V(CLV): Examining Variance in Models of Customer Lifetime Value

Daniel M. McCarthy

Peter S. Fader

Bruce G. S. Hardie *

February 28, 2016

Abstract

While accurate point estimation of customer lifetime value (CLV) has been the target of a large body of academic research, few have focused on the variance of CLV (V(CLV)), which represents the degree of uncertainty associated with a customer's expected CLV. This is ironic because academics have long known that V(CLV) is one of the most important characteristics that defines and differentiates customers from one another, affecting firms on many fundamental levels. No closed-form, forward-looking statistical procedures have been derived to estimate individual-level V(CLV). For the first time, the authors derive, predict, and validate V(CLV) using a powerful combination of stochastic models for the flow of transactions over time and the company's profit on each transaction. They provide these estimates for 561,100 customers of an omnichannel retailer tracked over a 2.25-year period, making this one of the

*Daniel McCarthy is a Ph.D. Candidate in the Department of Statistics at the Wharton School of the University of Pennsylvania (address: 400 Jon M. Huntsman Hal, 3730 Walnut Street, Philadelphia, PA 19104-6340; email: danielmc@wharton.upenn.edu, web: http://www-stat.wharton.upenn.edu/~danielmc/). Peter S. Fader is the Frances and Pei-Yuan Chia Professor of Marketing at The Wharton School of the University of Pennsylvania (address: 771 Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, PA 19104-6340; phone: (215) 898-1132; email: faderp@wharton.upenn.edu). Bruce G.S. Hardie is Professor of Marketing, London Business School (email: bhardie@london.edu; web: www.brucehardie.com).

We acknowledge the financial support of the Baker Retailing Center, thank Wharton Computing's Research Computing team for access to and support in the use of HPC resources, and thank the anonymous retailer that provided data used in this paper. We are grateful to Eric Schwartz for helpful input.

Electronic copy available at: http://ssrn.com/abstract=2739475

largest-scale CLV analyses to date. They highlight the importance of V(CLV), analyze its relationship to observable summary statistics such as recency, frequency, and monetary value, and uncover many substantive variance-related insights regarding customer segmentation, scoring, targeting, and more.

Keywords: customer lifetime value; CLV; RFM; customer-base analysis; uncertainty

Customer lifetime value (CLV), the net present value of the future profit associated with a customer, is increasingly being recognized within academia and industry as one of the most important metrics for a firm. Many stakeholders – executives, marketing managers, accountants, and investors – are waking up to the value-creating potential of CLV-driven business strategies. While countless papers have been written about accurate *point estimation* of expected CLV ($E(CLV)$), very few papers have focused on the *riskiness* or *variance* of CLV ($V(CLV)$). This is ironic because marketing and finance scholars have long known that $V(CLV)$ is one of the most important characteristics that defines and differentiates customers from one another, affecting how firms should think about / interact with their customers when implementing CLV-driven business strategies.

Many papers have touched upon the importance of $V(CLV)$. Zhang, Bradlow and Small (2014) note that differences in $V(CLV)$ across customers (in their paper, due to a so-called 'clumpiness' factor) has implications for firm investments: "From the investor perspective, clumpiness means high growth potential but large risk, thus a clumpiness measure might serve as an important factor that should be taken into account when firms make investment decisions." Srivastava, Shervani and Fahey (1998) note (and Kumar and Shah (2009) empirically support) that more stable and predictable cash flows have a higher net present value and thus increase shareholder value, necessitating a focus by marketing to reduce the variance of future cash flows. Decreasing the variance of future cash flows is mathematically equivalent to decreasing the variance of customer equity (CE, or the summation of the CLV's of all current and future expected customers of the firm). Rust, Kumar and Venkatesan (2011) note that firms may want to allocate additional resources towards customers with high $V(CLV)$: "a manager may want to pay extra attention to a "high-variance" customer who has a decent chance of being very profitable in the future, even if the customer's expected future profitability was only average, to avoid missing a potential opportunity." Gupta et al (2006) note that it may be optimal for firms to diversify their customer bases with respect to $V(CLV)$: "Current CLV measurement

Electronic copy available at: http://ssrn.com/abstract=2739475

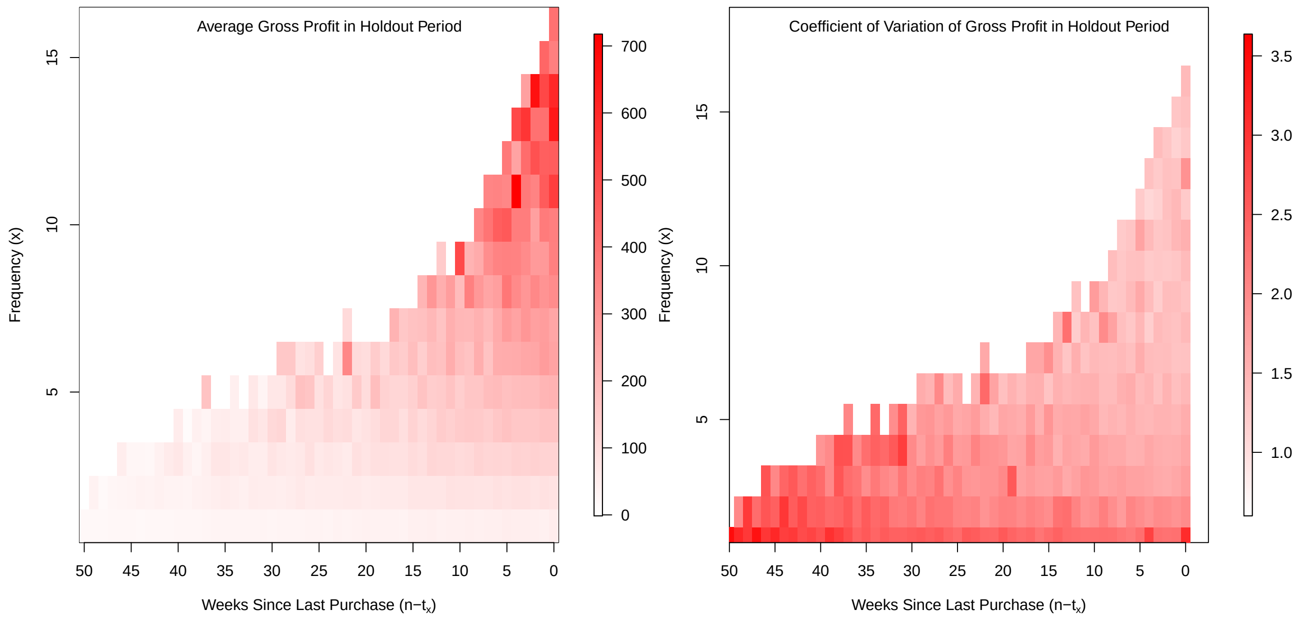
practices that focus on the expected value of a customer may predict that high-risk customers are more valuable than low-risk ones... However, the financial markets expect the firm to have a portfolio of customers that comprises a mix of low- and high-risk customers." Other papers framing customers as a portfolio to be optimized are Bolton, Lemon and Verhoef (2004), Dhar and Glazer (2003), Gupta and Lehmann (2005), and Tarasi et al. (2011). Schwartz, Bradlow and Fader (2015) note that the expected variance of future payoffs is a (positive) key factor within several sequential adaptive experiment recommendation policies. While they are primarily concerned with resource allocation to display advertisements, they note that the same policies could be applied to resource allocation decisions across customers. Rust, Lemon and Zeithaml (2004) note that parameter estimation uncertainty within their survey-based CLV model may cause the actual return associated with marketing drivers to deviate from expectations. While this is not $V(CLV)$ in the strictest sense, it nevertheless insightfully stresses the fact that marketing managers need to move beyond point estimation.

While all of these papers allude to the importance of $V(CLV)$, none of them make $V(CLV)$ central to their analyses. In most real-world applications, customer risk and return are estimated through the "rear-view mirror," i.e., assuming the future will be analogous to the past. Gupta et al (2006) mention $V(CLV)$ estimation as an important area of future research: "The developers of models used to compute customer lifetime value have focused on deriving expressions for expected lifetime value. To assess the risk of customers, we need to derive expressions for the distribution (or at least the variance) of CLV." Yet here we are, a decade after that paper was written, and no such closed-form expressions have been presented in the literature – until now.

Our aim is to fill this gap by deriving, predicting, diagnosing, and leveraging customer-level $V(CLV)$ in closed-form with a powerful combination of stochastic models for the flow of customer transactions over time and the profit earned by the company for each of those transactions. We perform our CLV analysis at full commercial-level scale, providing individual-level CLV estimates for an omnichannel retailer, covering 561,100 customers over a 2.25-year period using daily transaction-log profitability data. The size of the dataset used in this paper makes this one of the largest CLV analyses to date within the marketing literature. The size of this cohort makes it representative of the firm as a whole, diminishing concerns regarding the generalizability of the conclusions reached in this paper. Estimating our model and computing $E(CLV)$ and $V(CLV)$ predictions for all customers requires only a few minutes. This is important because it highlights the scalability and practical applicability of our modeling approach. In contrast, the two previous papers with methodologies that could approximate forward-looking $V(CLV)$, Rust, Kumar and Venkatesan (2011) and Kumar and Shah (2009), use a computationally intensive bootstrapping procedure which makes large-scale applications supporting daily marketing operations challenging. It is also for this reason that we focus upon $V(CLV)$ in lieu of other uncertainty metrics – while our proposed model provides us with the full distribution of individual-level CLV, marketing managers seeking a simple, straightforward, easily available summary statistic encoding the uncertainty of how much customers are worth would ask for $V(CLV)$.

Before turning to our analytical results, we provide a brief exploratory analysis to lay the groundwork for our model development. Consider the behavior of a cohort of 561,100 retail customers acquired in the third calendar quarter of 2013. We have detailed transaction-level data on their purchases and returns, as well as the profitability of each of these transactions, up to and including September 27th 2015. Imagine that we split this 117-week period into two periods – the first covering 1 year from July 1st 2013 through June 30th 2014, the second covering the remaining 1.25 years from July 1st 2014 through September 30th 2015. We group customers based upon the length of time which has elapsed since their last purchase ($n - t_x$, where n denotes the number of transaction opportunities they had in weeks and t_x denotes the week of their final purchase) and their purchase frequency (where x denotes the number of weeks in which they made a purchase)[1]. For all groups with more than 10 customers within them, we compute their average total gross profitability in the subsequent 65-week period, as well as the coefficient of variation (CV, or ratio of the standard deviation of holdout gross profit amounts to average gross profit) to examine the variability of future holdout gross profit while controlling for the level of expected profitability.[2] We plot the resulting heatmaps of frequency (x) against time since last purchase ($n - t_x$) in Figure 1. Darker shading is indicative of higher average total gross profits and higher dispersion around that average in the left and right plots, respectively.

Figure 1: Heatmaps of Week 52-117 Average Total Gross Profit and Coefficient of Variation of Total Spend By Time Since Last Purchase and Frequency



We know from Fader, Hardie, and Lee (2005b, FHL hereafter) that recency and frequency are both strongly associated with higher future gross profits – we can see this association again here through the increasing darkness of the plot of averages on the left in Figure 1 as we move from the lower-left side of the plot to the upper-right side of the plot. What may have been less obvious is what we see in the CV plot on the right. First, it is evident that the very best customers by future average gross profits also have the least amount of variability about that average, as evidenced by the the increasing lightness of the CV heatmap as we move from lower-left to upper-right. Second, we also see that this trend is less evident or not evident at all when recency is low – for example, when the customer's last purchase was more than 30 weeks ago (i.e., $n - t_x \geq 30$), it appears that CV actually goes up as frequency goes up. This previously undocumented pattern is indicative of meaningful, systematic differences in the uncertainty of future payoffs across customers, motivating a more thorough analysis and exploring its implications.

At the same time, this model-free view of the future is very noisy because of the high degree of sparseness (and general noise) in the data for many cells. To remove the noise from Figure 1 and uncover the true underlying relationships that exist in the data, we need to develop and validate a formal model. We need a model to obtain a clear view of $V(CLV)$ so that we may study it jointly with $E(CLV)$ to obtain a more informed view of the customer.

In the next section, we show how $V(CLV)$ can be expressed as a function of the first two moments of a transaction flow process that governs whether or not transactions will occur in the future, and a transaction profitability process that determines how profitable each of those future transactions will be. We propose and derive the first two moments of these processes, then bring these moments together into an expression for $V(CLV)$. After discussing model results and performing holdout validations, we conduct a series of analyses that shed light upon many aspects of $V(CLV)$ – how it maps to and from recency, frequency, and monetary value (RFM), what contributes to it, what it implies about the merits of RFM segmentation, and how it can be used to improve customer scoring and targeting for marketing managers.

MEASURING V(CLV)

Our first task is how to predict the forward-looking variance of CLV. To do so, we employ stochastic models for the flow of transactions over time and the profit earned by the company for each of those transactions.

We encode a customer's observed transaction process over time as an binary string, where $y_t = 1$ if a transaction occurred during the tth transaction opportunity, and 0 otherwise, $t = 0, 1, 2, \ldots$. We let Y_t denote the random variable corresponding to y_t. In the context considered here, a transaction opportunity is defined to be a well-defined time interval during which a transaction either occurs or does not occur. While this discreteness may at first seem unappealing, it is often theoretically sound (or even preferred):

1. Most CRM systems aggregate transaction data at the daily level, which imposes daily discretization by the recording process (see Fader, Hardie and Shang (2010), hereafter referred to as FHS).

2. If the natural periodicity with which transactions take place is at a relatively coarse temporal resolution (i.e., weekly), fits and forecasts for transaction flow may actually improve by modeling transaction flow at that coarser temporal resolution (Yoo, Hanssens and Kim 2012).

We encode the gross profit earned by the company on the tth transaction opportunity by W_t. While revenues must be strictly positive, profit may not always be positive. For example, retailers run seasonal clearance sales to monetize leftover inventory. In our dataset, 4,689 of the 561,100 customers within the cohort considered here had negative profitability in the calibration period.

Under the assumptions made above, letting d denote the per-period discount rate that accounts for the time value of money / cost of capital of the firm (McCarthy, Fader and Hardie 2015), CLV can be represented as an infinite sum of transaction flow indicators, multiplied by the profit accrued given a transaction has taken place:

$$CLV = \sum_{t=0}^{\infty} \frac{Y_t W_t}{(1+d)^t}, \tag{1}$$

where the customer is first acquired at $t = 0$.

Oftentimes, customer data has been observed up to and including a particular purchase opportunity. Hereafter, we assume that data has been observed up to/including transaction opportunity n. Denoting the net present value of historical profits earned up to and including purchase opportunity n as of time 0 by HV, and the net present value of profits earned after purchase opportunity n as of time n by RLV, Equation 1 can be re-written as

$$CLV = HV + \frac{1}{(1+d)^n} RLV, \quad \text{where} \tag{2}$$

$$HV = \sum_{t=0}^{n} \frac{Y_t W_t}{(1+d)^t} \quad \text{and} \tag{3}$$

$$RLV = \sum_{t=1}^{\infty} \frac{Y_{n+t} W_{n+t}}{(1+d)^t}. \tag{4}$$

Note that for a just-acquired customer, $n = 0$ (i.e., we have only observed his/her initial purchase).

Alternative customer profitability measures ignore spend and focus upon transaction incidence alone (Braun, Schweidel and Stein 2015; FHS). Most relevant to the discussion that follows is the discounted sum of residual transactions (DRT) of a customer:

$$DRT = \sum_{t=1}^{\infty} \frac{Y_{n+t}}{(1+d)^t}. \tag{5}$$

Going forward, let $DRT(r)$ denote DRT computed using a discount rate of r. If no discount rate is explicitly specified, assume a discount rate of d.

HV does not contribute to $V(CLV)$ because it is a known constant. As a result, if we were to assume that transaction flow and transaction profitability are independent of one another and that transaction profitability is time-invariant (which are empirically verifiable),

$$
\begin{aligned}
(6)\qquad V(CLV) &= V\left(\frac{1}{(1+d)^n}RLV\right) = \frac{1}{(1+d)^{2n}}(E(RLV^2) - E(RLV)^2), \\
E(RLV^2) &= E\left[\sum_{t=1}^{\infty}\frac{Y_{n+t}^2 W_{n+t}^2}{(1+d)^{2t}} + 2\sum_{t=2}^{\infty}\sum_{s=1}^{t-1}\frac{Y_{n+t}Y_{n+s}W_{n+t}W_{n+s}}{(1+d)^{s+t}}\right] \\
(7)\qquad &= E(W^2)E(DRT(2d+d^2)) \\
&\quad +2E(W)^2\left(E(DRT^2(d)) - E(DRT(2d+d^2))\right), \\
(8)\qquad E(RLV) &= E(W)E(DRT(d)).
\end{aligned}
$$

For details, please see Appendix. Equations 7 and 8 imply that $V(CLV)$ can be cleanly and conveniently separated into expressions involving the first two moments of DRT and profit per transaction, W. Therefore, we first develop a sub-model for transaction flow, $Y_t,\ t = 1, 2, \ldots$, and use it to derive closed-form expressions for the first two moments of $DRT(r)$. Thereafter, we build a separate sub-model for transaction profitability, $W_t,\ t = 1, 2, \ldots$, and derive closed-form expressions for its first two moments as well. We can then use the first two moments of both processes to derive $V(CLV)$. In the process, we also obtain $V(DRT)$ and $V(W)$ "for free." These additional variance expressions highlight the fact that variance is not a singular entity, and as we will see, these other variance expressions are interesting in their own right.

Modeling Transaction Flow

The Beta Geometric/Beta Binomial (BG/BB) model that FHS developed is a powerful framework for forecasting the flow of transactions over time. The BG/BB model has been well-validated in many applied settings (Dziurzynski et al. 2014; Sant'Anna and Ribeiro 2008; Schwartz, Bradlow and Fader 2014; Schweidel and Knox 2013; Yoo, Hanssens and Kim 2012; Zhang 2008). It is the discrete-time analog to the popular

Pareto/NBD model that Schmittlein et al. (1987) developed (Jerath, Fader and Hardie 2011; Reinartz and Kumar (2000, 2003)). In fact, the BG/BB model with increasingly short discrete-time increments converges in distribution to the Pareto/NBD model, so predictive results at the daily/weekly level are very comparable between the two models. In our data setting for example, the BG/BB model provides an excellent fit for transaction flow using daily and weekly temporal aggregation, so for computational and presentational simplicity we carry out our analysis using weekly temporal aggregation (i.e., $Y_t = 1$ if one or more purchases was made in week t, and the revenue derived from all purchases that week is summed). As a result, we use the phrase 'purchase-week' and 'purchase' interchangeably in the analysis that follows.

In contrast to the P/NBD model, the BG/BB model involves simpler analytical expressions and thus is much easier to implement. It can be run in Excel (Fader, Hardie and Berger 2011), and most importantly, the second moment of DRT is calculable in closed-form (it is unclear whether the same is true of the Pareto/NBD model).

The BG/BB model makes the following general assumptions about a customer's repeat buying behavior:

- A customer's relationship with the firm consists of two stages: he/she is "alive" for some period of time, then becomes permanently inactive ("dies").
- A "living" customer buys at each purchase opportunity with probability p: $P(Y_t = 1|p, alive_t) = p$, for $0 \le p \le 1$.
- A living customer dies at the beginning of each purchase opportunity with probability θ.
- Purchase and death propensities vary from customer to customer; heterogeneity in both across the customer base are modeled with independent beta distributions: $p \sim Beta(\alpha, \beta),\ \theta \sim Beta(\gamma, \delta)$.

The only individual-level data required to estimate the BG/BB model are *frequency* or the number of transactions in the calibration period (x), *recency* or the transaction opportunity at which the last observed transaction occurred (t_x), and the number of transaction opportunities in the calibration period (n).

Using this model, what are the first two moments of DRT which we may then plug in to Equation 6? The first moment of DRT for a customer with purchase history (x, t_x, n), purchase hyperparameters

$(\alpha, \beta, \gamma, \delta)$, and discount rate d, $E(DRT|d, \alpha, \beta, \gamma, \delta, x, t_x, n)$, was provided in Equation 14 in FHS. In the Appendix, we derive for the first time the corresponding second moment expression:

$$(9) \qquad E(DRT^2|d, \alpha, \beta, \gamma, \delta, x, t_x, n) = \frac{A_1A_2 + A_3A_4 + A_5A_6}{p(x, t_x|n, \alpha, \beta, \gamma, \delta)B(\alpha, \beta)B(\gamma, \delta)d^2(1+d)(2+d)},$$

where

$$\begin{aligned}
A_1 &= 2B(\alpha + x + 2, \beta + n - x)B(\gamma, 2 + \delta + n), \\
A_2 &= (2+d)_2F_1(1, \gamma; 2 + \delta + \gamma + n; \frac{-1}{d}) -_2 F_1(1, \gamma; 2 + \delta + \gamma + n; \frac{-1}{d(2+d)}), \\
A_3 &= d \times B(\alpha + x + 1, \beta + n - x)B(\gamma, 1 + \delta + n), \\
A_4 &= (2+d)_2F_1(1, \gamma; 1 + \delta + \gamma + n; \frac{-1}{d}) -_2 F_1(1, \gamma; 1 + \delta + \gamma + n; \frac{-1}{d(2+d)}), \\
A_5 &= B(\alpha + x + 1, \beta + n - x)B(1 + \gamma, 2 + \delta + n), \\
A_6 &= (2+d) \times_2 F_1(1, 1 + \gamma; 2 + \delta + \gamma + n; \frac{-1}{d}) -_2 F_1(1, 1 + \gamma; 2 + \delta + \gamma + n; \frac{-1}{d(2+d)}),
\end{aligned}$$

where ${}_2F_1(a, b; c; z)$ is the Gaussian hypergeometric function (Beukers 2007). This new derivation is central to our calculation of $V(CLV)$. The fact that we have a closed-form expression for $E(DRT^2|d, \alpha, \beta, \gamma, \delta, x, t_x, n)$ makes large-scale evaluation of it trivial. Efficient computation is a prerequisite for firms with large customer bases looking to adopt CLV-based strategies into ongoing operations.

Equation 9 also enables us to study uncertainty stemming from the transaction flow process in isolation, separate from the transaction profitability process:

$$(10) \quad V(DRT|d, \alpha, \beta, \gamma, \delta, x, t_x, n) = E(DRT^2|d, \alpha, \beta, \gamma, \delta, x, t_x, n) - E(DRT|d, \alpha, \beta, \gamma, \delta, x, t_x, n)^2$$

Marketing managers may be similarly interested in the variance of the total number of purchases a particular customer will make over a finite horizon of time conditional upon that customer's transaction history, $V(X(n, n{+}n^*)|\alpha, \beta, \gamma, \delta, x, t_x, n)$. Unlike DRT, $X(n, n{+}n^*)$ can be empirically validated without approximation/truncation. Furthermore, DRT's infinite horizon may not align with the planning horizon of some marketing managers. In the Appendix, we provide two derivations for the second moment of $X(n, n+n^*)$:

$$(11) \qquad E(X(n, n+n^*)^2 | \alpha, \beta, \gamma, \delta, x, t_x, n) = \frac{B_1 + B_2 B_3}{B(\alpha, \beta) B(\gamma, \delta) P(x, t_x | n, \alpha, \beta, \gamma, \delta)}$$

where

$$\begin{aligned}
B_1 &= B(\alpha + x + 1, \beta + n - x) \times B(\gamma - 1, \delta + n + 1) - B(\gamma - 1, \delta + n + n^* + 1), \\
B_2 &= 2B(\alpha + x + 2, \beta + n - x), \\
B_3 &= B(\gamma - 2, \delta + n + 2) - n^* B(\gamma - 2, \delta + n + n^* + 1) + (n^* - 1) B(\gamma - 2, \delta + n + n^* + 2).
\end{aligned}$$

This, in turn, provides us with an expression for the variance of the future number of holdout purchases:

$$(12) \qquad V(X(n, n+n^*) | \alpha, \beta, \gamma, \delta, x, t_x, n) = C_1 - C_2^2, \quad \text{where}$$

$$\begin{aligned}
C_1 &= E(X(n, n+n^*)^2 | \alpha, \beta, \gamma, \delta, x, t_x, n) \quad \text{and} \\
C_2 &= E(X(n, n+n^*) | \alpha, \beta, \gamma, \delta, x, t_x, n),
\end{aligned}$$

using Equation 11 above and Equation 13 from FHS, respectively. In the Validation section, we will test the BG/BB model's ability to predict the variance of future transactions through Equation 12.

Adding Transaction Profitability

In the previous section, we showed how uncertainty in customer-level future transaction flow can be estimated. While this is useful, marketing managers would like to know how uncertain each customer's *profitability* is. To do so, we need to model the profitability of each transaction. As Rust, Kumar and Venkatesan (2011) note, the existing literature shows that predicting future profitability from CRM data is surprisingly difficult. We will show that the combination of the transaction flow model proposed in the previous section and the transaction profitability model we propose below is very fast, requires nothing but CRM data, and does a remarkably good job of predicting future profitability.

One approach to modeling transaction profitability would be to model the revenue derived from each transaction, and then apply a constant margin to that number as in FHL and Reinartz and Kumar (2000, 2003). The gamma/gamma model for revenue per transaction used in FHL, adapted from Colombo and Jiang (1999), provided a particularly excellent goodness of fit. The model made the following assumptions:

1. Revenue derived from a transaction fluctuates randomly about the customer's mean revenue.

2. Mean revenue varies across customers but not over time according to a gamma distribution. Letting $Z_{i,j}$ denote the revenue associated with the ith transaction for customer j, $Z_{i,j} \sim \text{gamma}(\tilde{p}, \nu_j),\ i = 1, 2, \cdots, x_j,\ j = 1, 2, \cdots, J$.

3. The distribution of mean revenue across customers is independent of the volume of transactions.

4. Heterogeneity in mean revenue across the customer base is induced by making customer j's rate parameter, ν_j, gamma$(q, \tilde{\gamma})$-distributed across customers: $\nu_j \sim \text{gamma}(q, \tilde{\gamma})$.

This model was used to estimate/predict each customer's unobserved mean revenue, $E(M)$, as a function of his/her historical number of repeat purchases, x, and observed average revenue, $m_x = \sum_{i=1}^{x} z_i/x$. The gamma/gamma model has a number of favorable properties, including allowing for skewness and heterogeneity in mean revenue amounts across the customer base with only 3 parameters, and having simple closed-form expressions for key metrics including the expected mean revenue value given the spend hyperparameters $(\tilde{p}, q, \tilde{\gamma})$ and historical data (x, m_x). The model scales well as a result, lending itself well to large-scale analyses such as this one, making it a sensible way to model average revenue per transaction.

If we pursued this approach, we would then need to make a gross margin assumption to predict individual-level profit per transaction. FHL and Reinartz and Kumar (2000, 2003) assume a gross margin of 30% because transaction-level profitability figures were not available, and to err on the side of conservatism. In other words, the individual-level gross margin is assumed to be homogeneous and constant across the customer base. Is this a reasonable assumption? We have transaction-level gross margin data within our dataset, so we can examine its empirical validity. Let $W_{i,j}$ denote the gross margin associated with the ith transaction for customer j, respectively, in dollars. If margin were truly constant, then all transactions and thus all customers would have the observed same gross margin percentage:

$$\frac{\sum_{i=1}^{x_j} W_{i,j}}{\sum_{i=1}^{x_j} Z_{i,j}} = \frac{\sum_{i=1}^{x_{j'}} W_{i,j'}}{\sum_{i=1}^{x_{j'}} Z_{i,j'}} \ \forall j = 1, 2, \cdots, J. \tag{13}$$

If the gross margin were homogeneous across customers but randomly varied across transactions around a baseline average level, we would expect the historical average gross margin across customers to be reasonably tightly clustered about the overall firm-level gross margin. A highly disperse empirical historical average gross margin distribution would be indicative of heterogeneity in customer profitability which cannot be explained by average revenue volume alone (i.e., persistent bargain hunters mixed with list-price shoppers). To test this, we plot on the left in Figure 2 the empirical historical average gross margin across customers. This plot is suggestive of considerable heterogeneity (and left skewness) in the gross margin percentage across the customer base, about the firm-level average of 57.6%. Perhaps most interestingly, this strong heterogeneity in profitability would have been hard to detect had we focused upon the transactions themselves – indeed, the correlation between revenue per transaction and cost per transaction is very high, at 93.4%. This profitability heterogeneity is due to the existence of bargain hunters only purchasing when deals arise at one extreme, price insensitive list price shoppers at the other extreme, and a seemingly relatively continuous distribution of customers falling in between. It is not appropriate to assume a constant gross margin percentage across the customer base, and we suspect that this is a fairly general observation.

When the primary goal is prediction of CLV-related quantities and transaction-level profitability data is available, a better approach is to model gross profit directly in a very similar fashion to how we would have modeled revenue per transaction. Assume that the gross profit derived from a transaction fluctuates randomly about the customer's (unobserved) mean gross profit level, and that mean gross profit varies across customers but not over time, independent of the volume of transactions[3].

One complication worth noting is that gross profit may be negative. In the cohort of customers considered here for example, 4,689 customers had negative historical average gross margins. While these customers represent a small percentage of the total customer base, we do not want to ignore them because they are particularly adept at finding bargains. Furthermore, this problem is not specific to profitability modeling. In most retail transaction databases (including the one considered here), total revenue may also be negative as well for a variety of reasons (i.e., customers whose returns exceed their recorded purchases due to gifting, etc.). Schmittlein and Peterson (1994) offer one possible solution, proposing a model which assumes that spend amounts per transaction vary randomly about each customers' mean according to a normal distribution, with heterogeneity in mean spend amounts across the customer base varying according to a normal distribution. While this 'normal-normal' model allows for customers with negative mean gross profit per transaction, it also assumes that the resulting distribution of transaction values is normally distributed. This assumption is empirically violated by the data – revenue and gross profit per purchase are highly right skewed. For example, average skewness of gross margin per transaction-week is 5.3, resulting in an average gross margin (at \$119) which is far larger than the mode (at \$27).

We propose a variant of FHL's gamma-gamma model we call the *shifted* gamma-gamma model to parsimoniously account for heterogeneity in the mean gross profit level and right skewness in its distribution. Letting $W_{i,j}$ denote the gross profit associated with the ith transaction for customer j, the shifted gamma-gamma model posits that if $W_{i,j}$ were shifted rightwards by a shift parameter s, the resulting distribution is gamma($\tilde{p}, \tilde{\gamma}$)-distributed:

$$W_{i,j} + s \equiv \tilde{W}_{i,j} \sim \text{gamma}(\tilde{p}, \nu_j), \quad i = 1, 2, \cdots, x_j, \quad j = 1, 2, \cdots, J. \tag{14}$$

Analogous to the gamma-gamma model, different customers have different baseline average profitability levels, which is induced by making ν_j gamma($q, \tilde{\gamma}$)-distributed across customers:

$$\nu_j \sim \text{gamma}(q, \tilde{\gamma}). \tag{15}$$

Let $g_{x,j}$ and $\tilde{g}_{x,j}$ be the average observed gross profit per customer and its shifted counterpart, $g_{x,j} = \sum_{i=1}^{x_j} w_{i,j}/x_j$ and $\tilde{g}_{x,j} = \sum_{i=1}^{x_j} \tilde{w}_{i,j}/x_j$. It follows from the convolution and scaling properties of the gamma distribution that the distribution of $\tilde{g}_{x,j}$ is gamma($\tilde{p}x_j, \ \nu x_j$)-distributed:

$$f(\tilde{g}_{x,j}|s, \tilde{p}, \nu_j, x_j) = \frac{(\nu x_j)^{\tilde{p}x_j} \tilde{g}_{x,j}^{\tilde{p}x_j - 1} e^{-\nu x \tilde{g}_{x,j}}}{\Gamma(\tilde{p}x_j)}. \tag{16}$$

Taking the expectation of $f(\tilde{g}_{x,j}|s,\tilde{p},\nu_j,x_j)$ over the distribution of ν_j, we arrive at the following marginal distribution for $\tilde{g}_{x,j}$:

$$f(\tilde{g}_{x,j}|s,\tilde{p},q,\tilde{\gamma},x_j) = \frac{\Gamma(\tilde{p}x+q)}{\Gamma(\tilde{p}x)\Gamma(q)} \frac{\tilde{\gamma}^q \tilde{g}_x^{\tilde{p}x-1} x^{\tilde{p}x}}{(\tilde{\gamma}+\tilde{g}_x x)^{\tilde{p}x+q}} \tag{17}$$

Let GP_j be the random variable representing the average gross profit per transaction for customer j. Then the unconditional mean of GP_j conditional upon hyperparameters $(s, x\tilde{p}, q, \tilde{\gamma})$ and historical gross profit data $(g_{x,j}, x_j)$ is

$$E(GP_j|s,\tilde{p},q,\tilde{\gamma},g_{x,j},x) = \left(\frac{q-1}{\tilde{p}x+q-1}\right)\frac{\tilde{\gamma}\tilde{p}}{q-1} + \left(\frac{\tilde{p}x}{\tilde{p}x+q-1}\right)\tilde{g}_x - s, \tag{18}$$

a weighted average of the global average gross profit per transaction across the customer base, $\tilde{\gamma}\tilde{p}/(q-1)$, and the customer's own observed shifted gross profit, $\tilde{g}_{x,j}$, shifted back to the left by s.

Let $\mathbf{g}_x$ denote the resulting distribution of observed average gross profit per transaction figures across the customer base, $\mathbf{g}_x \equiv (g_{x,1}, g_{x,2}, \ldots, g_{x,J})$. We know that $s >$ -$\min(\mathbf{g}_x)$ to ensure that the support of $\tilde{g}_x$ is strictly positive. How much further to the right s should be is determined by the data – all hyperparameters, $(s,\tilde{p},q,\tilde{\gamma})$, are estimated via maximum likelihood.

As we will show when we validate the profitability submodel, the out-of-sample performance of the shifted gamma-gamma model is very impressive. While it is tempting to add further bells and whistles to this profitability model, we are likely to do more harm than good given this model's strong combination of out-of-sample goodness-of-fit, predictive performance, and model parsimony.

Returning to our original goal, what are the expressions we need to plug in to Equation 6 to derive $V(CLV)$? The first moment of GP has already been provided in Equation 18. In the Appendix, we derive the corresponding second moment expression by first deriving an expression for the variance, $V(GP|s,\tilde{p},q,\tilde{\gamma},\tilde{g}_x,x)$:

$$V(GP|s,\tilde{p},q,\tilde{\gamma},g_x,x) \;=\; \frac{\tilde{p}(\tilde{\gamma}+\tilde{g}_x x)^2}{\tilde{p}x+q-1}\left(\frac{\tilde{p}x+1}{x(\tilde{p}x+q-2)} - \frac{\tilde{p}}{\tilde{p}x+q-1}\right). \tag{19}$$

With this expression in hand, it follows that

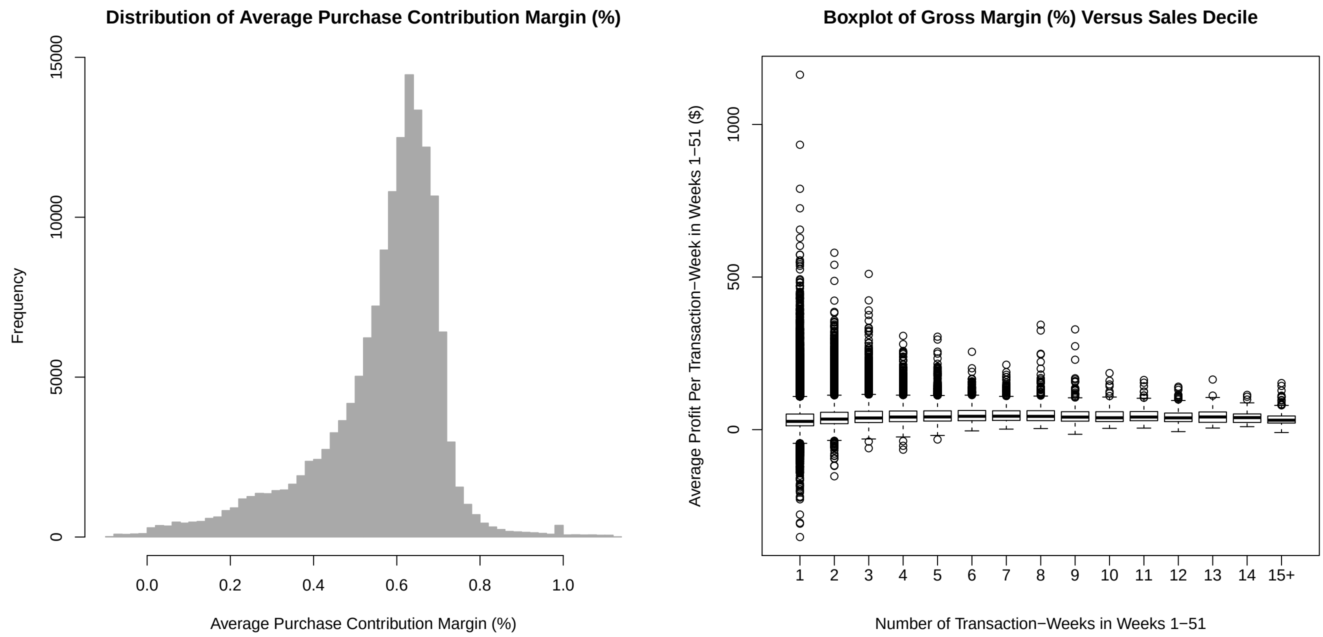
$$E(GP^2|s,\tilde{p},q,\tilde{\gamma},g_x,x) \quad = \quad V(GP|s,\tilde{p},q,\tilde{\gamma},\tilde{g}_x,x)+(E(GP|s,\tilde{p},q,\tilde{\gamma},\tilde{g}_x,x))^2, \tag{20}$$

where $E(GP|s,\tilde{p},q,\tilde{\gamma},g_x,x)$ is provided in Equation 18.

Assessing Independence of Purchase Profitability with Purchase Volume

Independence of the transaction flow process and the transaction profitability process is crucial to the model we have laid out. If this assumption were not valid, all of the closed form expressions derived above and in the Appendix would not be available. Using data from the 157,966 customers who had a repeat purchase in the calibration period, a simple correlation between average purchase profit and the number of purchases is only 6.5%. A box-and-whisker plot provided on the right in Figure 2 summarizing the distribution of average purchase profit by number of repeat purchases in the first 51 weeks shows that virtually all of the variation is within-group and not between-group. More pertinent to the concept of $V(CLV)$, we also see that within-group variability in average profit per purchase is generally declining as the number of purchases increases. This follows from Equation 19 and implies that all else equal, highly transacting customers have the favorable property of having less uncertain future gross profit per purchase.

Figure 2: Distribution of Average Contribution Margin (left) and Assessing Independence of Margin and Purchase Volume (right)



Bringing It All Together

We now have all the ingredients necessary to compute $V(CLV)$ for a customer with transaction history (x, t_x, n, g_x), having estimated the hyperparameters of the BG/BB and shifted gamma/gamma models $(\alpha, \beta, \gamma, \delta, s, \tilde{p}, q, \tilde{\gamma})$. For concision, we hereafter refer to all customer data collectively by $\mathbf{D}$ and all estimated hyperparameters by $\boldsymbol{\psi}$. Equation 6 is equivalent to

$$V\left(CLV|\boldsymbol{\psi}, \mathbf{D}, d\right) \;=\; \frac{1}{(1+d)^{2n}}\left(E(RLV^2|\boldsymbol{\psi}, \mathbf{D}, d) - E(RLV|\boldsymbol{\psi}, \mathbf{D}, d)^2\right) \tag{21}$$

Because of the independence of transaction flow and transaction profitability,

$$E(RLV|\boldsymbol{\psi}, \mathbf{D}, d) \;=\; E(GP|\boldsymbol{\psi}, \mathbf{D})E(DRT|\boldsymbol{\psi}, \mathbf{D}, d) \tag{22}$$

which involves known expressions (Equation 18 above and Equation 14 from FHS). We will show in the Appendix that the second moment of RLV for a particular customer is

$$(23) \qquad \begin{aligned} E(RLV^2|\boldsymbol{\psi}, \mathbf{D}, d) &= E(GP^2|\boldsymbol{\psi}, \mathbf{D})E(DRT|\boldsymbol{\psi}, \mathbf{D}, 2d + d^2) + D_1, \\ D_1 &= E(GP|\boldsymbol{\psi}, \mathbf{D})^2(E(DRT^2|\boldsymbol{\psi}, \mathbf{D}, d) - E(DRT|\boldsymbol{\psi}, \mathbf{D}, 2d + d^2)) \end{aligned}$$

Plugging Equations 22 and 23 into 21, it follows that

$$(24) \qquad \begin{aligned} V(RLV|\boldsymbol{\psi}, \mathbf{D}, d) &= E(GP^2|\boldsymbol{\psi}, \mathbf{D})E(DRT|\boldsymbol{\psi}, \mathbf{D}, 2d + d^2) + E(GP|\boldsymbol{\psi}, \mathbf{D})^2 \times E_1 \\ E_1 &= V(DRT|\boldsymbol{\psi}, \mathbf{D}, d) - E(DRT|\boldsymbol{\psi}, \mathbf{D}, 2d + d^2) \end{aligned}$$

We have expressions for all terms involved (Equation 20 above, Equation 14 in FHS, Equations 18 and 10 above, and Equation 14 in FHS, respectively). For the first time, then, we have an expression for $V(RLV)$ and thus $V(CLV)$ for individual customers.

Marketing managers may also be interested in an expression for the variance of total future profit generated by a customer over a finite horizon of time, conditional upon that customer's transaction history, which we denote by $V(V(n, n + n^*)|\boldsymbol{\psi}, \mathbf{D})$. To arrive at this expression, we first provide a derivation in the Appendix for the second moment of $V(n, n + n^*|\boldsymbol{\psi}, \mathbf{D})$:

$$(25) \qquad \begin{aligned} E(V(n, n + n^*)^2|\boldsymbol{\psi}, \mathbf{D}) &= E(X(n, n + n^*)|\boldsymbol{\psi}, \mathbf{D})E(GP^2|\boldsymbol{\psi}, \mathbf{D}) + E(GP|\boldsymbol{\psi}, \mathbf{D})^2 \times F_1, \\ F_1 &= E(X(n, n + n^*)^2|\boldsymbol{\psi}, \mathbf{D}) - E(X(n, n + n^*)|\boldsymbol{\psi}, \mathbf{D}), \end{aligned}$$

all terms of which we have closed-form expressions for (Equation 10 from FHS, Equations 18, 20, and 11 above, and Equation 10 from FHS, respectively). Subtracting the square of expected total future spend from Equation 25 gives us the following expression for the variance of total future spend for a particular customer:

$$(26) \qquad \begin{aligned} V(V(n, n + n^*)^2|\boldsymbol{\psi}, \mathbf{D}) &= E(X(n, n + n^*)|\boldsymbol{\psi}, \mathbf{D})E(GP^2|\boldsymbol{\psi}, \mathbf{D}) + E(GP|\boldsymbol{\psi}, \mathbf{D})^2 \times G_1, \\ G_1 &= V(X(n, n + n^*)^2|\boldsymbol{\psi}, \mathbf{D}) - E(X(n, n + n^*)|\boldsymbol{\psi}, \mathbf{D}), \end{aligned}$$

where $V(X(n, n+n^*)^2|\boldsymbol{\psi}, \mathbf{D})$ is defined Equation 12.

MODEL VALIDATION

Next, we validate how well we predict the mean and the variance of the sub-models we have built for transaction flow and transaction profitability. Only after we have rigorously established the predictive validity of these models, for both mean and variance predictions, will we turn to variance insights implied by these models. Before we begin, though, we provide a more thorough description of the dataset.

Data Description

The acquisition date of customers is defined to be the date of their first purchase with the company. As a result, the dataset is not left-censored. The dataset contains separate line items for returns, but does not link returns to previous purchases. Returns are not treated as purchase events because doing so would understate revenue and/or profit per transaction (i.e., if a customer were to purchase \$200 of goods, then return \$.01 of those goods, counting the second return as a purchase event would incorrectly suggest a historical revenue per purchase of approximately \$100). One of the appealing properties of the gamma/gamma model is that it allows us to model total revenues and/or profit per customer (see, for example, Equation 16). We exploit this property of the model, subtracting the profit associated with product returns from total profit in the numerator of $g_{x,j}$, while the denominator consists only of purchases, not of returns.

As is frequently the case in practice, our dataset does not include any demographic information, nor does it contain a historical log of corporate marketing activities. Therefore while we cannot incorporate demographics or marketing activities into our CLV model, we will show that the proposed CLV model does an excellent job of predicting CLV without this additional information.

Validating Means

We validate the BG/BB model for transaction flow. The maximum likelihood estimates of the BG/BB model are obtained by maximizing the marginal loglikelihood equation of the observed data (Equation 5 of FHS).

	α	β	γ	δ	LL
Estimate	.58	24.76	.26	2.22	-1529872
SE	.006	.176	.003	.068	

The model is well-identified with stable parameter estimates, as evidenced by the small standard errors in relation to hyperparameter estimates themselves. Next, we consider four common measures of in-sample and out-of-sample model goodness of fit which have been used in prior literature (Ascarza and Hardie 2013; Fader, Hardie and Lee 2005a; FHL; FHS).

Figure 3: Validating In-Sample Purchase Frequency (left), Incremental Purchases (right)

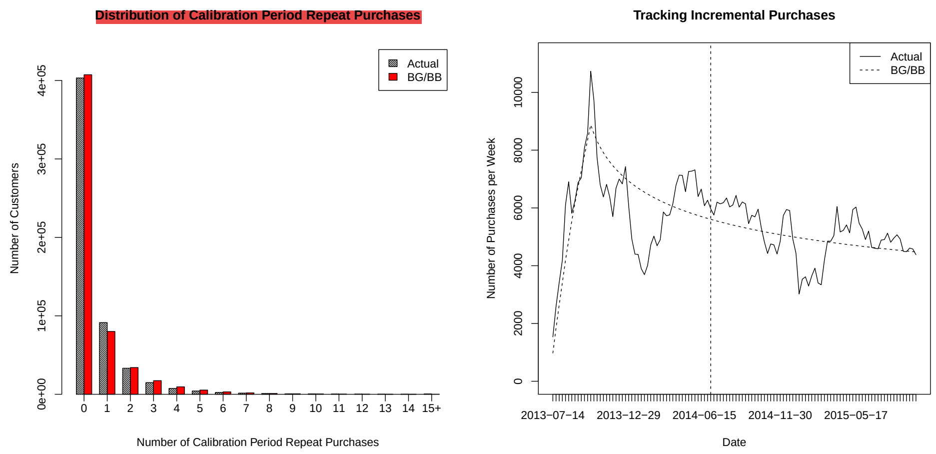


Figure 4: Validating Cumulative Purchases (left), Conditional Expected Purchases (right)

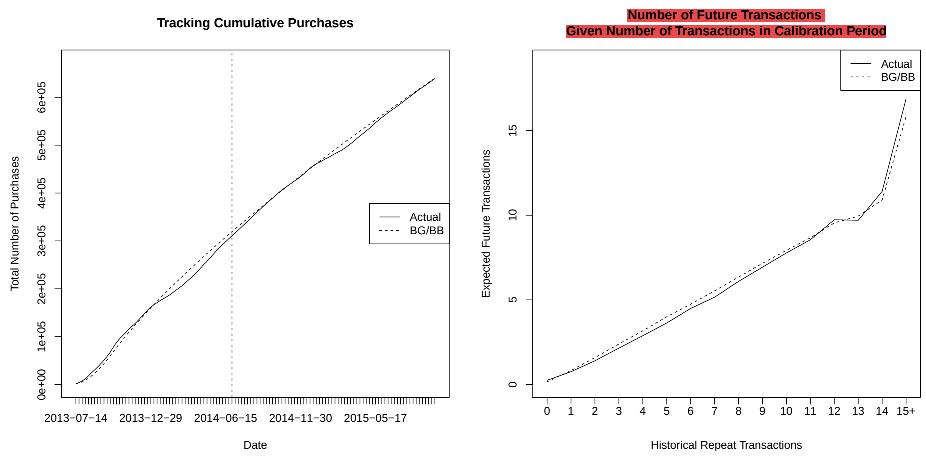
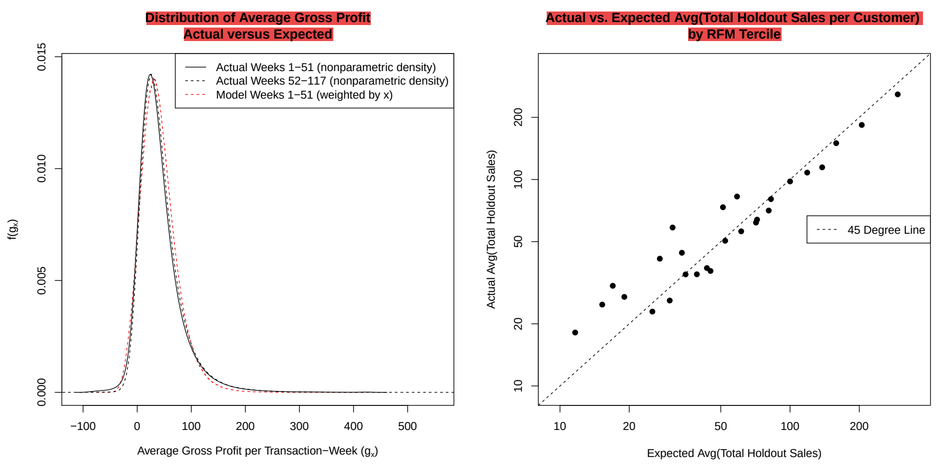


Figure 5: Validating Margin Distribution (left), Holdout Sales by RFM Group (right)



On the left in Figure 3 we compare the number of customers expected to purchase x times within the calibration period versus the number of customers who actually transaction x times, $x = 0, 1, 2, \ldots, 15+$. The frequency distribution is a simple byproduct of $P(X(n) = x|\boldsymbol{\psi})$, the probability that a customer with n purchase opportunities makes x purchases (Equation 7 in FHS), summing over customers with different values of n. We see near-perfect correspondence between the resulting distributions. The Pareto principle is valid for this cohort – 80% of all calibration-period repeat purchases were made by only 17.1% of the customers.

Next, we evaluate our ability to model and predict the flow of transactions over time, in-sample and out-of-sample. On the right in Figure 3, we plot the expected and actual number of purchases we expect to occur each week (i.e., the first differences of the total number of expected purchases over time shown on the left in Figure 4), where the expected number of transactions we expect to occur in week k is equal to $E(X(k)|\boldsymbol{\psi}) - E(X(k-1)|\boldsymbol{\psi})$, $k = 1, 2, \ldots, n + n^*$ (Equation 10 in FHS). While business is seasonally strong in the Spring and Fall (as is the case for many retailers and home sellers), this seasonality averages itself out over time and thus has little effect upon CLV. What is most important for CLV-related predictions is that we are capturing the long-run baseline level of demand accurately, which is evident from the left plot within Figure 4. In fact, adding an annual seasonal covariate when only a year of data is available to calibrate upon can be very dangerous: the effect of seasonality would not be well identified and noise during the seasonal period could be mistaken for signal, diminishing predictive accuracy in the holdout period.

Our final transaction flow validation plot evaluates our ability to make individual-level forecasts. On the right in Figure 4, we compute the expected and actual average number of purchases that will occur in the holdout period for a given customer, conditional upon the number of purchases that customer made in the calibration period. This is computed by evaluating $E(X(n, n+n^*)|\boldsymbol{\psi}, x, t_x, n)$ (Equation 13 in FHS) for all customers and averaging over t_x and n. We again see near-perfect correspondence between what expected and actual holdout purchasing.

Next, we validate the profitability submodel, after which we combine the transaction flow and monetary value models to validate $E(CLV)$. The maximum likelihood estimates of the shifted gamma/gamma model are obtained by maximizing the marginal loglikelihood equation for the shifted gamma/gamma model, in

Equation 17.

	s	$\tilde{p}$	q	$\tilde{\gamma}$	LL
Estimate	83.23	23.93	23.41	117.00	-772849
SE	.77	.09	.10	.48	

The estimated shift parameter s is $\$7.87$ larger than the minimum shift of $\$75.36$ needed to ensure the support of $\tilde{g}_x$ is strictly positive. The difference is also more than 10 times greater than the standard error associated with our estimate of s, implying that the shifted gamma-gamma model's goodness of fit is significantly improved over a gamma-gamma model which fixes s, implying that s is both statistically and practically greater than the minimum required shift. Parameter estimates are stable, as evidenced by the small magnitude of the standard errors relative to the MLE estimates themselves.

On the left in Figure 5 we plot the actual distribution of average gross profit per purchase across customers in the calibration period, alongside the theoretical distribution we would have expected under the shifted gamma/gamma model. For comparison purposes, we overlay the actual observed distribution of average gross profit per purchase in the holdout period, $\mathbf{g}_x^* = \{g_{x,1}^*, g_{x,2}^*, \ldots, g_{x,J}^*\}$, where $g_{x,j}^* = \sum_{i=1}^{x_j^*} w_{i,j}/x_j^*$, for $x_j^* > 0$, and x_j^* represents the number of transactions made by customer j in the holdout period.

The shifted gamma-gamma model provides excellent marginal in-sample and out-of-sample goodness of fit. Moreover, the fact that the calibration and holdout periods effectively share the same distribution provides strong support for the assumption that this distribution does not vary over time.

The model correctly suggests a very strong "regression-to-the-mean" effect. For example, customers with no prior transactions are regressed up to the global average across the customer base of \$41.69, which is within just 4% of the actual realized holdout gross profit per transaction for these customers of \$43.08.

Ultimately we are interested in how well the sub-models for flow and profitability combine to predict future spending for individual customers. To test this, we group customers on the basis of their RFM characteristics, as is often done in traditional RFM segmentation analyses. We assign each customer into one of 28 RFM groups after scoring all customers on the basis of R, F, and M. We construct the 28 groups as follows. Customers who made no repeat transactions are given scores of $R = F = M = 0$. We sort all remaining customers by recency, then score customers in the top (highest recency) tercile $R = 3$, customers in the second tercile by $R = 2$, and customers in the third tercile by $R = 1$. We repeat this process for frequency, sorting and then scoring customers in the top (highest) frequency tercile $F = 3$, and so on. We repeat this same process for monetary value, assigning scores of $M = 3$, $M = 2$, and $M = 1$ in the same manner. For each of these RFM groups, we obtain the empirical average total holdout spend per customer, as well as the variance of total future gross profit across all customers within each profit group:

$$E(V_{g(k)}((n, n+n^*))_{actual} = \frac{1}{n_{g(k)}} \sum_{j \in g(k)} x_j^* \times g_{x,j}^* \equiv \bar{v}_{g(k)} \tag{27}$$

where $k = 1, 2, \ldots, 28$, x_j^* is the total number of holdout purchases made by customer j, $g_{x,j}^*$ is the total gross profit per purchase in the holdout peirod for customer j, $n_{g(k)}$ is the number of customers within RFM group k, and $g(k)$ denotes the collection of customers within RFM group k. We compare these actual RFM group means to what we would have predicted them to be:

$$\widehat{E}(V_{g(k)}(n, n+n^*)|\boldsymbol{\psi}, \mathbf{D}) = \frac{1}{n_{g(k)}} \sum_{j \in g(k)} E(GP|\boldsymbol{\psi}, \mathbf{D}_j)E(X(n, n+n^*)|\boldsymbol{\psi}, \mathbf{D}_j) \tag{28}$$

using Equation 10 from FHS and Equation 18 above. The relationship between expected and actual holdout sales are most naturally visualized using a log-log scatterplot, which we provide on the right in Figure 5. Each dot corresponds to one of the RFM groups. A 45 degree line is added, so expected and actual holdout sales are equal for all points lying upon this line. There is a strong correspondence between actual and expected total holdout sales is strong in general across the 28 RFM groups, with a correlation of 98.4% between expected and actual across the groups.

Validating Variances

Having validated our ability to forecast what future purchases and spend will be, we must now validate our ability to predict the uncertainty about those forecasts. Only after we are comfortable with our ability to reliably predict variance will we be comfortable studying what our model implies for $V(DRT)/V(CLV)$.

We first validate transaction-flow variance longitudinally over time – in English, while we may know how many total purchases customers are expected to make over time, we would like to assess whether we also know how much variation there will be around our point estimate. That is, we compare the expected variance of cumulative transactions over time across all customers, $\sqrt{J \times V(X(0,t)|\boldsymbol{\psi})}$ (see Equation 12, letting $x = t_x = n = 0$ and $n^* = t$), to its empirical counterpart:

$$\sqrt{J \times V_{actual}(0,t)} = \sqrt{J/(J-1) \times \sum_{j=1}^{J} \left(y_{j,\sum 1:t} - \bar{y}_{\cdot,\sum 1:t}\right)^2}, \quad \text{where} \tag{29}$$

$$y_{j,\sum 1:t} = \sum_{i=1}^{t} y_{j,i}, \quad \text{and} \quad \bar{y}_{\cdot,\sum 1:t} = \frac{1}{J}\sum_{j=1}^{J} y_{j,1:t}. \tag{30}$$

The square root is taken so that the unit of measure used in this plot is dollars. This is the variance analog to the cumulative tracking plot provided in Figure 4. We plot the results on the left in Figure 6, along with its first differences on the right (analogous to the incremental tracking plot provided on the right in Figure 4) to examine the incremental contribution of each additional week to the variance of cumulative transactions over time.

Figure 6: Validating Cumulative (left) and Incremental (right) Purchase Variance over Time

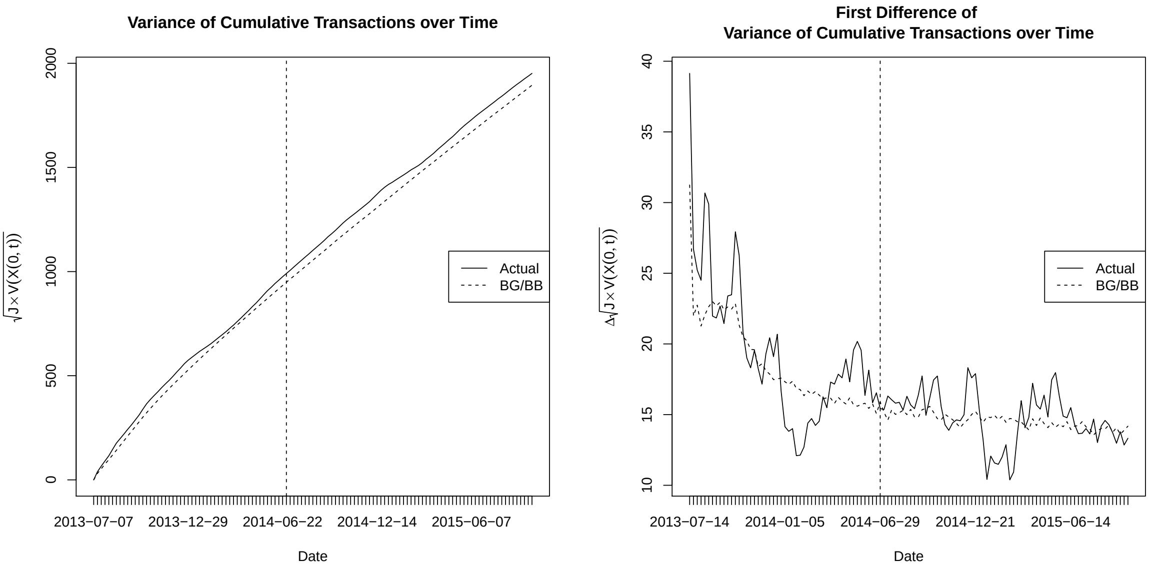
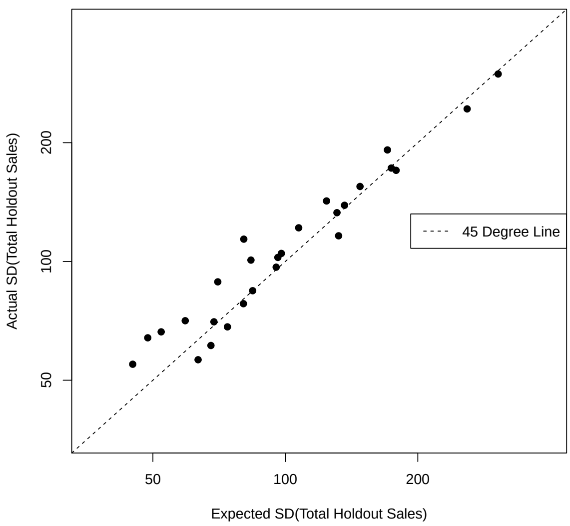


Figure 7: Total Standard Deviation of Sales in Holdout Period by RFM Group



While there is a small unexpected burst of variance at the beginning of the calibration period, the BG/BB model accurately predicts variance in cumulative transactions over time thereafter. The expected variance of cumulative transactions is generally within 4% of actual and the corresponding incremental plot shows that we accurately track the baseline incremental contribution of time to variance.

Studying the expected cumulative transactions plot on the left in Figure 4 alongside its variance counterpart on the left in Figure 6 is very relevant to marketing managers. If uncertainty were very high relative to expected total purchases, the future may be very different from what is expected. Conservative marketing managers faced with high uncertainty may want to apply a discount to their best guess of future total sales to account for this uncertainty (i.e., they may consider applying a one standard deviation "haircut" to their best guess). In the case considered here, uncertainty is small – by the very end of the holdout period on 9/27/2015, we expect approximately 640,000 purchases in total with an associated standard deviation of approximately 2,000, approximately .3% of the expectation. We have a (justified) high degree of confidence in the cumulative purchase behavior of the customer base as a whole.

Having established our ability to predict uncertainty over time, we now assess our ability to predict uncertainty for individual customers. We do so by forming RFM groups, as we had done when validating the means previously. After excluding the $R = F = M = 0$ group (because $V(GP|\boldsymbol{\psi}, \mathbf{D})$ does not exist when $x = 0$) this results in 27 groups. Our benchmark for comparison is the empirical variance of total future spend in the holdout period for each RFM group:

$$V(X_{g(k)}(n, n+n^*))_{actual} = \frac{1}{n_{g(k)} - 1} \sum_{j \in g(k)} \left(x_j^* g_{x,j}^* - \bar{v}_{g(k)}^*\right)^2, \tag{31}$$

where $k = 1, 2, \ldots, 27$, x_j^* is the total number of holdout purchases made by customer j, $g_{x,j}^*$ is the total gross profit per purchase in the holdout peirod for customer j, $n_{g(k)}$ is the number of customers within RFM group k, $g(k)$ denotes the collection of customers within RFM group k, and $\bar{v}_{g(k)}^*$ represents the average profit per customer in the holdout period for group k (Equation 27 above). We compare these actual RFM group holdout spend variances to what we would have predicted them to be, by the law of total variance:

$$\begin{aligned} \widehat{V}(V_{g(k)}(n, n+n^*)|\boldsymbol{\psi}, \mathbf{D}) &= \overline{V(V^*|\boldsymbol{\psi}, \mathbf{D})} \\ &\quad + \sum_{j \in g(k)} \frac{(E(V(n, n+n^*)|\boldsymbol{\psi}, \mathbf{D}_j) - \overline{E(V^*|\boldsymbol{\psi}, \mathbf{D})})^2}{n_{g(k)} - 1} \end{aligned} \tag{32}$$

$$\overline{V(V^*|\boldsymbol{\psi}, \mathbf{D})} = \frac{1}{n_{g(k)}} \sum_{j \in g(k)} V(V(n, n+n^*)|\boldsymbol{\psi}, \mathbf{D}_j), \quad \text{and} \tag{33}$$

$$\overline{E(V^*|\boldsymbol{\psi}, \mathbf{D})} = \frac{1}{n_{g(k)}} \sum_{j \in g(k)} E(V(n, n+n^*)|\boldsymbol{\psi}, \mathbf{D}_j). \tag{34}$$

where Equation 32 is evaluated using our expressions for the mean and variance of holdout sales in Equations 26 and 28 above. We visualize the results in Figure 7, plotting the expected and actual standard deviation of total spend for the 27 RFM groups using a log-log scatterplot, as in Figure 5 on the right. Correspondence between expected and actual continues to be very strong, with a 98% correlation between them.

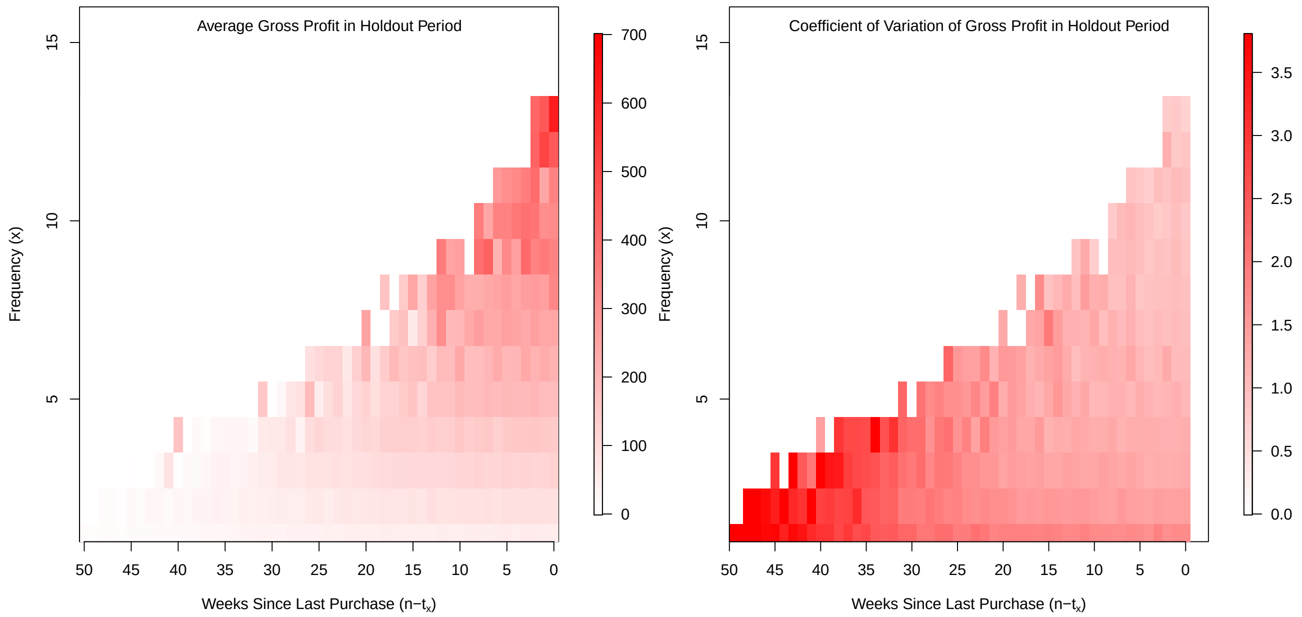
MODEL INSIGHTS

Now that we trust the predictions coming out of our CLV model, we explore some of the insights and implications that arise from it. We begin by studying the non-linear relationship between $E(RLV)$, $V(RLV)$ and RFM, and showing that transaction flow is a far larger driver of uncertainty in RLV than transaction profitability.

Mapping RFM to E(RLV) and V(RLV)

In Figure 1, we provided model-free evidence of how RFM maps to $E(RLV)$ and $CV(RLV)$. In Figure 8, we show the equivalent plot for an simulated dataset generated from our CLV model. That is, we generate a new dataset containing with the same number of customers having the same birth dates, but whose repeat purchases are drawn from a BG/BB repeat purchase process using the hyperparameters from the estimated transaction flow model, and whose gross profit amounts are drawn from a shifted gamma-gamma spend process using the hyperparameter estimates from the estimated transaction profitability model. Holdout average gross profit and coefficient of variation are then computed following a procedure identical to the one used to create Figure 1, computing the empirical mean and coefficient of variation of holdout profits for all RF cells with more than 10 customers within them.

Figure 8: Heatmaps of Week 52-117 Average Total Gross Profit and Coefficient of Variation of Total Spend By Time Since Last Purchase and Frequency; Data Simulated from Estimated Hyperparameters



Our simulated model-based plot is almost indistinguishable from the real data model-free evidence, and both share the same stylized facts noted in the introduction – that $CV(RLV)$ is generally an increasing function of recency and frequency, that this relationship breaks down when the time since the last purchase is over 30 weeks, but that the relationship is noisy. Next, we uncover the underlying relationships between RFM and $E(RLV)$ and $CV(RLV)$ systematically. We begin by moving from dollars to DRT, focusing on the relationship between the return and risk of DRT with recency and frequency. We will subsequently reintroduce profitability to complete the analysis. In this section, we characterize the risk of DRT through its coefficient of variation so that we can more easily study variation in uncertainty across customers without our analysis being confounded by $E(RLV)$. Thereafter, we will return to $V(RLV)$.

We visualize risk and return as a function of recency and frequency simultaneously through a heatmap-overlaid contour plot, in which a contour plot visualizes return and a heatmap overlaid on top visualizes risk. To obtain all values on the contour plot, we evaluate Equation 14 in FHS for all plotted recency/frequency combinations ($n - t_x = 0, 1, \ldots, 51; x = 0, 1, \ldots, 20$), fixing $n = 51$. To obtain all values on the overlaid heatmap, we separately evaluate $CV(DRT)$ by dividing Equation 14 in FHS by the square root of Equation

10 above. The results appear as a topographical heatmap in Figure 9[4].

The topographical heatmap provides context to marketing managers looking for insight not only into who their best customers are, but also which customers they know the least about. As we would have expected, the highest-value customers in the top-right of the plot are also generally the highest-certainty customers, as indicated by the lightness of the heatmap in that region. In general, certainty appears to be largely a recency effect – we are generally relatively certain about what high-recency customers will do in the future. For low-recency customers, we are generally much less certain what they will do, but this trend is attenuated for low-frequency customers.

There are many customers who have very similar means (i.e., they lie on the same iso-value "ridge") with dramatically different variances. Consider, for example, all customers residing on the $E(DRT) = 10$ ridge which begins in the bottom-right corner of Figure 9. As we wind our way from bottom-right to the left, then back upwards and to the right, we generally become more and more uncertain about DRT. These customers may share the same $E(DRT)$, but they are dramatically different from one another. Traditional value segmentation schemes (Hinshaw 2013) would put these customers into the same "low value" category, even though marketing managers knowing these facts would probably treat them very differently.

Customers who previously purchased very frequently and then suddenly stopped have very uncertain future value, as evidenced by the darkness of the heatmap along the diagonal of Figure 9. The high frequency with which they had purchased prior to going dormant implies they are most likely dead (and are unlikely to be worth much on average), but if they are not yet dead, they may resume purchasing frequently again. It may be profitable to invest in "probing" forms of marketing activities, with the intended goal of separating those who may still be alive from those who are certainly dead. If $CV(DRT)$ had been relatively low for these customers, such actions would not be sensible.

Having gained some insight into recency and frequency, we introduce average gross profit per purchase. As the frequency of purchases increases, we know intuitively (and can see empirically, on the left in Figure 3) that our uncertainty in the average gross profit per transaction will decline – but how quickly does our uncertainty decline, and is the magnitude of this effect the same for all historical gross profit per transaction amounts? We illustrate this in Figure 10, providing $E(GP|x, g_x, \boldsymbol{\psi})$ (on left) and $CV(GP|x, g_x, \boldsymbol{\psi})$ (on right) for three different values of historical average gross profit per transaction (g_x).

Figure 9: $E(DRT)$ and $CV(DRT)$ as a Function of Recency and Frequency

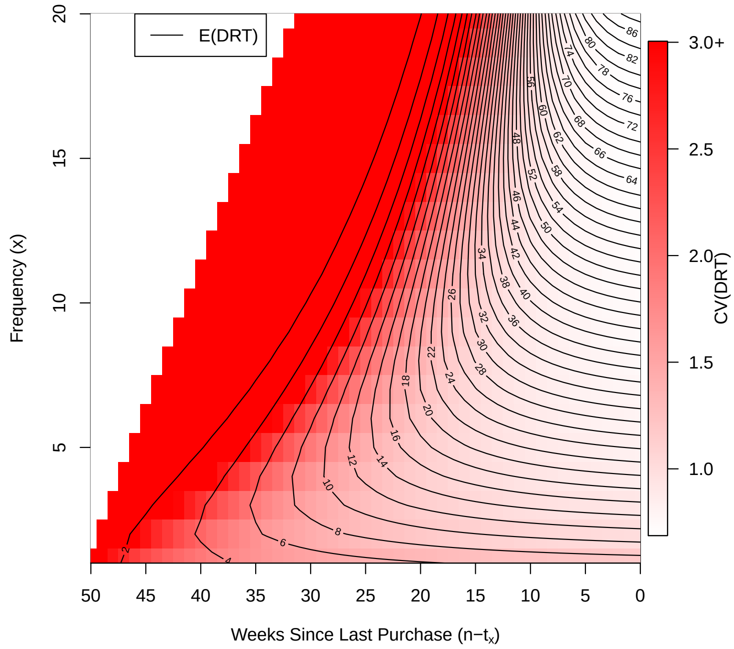
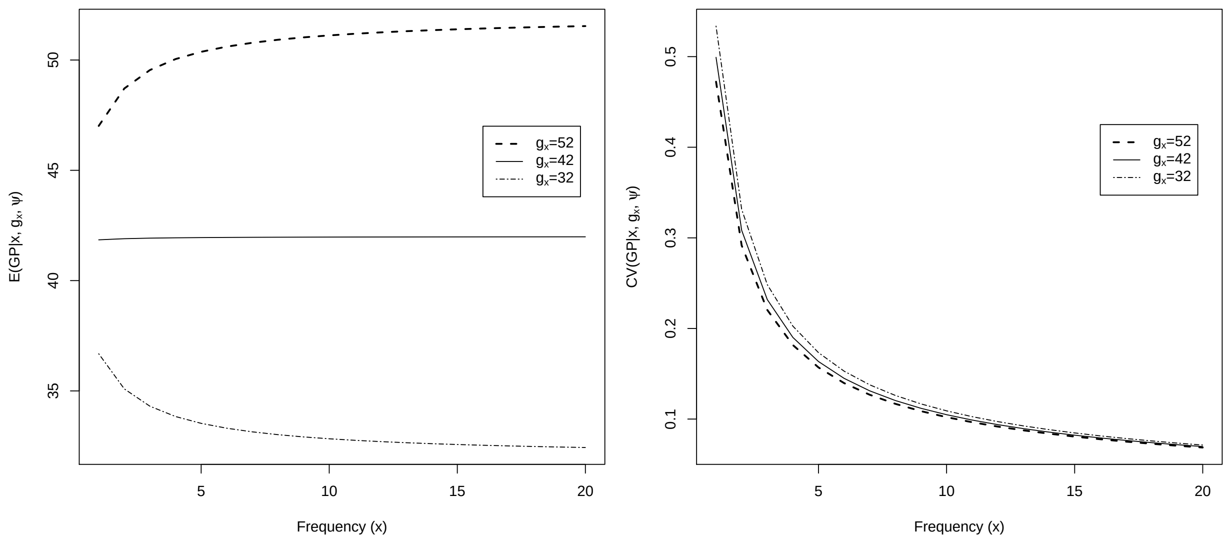


Figure 10: Expected Mean (left) and Coefficient of Variation (right) of Profit Per Transaction as a Function of Frequency and Historical Average Profit Per Transaction



The left-most plot in Figure 10 illustrates the "regression-to-the-mean" effect discussed in FHL – as a customer's number of repeat purchases increases, our best guess of that customer's future expected gross profitability per transaction gradually moves away from the global average of \$42, towards his/her own g_x. The right-most plot in Figure 10 highlights the relatively large reduction in uncertainty regarding future profit per transaction enjoyed by frequent buyers. $CV(GP)$ declines from .50 to .08 when x increases from 1 to 20 for customers with $g_x = \$42$, with most of the uncertainty reduction obtained by $x = 5$. There is also a small decline in $CV(GP)$ as g_x increases, pushing the line corresponding to $g_x = 52$ below the other two lines in the right-most plot in Figure 10 – this effect is a byproduct of the shift operation we performed, but does not have a practically meaningful effect upon $CV(GP)$.

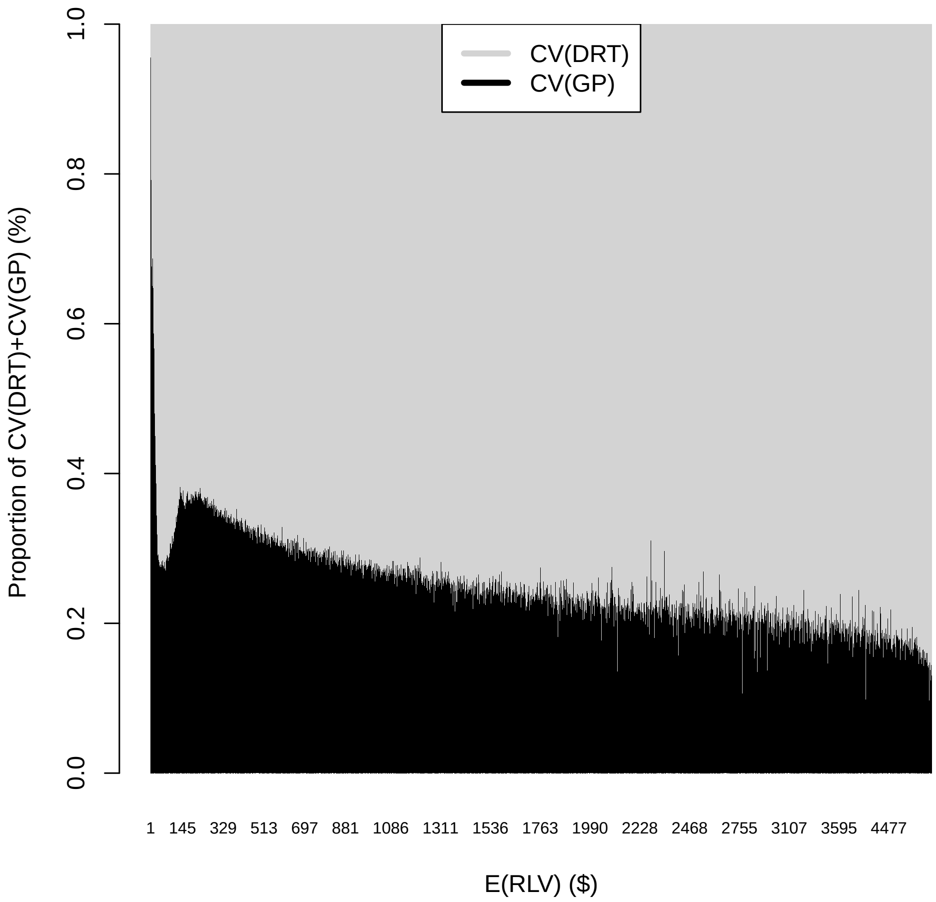
The transaction flow analysis suggests that higher recency strongly improves (decreases) $CV(DRT)$, whereas the effect of frequency is more complex. The transaction profitability analysis suggests that both higher frequency and higher average observed gross profit per purchase improve $CV(GP)$, but that frequency is by far the stronger of the two factors. Perhaps most importantly, though, what contributes the most to uncertainty in RLV – is it how many purchases the customers will make, or how profitable those purchases will be? We show that how many purchases the customer will make is the primary source of uncertainty next.

What Drives V(CLV)?

As with the preceding analysis, we characterize uncertainty through the coefficient of variation, which allows us to compare uncertainty due to transaction flow and uncertainty due to profitability in an apples-to-apples manner, and focus our attention upon repeat buyers ($x > 0$) for whom profitability variance is not infinite.

If $CV(DRT) > CV(GP)$, uncertainty in the how many purchases will be made in the future outweighs uncertainty in profit that will be earned on each of those transactions. In Figure 11 we plot for each customer $CV(DRT)$ and $CV(GP)$ as a percentage of their sum, $CV(DRT) + CV(GP)$. This allows us to study not only whether one factor tends to be larger than another, but also how this trend varies as a function of how valuable the customer is.

Figure 11: Proportion of $CV(DRT) + CV(GP)$ Arising From $CV(DRT)$ Versus $CV(GP)$, as a Function of $E(RLV)$



For the vast majority of repeat purchasers, transaction flow is the bigger of the two uncertainty drivers. Fully 96.9% of the customer base have a higher $CV(DRT)$ than $CV(GP)$. For the average customer, $CV(DRT)$ comprises approximately 2/3rds of $CV(DRT) + CV(GP)$. This implies that, empirically, transaction profitability is an important but secondary driver of $V(RLV)$. Knowing this, marketing managers may resolve more of their uncertainty regarding future profits by learning more about what prompts customers to make purchases, and perhaps less on how much these customers will spend while there. There is far less forward-looking variability in the latter process.

Figure 11 also suggests a well-behaved yet non-linear relationship between $E(RLV)$ and our uncertainty regarding that future value is coming from. Three notable inferences follow:

1. Not all customers have more uncertainty coming from transaction flow – for customers with very small expected values ($E(RLV < \$18)$), profitability is the larger driver. This segment consists of unprofitable (g_x between -\$30 and -\$60) but light-buying ($x = 1$) customers. The sparsity of their buying induces sharp "regression to the mean" in $E(GP)$, driving it to just north of \$0. We expect that these customers will make very few transactions in the future, so how profitable that next purchase will be is a larger driver of $E(RLV)$ than it usually is.

2. As $E(RLV)$ moves into the \$100-\$300 range, $CV(DRT)$ becomes a smaller source of total $CV(RLV)$. This is largely a recency effect – on average, $E(GP)$ and x remain approximately the same, with the bulk of the value increase coming from an increase in the average t_x from 7 to 22. As we know from Figure 9, improvement in recency has a strong beneficial (negative) effect upon uncertainty.

3. When $E(RLV)$ grows larger than \$300, $CV(DRT)$ generally becomes a larger and larger source of total $CV(RLV)$. Within this range, we see large increases in the average purchase frequency, x. Once x is reasonably large, increasing it further provides no material reduction in transaction profitability uncertainty, as we can see in the right-most plot of Figure 10, while uncertainty in the transaction flow process is far more persistent.

Having mapped RFM to $E(RLV)$ and $V(RLV)$ and studied the relative importance of flow and profitability, we now reverse the process and map $E(RLV)$ and $V(RLV)$ to RFM.

The Limitations of RFM Segmentation

We map $E(RLV)$ and $V(RLV)$ to RFM through the following process. We first score customers on the basis of R, F, and M. We create 28 RFM groups via the same process that we had followed in the Validation section. We then score customers on the basis of $E(RLV)$ and $SD(RLV)$ using the same process. We first rank all customers according to $E(RLV)$. Customers falling in the top quintile by $E(RLV)$ are given a score of 5, customers falling in the second highest quintile by $E(RLV)$ are given a score of 4, and so on. We then rank all customers according to $SD(RLV)$. Customers falling into the top quintile by $SD(RLV)$

are given a score of 5, customers falling in the second highest quintile a score of 4, and so on. Through this process, each customer has an E(RLV) score and a $SD(RLV)$ score between 1 and 5, resulting in 25 "EV cells". As a result, all repeat buyers are within one of 27 RFM groups, and one of 25 EV cells (customers making only one purchase during the calibration period are placed into one group under both segmentation schemes). In Table 1, we provide a 5 by 5 grid representing our EV cells, and then provide which RFM groups fall within the resulting 25 EV cells (each RFM group is placed into its most representative EV cell). Each RFM group is denoted by the concatenation of the R, F, and M scores (i.e., RFM group (123) represents all repeat buyers whose recency score R=1, frequency score F=2, and monetary value score M=3).

Table 1: Mapping of RFM Segments Within "EV Cells"

		SD Score				
		1	2	3	4	5
E Score	1	(111), (112), (121), (131)				
	2		(122), (132), (211), (311)	(113)		
	3		(221)	(212), (312)	(123)	
	4			(231), (321) (331)	(213), (222), (232), (313), (322)	(133)
	5					(223), (233), (323), (332), (333)

We see a very high degree of clustering of RFM groups within this 5 by 5 EV matrix, with 79% of the RFM groups falling into 20% of the EV cells and inadequate coverage of customers off the main diagonal. 6% of the customer base exists in EV cells which have no primary RFM group, and many EV cells are strongly over- and under-represented. For example, the (44) EV cell has 5 RFM groups within it and represents 13.7% of the customer base, while the (33) EV cell has only 2 RFM groups within it while also representing 13.7% of customers. Likewise, the (43) EV cell has 3 RFM groups within it but represents only 3.6% of customers. This would suggest that many RFM groups are redundant and can be collapsed together, while other groups need to be sub-divided to more effectively capture the variability on and around the diagonal of the EV matrix.

This has troubling implications for RFM customer segmentation, which has a long tradition and wide usage within the direct marketing community. RFM groups are created and then managed/targeted as distinct units (i.e., the business may perform A/B tests and/or run response models to evaluate the effect of marketing campaigns upon RFM groups). Doing so makes the strong assumption that customers are homogeneous in terms of what they will do in the future within each RFM group and are thus heterogeneous across RFM groups. Table 1 implies that there is low heterogeneity across RFM groups–many RFM groups share very similar expected means and variances.

A much simpler segmentation scheme, formed upon the basis of $E(RLV)$ and $SD(RLV)$ from a well-validated RLV model, provides superior customer segmentation results. Even though $E(RLV)$ and $SD(RLV)$ are not simple observables like RFM, computing them is fast and easy with models such as the one used in this analysis. If given the choice, most marketing managers would agree that it is preferable to segment customers on the basis of what they will likely do in the future and not on the basis of what customers had done in the past. Previous work has acknowledged this, proposing a related idea of segmenting the customer base purely upon $E(CLV)$ (i.e., through a "CLV pyramid" (Kumar et al. 2008; Rust, Zeithaml and Lemon 2001a; Zeithaml, Rust and Lemon 2001b). However, in the same way that recency and frequency represent different facets of the same customer, we need a multi-faceted forward-looking summary of the customer for reasons we have discussed at length. It would be challenging to come up with a uniform policy for all customers who share the same $E(RLV)$ because very different customer behaviors can result in the same $E(RLV)$. $V(RLV)$ is the most natural second dimension upon which to discriminate customers from one another.

We empirically test RFM segmentation's ability to create customer groups that are similar with respect to total profit in the holdout period against a simple, natural EV segmentation scheme. We evaluate both segmentations by holdout R^2. That is, we treat each customer's segment as a categorical predictor which is then regressed upon the total profit of the customer within the holdout period, then obtain the R^2 statistic associated with that regression. This is an intuitively appealing measure of a customer segmentation schemes ability to use historical data to explain variation in customers' future profitability. We compare the efficacy of RFM and EV-based segmentation by examining their performance in terms of this predictive performance measure.

While we could, in theory, operationalize EV segmentation into the 25 EV quintile-based groupings described earlier in this section, Table 1 makes it clear that this would be a highly inefficient segmentation scheme. Some EV cells are very sparsely populated (i.e., 6 EV cells have no customers within them and another 4 cells have less than 100 customers within them). If our goal is to create segments that are as homogeneous as possible in terms of future $E(RLV)$ and $SD(RLV)$, we should let the data tell us what our segmentation should be, optimizing how we "carve up" the customer base in $[E(RLV), SD(RLV)]$ space using a clustering algorithm. One popular clustering algorithm is k-means clustering (Hartigan and Wong 1979). That is, each customer j is represented by a 2-dimensional vector,

$$\mathbf{w}_j = [E(RLV|\boldsymbol{\psi}, \mathbf{D}_j), SD(RLV|\boldsymbol{\psi}, \mathbf{D}_j)]. \tag{35}$$

We partition repeat buyers into nG groups, $\mathbf{G} = G_1, G_2, \cdots, G_{nG}$, where these nG groups are formed using a standard procedure (we run Lloyd's algorithm, minimizing distortion with respect to $(\mathbf{w}_1, \ldots, \mathbf{w}_J)$ over 100 iterations with 200 different starting points for each iteration).

RFM segmentation is traditionally performed using tercile-based groupings as described in the Validation section, or the Hughes method (McCarty and Hastak 2007) which uses quintile-based groupings instead. While we could use one of these more traditional/popular RFM segmentation schemes, doing so results in subpar segmentation performance and is inconsistent with our EV segmentation approach. Instead, we let the data to tell us what our RFM segmentation scheme should be using k-means clustering, in the same way that we did so for EV segmentation. Instead of representing each customer j by a 2-dimensional vector on the basis of E(RLV) and $SD(RLV)$ (Equation 35), we represent each customer by a 3-dimensional vector on the basis of recency (t_x), frequency (x), and monetary value (g_x)[5]. This ensures that customers who are "closest" by RFM are placed into the same groups.

While we could simply set the number of customer groupings equal to a pre-specified number (i.e., 27), this would be wasteful if a smaller number of clusters achieves the same goodness of fit. Instead, we plot holdout R^2 as a function of the number of groups created. We let the number of groups vary from 1 to 27. This allows us to see how both methods perform in an absolute sense and relative to one another as a function of the number of groups created. We plot the results in Figure 12.

EV segmentation meaningfully outperforms RFM segmentation. EV segmentation with only 3 segments has a higher holdout R^2 than RFM segmentation with 27 segments. If we were to form customer segments on the basis of EV segmentation, it appears that we would not need more than 10 of them. This would make EV segmentation simpler to implement.

In summary, our mapping of $E(RLV)$ and $V(RLV)$ to RFM raised concerns regarding the ability of RFM segmentation to separate customers efficiently. These results suggest that a small number of segments formed using $E(RLV)$ and $V(RLV)$ will perform strictly better than RFM.

Scoring and Targeting Customers by the Return/Risk Ratio

Next, we discuss what our results imply for customer scoring and targeting. We begin by computing $E(RLV)$ as well as the ratio $E(RLV)/SD(RLV)$ (denoted hereafter by $E/SD(RLV)$) for each customer, which is a simple measure of customer return per unit of risk, analogous to the Sharpe ratio (Sharpe 1998) in finance when we assume there is no riskless (zero variance) customer. A scatterplot of the ranks of all customers along both metrics are provide in Figure 13.

Figure 12: Holdout R^2 as a Function of Number of Customer Segments

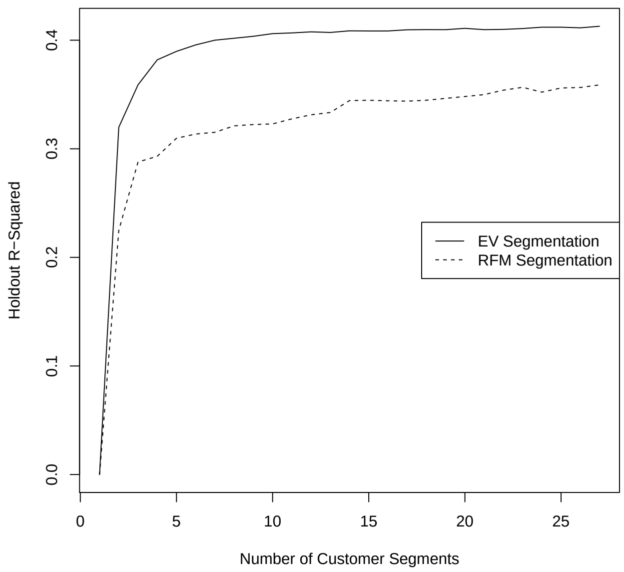


Figure 13: $E(RLV)$ vs. $E/SD(RLV)$, Arrows to Customers Discussed

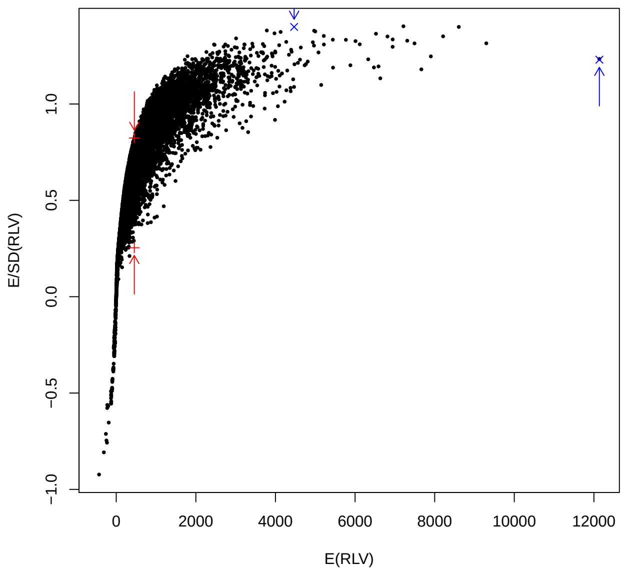


Figure 13 implies that while these two measures of customer performance measures are positively correlated (a raw correlation of the two measures is 79%), particularly at the lower extreme, the relationship is nonlinear. As we move towards the "middle of the pack," there is considerable variability in the return/risk ratio for a given level of $E(RLV)$. Finally, there seems to be no positive association at all between $E(RLV)$ and the return/risk ratio for customers who score the highest in terms of $E(RLV)$. The highest-value customers generally have especially high average profitability – some higher with lower recency, others lower with higher recency – causing large fluctuations in $E(RLV)$ without any real pattern in terms of $E/SD(RLV)$.

There is a sizable number of customers who are very similar in terms of E(RLV) who are quite different when we take into account $SD(RLV)$. Consider two customers, Variable Valerie and Steady Freddie ('+' symbols, pointed to with arrows, towards the left side of Figure 13). While they are identical in terms of $E(RLV)$ at $454, the similarities end there:

- Variable Valerie is a used to be reasonably profitable (heavy buying) customer. She made repeat purchases in 25 of her first 32 weeks, generating \$41 in average profit-per-purchase, but has not made any purchases in the past 12 weeks.

- Steady Freddie is a light but consistent buyer. He generates \$17 in profit, on average, when he buys, but he has bought 4 times and his last purchase was only 2 weeks ago.

Variable Valerie may have already ended her relationship with the firm, but if she hasn't, she could be worth much more in the future. In contrast, Steady Freddie is unlikely to ever be worth a fortune given how small and light his purchases are, but he will almost surely be back. As a result, Variable Valerie's $SD(RLV)$ is \$1788, much higher than Steady Freddie's \$552.

When thinking purely in terms of customer scoring, future expected risk-adjusted profitability is perhaps the most natural performance yardstick (Srivastava, Shervani and Fahey 1998), with the return/risk ratio being a natural operationalization. In this sense, Steady Freddie is clearly superior. All else being equal, customers that can provide the same return at lower risk should have higher scores.

When we move from customer scoring to opportunitistic customer targeting, however, it could very well be that the converse is true – for a given level of expected customer return, customers with more variability about that return could be better investment prospects. Rust, Kumar and Venkatesan (2011) would probably recommend that firms spend their re-targeting budgets upon Variable Valerie, trying to realize the sizable right tail to her future return distribution. Schwartz, Bradlow and Fader (2015) would also suggest investing in Variable Valerie, because firms have more prospect to learn (reduce uncertainty). They may curtail future investments in Steady Freddie on the premise that he will be coming back regardless.

To further emphasize the difference between $E(RLV)$ and $E/SD(RLV)$, it is informative to compare the very best customers along both measures. We highlight these customers towards the upper right in Figure 13 with '$\times$' symbols, with arrows pointing towards them.

The most valuable customer is a relatively infrequent buyer ($x = 9$) with a very high profit per transaction (\$329) who hasn't made a purchase in the past 5 weeks. He/she may be the highest ranked customer by $E(RLV)$, but his/her recency is creating uncertainty which pulls $E/SD(RLV)$ down to 1.23, worse than 100 other customers.

In contrast, the best customer by return-risk ratio is much "steadier", having made 25 purchases, including one in the most recent week. As a result, his/her $E/SD(RLV)$ was the highest ranked at 1.40. At the same time, his/her much lower average profit per transaction of \$40.25 pulled $E(RLV)$ down to \$4,469, far below the highest $E(RLV)$ of \$12,137. While both customers are clearly highly valuable by both performance measures, the differences between these two customers are substantial.

In summary, the return/risk ratio offers helpful dimension to customer scoring and targeting frameworks. $E(RLV)$ in isolation misses a key dimension of the forward-looking behavior of customers which may influence the service tier we should place them on, and the market actions we direct at them.

SUMMARY AND CONCLUSION

While point estimation of $E(CLV)$ has become very popular in recent academic literature, much less attention has been given to $V(CLV)$. Despite many brief mentions or analyses to this topic, very few have attempted to estimate $V(CLV)$ and none have validated specific $V(CLV)$ estimates or placed $V(CLV)$ at the center of their analyses. This paper offers several theoretically useful and substantively meaningful contributions to the CLV literature:

1. We derive closed-form expressions for the most important variance-related quantities, reducing considerably the computational effort needed to estimate $V(CLV)$.

2. We perform one of the largest-scale and most detailed CLV analysis to date within the marketing literature. Analysis at this scale makes fast, closed-form expressions for quantities of managerial interest particularly valuable to marketing managers.

3. We learn that the constant margin assumption is strongly violated by the data, uncovering an apparently continuous heterogeneity distribution for customer contribution margin.

4. We offer a number of diagnostic checks that we and other analysts can use to hold our models accountable to, and show that the predictive validity of our proposed CLV model with virtually no data pre-processing or modifications to the core model is surprisingly good.

5. We undertake the most thorough study of $V(CLV)$ in the marketing literature to date, learning, among other things, that (a) The most valuable customers are also the customers we are the most certain about, and vice versa, suggestive of customer-level "double jeopardy" effect (Ehrenberg, Goodhardt and Barwise 1990); (b) uncertainty in future value is largely a recency effect; (c) transaction flow is in general a much larger source of uncertainty than transaction profitability, implying that understanding the former process yields larger potential for reduction in forward-looking uncertainty.

6. We discover a non-linear relationship between recency, frequency, and $CV(RLV)$. When recency is low, increasing frequency greatly increases uncertainty.

7. We propose a risk/reward framework which exposes limitations of RFM segmentation. We propose a risk/reward-based segmentation scheme in its place which we show has meaningfully superior discriminant validity over RFM segmentation.

8. We propose the return-risk ratio ($E/SD(RLV)$) as an alternative scoring measure, and show that there are meaningful differences in the scores implied by the two measures. Examining these differences more closely reveals sharp differences in how firms may want to invest in/interact with customers.

Although these contributions are meaningful, this work is just the first step towards a full understanding of how firms should think about and leverage $V(CLV)$ to manage and grow custome value over time. Four natural extensions that could be considered by future work include (1) better understanding the nature of $V(CLV)$, (2) how $V(CLV)$ can be exploited (i.e., very carefully adding marketing mix variables to the model), (3) tying customer-level $V(CLV)$ to firm-wide optimization of the risk/return properties of the customer portfolio, and (4) moving from mean and variance to higher moments (i.e., skewness).

With respect to the first extension, an important question for marketing managers is whether the $V(CLV)$ is truly irreducible noise, as the model would suggest. If $V(CLV)$ is irreducible noise, then it should certainly be minimized (Srivastava, Shervani and Fahey 1998), for example by directing customer acquisition spend at customers proportionate to expected risk-adjusted returns. It may be (and marketing managers may hope), however, that customers/segments with particularly high $V(CLV)$ may actually represent multiple sub-segments which are more homogeneous within, some with higher value and others with lower value.

The firm could then invest in the low- and high-value customers commensurate to their future expected profitability, consistent with the approach taken for the rest of the customer base. It could be that these segments are identifiable in advance on the basis of observable attributes or latent customer propensities. To test this, marketing departments should pay very close attention to model diagnostics and maintain CRM systems replete with customer attribute data for validation purposes. If the adopted CLV model predicts $V(CLV)$ very well in terms of all diagnostic checks, the irreducible noise null hypothesis is more likely to be correct. One alternative hypothesis which may be particularly relevant for poor fitting models at firms in gaming/online media is that customer 'clumpiness' is present (Zhang, Bradlow and Small 2014), for example. Of course, if a stochastic RFM-C model was indeed the true model, then $V(CLV)$ for that model would again represent irreducible noise – whatever CLV model is being used, figuring out the nature of $V(CLV)$ is very important.

With respect to the second extension, one hypothesis that marketing managers may test through a properly designed field experiment is whether customers with very low uncertainty in terms of future value are on "auto-pilot." If the firm knows these customers will be coming back whether or not the firm offers them incentives to do so (i.e., coupons or discounts), the firm may save those resources, or redeploy them towards customers the firm is less sure about. This is a similar concept to the one proposed in Reinartz and Kumar (2000), except that latent customer trait driving firm decision making is uncertainty regarding future value and not the probability that the customer will make a purchase in the future.

As alluded to in the introduction, firms may also look to incorporate $V(CLV)$ or related quantities such as $V(E(CLV))$ to improve their customer acquisition decisions. Firms acquire customers who may be segmented by acquisition channel and/or customer attributes (i.e., demographics and socio-economic characteristics) over time. While these firms would like to invest in high-CLV customer segments, they will not know which segments have high CLV's until customers from each segment are acquired. This creates a so-called 'exploration versus exploitation' trade-off, and implies that customer segments with low historical returns may still be worth investing in if the firm knows very little about those segments. Algorithms to address simplified versions of this problem have been very well studied within the machine learning literature (see, for example, Auer, Cesa-Bianchi and Fischer (2002)'s `UCB1-Normal` algorithm and Kaufmann, Cappé and Garivier (2012)'s `Bayes-UCB` algorithm), generally proposing allocation policies which favor investment in segments with high historical means, high historical variances, and low historical investment. In our setting, with CLV as the objective function being maximized, $V(CLV)$ by segment is a natural proxy for how much is known about a particular customer segment, and may play a role in estimating how much learning benefit a firm may gain by investing in that segment. Framing the exploration/exploitation trade-off realistically and proposing a suitable allocation policy to address the trade-off is a topic for future research.

With respect to the third extension, it is natural to turn to financial portfolio theory for guidance regarding what companies should do with their customer bases – firms would like to optimize the risk / reward characteristics of their customer portfolios. We share these sentiments in theory, but we believe portfolio theory analogies should be made with caution as the assumptions underlying these analogies are important yet under-appreciated in the existing literature. For example, there are serious theoretical concerns with treating the overall customer base as a "market portfolio," and the underlying customer segments as "market portfolio constituents." Such an analysis often implies firms should "sell" or divest certain customer segments, but such a sale changes the composition of the market portfolio itself, which changes the covariances and thus values of all segments with the market portfolio (Buhl and Heinrich 2008). Moreover to calculate the beta of a segment in the first place its covariance with the market portfolio must be known, however the market portfolio is itself a weighted average of the unknown CLV's of all segments we are attempting to estimate. Thus, extending the framework we have constructed here into a financial portfolio theory context requires much more discussion and is beyond the scope of this paper.

With respect to the fourth extension, it may be that *skewness* or perhaps *kurtosis* are also relevant to consider – that is, for a given mean and variance, firms may favor investment in customers with higher right skewness when making opportunistic targeting decisions. While we believe this is the right direction to be moving in intellectually, there are a number of issues which must be considered. Perhaps least appreciated is whether or not those higher moments exist/are finite, a problem which is likely to be particularly acute for highly heterogeneous customer bases with many very light buying customers. In cases such as these, we may be able to obtain closed-form expressions but they may only be applicable to a small fraction of the customer base with relatively strong repeat purchasing. Holding this issue aside, there is also the issue of importance and operationalization. It is hard to deny that mean and variance are more important than skewness, but how much more? It is possible but not entirely clear how a marketing manager shoud construct a variance- and skewness-adjusted customer score which is then implemented at scale.

In summary, we believe risk (as measured by $V(CLV)$) is important and merits study alongside $E(CLV)$ on a regular basis: $V(CLV)$ should be modeled, tested, and applied with the same level of rigor that $E(CLV)$ has received. We view this paper as a call to action for further empirical studies to better understand $V(CLV)$ and apply it profitably in real-world settings.

WEB APPENDIX

In this appendix we present derivations for key results.

$\underline{E(X(n,n+n^*)^2|p,\theta,\mathbf{D})}$:

The second moment of $X(n,n+n^*)$ given individual-level parameters (p,θ) for a customer who is alive at time period n is

$$E(X(n,n+n^*)^2|p,\theta,alive_n) = E\left[(\sum_{j=1}^{n^*} Y_{n+j})^2|p,\theta,alive_n\right] \tag{36}$$

$$= E\left[\sum_{j=1}^{n^*} Y_{n+j}^2 + 2\sum_{j=2}^{n^*}\sum_{i=1}^{j-1} Y_{n+i}Y_{n+j}|p,\theta,alive_n\right] \tag{37}$$

$$= p\sum_{j=1}^{n^*}(1-\theta)^j + 2p^2\sum_{j=2}^{n^*}\sum_{i=1}^{j-1}(1-\theta)^j \tag{38}$$

Using the fact that

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x} \quad \text{and} \quad \sum_{j=n^*}^{\infty}\sum_{i=1}^{j-1} x^j = \frac{x^{n^*}(n^*-1-(n^*-2)x)}{(1-x)^2} \quad \text{for } |x|<1,$$

it follows that

$$p\sum_{j=1}^{n^*}(1-\theta)^j = \frac{p(1-\theta)(1-(1-\theta)^{n^*})}{\theta}, \quad \text{and} \tag{39}$$

$$2p^2\sum_{j=2}^{n^*}\sum_{i=1}^{j-1}(1-\theta)^j = 2p^2\sum_{j=2}^{\infty}\sum_{i=1}^{j-1}(1-\theta)^j - 2p^2\sum_{j=n^*+1}^{\infty}\sum_{i=1}^{j-1}(1-\theta)^j \tag{40}$$

$$= \frac{p^2}{\theta^2}\left((1-\theta)^2 - (1-\theta)^{n^*+1}(n^*-(n^*-1)(1-\theta))\right). \tag{41}$$

Plugging Equations 39 and 41 into Equation 38,

$$E(X(n,n+n^*)^2|p,\theta,alive_n) = \frac{p(1-\theta)}{\theta}\left[(1-(1-\theta)^{n^*}) + H_1\right] \tag{42}$$

$$H_1 = 2\frac{p}{\theta}(1-\theta)^2(1-(1-\theta)^{n^*-1}(n^*-(n^*-1)(1-\theta)) \tag{43}$$

Multiplying Equation 42 by the probability the customer is alive given individual-level parameters (p, θ) and data $\mathbf{D}$ (Equation A7 in FHS) gives us the desired expression:

$$
\begin{aligned}
&(44) \quad E(X(n, n+n^*)^2|p, \theta, \mathbf{D}) &=& \; E(X(n, n+n^*)^2|p, \theta, alive_n)P(alive_n|p, \theta, \mathbf{D}) \\
&(45) &=& \; I_1 \times I_2, \quad \text{where} \\
&(46) \quad I_1 &=& \; p^{x+1}(1-p)^{n-x}\theta^{-1}(1-\theta)^{n+1}, \quad \text{and} \\
&(47) \quad I_2 &=& \; (1-(1-\theta)^{n^*}) \\
&(48) && \; +2\frac{p}{\theta}(1-\theta)^2(1-(1-\theta)^{n^*-1}(n^*-(n^*-1)(1-\theta)).
\end{aligned}
$$

$\underline{E(X(n, n+n^*)^2|\boldsymbol{\psi}, \mathbf{D})}$:

The unconditional second moment of $X(n, n+n^*)$ can be computed by integrating Equation 45 against the joint posterior distribution of P and Θ (Equation A9 in FHS):

$$
(49) \qquad E(X(n, n+n^*)^2|\boldsymbol{\psi}, \mathbf{D}) = \int_0^1 \int_0^1 E(X(n, n+n^*)^2|p, \theta, \mathbf{D}) f(p, \theta|\boldsymbol{\psi}, \mathbf{D}) dp d\theta.
$$

Equation 11 follows.

We can also obtain an arbitrarily accurate approximation of it via averaging over Equation 45 with samples of (p, θ) obtained from the joint posterior distribution for P and Θ (Equation A9 in FHS). An exact, closed form sampling algorithm for $f(p, \theta|\boldsymbol{\psi}, \mathbf{D})$ is available using the following technique. Note that

$$
\begin{aligned}
&(50) \quad f(p, \theta|\mathbf{D}, \boldsymbol{\psi}) &=& \sum_{i=0}^{n-t_x} \frac{p^{\alpha+x-1}(1-p)^{\beta+t_x-x+i-1}\theta^{\gamma+\mathbb{1}_{i\neq n-t_x}-1}(1-\theta)^{\delta+t_x+i-1}}{B(\alpha,\beta)B(\gamma,\delta)P(\mathbf{D}|\boldsymbol{\psi})} \\
&(51) &=& \sum_{i=0}^{n-t_x} w_i \frac{p^{\alpha+x-1}(1-p)^{\beta+t_x-x+i-1}\theta^{\gamma+\mathbb{1}_{i\neq n-t_x}-1}(1-\theta)^{\delta+t_x+i-1}}{B(\alpha+x, \beta+t_x-x+i)B(\gamma+\mathbb{1}_{i\neq n-t_x}, \delta+t_x+i)} \\
&(52) \quad w_i &\equiv& \frac{B(\alpha+x, \beta+t_x-x+i)B(\gamma+\mathbb{1}_{i\neq n-t_x}, \delta+t_x+i)}{B(\alpha,\beta)B(\gamma,\delta)P(\mathbf{D}|\boldsymbol{\psi})}.
\end{aligned}
$$

Recognizing Equation 51 as the sum of Beta distributions with non-negative weights, the sum of the weights must equal 1 by construction. Therefore an exact sampling algorithm proceeds as follows:

1. Sample $i^{(k)} \sim$ Multinomial$(w_0, w_1, \ldots, w_{n-t_x})$.

2. Sample $p^{(k)} \sim$ Beta$(\alpha + x, \beta + t_x - x + i^{(k)})$.

3. Sample $\theta^{(k)} \sim$ Beta$(\gamma + \mathbb{1}_{i \neq n-t_x}, \delta + t_x + i^{(k)})$.

4. Repeat steps (1), (2) and (3) K times, where K is a large number (in practice, we set $K = 100,000$).

It then follows that

$$E(X(n, n+n^*)^2|\boldsymbol{\psi}, \mathbf{D}) = \int_0^1 \int_0^1 E(X(n, n+n^*)^2|p, \theta, \mathbf{D}) f(p, \theta|\boldsymbol{\psi}, \mathbf{D}) dp d\theta \tag{53}$$

$$\approx \frac{1}{K} \sum_{k=1}^{K} E(X(n, n+n^*)^2|p^{(k)}, \theta^{(k)}, \mathbf{D}), \tag{54}$$

where $(p^{(k)}, \theta^{(k)})$ is the kth sample obtained using the sampling procedure defined above.

$\underline{V(X(n, n+n^*)|\boldsymbol{\psi}, \mathbf{D})}$:

We know that

$$V(X(n, n+n^*)|\boldsymbol{\psi}, \mathbf{D}) = E(X(n, n+n^*)^2|\boldsymbol{\psi}, \mathbf{D}) - E(X(n, n+n^*)|\boldsymbol{\psi}, \mathbf{D})^2. \tag{55}$$

We computed the first term in Equation 11, while the second term in Equation 55 was derived in FHS (Equation 13).

$\underline{E(DRT^2|\boldsymbol{\psi}, \mathbf{D}, d)}$:

We know that

$$DRT^2 = \sum_{j=1}^{\infty} \frac{Y_{n+j}^2}{(1+d)^{2j}} + 2 \sum_{j=2}^{\infty} \sum_{i=1}^{j-1} \frac{Y_{n+i} Y_{n+j}}{(1+d)^{i+j}}, \tag{56}$$

which is the sum of on and off-diagonal terms. We separately calculate the expectation of the on and off diagonal terms conditional upon p, θ, and the customer being alive at time point n. Taking one generic on-diagonal term from Equation 56,

$$
\begin{aligned}
(57) \qquad E\left(\frac{Y_{n+j}^2}{(1+d)^{2j}} \,\middle|\, p,\theta,alive_n,d\right) &= \frac{1}{(1+d)^{2j}} P(Y_{n+j}=1|p,\theta,alive_n) \\
&= \frac{1}{(1+d)^{2j}}(1-\theta)^j p,
\end{aligned}
$$

which implies that

$$
(58) \qquad E\left(\sum_{j=1}^{\infty} \frac{Y_{n+j}^2}{(1+d)^{2j}} \,\middle|\, p,\theta,alive_n,d\right) = p\sum_{j=1}^{\infty}\left(\frac{1-\theta}{(1+d)^2}\right)^j = \frac{p(1-\theta)}{d^2+2d+\theta},
$$

Turning to a generic off-diagonal term from Equation 56 as using the same logic,

$$
(59) \qquad E\left(\frac{Y_{n+j}Y_{n+i}}{(1+d)^{i+j}} \,\middle|\, p,\theta,alive_n,i<j,d\right) = \frac{p^2(1-\theta)^j}{(1+d)^{i+j}},
$$

which implies that

$$
(60) \qquad E\left(\sum_{j=2}^{\infty}\sum_{i=1}^{j-1} \frac{Y_{n+j}Y_{n+i}}{(1+d)^{i+j}} \,\middle|\, p,\theta,alive_n,d\right) = \frac{2p^2(1-\theta)^2}{(d+\theta)(\theta+2d+d^2)},
$$

a double power series. Adding Equation 58 to Equation 60, we obtain DRT^2 for a customer with parameters (p, θ) who is known to be alive at time point n:

$$
(61) \qquad E(DRT^2|p,\theta,alive_n,d) = \frac{(1-\theta)p(2p(1-\theta)+d+\theta)}{(d^2+2d+\theta)(d+\theta)}.
$$

Multiplying Equation 61 by the probability the customer is alive at time point n given parameters (p,θ) and transaction history $\mathbf{D}$ gives us the desired expression for DRT^2 conditional upon the parameters for a customer with transaction history $\mathbf{D}$:

$$
\begin{aligned}
(62) \qquad E(DRT^2|p,\theta,\mathbf{D},d) &= E(DRT^2|p,\theta,alive_n)p(alive_n|p,\theta,\mathbf{D}) \\
(63) \qquad &= \frac{p^{x+1}(1-p)^{n-x}(1-\theta)^{n+1}(2p(1-\theta)+d+\theta)}{(d+\theta)(d^2+2d+\theta)p(x,t_x,n|p,\theta)}
\end{aligned}
$$

Integrating Equation 63 against the joint posterior distribution for (P, Θ),

$$E(DRT^2|\boldsymbol{\psi}, \mathbf{D}, d) = \int E(DRT^2|p, \theta, \mathbf{D}) f(p, \theta|\mathbf{D}, \boldsymbol{\psi}) dp d\theta \tag{64}$$

$$= \int \frac{p^{x+\alpha}(1-p)^{n-x+\beta-1}\theta^{\gamma-1}(1-\theta)^{n+\delta}(2p(1-\theta)+d+\theta)}{(d+\theta)(d^2+2d+\theta)P(\mathbf{D}|\boldsymbol{\psi})B(\alpha,\beta)B(\gamma,\delta)} dp d\theta \tag{65}$$

This separates into two terms. Integrating with respect to p,

$$E(DRT^2|\boldsymbol{\psi}, \mathbf{D}, d) = J_1 + J_2, \tag{66}$$

$$J_1 = \frac{2B(x+\alpha+2, n-x+\beta)}{P(x, t_x|n, \boldsymbol{\psi})B(\alpha,\beta)B(\gamma,\delta)} \int \frac{\theta^{\gamma-1}(1-\theta)^{n+1+\delta}}{(d+\theta)(d^2+2d+\theta)} d\theta, \tag{67}$$

$$J_2 = \frac{B(x+\alpha+1, n-x+\beta)}{P(x, t_x|n, \boldsymbol{\psi})B(\alpha,\beta)B(\gamma,\delta)} \int \frac{\theta^{\gamma-1}(1-\theta)^{n+\delta}}{(d^2+2d+\theta)} d\theta. \tag{68}$$

Letting $s = 1 - \theta$ as in FHS, and recalling the integral representation of the Gaussian hypergeometric function,

$$\int_0^1 t^{b-1}(1-t)^{c-b-1}(1-zt)^{-a} dt = B(b, c-b)\,{}_2F_1(a, b; c; z), \quad c > b \tag{69}$$

gives us 9.

$\underline{V(DRT|\boldsymbol{\psi}, \mathbf{D}, d)}$:

We know that

$$V(DRT|\boldsymbol{\psi}, \mathbf{D}, d) = E(DRT^2|\boldsymbol{\psi}, \mathbf{D}, d) - E(DRT|\boldsymbol{\psi}, \mathbf{D}, d)^2. \tag{70}$$

We computed the first term in Equation 9, while the second term in Equation 70 was derived in FHS.

$\underline{E(\tilde{GP}^2|s, \tilde{p}, \nu, x)}$:

Integrating $\tilde{GP}^2$ against its individual-level density,

$$E(\tilde{GP}^2|s,\tilde{p},\nu,v) = \int_0^\infty \tilde{g}^2 f(\tilde{g}|s,p,\nu,x)d\tilde{g} \tag{71}$$

$$= \frac{(\nu x)^{\tilde{p}x}}{\Gamma(\tilde{p}x)} \int_0^\infty \tilde{g}^{\tilde{p}x+2-1} e^{-\nu\tilde{g}} d\tilde{g} \tag{72}$$

$$= \frac{(\tilde{p}x+1)\tilde{p}}{\nu^2 x}. \tag{73}$$

This implies that at the individual level, the second moment of $\tilde{GP}$ does not exist when $x = 0$.

$\underline{E(\tilde{GP}^2|\boldsymbol{\psi}, g_x, x)}$:

Integrating $E(\tilde{GP}^2|s,\tilde{p},\nu,x)$ (Equation 73) against the density of ν conditional upon $(\boldsymbol{\psi}, g_x, x)$ (FHL, p. 419),

$$E(\tilde{GP}^2|\boldsymbol{\psi}, g_x, x) = \int_0^\infty E(\tilde{GP}^2|s,\tilde{p},\nu,x)\tilde{p}(\nu|\boldsymbol{\psi}, g_x, x)d\nu \tag{74}$$

$$= \int_0^\infty \frac{(\tilde{p}x+1)\tilde{p}}{\nu^2 x} \frac{(\tilde{\gamma}+\tilde{g}_x x)^{\tilde{p}x+q}}{\Gamma(\tilde{p}x+q)} \nu^{\tilde{p}x+q-1} e^{-\nu(\tilde{\gamma}+\tilde{g}_x x)} d\nu \tag{75}$$

$$= \frac{(\tilde{p}x+1)\tilde{p}(\tilde{\gamma}+\tilde{g}_x x)^{\tilde{p}x+q}}{x\Gamma(\tilde{p}x+q)} \int_0^\infty \nu^{\tilde{p}x+q-2-1} e^{-\nu(\tilde{\gamma}+\tilde{g}_x x)} d\nu \tag{76}$$

$$= \frac{(\tilde{p}x+1)\tilde{p}(\tilde{\gamma}+\tilde{g}_x x)^{\tilde{p}x+q}}{x\Gamma(\tilde{p}x+q)} \frac{\Gamma(\tilde{p}x+q-2)}{(\tilde{\gamma}+\tilde{g}_x x)^{\tilde{p}x+q-2}} \tag{77}$$

$$= \frac{(\tilde{p}x+1)\tilde{p}(\tilde{\gamma}+\tilde{g}_x x)^2}{x(\tilde{p}x+q-1)(\tilde{p}x+q-2)}. \tag{78}$$

$\underline{V(GP|\boldsymbol{\psi}, g_x, x)}$:

We know that

$$V(GP|\boldsymbol{\psi}, g_x, x) = V(\tilde{GP}|\boldsymbol{\psi}, g_x, x),$$

because adding a constant to a random variable does not change its variance. Therefore,

$$V(GP|\boldsymbol{\psi}, g_x, x) = E(\tilde{GP}^2|\boldsymbol{\psi}, g_x, x) - (E(\tilde{GP}|\boldsymbol{\psi}, g_x, x))^2.$$

As a result, because we already have expressions for $E(\tilde{GP}^2|\boldsymbol{\psi}, g_x, x)$ (Equation 78) and $E(\tilde{GP}|\boldsymbol{\psi}, g_x, x)$

(Equation 4 in FHL), we have the desired expression for $V(\tilde{GP}^2|\boldsymbol{\psi}, g, x)$, and 19 follows.

<u>$V(V(n+n^*)|\boldsymbol{\psi}, \mathbf{D})$ and $V(V(n, n+n^*)|\boldsymbol{\psi}, \mathbf{D})$</u>:

We know that

$$V\left(V(n+n^*)|\boldsymbol{\psi}, \mathbf{D}\right) = V(V(n, n+n^*)|\boldsymbol{\psi}, \mathbf{D}) \tag{79}$$

$$= E(V(n, n+n^*)^2|\boldsymbol{\psi}, \mathbf{D}) - (E(V(n, n+n^*)|\boldsymbol{\psi}, \mathbf{D}))^2 \tag{80}$$

We derive the latter term first before turning to the former. Decomposing $V(n, n+n^*)$ into a stream of transaction and profit variables,

$$E(V(n, n+n^*)|\boldsymbol{\psi}, \mathbf{D}) = E(\sum_{j=1}^{n^*} Y_{n+j} W_{n+j}|\boldsymbol{\psi}, \mathbf{D}) \tag{81}$$

$$= \sum_{j=1}^{n^*} E(Y_{n+j}|\boldsymbol{\psi}, \mathbf{D}) E(W_{n+j}|\boldsymbol{\psi}, \mathbf{D}) \tag{82}$$

$$= E(X(n, n+n^*)|\boldsymbol{\psi}, \mathbf{D}) E(GP|\boldsymbol{\psi}, \mathbf{D}). \tag{83}$$

We have closed form expressions for both terms (Equation 13 from FHS above and Equation 18).

Turning to the former term,

$$E(V(n, n+n^*)^2|\boldsymbol{\psi}, \mathbf{D}) = E(\sum_{j=1}^{n^*} \left(Y_{n+j} W_{n+j}\right)^2 |\boldsymbol{\psi}, \mathbf{D}) \tag{84}$$

$$= E(\sum_{j=1}^{n^*} Y_{n+j} W_{n+j}^2 + 2 \sum_{j=2}^{n^*} \sum_{i=1}^{j-1} Y_{n+i} Y_{n+j} W_{n+i} W_{n+j}|\boldsymbol{\psi}, \mathbf{D}) \tag{85}$$

$$\begin{aligned} &= E(X(n, n+n^*)|\boldsymbol{\psi}, \mathbf{D}) E(GP^2|\boldsymbol{\psi}, \mathbf{D}) \\ &\quad + E(GP|\boldsymbol{\psi}, \mathbf{D})^2 E(2 \sum_{j=2}^{n^*} \sum_{i=1}^{j-1} Y_{n+i} Y_{n+j}|\boldsymbol{\psi}, \mathbf{D}) \end{aligned} \tag{86}$$

Recognizing that

$$\left(\sum_{j=1}^{n^*} Y_{n+j} \right)^2 = \sum_{j=1}^{n^*} Y_{n+j} + 2 \sum_{j=2}^{n^*} \sum_{i=1}^{j-1} Y_{n+i} Y_{n+j},$$

it follows that $E(2\sum_{j=2}^{n^*}\sum_{i=1}^{j-1} Y_{n+i}Y_{n+j}|\boldsymbol{\psi},\mathbf{D})$ is equal to

$$E(2\sum_{j=2}^{n^*}\sum_{i=1}^{j-1} Y_{n+i}Y_{n+j}|\boldsymbol{\psi},\mathbf{D}) \;=\; E(X(n,n+n^*)^2|\boldsymbol{\psi},\mathbf{D}) - E(X(n,n+n^*)|\boldsymbol{\psi},\mathbf{D}), \tag{87}$$

giving us the desired result in Equation 25.

<u>$V(CLV|\boldsymbol{\psi},\mathbf{D})$ and $V(RLV|\boldsymbol{\psi},\mathbf{D})$</u>:

$$E(RLV^2|\boldsymbol{\psi},\mathbf{D},d) \;=\; E(\sum_{j=1}^{\infty}\frac{Y_{n+j}W_{n+j}^2}{(1+d)^{2j}} + 2\sum_{j=2}^{\infty}\sum_{i=1}^{j-1}\frac{Y_{n+i}Y_{n+j}W_{n+i}W_{n+j}}{(1+d)^{i+j}}|\boldsymbol{\psi},\mathbf{D},d) \tag{88}$$

$$= K_1 + K_2, \quad \text{where} \tag{89}$$

$$K_1 \;=\; E(GP^2|\boldsymbol{\psi},\mathbf{D})E(\sum_{j=1}^{\infty}\frac{Y_{n+j}}{(1+d)^{2j}}|\boldsymbol{\psi},\mathbf{D},d) \quad \text{and} \tag{90}$$

$$K_2 \;=\; E(GP|\boldsymbol{\psi},\mathbf{D})^2E(2\sum_{j=2}^{\infty}\sum_{i=1}^{j-1}\frac{Y_{n+i}Y_{n+j}}{(1+d)^{i+j}}|\boldsymbol{\psi},\mathbf{D},d) \tag{91}$$

Using the fact that

$$(1+d)^{2j} = (1+2d+d^2)^j, \quad \text{and} \quad 2\sum_{j=2}^{\infty}\sum_{i=1}^{j-1}\frac{Y_{n+i}Y_{n+j}}{(1+d)^{i+j}} = DRT^2 - \sum_{j=1}^{\infty}\frac{Y_{n+j}}{(1+d)^{2j}}, \tag{92}$$

Then letting

$$\tilde{d} \equiv 2d + d^2, \tag{93}$$

Equation 23 follows.

References

Ascarza, Eva, Bruce GS Hardie. 2013. A joint model of usage and churn in contractual settings. *Marketing Science* **32**(4) 570–590.

Auer, Peter, Nicolo Cesa-Bianchi, Paul Fischer. 2002. Finite-time analysis of the multiarmed bandit problem. *Machine Learning* **47**(2) 235–256.

Beukers, Frits. 2007. Gauss hypergeometric function. *Arithmetic and geometry around hypergeometric functions*, vol. 260. Springer, 23–42.

Bolton, Ruth N, Katherine N Lemon, Peter C Verhoef. 2004. The theoretical underpinnings of customer asset management: A framework and propositions for future research. *Journal of the Academy of Marketing Science* **32**(3) 271–292.

Braun, Michael, David A Schweidel, Eli Stein. 2015. Transaction attributes and customer valuation. *Journal of Marketing Research* **52**(6) 848–864.

Buhl, Hans Ulrich, Bernd Heinrich. 2008. Valuing customer portfolios under risk-return-aspects: a model-based approach and its application in the financial services industry. *Academy of Marketing Science Review* **12**(5) 1–32.

Colombo, Richard, Weina Jiang. 1999. A stochastic RFM model. *Journal of Interactive Marketing* **13**(3) 2–12.

Dhar, Ravi, Rashi Glazer. 2003. Hedging customers. *Harvard business review* **81**(5) 86–92.

Dziurzynski, Lukasz, Edward Wadsworth, Peter Fader, Elea McDonnell Feit, Daniel McCarthy, Bruce Hardie, Arun Gopalakrishnan, Eric Schwartz, Yao Zhang. 2014. Package 'BTYD'. Available at `https://cran.r-project.org/web/packages/BTYD/index.html`. Online; accessed February 27, 2016.

Ehrenberg, Andrew SC, Gerald J Goodhardt, T Patrick Barwise. 1990. Double jeopardy revisited. *The Journal of Marketing* **54**(3) 82–91.

Fader, Peter S, Bruce GS Hardie, Paul D Beger. 2011. Implementing the BG/BB model for customer-base analysis in excel. Available at `http://www.brucehardie.com/notes/010/`. Online; accessed February 27, 2016.

Fader, Peter S, Bruce GS Hardie, Ka Lok Lee. 2005a. Counting your customers the easy way: an alternative to the Pareto/NBD model. *Marketing Science* **24**(2) 275–284.

Fader, Peter S, Bruce GS Hardie, Ka Lok Lee. 2005b. RFM and CLV: Using iso-value curves for customer base analysis. *Journal of Marketing Research* **42**(4) 415–430.

Fader, Peter S, Bruce GS Hardie, Jen Shang. 2010. Customer-base analysis in a discrete-time noncontractual setting. *Marketing Science* **29**(6) 1086–1108.

Gupta, Sunil, Dominique Hanssens, Bruce Hardie, Wiliam Kahn, V Kumar, Nathaniel Lin, Nalini Ravishanker, S Sriram. 2006. Modeling customer lifetime value. *Journal of Service Research* **9**(2) 139–155.

Gupta, Sunil, Donald Lehmann. 2005. *Managing customers as investments: the strategic value of customers in the long run*. Wharton School Publishing, Upper Saddle River, NJ.

Hartigan, John A, Manchek A Wong. 1979. Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **28**(1) 100–108.

Hinshaw, Michael. 2013. 5 segmentation lessons from CVS. *CMO* Online; accessed February 27, 2016.

Jerath, Kinshuk, Peter S Fader, Bruce GS Hardie. 2011. New perspectives on customer "death" using a generalization of the Pareto/NBD model. *Marketing Science* **30**(5) 866–880.

Kaufmann, Emilie, Olivier Cappé, Aurélien Garivier. 2012. On bayesian upper confidence bounds for bandit problems. *International Conference on Artificial Intelligence and Statistics*. 592–600.

Kumar, V, Denish Shah. 2009. Expanding the role of marketing: from customer equity to market capitalization. *Journal of Marketing* **73**(6) 119–136.

Kumar, V, Rajkumar Venkatesan, Tim Bohling, Denise Beckmann. 2008. The power of CLV: Managing customer lifetime value at IBM. *Marketing Science* **27**(4) 585–599.

McCarthy, Daniel, Peter Fader, Bruce Hardie. 2015. Valuing subscription-based businesses using publicly disclosed customer data. *Available at SSRN 2701093* .

McCarty, John A, Manoj Hastak. 2007. Segmentation approaches in data-mining: A comparison of rfm, chaid, and logistic regression. *Journal of business research* **60**(6) 656–662.

Reinartz, Werner J, Vijay Kumar. 2000. On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing. *Journal of marketing* **64**(4) 17–35.

Reinartz, Werner J, Vita Kumar. 2003. The impact of customer relationship characteristics on profitable lifetime duration. *Journal of marketing* **67**(1) 77–99.

Rust, Roland T, V Kumar, Rajkumar Venkatesan. 2011. Will the frog change into a prince? Predicting future customer profitability. *International Journal of Research in Marketing* **28**(4) 281–294.

Rust, Roland T, Katherine N Lemon, Valarie A Zeithaml. 2004. Return on marketing: using customer equity to focus marketing strategy. *Journal of marketing* **68**(1) 109–127.

Sant'Anna, Annibal Parracho, Rodrigo Otavio de Araujo Ribeiro. 2008. Application of stochastic models to determine customers lifetime value for a brazilian supermarkets network. *Pesquisa Operacional* **28**(3) 563–576.

Schmittlein, David C, Donald G Morrison, Richard Colombo. 1987. Counting your customers: Who are they and what will they do next? *Management Science* **33**(1) 1–24.

Schmittlein, David C, Robert A Peterson. 1994. Customer base analysis: an industrial purchase process application. *Marketing Science* **13**(1) 41–67.

Schwartz, Eric M, Eric Bradlow, Peter Fader. 2015. Customer acquisition via display advertising using multi-armed bandit experiments Available at SSRN: http://ssrn.com/abstract=2368523.

Schwartz, Eric M, Eric T Bradlow, Peter S Fader. 2014. Model selection using database characteristics: Developing a classification tree for longitudinal incidence data. *Marketing Science* **33**(2) 188–205.

Schweidel, David A, George Knox. 2013. Incorporating direct marketing activity into latent attrition models. *Marketing Science* **32**(3) 471–487.

Sharpe, William F. 1998. The sharpe ratio. *Journal of Portfolio Management* **21**(1) 49–58.

Srivastava, Rajendra K, Tasadduq A Shervani, Liam Fahey. 1998. Market-based assets and shareholder value: a framework for analysis. *The Journal of Marketing* **62**(1) 2–18.

Tarasi, Crina O, Ruth N Bolton, Michael D Hutt, Beth A Walker. 2011. Balancing risk and return in a customer portfolio. *Journal of Marketing* **75**(3) 1–17.

Yoo, Shijin, Dominique M Hanssens, Ho Kim. 2012. Marketing and the evolution of customer equity of frequently purchased brands. *Working Paper* .

Zeithaml, Valarie A, Katherine N Lemon, Roland T Rust. 2001a. *Driving customer equity: How customer lifetime value is reshaping corporate strategy*. Simon and Schuster, New York, NY.

Zeithaml, Valarie A, Roland T Rust, Katherine N Lemon. 2001b. The customer pyramid: creating and serving profitable customers. *California Management Review* **43**(4) 118–142.

Zhang, Harvey Yang. 2008. Modeling discrete-time transactions using the BG/BB model. *Wharton Research Scholars Journal* (53).

Zhang, Yao, Eric T Bradlow, Dylan S Small. 2014. Predicting customer value using clumpiness: from RFM to RFMC. *Marketing Science* **34**(2) 195–208.

Notes

[1]We plot time since final purchase ($n - t_x$) instead of recency (t_x) on the x-axis of Figure 1 because our cohort consists of customers acquired over a 13 week period. This creates variability in n across customers, so the very best customers by t_x (but not by $n - t_x$) vary depending upon the customer acquisition date.

[2]Had we focused upon variance instead of CV, the right-most graph in Figure 1 would be confounded by the strong positive correlation between mean and variance. We will switch back and forth between V and CV as appropriate for the analysis being performed.

[3]In very large datasets, very slight windsorization is needed to blunt the effect of data errors or unauthorized distributors/resellers who should not be analyzed alongside regular customers. A very small number of customers (.1% of customers representing only .05% of purchases) were windsorized.

[4]To more easily visualize variation in $CV(DRT)$ for the majority of customers, we windsorize $CV(DRT)$ at 3 within this graph because $CV(DRT)$ grows arbitrarily large when $x = t_x$, skewing the coloration of the heatmap.

[5]To ensure that all variables are scaled consistently, we standardize E(RLV), $SD(RLV)$, x, t_x, and g_x by subtracting their sample means and dividing by their sample standard deviation before running the clustering algorithms.