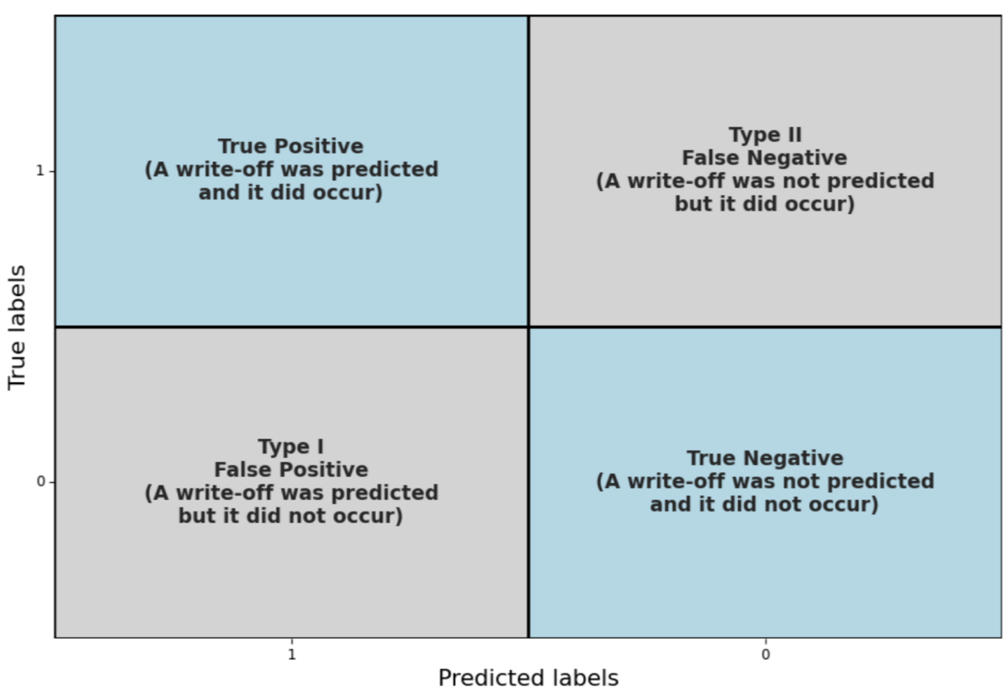
This model calculates the probability of each shipment being written off, based on various input features. It then classifies each shipment as either a non-write-off (0) or a write-off (1) by applying a threshold value, which can range from 0 to 1 (Figure 1). To demonstrate and evaluate the results of a binary classification model, the output will be classified using a confusion matrix. A confusion matrix is a diagram that maps the results into four quadrants. Those predicted correctly are either in the True Positive or the True Negative quadrant. And those wrongly forecasted lie in the False Positive or False Negative quadrants. The False Positives (also known as Type I errors) are those misclassified as having a write-off while they actually did not get written off. The False Negatives (Type II errors) are those misclassified as not being written off but actually have been written off.

Figure 1 and the bullet points below explain how a confusion matrix is interpreted for each quadrant:

- **The upper left quadrant** represents a *True Positive* where a write-off was predicted to occur, and it did occur.
- **The lower right quadrant** represents a *True Negative* where a write-off was not predicted to occur, and it did not occur.
- **The upper right quadrant** represents a *False Negative* *(Type II Error)*, where a write-off was not predicted to occur, but a write-off did occur.
- **The lower left quadrant** represents a *False Positive* *(Type I Error)*, where a write-off was predicted to occur, and it did not occur.

Figure 1

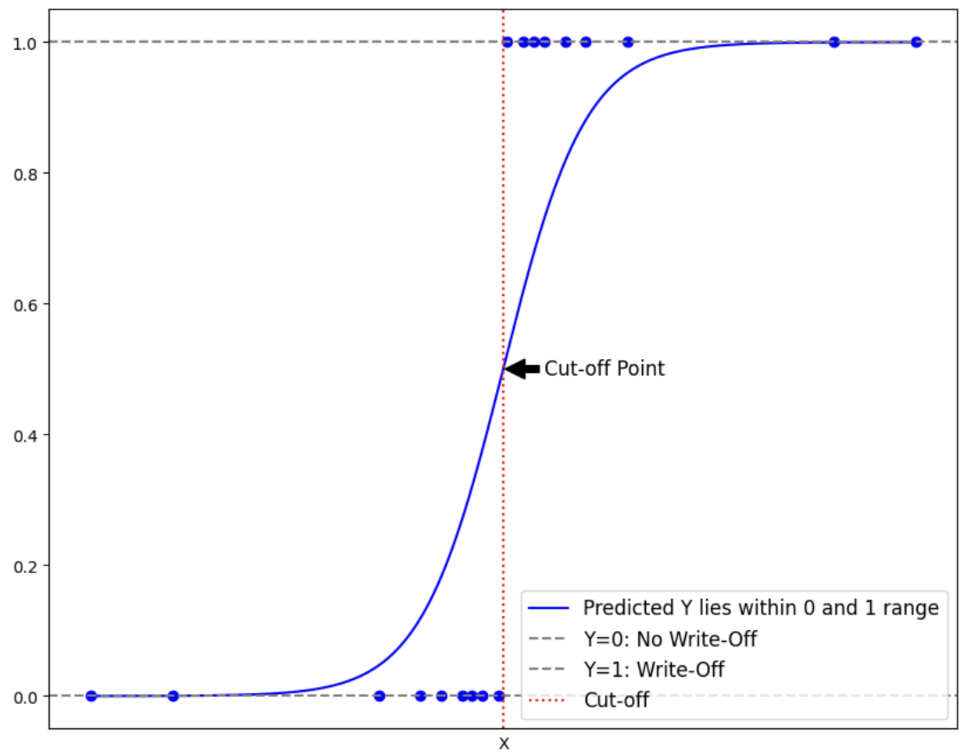
Example of Confusion Matrix for Write-Off Predictions



For this analysis, a threshold or cutoff is adapted to each model, aiming to parametrize the ability of the model to capture high-risk shipments while minimizing the risk of Type II errors (false negatives). This is due to the business need of the sponsor to ensure the model is able to flag most of the potential shipments susceptible to leading to a write-off. Figure 2 illustrates how the model rounds the output to 1 or 0 based on a given threshold and the computed value for each shipment.

Figure 2

Binary Logistic Regression Model Example



This specific threshold was selected balancing the trade-off of the model's accuracy and recall. Accuracy measures the number of correct predictions divided by the total number of predictions as referred in formula (1), while recall, also known as sensitivity or true positive rate, measures the proportion of actual positives that are correctly identified by the model as seen in formula (2). Recall was considered with the same weight of importance as accuracy, as the sponsor company considered Type II errors (mislabeling a write-off as a non-write-off) as higher risk than Type I errors (labeling a non-written-off shipment as a write-off).

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + False\ Positive + True\ Negative + False\ Negative} \quad (1)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (2)$$

Following the initial descriptive analysis, two binary models were developed for both partial and full write-offs. These models incorporated various predictors, including industry type, shipment characteristics, and historical payment behaviors. These characteristics were evaluated through correlation analysis, and certain factors were excluded to 1) avoid multicollinearity, 2) remove insignificant factors based on p-values, and 3) simplify model inputs to reduce the likelihood of over-fitting.

Next, a linear regression model was developed to quantify the magnitude of partial write-offs. This model considered similar factors, such as distance of haul, industry type, and equipment type in assessing the financial risk associated with each shipment. However, certain characteristics were excluded due to insignificant p-values. This continuous model was used only to evaluate partial write-off magnitudes, as full write-offs already comprise the full amount of the customer rate, or the total amount charged to the customer for a shipment. This model was evaluated using R-squared and mean squared error (MSE).

Finally, by analyzing the coefficients derived from these models, specific problem areas were identified where the sponsor company can focus efforts to mitigate the risk of write-offs and realize expected positive financial impacts. Overall, the methodology facilitated a comprehensive understanding of write-off risk factors and enabled the development of practical strategies for reducing their occurrence.

4.4.2 Machine Learning Models

To enhance the robustness of the methodology, the machine learning models XGBoost and Random Forest were evaluated for predictive accuracy, and their feature importance was considered—a technique that assigns scores to model features based on their influence on outcomes. Despite the predictive strengths of these models, binary logistic and linear regression models were selected for their coefficient interpretability, which is crucial for the sponsor company. Although Random Forest and XGBoost, which improve prediction by combining decision trees and optimizing weighted models, respectively, offer significant predictive power, their "black box" nature complicates understanding the effects of individual features. This lack of interpretability, especially in determining the positive or negative impact of variables, does not meet the sponsor company's need for actionable insights for root-cause analysis. However, in terms of accuracy and recall, these models slightly outperform logistic regression.

5 Results

The findings from this study are presented in two sections: 1) Binary Logistic Regression Models, and 2) Continuous Regression Model. The section on binary logistic regression models discusses the performance similarities and differences between the full and partial write-off models. Meanwhile, the section on the continuous regression model outlines key takeaways from the continuous model analysis.

5.1 Binary Logistic Regression

The logistic regression models developed to predict the likelihood of full and partial write-offs for shipments yielded several significant predictors, as indicated by their coefficients and associated p-values in sections 5.1.1 through 5.1.3.

5.1.1 Full Write-Offs

This section reports on the results from running a binary logit model to predict the likelihood of a full write-off occurring based on the following table of predictor variables. The feature name, resulting coefficients, and resulting p-values were included below (Table 3). All statistically insignificant factors were excluded from the model.

Table 3

Features with Coefficients and P-values for Logistic Regression Full Write-Off Model

Feature	Coefficient
Customer Segment: Enterprise	0.47***
AutoBuilder	-0.95***
Terms of Payment 0 days	-1.27***
Terms of Payment 10 days	-0.81***
Terms of Payment 15 & 30 days	-0.97***
Terms of Payment 45	-1.27***
Terms of Payment 75 days	-1.39***
Terms of Payment 90 days	-0.75***

Mode: Drayage	1.03***
Mode: Final Mile	0.77***
Mode: FTL	0.07***
Miles of Shipment: (0-10)	0.74***
Miles of Shipment: (50-100)	0.20***
Miles of Shipment: (250-1000)	-0.20***
Miles of Shipment: (1000+)	-0.41***
Days Since Last Shipment: 0-15 days	0.17***
Days Since Last Shipment: 30-180 days	0.63***
Days Since Last Shipment: >365+ days	0.31***
Credit Limit: $0-$25K	0.92***
Credit Limit: $25K-$100K	0.42***
Credit Limit: $800K-2.175M	-0.50***
Credit Limit: $2.175M+	-0.72***
Days As a Customer: 0-90 & 365+ days	0.43***
Agent Tenure: First Year (90-365 days)	-0.22***
Agent Tenure: Veteran (2+ years)	-0.07***
** for p-values < 0.05 (significant at the 5% level), ** for p-values < 0.01 (significant at the 1% level), and *** for p-values < 0.001 (significant at the 0.1% level)*	

To interpret the coefficients, it is crucial to note that a one-unit increase in the predictor variable leads to a β change in the log odds of the dependent variable. A one-unit increase in a feature results in a β increase in the log odds of a write-off occurring. For example, the odds of a write-off occurring when the customer segment is an enterprise customer is $e^{.47} \approx 1.60$ or approximately 60% higher than other customer segments, keeping other factors as fixed. To convert the log odds to probability, the following logistic function is utilized in formula (3):

$$P(Y = 1) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_n X_n)}} \quad (3)$$

Where:

- *P(Y=1)* is the probability of the event occurring
- β is the estimated coefficient for each feature
- *X* is the predictor variable

In logistic regression, reference categories establish the baseline for comparisons. For the Full Write-Off Binary Logit Regression Model, reference categories were chosen based on 1) the most frequently occurring categories of each feature and 2) preliminary findings indicating a lack of statistical significance when treated separately, thereby simplifying the model and enhancing interpretability. Here are the selected reference categories for each feature:

- **Customer Segment:** Small and Medium-Sized Business (SMB) and Freight Forwarder
- **Customer Tenure:** First Year (90-365 days)
- **Miles Driven:** 50-250 miles
- **Mode of Transportation:** Expedited and Others
- **Terms of Payment:** 21 days and 60 days
- **Agent Tenure:** Second Year and New Agent (0-90 days)
- **Days Since Last Shipment:** 15-30 days and 180-365 days
- **Credit Limit:** $100K-285K

These categories serve as the comparative baseline for each variable in the model, facilitating a focused analysis on significant differences across other categories. Customer segments, payment terms, shipping modes, shipment mileage, days since the last shipment, credit limits, days as a customer, and agent tenure were all relevant factors in predicting shipment write-offs:

- **Customer Segment:** Refers to the classification of customers based on company size, which influences transaction dynamics. Enterprise customers are defined as large organizations that typically engage in higher volumes of transactions with the sponsor company. Enterprises are more likely to experience full write-offs, as indicated by a positive coefficient of 0.47. This suggests that shipments involving enterprise customers carry a higher risk of write-off, which may occur due to the complex and large-scale nature of their operations.
- **AutoBuilder**: AutoBuilder refers to the system that the sponsor company utilizes to automatically broker agreements between shippers and carriers. The coefficient for AutoBuilder was negative and statistically significant ($\beta = -0.95$, $p < 0.001$), suggesting that shipments processed using the system to automatically create shipments are associated with a decreased probability of a full write-off compared to manually processed shipments. This suggests that this system mitigates the risk of a full-write off occurring.
- **Payment Terms:** Indicates the agreed upon time frame within which customers must settle their invoices. Shorter payment terms generally correlate with a lower likelihood of write-offs. The coefficients become increasingly negative as the payment term

lengthens, from -0.81 for 10 days to -1.39 for 75 days. This pattern indicates that longer payment terms exceeding 75 days may increase the financial risk associated with the shipment. Payment terms have a low correlation with the enterprise customer category or the days as customer as the correlation coefficient is lower than 0.40.

- **Shipping Mode:** Relates to the method of transport used for shipments. Different shipping modes exhibit varying risks. Drayage (delivery over a short distance, typically from a port to a nearby railyard, warehouse, or other destination) shows a significantly higher risk (coefficient of 1.03), whereas full truckload (FTL) shipping shows a minimal increase in risk (coefficient of 0.07). Each shipment can only have one mode assigned to it, so each are mutually exclusive.
- **Mileage of Shipment:** Shorter shipments, specifically those covering distances of 0-10 miles, are associated with an increased likelihood of being written off, as indicated by a positive coefficient of 0.74. Correlations with shipments of 0-10 miles and 10-50 miles were -.01 and .19 respectively. Conversely, longer shipments, particularly those exceeding 1,000 miles, show a decreased probability of write-off, demonstrated by a negative coefficient of -0.41.
- **Days Since Last Shipment:** Measures the time interval since the last shipment was processed for a customer. More frequent shipments (days since last shipment <15) show a minimal increase in risk (coefficient of 0.17). Customers that recurrently ship within 30 to 180 days show the highest risk of a full write-off, with a coefficient of 0.63. And those shipments that are sporadic, with less than 1 shipment per year, drive an increase in the probability of a write-off with a 0.31 coefficient.
- **Credit Limit:** Indicates the maximum amount of credit extended to a customer, reflecting trust and financial risk. Lower credit limits (e.g., ≤$25K) are associated with higher risk (coefficient of 0.92), while very high credit limits (e.g., $2.175M+) correlate with lower risk (coefficient of -0.72).
- **Days as a Customer:** Represents the duration of the business relationship with a customer, affecting familiarity and operational predictiveness. Newer relationships (customer for <90 days) show slightly increased risk factors (coefficients of 0.43), suggesting that established relationships may reduce the risk of write-offs.
- **Agent Tenure:** The length of time an agent has been with the sponsor company influences experience and expertise in managing shipments. The analysis indicates that agents with two years of experience have a higher likelihood of creating a load with a full write-off, as evidenced by their positive coefficient relative to other agent tenures

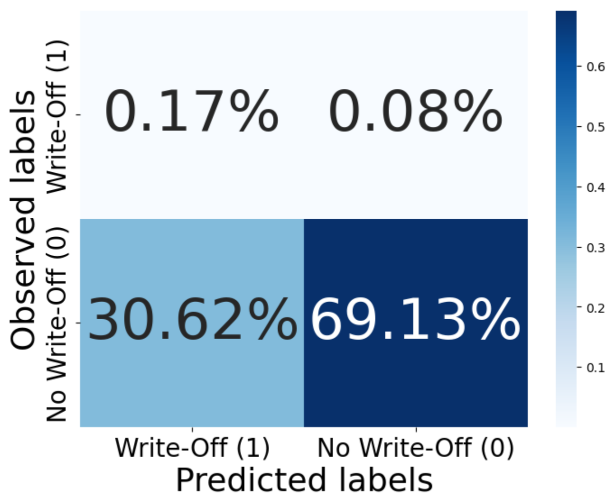
(both new and veteran), which display negative coefficients. New agents may have better performance in terms of write-offs as they are trained more recently and have more direct oversight from more veteran employees.

5.1.1.1 Confusion Matrix Analysis

Additionally, a confusion matrix was generated to evaluate the performance of the binary logit model (Figure 3). The confusion matrix is structured as follows:

Figure 3

Full Write-Off Binary Logit Confusion Matrix



The evaluated metrics for the confusion matrix above were recall and accuracy, which are calculated as follows in formula (4) and (5):

$$Recall: 68.00\% = \frac{0.17\%}{0.17\% + 0.08\%} \tag{4}$$

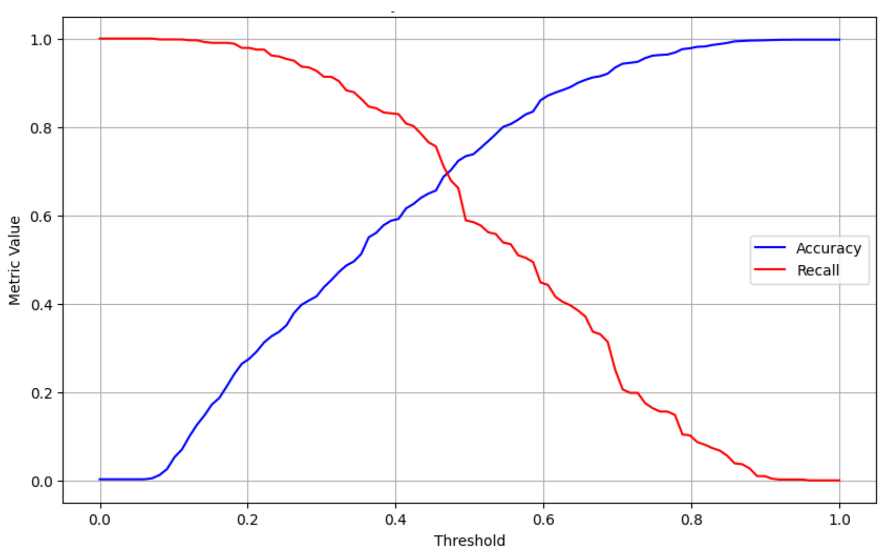
$$Accuracy: 69.30\% = 0.17\% + 69.13\% \tag{5}$$

For this model, the threshold selected is 0.4705, to balance the accuracy and recall. This means that any combination of characteristics that yields a probability over 0.4705 will be predicted as a write-off, while anything under that will be categorized as “not a write-off.”

The model assigns probabilities of write-off occurrence based on the characteristics of the shipments within the test dataset. To evaluate the model's performance, a threshold ranging from 0 to 1 is employed to determine whether a given probability should be rounded to 1 (indicating a write-off) or to 0 (indicating no write-off), thus converting the probabilistic output into a binary classification. Figure 4 illustrates the inherent trade-off between model accuracy and recall as the threshold is adjusted. A heightened threshold mandates a greater computed probability for a load to be classified as a write-off, enhancing model precision—particularly in light of the dataset's imbalance—but concurrently diminishing recall. Conversely, a diminished threshold lowers the probability requisite for a write-off classification, potentially increasing recall at the expense of accuracy. The intersection of the accuracy and recall curves represents an equilibrium point where the capacity of the model to predict full write-off over the total full write-off occurrence is matched with the general accuracy of the model to successfully categorize a shipment in either a write-off or not a write-off.

Figure 4

Trade-off between the model's accuracy and recall with different thresholds



5.1.2 Partial Write-Offs

This section highlights the results from running a binary logit model to predict the likelihood of a partial write-off occurring based on the following table of leading indicators. The feature name, resulting coefficients, and resulting p-values were included below (Table 4). All statistically insignificant factors were excluded from the model.

Table 4

Features with Coefficients and P-values for Logistic Regression Partial Write-Off Model

Feature	Coefficient
Customer Segment: Enterprise	0.16***
AutoBuilder	1.15***
Terms of Payment 0-10	-0.60***
Terms of Payment 15 days	-0.37***
Terms of Payment 30 & 45 days	-0.15***
Terms of Payment 75 days	-0.60***
Terms of Payment 90 days	0.19***
Mode: Drayage	1.75***
Mode: Final Mile	0.67***
Mode: FTL	0.44***
Miles of Shipment: (0-10)	-0.28***
Miles of Shipment: (50-100)	0.18***
Miles of Shipment: (250-1000)	-0.06***
Miles of Shipment: (1000+)	-0.18***
Credit Limit: $0-$25K	0.10***
Credit Limit: $285K-$800K	-0.35***
Credit Limit: $800K-$2.175M	0.32***
Credit Limit: $2.175M+	-0.62***
Agent Tenure: New (0-90 days)	-0.39***
Agent Tenure: First Year (90-365 days) and Veteran (2+ years)	-0.18***
Days As Customer: New (0-90 days)	0.36***
Days As Customer: 2nd Year	0.16***
Days As Customer: Legacy (2+ years)	-0.29***
Days Since Last Shipment: 60-90 days	-1.03***
Days Since Last Shipment: 0-15, 30-60, 90-180 days	-0.84***
Days Since Last Shipment: 365+ days	-0.46***
* for p-values < 0.05 (significant at the 5% level), ** for p-values < 0.01 (significant at the 1% level), and *** for p-values < 0.001 (significant at the 0.1% level)	

For the Partial Write-Off Model, reference categories were chosen based on 1) the most commonly occurring categories of each feature and 2) preliminary findings indicating a lack of statistical significance when treated separately, thereby simplifying the model. Here are the selected reference categories for each feature:

- **Customer Segment:** SMB and Freight Forwarder

- **Customer Tenure:** First Year (90-365 days)
- **Miles Driven:** 10-50 and 100-250 miles
- **Mode of Transportation:** Expedited and Others
- **Terms of Payment:** 21 days and 60 days
- **Agent Tenure:** Second Year and New Agent (0-90 days)
- **Days Since Last Shipment:** 15-30 days and 180-365 days
- **Credit Limit:** $100K-285K

These categories serve as the comparative baseline for each variable in the model, facilitating a focused analysis on significant differences across other categories.

This regression analysis aimed to predict the likelihood of partial write-offs in shipments. The model included various categorical predictors, each represented through dummy variables, with coefficients indicating their relative influence on the outcome when other variables are held constant. All predictors showed significant results ($p < 0.001$). The results revealed significant coefficients for several predictors. The interpretation for the largest influenced factors is included below:

- **Customer Segment:** Enterprises are more likely to have partial write-offs, with a coefficient of 0.16, suggesting that larger organizations have slightly higher risk factors possibly due to the scale and complexity of their operations.
- **AutoBuilder:** The shipments that were automatically created through the Autobuilder system had a strong positive coefficient of 1.15, indicating a significant increase in the likelihood of partial write-offs. This might reflect specific vulnerabilities or challenges in shipments involved in the automated building system such as marginal inaccurate quotes to customers.
- **Terms of Payment:** Different payment terms show varying impacts. Shorter (0-10 and 15 days) payment terms are negatively associated with partial write-offs, with coefficients of -0.60 and -0.37, respectively, suggesting quicker payment terms are less likely to incur a partial write-off. Longer terms (90 days), with a coefficient of 0.19, slightly increase the risk.
- **Mode of Shipment:** Drayage and Final Mile shipping modes increase the likelihood of partial write-offs, with coefficients of 1.75 and 0.67, respectively, highlighting higher risk associated with these types of transport when compared to others.
- **Miles of Shipment:** The distance covered by shipments also affects the likelihood of

partial write-offs. Short distances (0-10 miles) decrease the risk (coefficient -0.28), while intermediate distances (50-100 miles) have relatively higher coefficients when compared with longer-mileage shipments.

- **Credit Limit:** Lower credit limits (0-$25K) are associated with a slight increase in risk (coefficient 0.10), while very high limits ($2.175M+) show a strong negative association (coefficient -0.62), suggesting financial stability as a protective factor.
- **Agent Tenure:** New agents (coefficient -0.39) and those with more than two years of experience (coefficient -0.18) show lower risks of partial write-offs, indicating that agents with one to two years of experience (coefficient of 0) are most likely to create a shipment with a partial write-off.
- **Days as Customer:** New customers (0-90 days) exhibit a higher risk of partial write-offs (coefficient 0.36). However, customers with a longer relationship (legacy, 2+ years) show a reduced risk (coefficient -0.29), suggesting that longer established relationships reduce the likelihood of a partial write-off occurring.
- **Days Since Last Shipment**: Customers that ship with a frequency between 15-30 days and 180-365 days served as the base case for this feature, indicating the coefficient for these values is 0. Shipping recurrence of over 365 days reduces the likelihood of a partial write-off with a coefficient value of -.46. Meanwhile, shipments with a recent shipment of less than 15 days previously or between 30-180 days demonstrate the lowest probability of a partial write-off occurring (with a coefficient between -0.84 and -1.03).

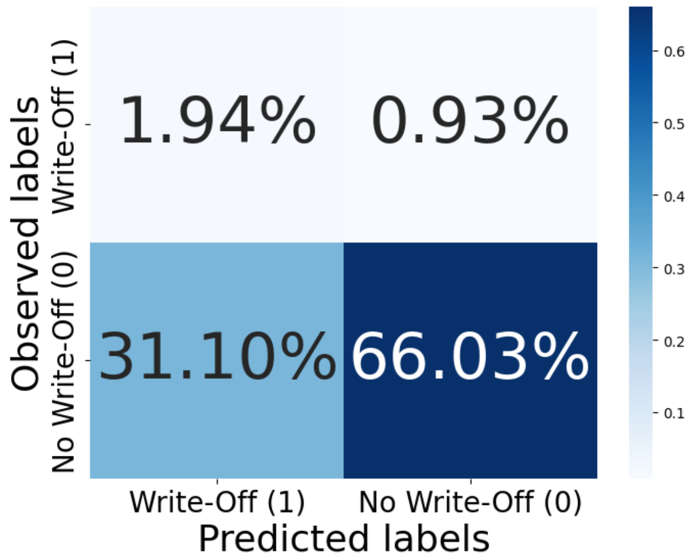
The analysis demonstrates that several factors significantly affect the likelihood of partial write-offs in shipments. The model suggests that particular attention should be paid to the type of customer, shipment mode, payment terms, and the temporal dimensions of customer engagement and agent experience.

5.1.2.1 Confusion Matrix Analysis

The following confusion matrix was generated to evaluate the performance of the binary logit model (Figure 5):

Figure 5

Partial Write-Off Binary Logit Confusion Matrix



The evaluated metrics for the confusion matrix above were recall and accuracy, which are calculated as follows:

$$Recall: 67.60\% = \frac{1.94\%}{1.94\% + .93\%} \tag{6}$$

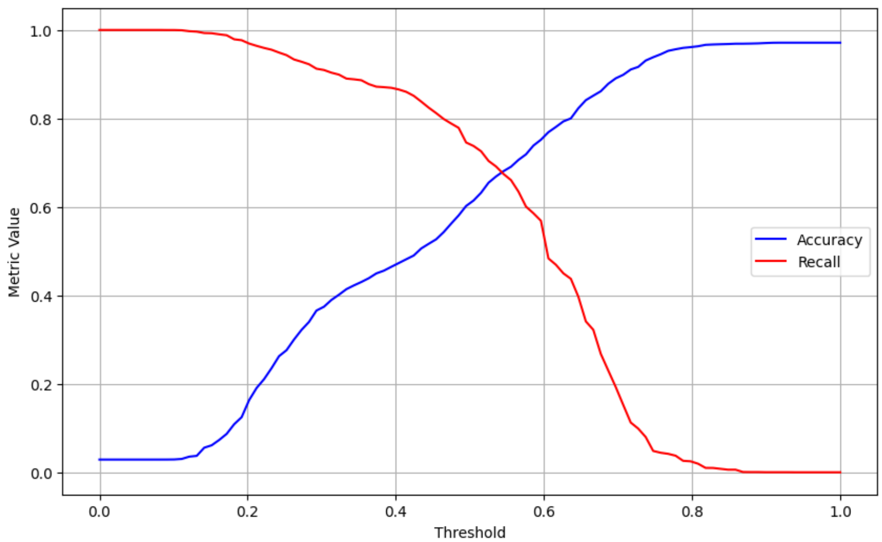
$$Accuracy: 67.97\% = 1.94\% + 66.03\% \tag{7}$$

The performance of the predictive model, as measured by accuracy and recall, was observed to be lower for predicting partial write-offs compared to full write-offs. A key factor contributing to this disparity is the distribution of the data set. Specifically, partial write-offs are more prevalent than full write-offs within the data. This imbalance leads to a greater opportunity for the model to correctly predict full write-offs simply because there are fewer instances for it to identify. In other words, the probability of making a correct prediction by chance is higher for full write-offs due to their lower occurrence rate. Consequently, while the model demonstrates a commendable capacity for identifying full write-offs, its performance metrics for partial write-offs are somewhat diminished due to the higher number of cases it must correctly classify, thus providing a sterner test of its predictive abilities.

For this model, the threshold selected is 0.54, to balance the accuracy and recall (Figure 6). This means that any combination of characteristics that yields a probability over 0.54 will be predicted as a write-off, while anything under that will be categorized as “not a write-off.”

Figure 6

Trade-off between the model's accuracy and recall with different thresholds



5.1.3 Contrasting Output of Partial vs. Full Binary Logit Model

The following section will outline the similarities and differences between the Partial and Full Write-Off Binary Logit Models. Table A3 in the appendix compares the coefficients assigned by each model side by side, illustrating the differences between the partial and full write-off models. In comparing the two models for full and partial write-offs, several features exhibit notably different behaviors in terms of the coefficients' signs and magnitudes:

- **AutoBuilder** shows a complete reversal between the models, with a strong negative impact (-0.95) for full write-offs and a strong positive impact (1.15) for partial write-offs. This illustrates that the systematic creation of brokered shipments may be effective in preventing full write-offs which are more costly in terms of magnitude. However, due to the strong positive impact for partial write-offs, it is clear that this system may have issues in smaller discrepancies in quoting customers their respective rates.
- **Terms of Payment of 90 days** also reverses, negatively influencing full write-offs (-0.75) and positively influencing partial write-offs (0.19). This implies extending payment terms to 90 days decreases the occurrence or amount of full write-offs, possibly by giving customers more time to fulfill their payment obligations and thus reducing total defaults.

The positive coefficient for the partial model (0.19), although weaker, suggests that longer payment terms marginally increase the occurrence or amount of partial write-offs.

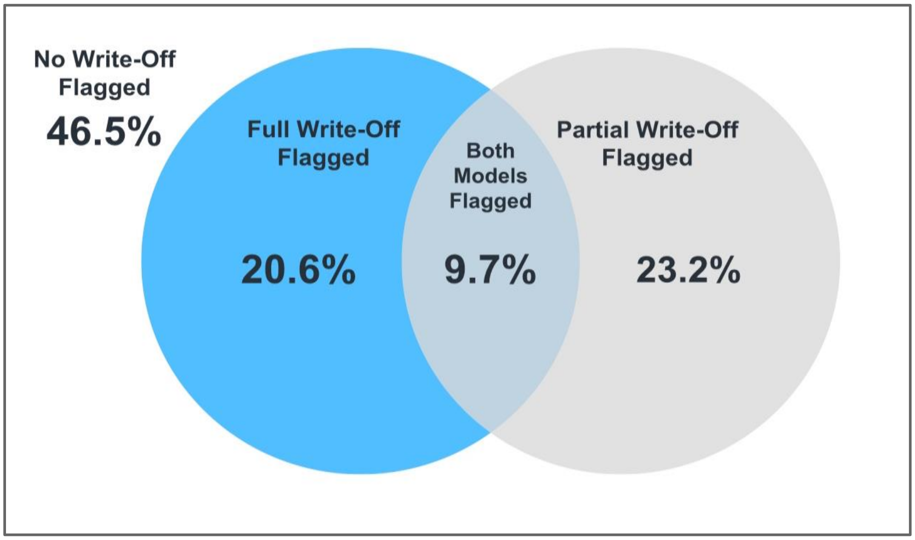
- **Drayage mode** has positive impacts in both models but is significantly stronger in the partial write-offs model (1.75 in partial model vs 1.03 in full model). This suggests that employing drayage mode increases the likelihood of both full and partial write-offs. The increased likelihood of write-offs with drayage mode might stem from the complexities of coordinating this transport type, which involves moving goods over short distances often between different transportation modes. Additionally, drayage operations often face stringent regulatory requirements, especially when moving goods through ports or across international borders.
- **Credit Limit of $800K-2.175M** shifts from a negative impact in full write-offs (-0.50) to a positive impact in partial write-offs (0.32). The negative impact for full write-offs (-0.50) suggests that higher credit limits reduce the likelihood or amount of full write-offs, perhaps because these customers have been evaluated using a credit score, implying strong financials.

These differences highlight some of the potential ways that partial and full write-offs behave within the provided dataset. Both models, for full and partial write-offs, were applied to the shipment data.

As depicted in Figure 7, the Venn Diagram illustrates the distribution of shipments flagged by the models: 20.6% were predicted as full write-offs, 23.2% as partial write-offs, 9.7% were flagged by both models simultaneously, and 46.5% of shipments were not identified as potential write-offs by either model. This bimodal distribution substantiates the necessity for employing two distinct models.

Figure 7

Venn Diagram showing the predictions of full and partial write-off models



5.2 Continuous Regression

The analysis presented focuses on the outcomes of employing a continuous Ordinary Least Squares (OLS) model to predict the percentage of a shipment that will be written off for partial write-offs, utilizing a dataset characterized by various operational and customer-related factors. The evaluation analyzes the model's coefficients, their significance levels, and overall predictive reliability in a testing dataset.

5.2.1 Ordinary Least Squares (OLS) Model

The continuous OLS model incorporated factors ranging from customer segmentation, duration of customer relationship, distance covered by the shipment, mode of transportation, terms of payment, and credit limits. As detailed in Table A2 (Appendix), coefficients for these factors varied considerably, indicating differing impacts on the probability of a shipment being written off.

Key predictors such as 'Terms of Payment: 0 days' and 'Miles (0-10)' exhibited strong positive coefficients, suggesting higher risks associated with immediate payment terms and shorter travel distances. Conversely, factors like 'Mode: Final Mile' and 'Credit Limit: $2.17M+' showed significant negative coefficients, implying lower write-off rates under these conditions.

The statistical significance of these coefficients was confirmed with p-values < 0.001, denoting robust associations within the dataset.

5.2.2 Error Analysis

In assessing the model's performance on the testing dataset, the results indicated a mean squared error (MSE) of 101.8 and an R-squared value of 12.3%. The high MSE points to considerable average errors in the prediction of write-off percentages, suggesting that the model predictions deviate significantly from actual outcomes. Figure A1 in the Appendix shows the distribution of the residuals. Moreover, the low R-squared value underscores a weak explanatory power of the model, as it accounts for only a small fraction of the variance in the response variable observed in the data.

5.2.3 Conclusions

The analysis of the continuous OLS model reveals that despite the statistical significance of individual predictors, the overall model performance is suboptimal in terms of predictive accuracy and reliability. The high error rates and low explanatory power necessitate a reconsideration of the modeling approach for this application. Given these limitations, further research should explore alternative modeling techniques that could potentially enhance predictive accuracy. Additionally, expanding the feature set or revisiting the data preprocessing stages might provide improvements in model outcomes.

In summary, while the continuous OLS model provided valuable insights into the factors influencing shipment write-offs, its application in operational settings is challenged by inadequate predictive performance. A strategic pivot to more sophisticated analytical methods is recommended to achieve more reliable and actionable results in future studies.

6 Discussion and Recommendations

In this section, the implications and significance of the findings from the predictive models are explored. These results are related to the specific challenges faced by the sponsoring company and consider broader impacts within the freight brokerage industry. Recommendations are offered to address identified risks, enhancing the sponsor's financial management strategies. Additionally, the limitations of the study, which may influence the interpretation and applicability of the results, are acknowledged.

6.1 Implications of the Results

The results of this series of predictive models provide critical insights into the factors that contribute to shipment write-offs within the freight brokerage industry, particularly for the sponsoring company. Several key variables have been identified that impact the likelihood of both partial and full write-offs. These include customer segment, terms of payment, shipping mode, mileage, days since last shipment, credit limits, agent tenure, and customer tenure.

The variable, ***Days Since Last Shipment***, has a correlation where customers with a last shipment greater than 60 days are more prone to full write-offs but are less likely to incur partial write-offs. This suggests that while these customers may engage with a broker less frequently, when they do, the transactions carry higher risk levels. Conversely, customers who have shipped a load within the last 15 days appear to mitigate risk for both full and partial write-offs, pointing to the benefits of continuous engagement from customers and shorter transaction cycles in reducing write-off risks.

Credit Limits showed a consistent pattern, aligning with risk levels where lower credit limits correlate with higher likelihoods of full write-offs, and higher credit limits correspond with reduced risks of partial write-offs. This finding underscores the importance of thorough credit assessments in managing financial risk. The methods in place at the sponsoring company for determining the credit risk of each shipper accurately align with the findings from these models.

The use of ***AutoBuilder*** in shipment processing revealed a dual impact: it lowers the risk of full write-offs but increases the risk of partial write-offs. This might suggest that while automatically created shipments help avoid issues that create full write-offs and reduce the chance of significant losses, they may introduce minor discrepancies or errors leading to smaller financial setbacks. Table 5 presents the summary statistics for the write-off percentages associated with manual versus automated shipments. It includes the percentage of shipments with partial write-offs generated automatically, along with the mean, 25th percentile, median, and 75th percentile of the written-off amounts. The data indicate that the use of AutoBuilder results in a reduced financial impact.

Table 5

Summary statistics for shipments with Partial Write-offs generated manually or with Autobuilder

AutoBuilder Flag	Shipments with Partial Write-Off (%)	Avg. Write-Off (%)	25% Q	Median	75% Q
Manual	49.7%	11.5%	1.2%	5.4%	15.3%
Automated	50.3%	3.9%	0.3%	1.0%	2.8%

Agent Tenure and ***Customer Tenure*** also play significant roles. Newer agents and customers show higher risks than veteran agents and legacy customers, indicating the necessity for rigorous training for the agents, and onboarding processes for the customers to mitigate early errors and misunderstandings that could lead to write-offs.

In the broader context of the freight brokerage and logistics industries, these findings contribute to the ongoing discussions around risk assessment and financial management. The study demonstrates the applicability of predictive modeling techniques such as logistic and linear regression in a practical, industry-specific setting, providing a roadmap for other firms facing similar challenges. It highlights the potential of data-driven decision-making in improving financial outcomes and customer relationships in logistics.

6.2 Potential for Realized Savings

Predictive modeling can serve as a strategic tool in the freight brokerage industry by enabling the preemptive identification of potential write-offs. The model facilitates this by flagging shipments that exhibit a high likelihood of write-off, allowing agents to reevaluate and possibly change these shipments to mitigate risks. Adjustments to the model's recall—or sensitivity rate—affect the number of shipments flagged; higher sensitivity increases the number of reviews required but captures more at-risk shipments, while prioritizing accuracy reduces false positives, though some write-offs may be missed.

Balancing the model's threshold involves weighing the cost of reviewing additional shipments against the potential financial repercussions of overlooked write-offs. Simulation data from the first quarter of 2023 illustrates this balance: a 50% probability threshold led to identifying 61.8% of full write-offs while flagging 26.3% of shipments. If this threshold is increased to 60%, then the model identifies 51.2% of the fully written-off amount while only flagging 13.6% of the shipments. A simulation tool has been developed so that the sponsor company can optimize this threshold to maximize savings based on the costs of reviewing shipments and the potential savings in avoiding write-offs.

If the assumption is made that a portion of each of these flagged write-offs can be corrected before the shipment is finalized, then the sponsor company would potentially realize savings by avoiding financial risk from these write-offs. However, it is important to recognize that not all losses can be prevented due to external factors like market fluctuations and logistical issues.

6.3 Limitations

This study has limitations that are important to acknowledge when considering the findings. The analysis relies solely on data from one brokerage, which could limit the broader applicability of the results across the entire freight brokerage industry. Additionally, the exclusion of Less-than-Truckload (LTL) shipping data omits potentially valuable insights into the dynamics of smaller shipment operations. The models also do not include exogenous variables like macroeconomic indicators or industry-specific trends that could also drive the likelihood of write-offs. Addressing these limitations is crucial for the sponsoring company to enhance its risk management strategies.

6.4 Recommendations

Based on these findings, the following six recommendations and actions were provided to the sponsoring company:

- **Optimize Use of *Auto-Builder*:** Automation protects the sponsoring firm from larger financial setbacks associated with full write-offs. Expanding this automation program could lead to greater protection against these larger financial risks. However, while automation is beneficial, it is crucial to continually refine and monitor automated processes to reduce the incidence of partial write-offs, as these automated models led to a higher likelihood of partial write-offs associated with rate discrepancies and fuel rate discrepancies. Increasing the accuracy of this tool in determining rates for customers could further improve the efficacy of this automation program.
- **Agent Management:** Analysis indicates that agents in their second year demonstrate a higher propensity for incurring both full and partial write-offs, suggesting potential issues related to job fatigue or diminished job engagement. In contrast, the performance of new

agents aligns closely with that of seasoned agents, likely due to their recent exposure to training and onboarding. Therefore, establishing a specialized continuous training program for agents transitioning into their second year could help reduce the likelihood of write-offs by these agents. This initiative could refresh the expertise of second-year agents and cultivate an environment of sustained professional growth.

- **Model Integration with Daily Operations:** These predictive models can be systematically integrated as tools for agents who facilitate agreements between shippers and carriers in their daily operations. By embedding these models within the sponsor company's shipment system, it is possible to evaluate and flag risky shipments before an agreement is signed. Upon flagging, agents would have the opportunity to reassess these potential shipments. Furthermore, the coefficients derived from these models can highlight the specific characteristics of a shipment that triggered the flag, providing agents with actionable insights for each flagged shipment.
- **Estimate ROI of Investments in Enterprise Customers:** The analysis shows that enterprise customers (large clients of the sponsor company) are more prone to write-offs than small and medium-sized businesses or freight forwarders. The sponsor company's current strategy involves 'investment in continuing business' to retain these large clients. This strategy is activated during payment disputes with a customer, where the sponsor company opts to concede favorably to the customer to maintain the business relationship. Such concessions are recorded as write-offs.

  It is advised that the company performs a detailed assessment to evaluate the return on investment (ROI) from these expenditures. Should these investments fail to yield a positive ROI, it would be wise for the company to reconsider or diminish these financial engagements with such customers, particularly if they do not demonstrably benefit the company. Adopting this approach will ensure more effective resource allocation and alignment of investment strategies with the company's financial health and risk management goals.
- **Improve Data Collection Practices:** While the sponsoring company collects exhaustive records of historical shipments, there are opportunities to improve the quality of data collection for certain characteristics such as the customer's industry. Collecting more data features of shipments could improve the quality of these models by capturing other potential causes or indicators for write-offs.
- **Develop Tracking for Write-Off Process:** The current process deems a shipment as uncollectible after 120 days with no payment from a shipper. While this process ensures

that the sponsor company does not carry uncollectibles in accounts receivable, there is not currently a key performance indicator (KPI) in place to track adherence to this policy. Implementing this KPI could create benefits such as improved cash flow management by reducing the accumulation of aged receivables, enhanced visibility into the financial health of the company, and the ability to identify patterns or trends in payment delays.

7 Conclusion

This study developed predictive models to anticipate and quantify the risk of write-offs in the freight brokerage industry, providing a valuable tool for mitigating financial risks associated with unpaid receivables. By utilizing logistic and linear regression techniques on anonymized shipment data, the research has identified key predictors of write-offs, enabling freight brokerages to make informed decisions to safeguard their operations.

Significant insights emerged from the analysis, demonstrating that specific shipment characteristics like payment terms, shipment mode, and customer credit limits play pivotal roles in influencing the likelihood and magnitude of write-offs. These findings equip brokerages with the ability to preemptively address high-risk transactions, enhancing their financial stability and operational efficiency.

However, the study's reliance on data from a single brokerage and the exclusion of Less-than-Truckload (LTL) shipping data represent limitations that could impact the generalizability of the results. Future research should aim to include a broader dataset encompassing multiple brokerages and additional shipping modes to validate and refine the predictive models.

The application of predictive modeling in this context not only supports the sponsoring entity in reducing financial losses but also advances the broader freight logistics sector by highlighting the importance of data-driven decision-making. This study highlights the potential of analytics to enhance the economic resilience and operational effectiveness of freight brokerages.

8 References

Cheng, C.-B., Chen, C.-L., & Fu, C.-J. (2006, May 22). Financial distress prediction by a radial basis function network with logit analysis learning. *Computers & Mathematics with Applications.* https://www.sciencedirect.com/science/article/pii/S0898122105005237

Coyote Logistics. (n.d.). *Understanding the U.S. Truckload Market: the Complete Guide for Logistics Professionals*. Coyote Logistics Resource Center. Retrieved October 12, 2023, from https://resources.coyote.com/download/us-truckload-market

Denim Corporation. (n.d.). *Understanding the freight payment process for freight brokers*. Freight Payment System. https://www.denim.com/blog/understanding-freight-payment-process-for-freight-brokers

DAT. (2023, September 21). What Is a Freight Broker and What Do They Do? DAT. Retrieved October 10, 2023, from https://www.dat.com/solutions/what-is-a-freight-broker

Elgin, S. (2024, January 16). *Trucking Industry Trends, Statistics, & Forecast – 2023 Edition*. TruckInfo.net. Retrieved April 18, 2024, from https://www.truckinfo.net/research/trucking-statistics

I. Han, J. Chandler and T. Liang, The impact of measurement scale and correlation structure on the performance of statistical and inductive learning approaches, Expert Systems with Applications 10, 209-221, (1996).

Kalnins, A. (2022, June 2). When does multicollinearity bias coefficients and cause type 1 errors? A reconciliation of Lindner, Puck, and Verbeke (2020) with Kalnins (2018). *Journal of International Business Studies*, *53*(536–1548), 13.

Kearney. (2023, June 20). The great reset. *CSCMP's Annual State of Logistics Report*, pp. *34*, 63. https://epages.gopenske.com/rs/004-RHK-737/images/2023-CSCMP-State-of-Logistics-Report.pdf

Li, S. (2017, September 28). *Building A Logistic Regression in Python, Step by Step*. Towards Data Science. Retrieved November 27, 2023, from https://towardsdatascience.com/building-a-logistic-regression-in-python-step-by-step-becd4d56c9c8

Placek, M. (2022, April 20). Trucking industry in the U.S. - statistics & facts. https://www.statista.com/topics/4912/trucking-industry-in-the-us/#topicOverview

US Bureau of Labor Statistics. (2023, May). Heavy and Tractor-Trailer Truck Drivers. *Bureau of Labor Statistics.* Retrieved April 24, 2024, from https://www.bls.gov/oes/current/oes533032.htm

US Department of Transportation. (n.d.). Number of Air Carriers, Railroads, Interstate Motor Carriers, Marine Vessel Operators, and Pipeline Operators. *Bureau of Transportation Statistics*. Retrieved April 24, 2024, from https://www.bts.gov/content/number-air-carriers-railroads-interstate-motor-carriers-marine-vessel-operators-and-pipeli-0

Wanner, J., Herm, L.-V., Heinrich, K., & Janiesch, C. (1970, January 1). Stop ordering machine learning algorithms by their explainability! An empirical investigation of the tradeoff between performance and explainability. SpringerLink. https://link.springer.com/chapter/10.1007/978-3-030-85447-8_22

Wu, C.-W., Chen, C.-L., & Cheng, C.-B. (2021, January 19). *Asset Write-Offs Prediction by Support Vector Machine and Logistic Regression*. International Journal of Applied Science and Engineering. https://gigvvy.com/journals/ijase/articles/ijase-201010-8-1-047

9 Appendix

Table A1

Description of the features considered in the modeling

Feature	Description
LoadId	Identification number for a specific shipment.
PickupDate	Scheduled date and time for the freight to be picked up.
Industry	Sector of the economy the freight belongs to.
CustomerFirstLoad	Date and time when the customer's first shipment was scheduled.
CarrierId	Unique identifier for the freight carrier.
CSRepId	Carrier Sales Representative ID
BuilderId	Identification number for the individual or system that compiled the shipment.
InvoicingAgentId	Unique identifier for the agent responsible for invoicing.
SalespersonId	Identification number for the sales representative.
CustomerOwnerId	This is the ID of the person who currently owns the account, as opposed to the SalespersonId that is relative to when the shipment shipped
TermsOfPay	Payment terms agreed upon, in days
CustomerSegment	Category of the customer business type.
Mode	Transportation mode for the freight
EquipmentGroup	Type of equipment required for the freight
CustomerRate	Rate charged to the customer for freight transportation.
TotalCustomerRate	Total rate charged to the customer for the shipment (including accessorials).
CarrierRate	Rate paid to the carrier for freight transportation.
TotalCarrierRate	Total rate paid to the carrier for the shipment (including accessorials).
InvoicingAgentHireDate	The hire date of the agent responsible for invoicing

WriteOffAmount	Amount written off due to non-payment or other issues.
DiscountReason	Reason for any write-offs applied, due to bad debt or customer relations.
WO_percent	Percentage of the invoice written off.
AutoBuilder	Indicates if the shipment was auto-generated by a system (0 for no, 1 for yes).
Miles	Total miles the freight will travel from pickup to delivery location.

Table A2

Features with Coefficients and P-values for Continuous Model

Feature	Coefficients
Intercept	18.97***
AutoBuilder	-5.18***
Customer Segment: Enterprise	4.51***
Days As Customer: First Year (90-365 days)	1.93***
Days As Customer: Second Year	-1.05***
Miles (0-10)	4.83***
Miles: (10-50)	3.07***
Miles: (50-100)	0.89***
Miles: (250-1000)	-1.42***
Miles: (1000+)	-2.29***
Mode: Drayage	-1.81***
Mode: Final Mile	-9.52***
Mode: FTL	-6.06***
Terms of Payment: 0 days	7.88***
Terms of Payment: 10 days	6.76***
Terms of Payment: 15 days	6.74***
Terms of Payment: 20 days	3.99***
Terms of Payment: 30 days	-0.87***
Terms of Payment: 90 days	5.53***
Credit Limit: <$800K	-1.85***

Credit Limit: $800K-$2.17M	-5.54***
Credit Limit: $2.17M+	-6.14***
* for p-values < 0.05 (significant at the 5% level), ** for p-values < 0.01 (significant at the 1% level), and *** for p-values < 0.001 (significant at the 0.1% level)	

Figure A1

Distribution of the errors in the forecast in the continuous model

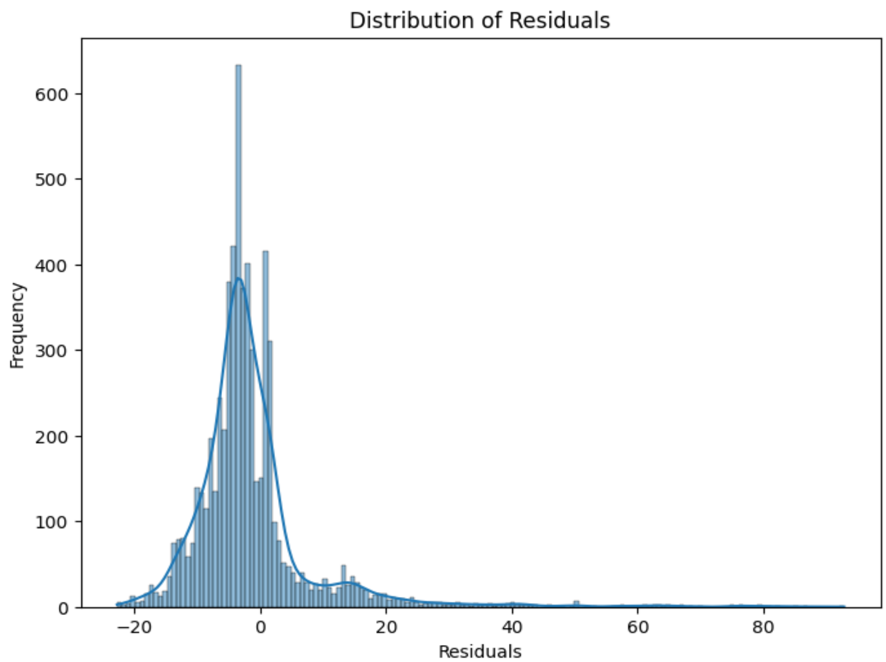


Table A3

Features with Coefficients and P-values for Full Write-Off and Partial Write-Off Model

Feature	Full Write-Off Model: Coefficient	Partial Write-Off Model: Coefficient
Customer Segment: Enterprise	0.47***	0.16***
AutoBuilder	-0.95***	1.15***
Terms of Payment 0 days	-1.27***	-0.60***
Terms of Payment 10 days	-0.81***	-0.60***
Terms of Payment 15 days	-0.97***	-0.37***

Terms of Payment 30 days	-0.97***	-0.15***
Terms of Payment 45	-1.27***	-0.15***
Terms of Payment 75 days	-1.39***	-0.60***
Terms of Payment 90 days	-0.75***	0.19***
Mode: Drayage	1.03***	1.75***
Mode: Final Mile	0.77***	0.67***
Mode: FTL	0.07***	0.44***
Miles of Shipment: (0-10)	0.74***	-0.28***
Miles of Shipment: (50-100)	0.20***	0.18***
Miles of Shipment: (250-1000)	-0.20***	-0.06***
Miles of Shipment: (1000+)	-0.41***	-0.18***
Days Since Last Shipment: 0-15 days	0.17***	-0.84***
Days Since Last Shipment: 30-60 days	0.63***	-0.84***
Days Since Last Shipment: 60-90 days	0.63***	-1.03***
Days Since Last Shipment: 90-180 days	0.63***	-0.84***
Days Since Last Shipment: 365+ days	0.31***	-0.46***
Credit Limit: $0-$25K	0.92***	0.10***
Credit Limit: $25K-$100K	0.42***	N/A
Credit Limit: $285K-$800K	N/A	-0.35***
Credit Limit: $800K-2.175M	-0.50***	0.32***
Credit Limit: $2.175M+	-0.72***	-0.62***
Days As Customer: New (0-90 days)	0.43***	0.36***
Days As Customer: 2nd Year	0.43***	0.16***
Days As Customer: Legacy (2+ years)	0.43***	-0.29***
Agent Tenure: New (0-90 days)	N/A	-0.39***
Agent Tenure: First Year (90-365 days)	-0.22***	-0.18***
Agent Tenure: Veteran (2+ years)	-0.07***	-0.18***

** for p-values < 0.05 (significant at the 5% level), ** for p-values < 0.01 (significant at the 1% level), and *** for p-values < 0.001 (significant at the 0.1% level)*

Factors with N/A values were excluded from the respective models due to insignificant p-values.