Carnegie Mellon University
Tepper School of Business

ACCOUNTING FOR SEASONALITY AND COVID IN CLV MODEL PREDICTIONS FOR RETAIL SPONSOR

MS Business Analytic Program

CMU MSBA 2022 Students:
Andrea Levy
Michael Zakin
Vivek Jayaprakasan

ABSTRACT

Customer Lifetime Value ("CLV") modeling focuses on measuring and understanding the value of a customer based on past customer transaction data. Reliable models that accurately predict CLV is a critical tool that helps a company understand customer purchasing behavior, enabling managers to make focused marketing decisions.

Two common "buy-'til-you-die" probability models are used to predict CLV in non-contractual customer retail environments: the Pareto-NBD model (Reference 1), and the BG-NBD (Reference 2) model. One of the primary benefits of these types of models is that they work with aggregate data and are easy to implement.

INTRODUCTION

The retail sponsor we worked with is a seasonal business with customers purchasing at a higher frequency and order value during the peak season in months 11 and 12. One of the downsides to traditional "buy-'til-you-die' models for a seasonal business such as the retail sponsor is the assumption that purchasing rates occur evenly throughout time.

We were tasked to explore how to implement seasonal purchasing elements into these traditional CLV models. In addition, being able to account for the impact of COVID has meaningful implications for forecasting expected customer behavior.

Recent literature was consulted to tackle these limitations. An approach to including time-varying covariates to the BG-NBD model (in the context of distinguishing between "high" season and "regular" season) was proposed by Fader and Hardie (Reference 3). There were no known implementations of this current model available for broader consumption.

Another approach to adding contextual features to traditional latent attrition models was proposed by Bachmann, et al (Reference 4) in which consideration was given for incorporating seasonality, tactical marketing initiatives and customer-specific purchasing patterns as features into the Pareto-NBD model. The authors supplemented the paper with a package that implemented the method in R, the CLVTools package.

METHODS AND MATERIALS

We utilized the CLVTools package in R to incorporate time-varying covariates into the traditional Pareto-NBD model to perform our analysis.

The retail sponsor provided transactions from a sample of customers for a 5 year period. The dataset included purchase date, total sales amount, and purchase channel for each customer. We trained our model using the first 4 years of data and tested the accuracy of CLV predictions for the following year. The key metrics used to compare the performance of our models were the total error rate and the weekly mean absolute error rate. Afterwards, we extrapolated our CLV prediction to future unseen time periods - predicting unseen year 6 behavior and extending the model further to 3-year and 5-year CLV predictions.

We experimented with incorporating time-varying covariates into both the transaction process and the attrition process, using the different optimization routines and parameter tuning methods available. Five main covariates were used to capture seasonality and the impact of COVID: (1) month 11, (2) month 12, (3) week 4, month 11, (4) months 3-5 in year 4, and (5) months 6-12 in year 4.

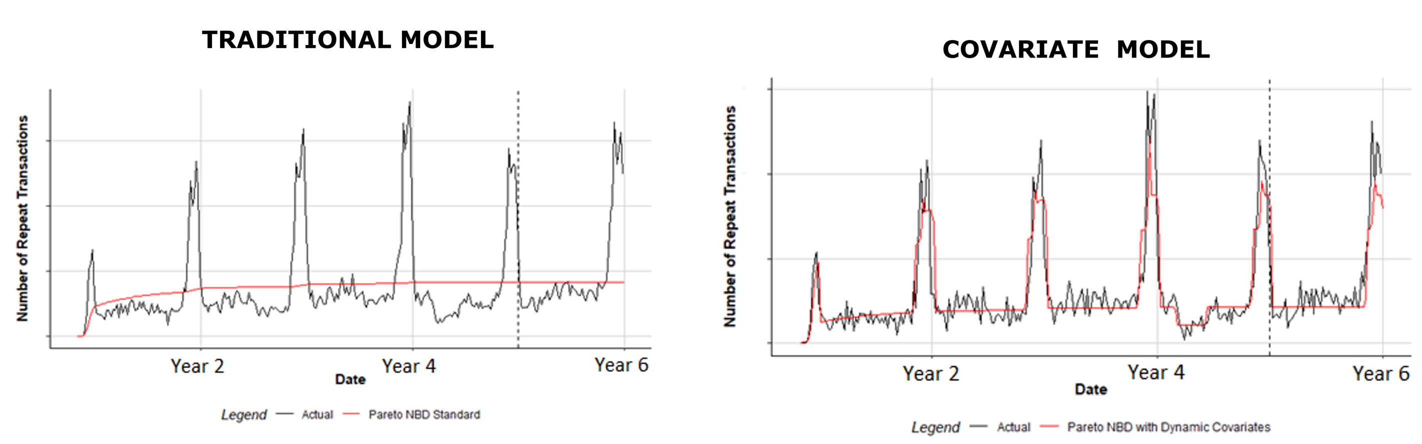
We tested the performance of the traditional Pareto-NBD model against the time-varying Pareto-NBD covariate model in aggregate and for customer clusters.

RESULTS

The traditional pareto-NBD model did a good job of predicting overall transaction rates at an aggregate level over a wide time horizon, especially for less seasonal customers.

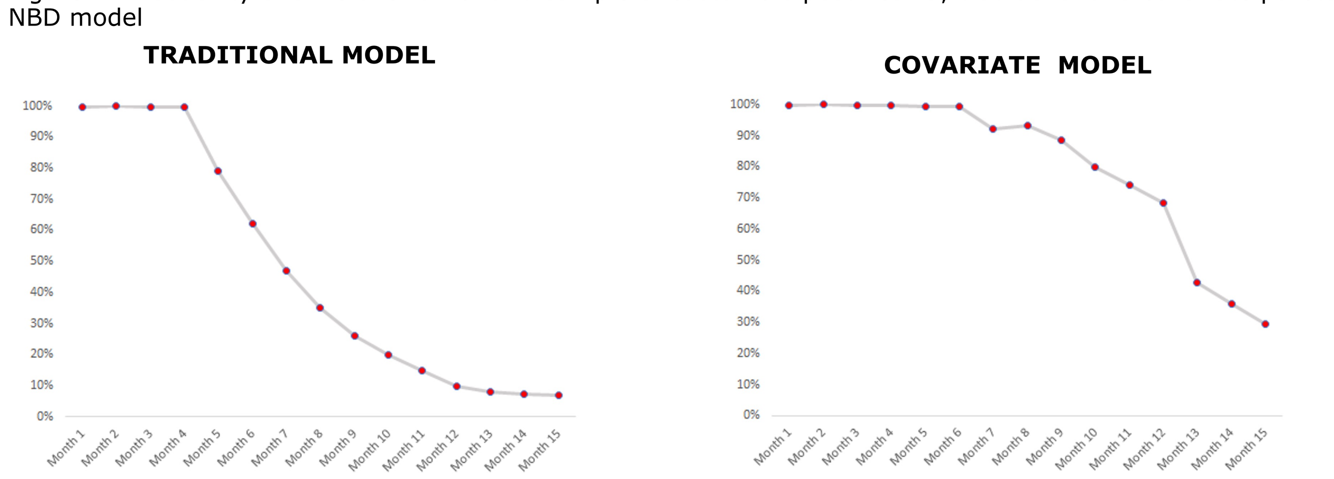
Adding covariates into the pareto-NBD model for highly seasonal customers had benefits in terms of accuracy of prediction and was an overall better fit as seen in the graphs below.

Figure 1: Transaction volume over time for a sample of customers; traditional and covariate pareto-NBD model



The covariate model did a better job of retaining customer value in terms of probability of alive prediction, especially for seasonal customers - customers are predicted to stay alive until the next seasonal period.

Figure 2: Probability of alive over time for a sample of customers post Year 5; traditional and covariate pareto-NBD model



The pareto-NBD model with covariates could also better control for the impact of COVID, when looking at year 5 predictions and extrapolating CLV predictions to 3-year and 5-year time horizons.

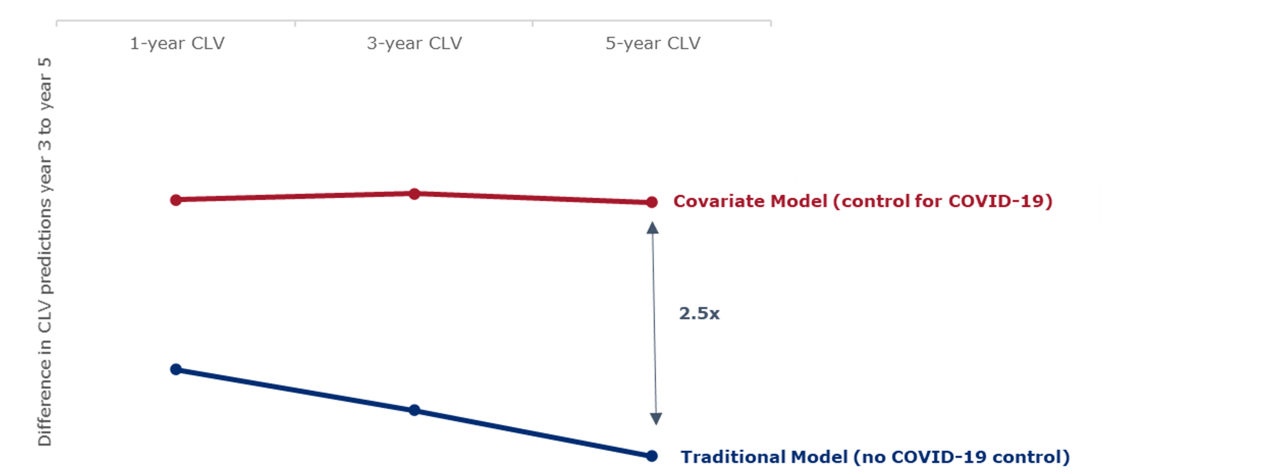
When adding the time-varying covariates to the Pareto-NBD model, the algorithm became computationally expensive. Optimal convergence was sensitive to the optimization routine, starting values for parameters, and homogeneity of customers.

BUSINESS IMPLICATIONS

After completing the data modeling process and implementing an improved CLV model over traditional methods, we connected the results of 5-year CLV predictions to practical business application. Analyzing a customer's CLV as of a single point in time (e.g., year 5) does not tell the entire story. Two customers with similar forecasts for CLV may have different growth trajectories.

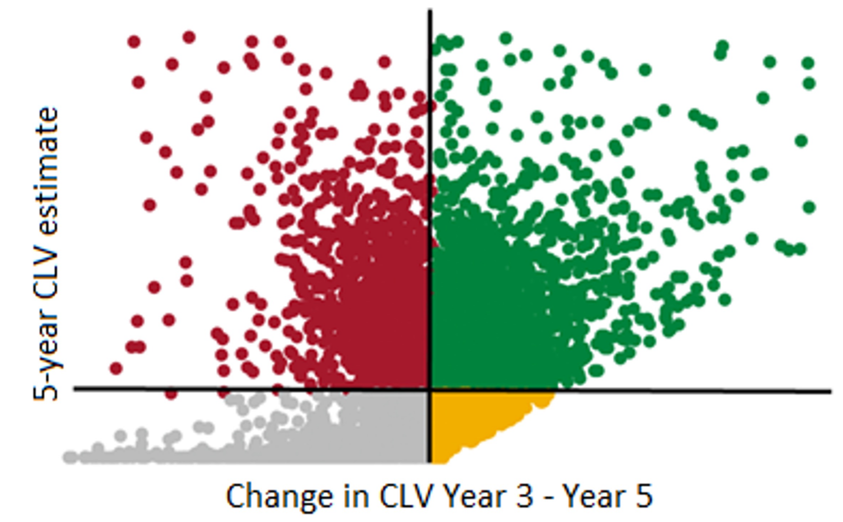
To combat this limitation, we used the improved model to calculate 5-year CLV predictions pre and post pandemic to incorporate the change in CLV into our analysis. By controlling for the impact of COVID in the covariate model, customers that returned to pre pandemic purchasing volumes also returned to similar levels of CLV. This was a novel way to understand changes in purchasing rates.

Figure 3: Change in CLV predictions from year 3 to year 5; traditional and covariate pareto-NBD model



We presented a simple segmentation of the retail sponsor's customers into quadrants based on size and trajectory of the forecast. The evolution of customer behavior was highlighted in the chart below.

Figure 4: 5-year CLV estimate from covariate model plotted against change in CLV



CONCLUSION

The covariate Pareto-NBD model accomplished the goal of improving CLV predictions for seasonal customers in both shorter and longer time horizons. While we cannot predict such transformative events like COVID, the covariate model neutralized significant penalization for the drop-off in year 4 purchasing volume in future customer forecasts.

We were able to enhance CLV predictions and uncover insights on customer purchasing behavior with just a couple columns of data. The results of this covariate model can be used as a mechanism to drive and measure the impact of business decisions.

CONTACT

Andrea Levy (aplevy@tepper.cmu.edu)
Michael Zakin (mzakin@tepper.cmu.edu)
Vivek Jayaprakasan (vjayapra@tepper.cmu.edu)

REFERENCES

1. Schmittlein, David C., Morrison, Donald G., and Colombo, Richard (1987), "Counting Your Customers: Who are They and What Will They Do Next?"
2. Fader, Peter S., Bruce G.S. Hardie, and Ka Lok Lee (2004), "'Counting Your Customers' the Easy Way: An Alternative to the Pareto/NBD Model."
3. Fader, Peter S., Bruce G.S. Hardie, and Chun-Yao Huang (2020), "Incorporating Time-Varying Covariates in the BG/NBD Model."
4. Bachmann, Meierer, and Naf (2021), "The Role of Time-Varying Contextual Factors in Latent Attrition Models for Customer Base Analysis."