Probabilistic Models for Analyzing Customer Purchase Behavior: A Primer

Markus Meierer Patrik Schilter Jeffrey Näf Patrick Bachmann

April 15, 2025

Abstract

This document provides an intuitive, high-level overview of probabilistic models for analyzing customer purchase behavior. It focuses particularly on non-contractual settings. We explain the core concepts behind modeling purchase frequency, customer attrition, and spending amount. Models like the Pareto/NBD model and Gamma-Gamma model serve as key examples. Important ideas such as customer heterogeneity and covariates are introduced. Additionally, the process for calculating financial metrics like Residual Customer Lifetime Value (RCLV) by combining model outputs is discussed. The text avoids complex mathematics for accessibility. The difference between using covariates for inference versus prediction is also highlighted.

Contents

1 Three Key Processes of Customer Purchase Behavior

Understanding and predicting future customer behavior is essential for making informed business decisions, from resource allocation and marketing targeting to inventory management. A particular challenge arises in non-contractual business settings (like retail, e-commerce, or streaming services) where customers do not have formal contracts – they simply stop interacting when they decide to leave. This makes the event of a customer becoming permanently inactive unobservable; we don't know if a long period of silence means the customer relationship has ended or if they are just in a long pause between purchases. This process of customers lapsing is often referred to as customer attrition. Because it is unobserved, it is also called latent attrition.

Probabilistic models, sometimes called "Buy 'Til You Die" (BTYD) models, provide a robust framework to analyze transaction histories and project future engagement in these settings. They model three fundamental dimensions of behavior:

1. **Customer Attrition/Activity:** Is the customer currently active (still a potential buyer), or have they likely become permanently inactive (undergone attrition)?
2. **Purchase Frequency:** *Given* that a customer is active, how often do they tend to make purchases?
3. **Purchase Amount:** When the customer makes a purchase, how much do they typically spend?

Typically, distinct but complementary models address these dimensions. Models like the **Pareto/NBD model or BG/NBD model** focus on the first two dimensions (attrition/activity and frequency), while models like the **Gamma-Gamma model** focus on the third (spending amount). By integrating predictions from these individual models, as discussed later (Section 4), we can estimate key forward-looking financial metrics. The overall conceptual structure is shown in Figure 1.

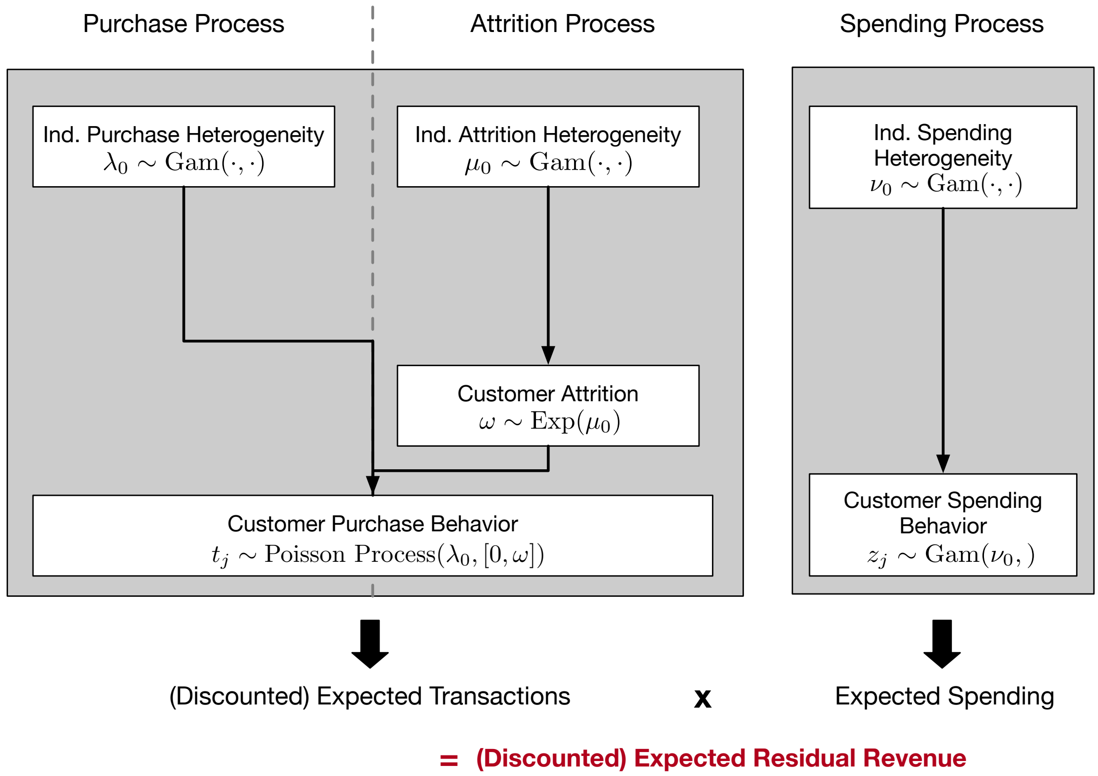

Figure 1: Conceptual Overview: Combining purchase/activity models (e.g., Pareto/NBD model, handling frequency and customer activity/attrition) with spending models (e.g., Gamma-Gamma model, handling transaction value) enables forecasts of future revenue, leading ultimately to metrics like DERR and RCLV.

These models share two key strengths. On the one hand, the models explicitly handle **customer heterogeneity**. They formally recognize customers differ in buying rates, propensity to become inactive, and spending levels, using probability distributions to capture this variation. On the other hand, these models excel at **long-term predictions**. They capture underlying *stochastic processes*, allowing reliable extrapolation.

Going even further into the modeling details, **the models can be divided into two categories**, depending on which information they leverage:

- **Standard versions**, like the Pareto/NBD or Gamma-Gamma model, use micro-level behavioral summaries of their purchase history for individual predictions. Predictions are informed by the resulting aggregate-level pattern estimated across the *entire customer base*. Despite being extremely data parsimonious, these standard versions have proven to provide fairly accurate predictions.
- **Extensions using covariates add additional micro-level information (e.g., demographics) and potentially macro-level context** (e.g., seasonality) beyond basic summaries, refining forecasts. This comes with higher data requirements, but can result in better predictive accuracy. An example for such a model is the Extended Pareto/NBD model (Bachmann et al., 2021).

As Figure 1 also illustrates, combining outputs enables the estimation of important financial projections. It is crucial to distinguish between revenue-based and profit-based forecasts:

- **Residual Expected Future Revenue:** The forecast of total revenue expected from a customer *from now onwards*. "Residual" emphasizes this remaining, future potential.
- **Discounted Expected Residual Revenue (DERR):** The present value of the residual expected future revenue, after discounting. This is still a revenue-based metric.
- **Residual Customer Lifetime Value (RCLV):** This metric represents the expected net profit from the customer going forward, discounted to its present value. Calculating RCLV requires subtracting the present value of expected future **costs** from the DERR (i.e., RCLV = DERR - Discounted Costs). RCLV is the key metric representing customer value in terms of profitability.
- **"Lifetime" vs. "Truncated":** "Lifetime" implies an indefinite forecast horizon. Predictions for a finite period yield "Truncated" versions of these metrics. This choice depends on the application and is discussed further in Section 4.

In the remainder, this document offers a high-level overview of probabilistic models for analyzing customer purchase behavior, focusing on logic without deep mathematics. For a technical overview, refer to Meierer et al. (2025) and the references within this article. For readers interested in the historical development and evolution of these probabilistic modeling approaches, Fader et al. (2014) provides a valuable perspective.

2 Step 1: Modeling Customer Attrition and Purchase Frequency

Let us first focus on modeling the intertwined patterns of purchase timing (**frequency**) and the hidden customer **activity** state (specifically, inferring when customers become permanently **inactive**), addressing the first two dimensions. When looking at transaction data, two realities stand out:

1. Customers buy at different rates – some shop frequently, others occasionally.
2. Customers eventually stop buying, but we don't know precisely when they decided to become inactive. A long silence could indicate permanent inactivity (attrition), or just a long pause for an active customer.

The Core Idea: Models like the **Pareto/NBD model**, a foundational example, tackle this uncertainty by simultaneously describing two processes for each customer:

- **Purchase Process:** While a customer is considered active (has not yet become permanently inactive), they make purchases randomly around their characteristic purchasing tendency (parameter λ). This governs purchase frequency.

- **Attrition Process:** This core process governs how long a customer remains active before the unobserved event of becoming permanently inactive occurs. Each customer has an unobserved duration or "lifetime" as an active customer, determined by their individual propensity related to attrition (parameter μ).

By modeling these two processes together, the model can infer the likelihood that a quiet customer is still active versus having become inactive.

Handling Customer Differences (Heterogeneity): The model explicitly recognizes customer heterogeneity. There's variation across the customer base in latent purchasing tendencies (λ) and propensities related to attrition (μ). The Pareto/NBD model uses probability distributions (Gamma distributions) to describe how these characteristics vary across the population, based on observed histories. This yields the aggregate-level pattern discussed earlier.

What Data are Needed? The standard Pareto/NBD model uses specific micro-level behavioral summaries:

- **Recency (t_x):** Time between first and last observed purchase.
- **Frequency (x):** Number of repeat purchases during observation.
- **Observed Time (T):** Total time elapsed between the customer's *first observed* purchase and the *end* of the observation period (sometimes referred to as customer "age" in the literature).

What Does the Model Tell Us? Fitting the model estimates parameters describing heterogeneity. Conditioning on an individual's x, t_x, T allows us to estimate two key outputs:

- The probability that a specific customer is still active (has not yet become permanently inactive) at the end of the observation period. This is often referred to $P(Alive)$.
- The expected number of future purchases over a given period, accounting for the fact that the customer might not be active anymore.

These outputs are the crucial inputs from this type of model needed for forecasting future revenue streams (as detailed in Section 4). The related **BG/NBD model** provides similar outputs but makes different assumptions about when inactivity can occur.

3 Step 2: Modeling Purchase Amount

To incorporate monetary value into forecasts, we need to model the third dimension: spending per transaction. A commonly used approach is the **Gamma-Gamma model**. Note that using a separate model such as the Gamma-Gamma model assumes: (1) independence between purchase frequency and average spending, and (2) stable average spending over the customer's active lifetime. In plain English, this means the model expects that how often someone buys is unrelated to how much they spend per purchase, and that their typical purchase amount doesn't systematically increase or decrease while they remain an active customer. While these assumptions might not always hold in reality, the approach has been shown to be relatively robust across many industries.

The Core Idea: The Gamma-Gamma model focuses specifically on the monetary value of transactions, assuming:

- Each customer has their own underlying typical (or average) transaction value (ζ).
- Customers differ in this value (heterogeneity).
- Individual transactions fluctuate randomly around the customer's typical value.

Probabilistically, it assumes the typical value varies across customers according to a Gamma distribution, and individual transactions for a customer also follow a Gamma distribution pattern.

Handling Customer Differences (Heterogeneity): Heterogeneity in spending is captured directly by assuming the typical transaction value (ζ) varies across customers according to a Gamma distribution.

What Data are Needed? Requires historical average transaction value ($\bar{z}$) and number of transactions (x) per customer.

What Does the Model Tell Us? Fitting the Gamma-Gamma model estimates the overall patterns of spending heterogeneity. For an individual customer with historical average spend $\bar{z}$ and transaction count x, the model allows us to make inference about their *expected average transaction value*, let's call it $E[\text{Spend}]$. This prediction smartly combines the individual's observed spending with the spending patterns learned from the entire customer base.

This expected spend per transaction is the key output needed for financial forecasting (Section 4). An alternative, the Normal-Normal model for spending, exists but its distributional assumptions may make it less flexible for typically skewed monetary data compared to the Gamma-Gamma model.

4 Step 3: Deriving Financial Projections

The real power of these models comes from combining the outputs of the attrition/frequency models (Section 2) and the purchase amount models (Section 3) to generate forward-looking financial projections like DERR and RCLV. This involves the following steps:

1. **Forecast Future Purchases:** Use the attrition/frequency model (e.g., **Pareto/NBD model**) to predict the expected number of transactions for a customer over a chosen future period T_{pred}, conditional on their history (x, t_x, T). Let this be $E[\text{Purchases}(T_{pred})]$. (Note: If using time-varying covariates for prediction here, they must be of the forecastable type discussed in Section 5).

2. **Forecast Average Spend:** Use the purchase amount model (e.g., **Gamma-Gamma model**) to infer the expected average spend per future transaction $E[\text{Spend}]$, based on past spending $(x, \bar{z})$.

3. **Calculate Residual Expected Future Revenue:** Multiply expected purchases by expected spend: $E[\text{Revenue}(T_{pred})] = E[\text{Purchases}(T_{pred})] \times E[\text{Spend}]$. This provides the foundational revenue forecast.

4. **Apply Discounting for DERR:** Convert the future revenue stream to present value using a discount rate (d) to get the DERR. This represents the present value of expected future revenues.

5. **Incorporate Costs for RCLV:** To estimate customer value (profitability), subtract the present value of expected future costs from DERR to arrive at the RCLV. RCLV = DERR - Discounted Costs. This final step moves from a revenue prediction to a profit-based value metric.

These calculations allow managers to rank customers based on expected future profitability and make more informed decisions about resource allocation.

A key decision in these calculations is the choice of the forecast horizon, T_{pred}. As mentioned in the introduction, this determines whether we calculate "Lifetime" or "Truncated" metrics.

- Calculating metrics over an indefinite or infinite horizon ("**Lifetime**") is often mathematically convenient within the standard BTYD framework (i.e., without time-varying covariates) and represents the theoretical total future potential. However, models that incorporate time-varying covariates typically require projecting these covariates into the future, making infinite-horizon forecasts computationally complex or currently impractical.

- Calculating metrics over a finite, fixed period ("**Truncated**"), such as the next year ($T_{pred} = 1$) or the next three years ($T_{pred} = 3$), is often more practical for managerial purposes like short- to medium-term planning, target setting, or evaluating specific interventions, and is readily achievable with both standard models and those including covariates (provided future covariate values are known or forecastable for that period).

The choice depends on the specific business question being addressed and the modeling approach used. A manager might use a truncated 1-year RCLV for budget allocation but consider the lifetime RCLV (if feasible) for strategic customer segmentation. Naturally, truncated values will be lower than lifetime values.

5 Optional: Enhancing Understanding and Predictions with Covariates

While standard models use behavioral summaries (x, t_x, T), incorporating **additional information** via *covariates* can refine understanding and potentially improve predictions. Covariates link observable characteristics to a specific latent behavioral process, e.g., the attrition process.

There are two main types:

- **Time-Invariant Covariates:** Stable attributes (demographics, acquisition source) explaining baseline differences in, for example, attrition propensities or purchasing tendencies across different customer segments.
- **Time-Varying Covariates:** Dynamic factors (marketing, seasonality) modulating the likelihood of purchasing or becoming inactive over time.

Using covariates increases data needs and complexity, but offers richer insights. However, distinguish between using covariates for **inference** (understanding the past) versus **prediction** (forecasting the future):

- **For Inference:** Any historical covariate is valuable to quantify past drivers of purchasing behavior and attrition.
- **For Prediction:** Time-invariant covariates are usable for prediction. Time-varying covariates require known/forecastable future values. This limits their use for prediction (e.g., for DERR or, subsequently, RCLV calculation) to predictable patterns (e.g., seasonality). As noted in Section 4, the need to project time-varying covariates also makes infinite-horizon forecasts challenging for these models.

Thus, only a subset of time-varying covariates that enhance inference can typically be used for forecasting, especially over long or indefinite horizons.

6 Conclusion

Probabilistic models offer a structured framework for dissecting transaction histories capable of handling customer heterogeneity. They enable robust **long-term forecasting** of future activity and purchases based on past purchase behavior and potentially covariates (with limitations for prediction, especially regarding time-varying covariates and infinite horizons).

Combining models allows estimating **Residual Expected Future Revenue**. Discounting yields DERR, representing discounted future revenue. Further incorporating **costs** leads to the profit-based RCLV. Understanding the distinction between these revenue and value metrics, model assumptions, covariate nuances, and the implications of the chosen forecast horizon is key for applying these tools effectively for customer valuation and management. In R, the `CLVTools` package (Bachmann et al., 2023) provides a way to estimate these models and predict customers' future value to a firm.

References

Patrick Bachmann, Markus Meierer, and Jeffrey Näf. The Role of Time-Varying Contextual Factors in Latent Attrition Models for Customer Base Analysis. *Marketing Science*, 40(4):783–809, 2021.

Patrick Bachmann, Markus Meierer, Jeffrey Näf, and Patrik Schilter. *Walkthrough for the CLVTools Package*, 2023. URL `https://cran.r-project.org/web/packages/CLVTools/vignettes/CLVTools.pdf`.

Peter S Fader, Bruce GS Hardie, and Subrata Sen. Stochastic Models of Buyer Behavior. In Russell S Winer and Scott A Neslin, editors, *The History of Marketing Science*, pages 165–205. World Scientific, Singapore, 2014.

Markus Meierer, Patrik Schilter, Jeffrey Näf, and Patrick Bachmann. Estimating Individual Customer Lifetime Values with R: The CLVTools Package. *Working Paper*, 2025.