2017

Customer-Based Corporate Valuation: Modeling With Missing, Aggregated Data Summaries

Daniel Minh Mccarthy

University of Pennsylvania, danielmc999@gmail.com

Customer-Based Corporate Valuation: Modeling With Missing, Aggregated Data Summaries

Abstract

There is growing interest in "customer-based corporate valuation," explicitly tying the value of a firm's customer base to its financial valuation. This dissertation studies the theory and application of customer-based corporate valuation. The dissertation is comprised of three essays, each of which studies a different aspect of the topic. In the first essay, we develop a general customer-based corporate valuation framework. In doing so, we enumerate the determinants of corporate value and how predictions of customer base activity can be used to inform these determinants. In the second essay, we develop a customer-based corporate valuation model that is specifically suited to contractual (or subscription-based) businesses. We apply this model to publicly-disclosed data from two companies, DISH Network and Sirius XM Holdings. In the third essay, we develop a customer-based corporate valuation model for non-contractual (or non-subscription-based) businesses. This is a more challenging problem, because non-contractual businesses have more complex transactional patterns -- they are characterized by latent attrition instead of observable churn behavior, and often have irregular purchase timing and spend amounts. We apply this methodology to data from a large business unit of an e-commerce retailer, valuing the business unit as a whole, decomposing this valuation into existing and yet-to-be-acquired customers, and analyzing customer profitability. In both essays two and three, we assume that the modeler is an external stakeholder, and thus only has the ability to observe a very limited, possibly incomplete, periodically disclosed collection of customer data summaries, unlike a situation in which the granular data is observed. We conclude with a short chapter which describes areas for future research.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Statistics

First Advisor
Eric T. Bradlow

Second Advisor
Shane T. Jensen

Keywords
customer equity, customer lifetime value, indirect inference, marketing metrics, valuation

CUSTOMER-BASED CORPORATE VALUATION:
MODELING WITH MISSING, AGGREGATED DATA SUMMARIES

Daniel M. McCarthy

A DISSERTATION

in

Statistics

For the Graduate Group in
Managerial Science and Applied Economics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2017

Supervisor of Dissertation

Co-Supervisor of Dissertation

Eric T. Bradlow
K.P. Chao Professor; Professor of Marketing,
Statistics, Education and Economics

Shane T. Jensen
Associate Professor of Statistics

Graduate Group Chairperson

Catherine Schrand, Celia Z. Moh Professor; Professor of Accounting

Dissertation Committee:

Lawrence Brown, Miers Busch Professor; Professor of Statistics
Bruce Hardie, Professor of Marketing, London Business School

CUSTOMER-BASED CORPORATE VALUATION:
MODELING WITH MISSING, AGGREGATED DATA SUMMARIES

Daniel M. McCarthy

Acknowledgments

I cannot be more grateful to my advisors, Eric Bradlow and Shane Jensen. I am also thankful to have Lawrence Brown and Bruce Hardie on my dissertation committee. My committee's ability to balance support and encouragement with constructive feedback greatly improved the quality of my dissertation.

My greatest thanks go to Peter Fader, to my wife Elica Dhundia McCarthy, and to my family. The fact that I have learned more academically from Peter than anyone else at Wharton goes without saying. At least as important is what I have learned non-academically. When I first began working with Peter, I wondered why his students were called "the family." As I complete this dissertation, it is now obvious to me – I truly feel like a family member. I can only hope to come anywhere close to being as good of a mentor to others as he has been to me. Finally, and most importantly, the constant love, support and encouragement of my wife Elica, my immediate (non-academic) family James, Madeleine, Marc, and Jenn, and my extended family Suresh, Ravi, Aprajita, and Chandni has been invaluable. I am blessed to have them in my life.

ABSTRACT

CUSTOMER-BASED CORPORATE VALUATION: MODELING WITH MISSING, AGGREGATED DATA SUMMARIES

Daniel M. McCarthy

Eric T. Bradlow

Shane T. Jensen

There is growing interest in "customer-based corporate valuation," explicitly tying the value of a firm's customer base to its financial valuation. This dissertation studies the theory and application of customer-based corporate valuation. The dissertation is comprised of three essays, each of which studies a different aspect of the topic. In the first essay, we develop a general customer-based corporate valuation framework. In doing so, we enumerate the determinants of corporate value and how predictions of customer base activity can be used to inform these determinants. In the second essay, we develop a customer-based corporate valuation model that is specifically suited to contractual (or subscription-based) businesses. We apply this model to publicly-disclosed data from two companies, DISH Network and Sirius XM Holdings. In the third essay, we develop a customer-based corporate valuation model for non-contractual (or non-subscription-based) businesses. This is a more challenging problem, because non-contractual businesses have more complex transactional patterns – they are characterized by latent attrition instead of observable churn behavior, and often have irregular purchase timing and spend amounts. We apply this methodology to data from a large business unit of an e-commerce retailer, valuing the business unit as a whole, decomposing this valuation into existing and yet-to-be-acquired customers, and analyzing customer profitability. In both essays two and three, we assume that the modeler is an external stakeholder, and thus only has the ability to observe a very limited, possibly incomplete, periodically disclosed collection of customer data summaries, unlike a situation in which the granular data is observed. We conclude with a short chapter which describes areas for future research.

Contents

List of Tables

List of Figures

1

Introduction

On February 22nd, 2017, DISH Network (DISH), a large satellite pay-TV provider, releases its 2016 annual report. DISH reports that it has 13.7MM active customers at the end of the year, down from 13.9MM the year before. While DISH acquired 2.6MM new customers during the year, they lost another 2.8MM customers, as the churn rate increased to 1.83% this year from 1.71% in the prior year. DISH's average revenue per user (ARPU) increased by 2.2% to $88.7, and its cost to acquire new users fell by 11.1% to $643 per gross customer acquisition.

Some may point to the declining size of the customer base and the rising churn rate and conclude that the company is unable to effectively retain its existing customer base in the face of internet streaming. These individuals may believe that the underlying financial condition of the firm has deteriorated and will continue to deteriorate into the future, so the valuation of the firm should be low. Others may point to the rising ARPU and the falling cost of new customer acquisition and conclude that while DISH is losing customers, it has nevertheless been able to hold on to its higher-value customers and maintain the profitability of new customers by sharply reducing the cost to acquire them. These individuals may believe that the financial condition of the firm has remained stable, or maybe even improved, during the period.

While both constituencies would be making their respective investment decisions

by linking customer metrics to the value of the firm, their mental models would not be grounded in the principles of corporate valuation theory. Our goal with this dissertation is to make this connection correctly. This process is called "customer-based corporate valuation" (CBCV), which we formally define to be the act of valuing the firm by forecasting current and future customer behavior using customer data in conjunction with traditional financial data. We posit a customer-driven model for the overall valuation of the firm – how customers will be acquired over time, how long those customers will remain with the firm, how much those customers will spend while they are alive, and how the resulting overall future revenue estimates can be inserted into a financial valuation model. After positing such a model, we use the observed data to estimate the model's underlying parameters and in turn the valuation of the firm.

The understandings which arise from the model have important implications for marketing departments, who are responsible for managing the value of customer assets over time. While marketing departments devote much of their attention day-to-day to more tactical initiatives (e.g., who to target in an upcoming Fall season catalog mailing to optimize expected lift, or how to allocate marketing budget across channels), their ultimate objective should be to increase the value of the firm by (1) improving the acquisition of new customers and the lifetime value (CLV) of new and existing customers, and (2) decreasing the riskiness of the customer portfolio. Marketing departments must not lose sight of these higher-level overarching objectives.

CBCV can be used to decompose firm value into the value arising from already-acquired and yet-to-be-acquired customers, estimate the proportion of newly-acquired customers who will be unprofitable, and more. In doing so, CBCV summarizes the cumulative effect of the marketing department's activities each period on measures of the health of the overall customer base, and in turn, on shareholder value (SHV)

2

and its riskiness. CBCV is an important bridge between the marketing department and the C-suite, and an important source of marketing accountability.

CBCV is also a promising use case for common customer metrics (Farris et al. 2010). These metrics need not be ends in their own right (i.e., as standalone firm-wide key performance indicators) – they can be leveraged to better understand the true underlying propensity of customers to acquire services, make purchases, and spend, and how these propensities vary across customers. The proposed methodology turns backward-looking customer metrics into important forward-looking measures, which decrease investor uncertainty regarding future cash flows and thus increase the valuation of the firm (Bayer et al. 2017).

Several researchers in the fields of marketing and accounting have explored this area. However, progress to date has been impeded by three main challenges: (1) the underlying models of customer acquisition, retention, and spend used do not reflect important empirical realities associated with these behaviors (e.g., that churn, purchase, and spend propensities may differ across customers and over time), (2) the customer data which best identifies these empirical realities are not clear, and (3) the estimation methodology which links the aggregated customer data to the parameters underlying the models can be complex. Because of these challenges, current CBCV models do not meet the standards of finance professionals. The goal of this dissertation is to develop the necessary framework and methodology to address these challenges.

The rest of the dissertation is organized as follows. In chapter two, we create a general customer-based corporate valuation framework which can be applied in any business setting, then discuss model building considerations as this framework is applied to specific problem settings. In chapter three, we develop a customer-based corporate valuation model that is specifically suited to contractual (or subscription-

based) businesses, and apply the model to publicly-disclosed data from two companies, DISH Network and Sirius XM Holdings. In chapter four, we develop a customer-based corporate valuation model for non-contractual (or non-subscription-based) businesses, then apply this methodology to data from a large business unit of an e-commerce retailer. We value the business unit as a whole, decompose this valuation into existing and yet-to-be-acquired customers, and analyze customer profitability. Finally, in chapter five, we conclude with areas for future research.

2

Customer-based Corporate Valuation: A Framework and Considerations for Model Building

2.1 Introduction

As we had noted in the previous chapter, customer-based corporate valuation (CBCV) is the process of valuing the firm by forecasting current and future customer behavior using customer data in conjunction with traditional financial data. A proper framework for CBCV, then, consists of two parts: (1) a general corporate valuation framework, and (2) a process through which customer behaviors are incorporated into the general corporate valuation framework. We discuss each of these parts next.

Corporate valuation has been extensively studied and implemented by academics and practitioners within finance and accounting, making their techniques central to our work. Furthermore, for CBCV to be more widely adopted by firms, it is important that the proposed methodology be consistent with the corporate valuation methodologies that finance professionals are currently using. Therefore, we borrow from the finance and accounting literatures, making their standard financial valuation models the cornerstone of our overall valuation framework.

Our proposed approach to insert customer behaviors into the overall valuation

equation rests upon the following simple fact: every dollar of revenue that a company generates must come from its customers. If we have perfect foresight into the how much customers will spend, this must dictate what the firm's overall revenues will be. Therefore, we posit a model for how customers are acquired over time, how long they will remain with the firm, and how much they will spend while they are alive. We use this model to decompose total revenues into total purchases and the amount spent upon each of those purchases, and further decompose total purchases into total "alive" customers (i.e., customers who have not churned yet) and the number of purchases made by those alive customers. This approach allows us to incorporate a customer-driven revenue model directly "on top of" standard corporate valuation models, which traditionally begin with future revenue projections. This is the framework through which customer data can provide additional structure and dimension to future revenue forecasts. To the extent that we can predict future revenues more accurately by leveraging this decomposition than we would have if we could only observe raw sales data alone, which is the standard operating procedure for non-CBCV models, the resulting overall valuation forecasts must be more accurate as well.

In the next section, we discuss the most popular valuation models within finance and accounting. We specify in detail how customer modeling is used to augment many of these most popular valuation models across a number of different problem settings. We discuss the limitations of the available data, a key aspect of our problem setting, which creates unique statistical challenges. Thereafter, we discuss context-specific modeling issues which will commonly arise for researchers performing CBCV in other settings. We then survey the literature related to CBCV and to estimating individual-level models with aggregated data before providing concluding remarks.

2.2 Traditional Financial Valuation

According to standard corporate valuation theory (Damodaran 2012; Greenwald et al. 2004; Holthausen & Zmijewski 2013; Koller et al. 2010), the value of a firm equals the value of the operating assets (OA) plus the non-operating assets (NOA), minus the net debt (ND) of the firm. Denoting the shareholder value of the firm at time T by SHV(T), we have

$$\text{SHV(T)} = \text{OA(T)} + \text{NOA(T)} - \text{ND(T)} \,. \tag{2.1}$$

For publicly-traded firms, SHV is observed in effectively continuous time. OA, NOA, and ND are observed at the end of each quarter by external stakeholders through mandatory quarterly filing disclosures.

Two of the three determinants of SHV are relatively straightforward to value:

- ND is equal to the market value of the sum of all outstanding debt obligations of the firm (e.g., short term debt, long term debt, and off-balance sheet liabilities such as operating leases, synthetic leases, and the liabilities associated with special purpose entities), less cash and cash equivalents.

- NOA is equal to the market value of all assets which are unrelated to the running of the day-to-day operations of the business (e.g., excess real estate and other illiquid non-cash assets). The existence of these assets should have no effect upon the firm's ability to generate revenues and operating profits.

Typically, the most important yet challenging determinant of SHV is OA. OA is the market value of the core revenue-generating operations of the firm. The market valuation of OA is driven by many factors, including but not limited to competition, brand value, distribution channel relationships/management, pricing, intellectual property

protections, new product innovation, customer service, human resources, and more (Farris et al. 2010; Kamakura & Russell 1993; Srinivasan & Hanssens 2009).

There are many financial valuation models that are traditionally used to value OA. Damodaran (2012, p. 925) argues that there are too many: "[t]he problem in valuation is not that there are not enough models to value an asset, it is that there are too many. Choosing the right model to use in valuation is as critical to arriving at a reasonable value as understanding how to use the model." We use the decision framework developed in Damodaran (2012, Chapter 34) to select the valuation methodology most appropriate to the customer-based valuations we perform in essays two and three. In other problem settings, a different valuation methodology may be more appropriate, so we recommend that researchers refer to the decision framework in Damodaran (2012) before beginning the valuation process. That being said, while there are many different valuation methodologies, the two most popular methodologies for going concern businesses are (1) discounted cash flow (DCF) modeling, and (2) relative valuation (Damodaran et al. 2005). We elaborate upon these valuation methods next, before showing how customer modeling can be used to improve them.

2.2.1 Discounted Cash Flow Valuation

DCF valuation is the de-facto industry standard way in which operating assets are valued within the financial industry. Central to DCF valuation is the fact that OA is equal to the sum of all future free cash flows (FCFs) the firm will generate,[1] discounted at the weighted average cost of capital (WACC):

$$OA(T) = \sum_{t=0}^{\infty} \frac{FCF(T+t)}{(1 + WACC)^t}.$$ (2.2)

---

[1]Strictly speaking, we are referring to *expected* free cash flows.

8

The unit of time chosen (e.g., day, week, month, or quarter) may differ depending upon the problem setting. FCF is equal to the net operating profit after taxes (NOPAT) minus the difference between capital expenditures (CAPEX) and depreciation and amortization (D&A), minus the change in non-financial working capital (ΔNFWC):

$$\text{FCF}(t) = \text{NOPAT}(t) - (\text{CAPEX}(t) - \text{D\&A}(t)) - \Delta\text{NFWC}(t). \qquad (2.3)$$

The most important ingredient of FCF is NOPAT, which is a measure of the underlying profitability of the operating assets of the firm. NOPAT is equal to revenues (REV) multiplied by the contribution margin ratio $(1-\text{VC})$ minus total customer acquisition expenses (cost per acquired customer multiplied by the number of acquired customers, or CAC$\times$A) and fixed operating costs (FC), after taxes (where TR is the corporate tax rate for the firm):

$$\text{NOPAT}(t) = \{\text{R}(t) \times [1 - \text{VC}(t)] - \text{FC}(t) - \text{CAC}(t) \times \text{A}(t)\} \times [1 - \text{TR}]. \qquad (2.4)$$

The other elements of Equation 2.4 make adjustments for balance sheet-related cash flow effects, and are generally of secondary importance to the value of the firm. At the heart of the DCF valuation exercise is the estimation of period-by-period FCF, central to which are estimates of period-by-period revenue (Equations 2.3 and 2.4). The task of generating accurate revenue projections has received surprisingly little attention in the finance community (Damodaran et al. 2005). We will revisit these equations in the valuation analyses of essays two and three.

2.2.2 Relative Valuation

Relative valuation entails valuing an asset using the price of other "comparable" assets (Alford 1992). Comparable assets are traditionally firms operating within the same industry, or firms with similar margins, growth profiles, and market structures as the focal firm being valued. Firms are usually compared using a common variable, such as revenues, cash flows (e.g., EBITDA, or FCF), profits (e.g., operating profit or net profit), or assets (e.g., book value or another measure of asset value).

Relative valuation methods study the ratio of the value of the firm (e.g., enterprise value or market capitalization) to the common variable. Common valuation ratios (also known as multiples) include price-to-earnings (Campbell & Shiller 1988b), PE-to-growth (Damodaran 2010), price-to-dividends (Campbell & Shiller 1988a), price-to-book value (Fama & French 1993), enterprise value to sales, and enterprise value to EBITDA. We use this ratio across comparable firms in conjunction with the common variable of the focal firm being valued to obtain the focal firm's fair valuation[2].

Suppose we are valuing a beverage company, for example, which generated $10MM in earnings over the previous 12 months. We may perform a relative valuation analysis by examining the P/E ratios of other beverage companies. Table 2.1 contains valuation ratios for the beverage sector from Damodaran (2010). The average P/E ratio across these 16 beverage companies is 22.66. If it is reasonable to assume that the firm being valued is "similar" to the average of its peers, we may multiply this average P/E ratio by the firm's trailing net profits of $10MM to obtain a fair valuation for the focal firm of $226.6MM.

Relative valuation is also very commonly used to perform scenario analysis for firms that are undergoing change. For example, we may be interested in valuing a young, fast-growing apparel retailer, or a large e-commerce company in the middle

---

[2]The distribution of ratios across comparable firms can also be used to perform sensitivity analysis.

Table 2.1: Price-to-Earnings and PE-Growth Ratios: The Beverage Sector

Company Name	P/E Ratio	PEG Ratio
Coca-Cola Bottling	29.18	3.07
Molson Inc. Ltd. 'A'	43.65	2.82
Anheuser-Busch	24.31	2.21
Corby Distilleries Ltd.	16.24	2.16
Chalone Wine Group Ltd.	21.76	1.55
Andres Wines Ltd. 'A'	8.96	2.56
Todhunter Int'l	8.94	2.98
Brown-Forman 'B'	10.07	0.88
Coors (Adolph) 'B'	23.02	25.30
PepsiCo, Inc.	33.00	3.14
Coca-Cola	44.33	2.33
Boston Beer 'A'	10.59	0.62
Whitman Corp.	25.19	2.19
Mondavi (Robert) 'A'	16.47	1.18
Coca-Cola Enterprises	37.14	1.38
Hansen Natural Corp	9.70	0.57

of an operational restructuring. Common financial variables such as trailing sales and/or earnings may be be artificially depressed, so traditional application of a multiples analysis may depress fair valuation estimates. Instead, we may project what a key future financial variable will be (e.g., we may estimate what sales at the apparel retailer will be in five years, after the business has matured and has more fully penetrated its target market), then multiply that future expected financial variable by an appropriate multiple to arrive at a fair valuation estimate for that scenario. We then weigh all possible scenarios by their respective probabilities of occurring, sum them up, and discount the resulting expected future value back to account for the time value of money and the non-diversifiable riskiness of the business to arrive at fair valuation in current dollar terms.

Both DCF and relative valuation methodologies rely upon revenue and thus free

cash flow projections. However, DCF valuation provides modelers with all of the above multiples "for free," in the sense that we may simply take the implied SHV or enterprise value estimate from the DCF valuation and divide it by trailing revenues, earnings, dividends, book value, or EBITDA. The same is not true in reverse – we cannot obtain all future revenue and free cash flow projections from a multiples valuation. While it is informative to take into account information regarding comparable firms, assessing the exact comparability of the focal firm being valued to its peers requires a comparison of all firms' current and future revenue and margin growth potential, which can be both subjective and distracting. For these reasons, we will perform our valuations using the DCF valuation methodology in the sections and chapters that follow, and discuss a proper multiple company valuation analysis as future work in chapter five.

2.3   The Customer Base Model

Generating accurate revenue projections is at the heart of both of the most popular going concern valuation methodologies. Financial professionals typically model and forecast revenues and expenses using time-series models. This may be sensible when firm financial disclosures do not include customer data, where customer data are broadly defined to be measures that are derived from a customer database that are not reported in a typical profit/loss statement. If customer data are available, however, forecasting accuracy can be improved by decomposing customer-driven financial line items into their constituent parts, estimating models and then forecasting these parts into the future, then aggregating these parts together again to form projections of future customer-driven financial line items [3]. In the next section, we propose a data

---

[3]There is an obvious parallel with the new product sales forecasting literature (e.g., Fourt & Woodlock 1960; Eskin 1973; Fader et al. 2004), where we first decompose total sales into its trial

structure for customer base prediction which allows us to perform this decomposition.

2.3.1 A Disaggregated Data Structure for Customer Base Prediction

The behavior of a customer can be summarized by the following five processes:

(a) Prospect[4] acquisition (i.e., when an individual becomes a prospect)

(b) Customer acquisition (i.e., when a prospect becomes a customer)

(c) Life (i.e., whether the customer has not churned yet after being acquired)

(d) Purchase (i.e., whether purchases were made during the period)

(e) Spend

Prospect birth, customer acquisition, and lifetime duration are timing processes, purchase is a choice process, while spend determines magnitude (i.e., at the individual-level, it is distributed gamma, lognormal, or another analogous distribution). Hereafter, we refer to this collection of five processes as the "customer base model." This general framework provides modelers considerable leeway in specifying the processes governing these behaviors depending upon the company being analyzed and the nature of the data which is available to train the model upon[5]. We provide a simple visual schema of the five processes over the lifetime of a customer in Figure 2.1.

---

and repeat components, ii) develop separate models for trial and repeat sales, and ii) then combine the associated forecasts of these components to arrive at a forecast of total sales.

[4]Conceptually, a prospect is a potential customer qualified on the basis of his/her buying authority, financial capacity, and willingness to buy. Given the particular problem setting the modeler is facing, it may be appropriate to broaden or narrow this definition.

[5]In most cases, the modeler will not know exactly when individuals become prospects. Instead, the modeler will typically use a deterministic decision rule (e.g., all households become prospects at the time those households are formed).

Figure 2.1: Visual Schema of Customer Base Model
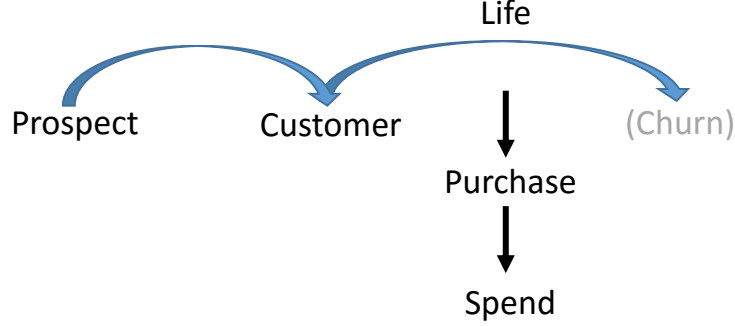
The customer base model summarizes customer behavior at both contractual (or subscription-based) and non-contractual (or non-subscription-based) firms. Contractual customer behavior is a special case of non-contractual customer behavior. A contractual customer must purchase at least once on a periodic basis (e.g., monthly) if and only if the customer is "alive." The observability of the retention process at contractual firms is a byproduct of this fact regarding the purchase process.

The disaggregated data associated with an individual is encoded through a $[5 \times \mathrm{T}]$ matrix, in which each row corresponds to the above processes (prospect, acquisition, life, purchase, and spend) and each column corresponds to a particular time point. We assume time is recorded discretely. This is a reasonable assumption, because all firms' transaction logs record time discretely (typically daily)[6]. The first four rows of the matrix are binary, equal to one if the event of interest occurred at a particular time point, and zero otherwise. The final row is a numeric vector corresponding to the total amount spent at each time point.

For illustrative purposes, Figure 2.2 depicts the behavior of an individual at a non-contractual firm over $\mathrm{T} = 10$ time periods after the beginning of commercial operations. This individual became a prospect at $t = 2$. The individual was acquired

---

[6]The data structure associated with a continuous time customer base model is a ragged array with five elements.

as a customer four periods later, at $t = 6$. A customer is acquired when he/she makes his/her first purchase. Therefore, the customer also made his/her first purchase at $t = 6$, spending \$29. The customer was alive from $t = 6$ to $t = 9$. The customer made a repeat purchase at $t = 9$, spending \$51, before churning at $t = 10$.

Table 2.2: Numerical Example of Customer Behavior

Time period	1	2	3	4	5	6	7	8	9	10
Prospect birth	0	1	0	0	0	0	0	0	0	0
Customer acquisition	0	0	0	0	0	1	0	0	0	0
Life	0	0	0	0	0	1	1	1	1	0
Purchase	0	0	0	0	0	1	0	0	1	0
Spend	0	0	0	0	0	29	0	0	51	0

The customer base model summarized in this section provides modelers with considerable flexibility and can be extended as needed. For example, the customer base model for so-called "freemium" games may involve an additional visit process, while the corresponding model for a telecommunications firm may involve disaggregate usage activity for cable, internet, and phone activity separately (Schweidel et al. 2011).

2.3.2 Incorporating Customer Base Predictions into Traditional Financial Valuation

Assume that the modeler has specified the customer-base model, which is parameterized by a parameter vector $\boldsymbol{\psi}$. After estimating the parameters of the model given the observed data, call it $\boldsymbol{D}$, the resulting parameter estimates $\widehat{\boldsymbol{\psi}}$ summarize what the disaggregated data are expected to be. For example, they determine how many prospects are available to be acquired, the flow of acquisitions, how many purchases those customers will make while they are alive, and how much they will spend on each of those purchases.

The total current and future expected revenues of the firm (R in Equation 2.4)

can be obtained by creating an "expected spend" matrix, which is formed by stacking each customers' expected spend vectors row-by-row (i.e., the final row in Table 2.2), then taking the column sums of this matrix[7]. Instead of using revenue projections from a time series model, we use revenue projections from the customer base model within our DCF model or relative valuation model to perform corporate valuation.

The total current and future expected customers acquired (A in Equation 2.4) can be obtained in a similar manner by creating an "expected acquisitions" matrix. The explicit dependence of future acquisition expenses upon customer acquisitions within DCF models is typically ignored in traditional financial valuation models.

Figure 2.2 depicts a flow diagram summarizing how the customer base model informs the DCF model through Equations 2.3 and 2.4, denoting by I the total number of customers that are acquired by the end of the forecasting horizon.

Other figures which provide additional insight into the health of a company's customer base (and thus the riskiness of the company's valuation) also follow easily from a particular set of customer base model parameter estimates $\widehat{\psi}$ in conjunction with the financial valuation model, including but not limited to: (1) the proportion of overall firm value coming from already-acquired customers versus yet-to-be-acquired customers, (2) the proportion of total company value derived from customers within particular tenure bands (e.g., customers who have been with the firm for over ten years, five to ten years, and so on), (3) the distribution of customer lifetime value across customers, and (4) the proportion of newly-acquired customers who will generate $x\%$ of the overall value of the overall cohort.

While we may predict what the disaggregate data should be (i.e., Table 2.2 for all individuals across all time periods), given a set of parameter estimates $\widehat{\psi}$, we will rarely, if ever, be able to directly observe all of the granular data. We discuss the

---

[7]In a similar fashion, we can easily obtain projections for the firm's total customer acquisitions, customer losses, and total purchases.

Figure 2.2: Schematic of Relationship Between Customer Base Model and DCF Model
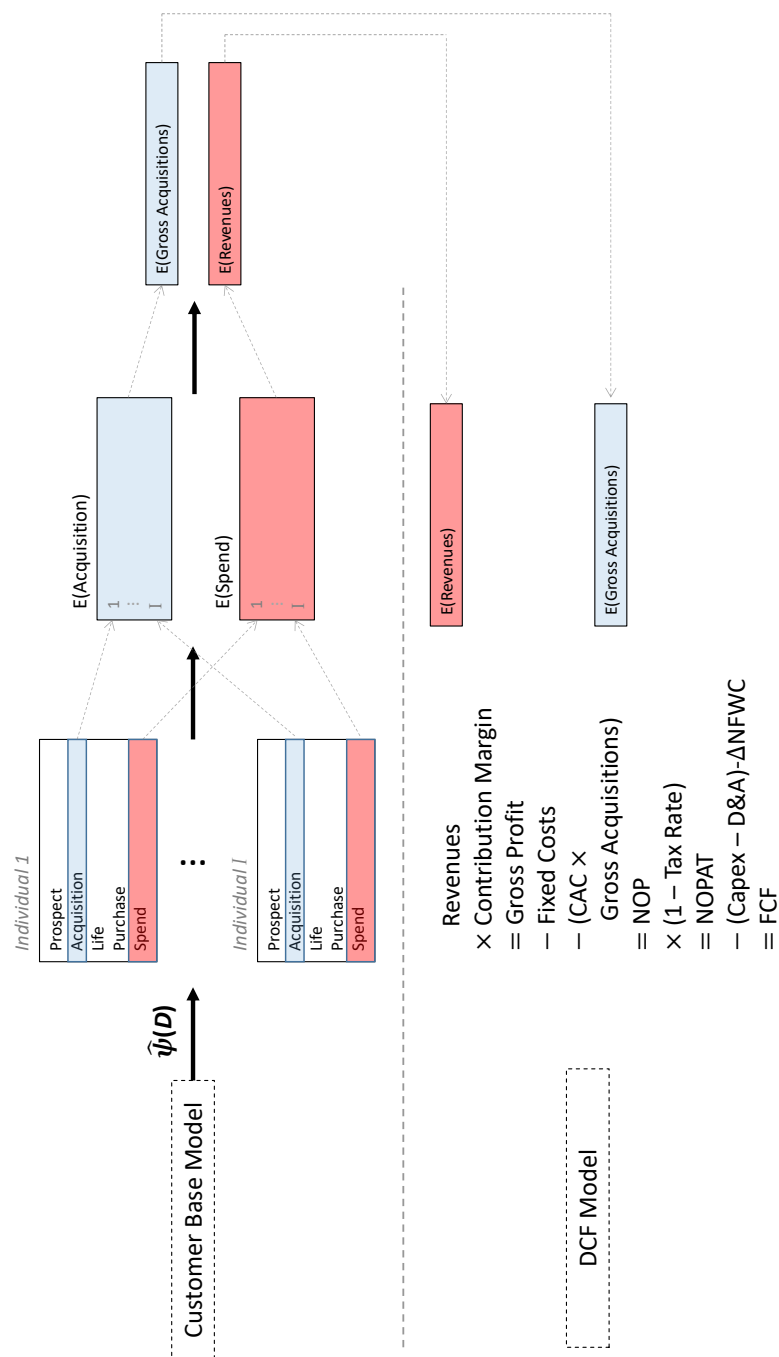
available data next.

2.4 Data Disclosure and Limitations

Even if we are within the firm itself and thus have access to the company's internal transaction logs, we are likely to have imperfect information regarding the incidence of individual prospects (Reinartz et al. 2005; Schweidel et al. 2008a). Internal transaction logs are usually left-censored because of data roll-off. At non-contractual firms, customer attrition is not observed[8].

Data available to external stakeholders (e.g., shareholders, bondholders, customers, suppliers, competitors, or vendors) are far more limited than data available to internal stakeholders. External stakeholders are typically unable to observe the raw transaction logs of the firm. It is usually the case that external stakeholders can only observe common customer metrics which are periodically disclosed by firms (e.g., customers acquired each quarter, the number of active customers, etc.). These periodically disclosed common customer metrics are marginalized statistics which summarize the acquisition, retention, purchase, and spend behavior across customers and over time. It is frequently the case that some metrics are disclosed on a regular basis before others are (e.g., for subscription-based firms, data regarding the total size of the customer base is often regularly disclosed for some period of time before data regarding customer acquisitions). Because firms do not disclose all common customer metrics back to the beginning of their commercial operations, the aggregated data are almost always left-censored as well.

Customer data availability varies considerably within and across business type and industry. We conducted a large-scale search for all companies which have dis-

---

[8]Contractual and non-contractual firms represent two opposing extremes in terms of the observability of customer attrition. Some firms may have multiple causes of churn, some of which are observable and others of which are not. For example, at social networking company Facebook and transportation network company Uber, customers may delete their accounts, or permanently cease activity at the firm without deleting their accounts. Churn is observable in the former scenario but not in the latter scenario (Ascarza et al. 2017).

closed customer data in their public filings (10-Q and 10-K reports) between 1994 and 2015 using data provider WRDS SEC Analytics Suite. Through this process, we identified 106 such companies. The industry with the largest number of companies disclosing customer metrics was the telecommunications industry. We identified 72 telecommunications companies which disclosed customer metrics. This finding is consistent with Bayer et al. (2017), who perform a large-scale customer metric disclosure analysis upon the telecommunications and airline industries and note that the telecommunications industry is by far the larger of the two in terms of disclosure frequency.

The majority of firms disclosing customer metrics within our sample were contractual firms, even if we were to exclude all telecommunications companies (e.g., firms such as ADT, AOL, DirecTV, DISH Network, Netflix, and SiriusXM). This is likely due to the fact that contractual firms are better able to identify their customers than non-contractual firms. This is also the likely explanation for much of the variation in disclosure frequency among non-contractual firms, both within and across industries. While customers are required by law to identify themselves when they purchase airline tickets, for example, they do not need to do so when they make purchases at a retailer. This makes it easier for airliners to collect (and thus disclose) detailed customer data than retailers. Similarly, online-only retailers (e.g., Amazon) are more able to identify their end customers than retailers with a large proportion of purchases coming through brick-and-mortar stores and/or wholesale channels (e.g., JCPenney). Even within brick-and-mortar-only businesses, some firms are better able to identify their customers. For example, restaurant chain SweetGreen stopped accepting cash payments in January 2017 (Dawson 2017). Cash payments are harder to track than payments through credit card or app. As an increasing proportion of purchases are made online instead of in-store, and through store apps and credit cards instead of

cash, we anticipate that the cost of disclosure will continue to decline and the number of companies disclosing customer metrics will increase.

In summary, data are typically limited in our CBCV setting. The nature and availability of the observed data vary considerably by business type, industry, and company. In the next section, we will discuss the implications that data availability has for modelers, as part of a broader discussion on model building considerations.

2.5   Model Building Considerations

There are two main contextual factors which drive what customer base model specification is appropriate in a particular problem setting: (i) the business type of the firm, and (ii) the nature of the available data. While we alluded to the importance of these factors in previous sections, we did not discuss how these factors influence what model is most appropriate in which setting. In this section, we address the modeling implications of these contextual factors, which may be helpful as future researchers build their own CBCV models.

Business type

Perhaps the most important contextual factor is whether the firm being valued is contractual (subscription-based) or non-contractual. Expected customer lifetime value by sales (call it E(SLV)) is the net present value of all future spend associated with a customer[9]:

$$E(SLV) = E\left[\int_0^\infty t(t) \; r(t) \; S(t) \; d(t) \; dt\right], \tag{2.5}$$

where $t(t)$ is the transaction rate at time t, $r(t)$ is the revenue associated with a

---

[9]In contrast, E(CLV) represents the net present value of all future after-tax marginal profits associated with a customer.

transaction at time t, S(t) is the survivor function, and d(t) is the discount factor which accounts for the time value of money and the non-diversifiable riskiness of the customer and/or firm.

While the empirical survivor function S(t) in Equation 2.5 is observable for contractual firms, it is *unobservable* for non-contractual firms. When customers of a cell phone provider would like to end their relationships with the firm, they must let the firm know. In contrast, if customers of an e-commerce retailer decide to end their relationships with the firm, they simply discontinue purchasing from the firm. This complicates the underlying model required to predict future customer activity because non-contractual firms cannot report customers lost.

Even if a non-contractual customer were known to be alive, there are additional complexities associated with her repeat purchase and spend behavior that must be taken into account. Transaction rates t(t) and spend amounts associated with each of those purchases r(t) can be highly variable over time and across customers for non-contractual firms. An accurate model for the future purchase and spend behavior for non-contractual customers must be able to estimate heterogeneity in these behaviors. In contrast, customers of contractual firms traditionally pay a fixed subscription fee each period, which is equivalent to exactly one purchase each period for all customers (i.e., t(t)=1) with a spend amount that is relatively constant over time and relatively homogeneous across customers.

In summary, CBCV models for non-contractual firms must account for (1) latent attrition and (2) heterogeneity in acquisition, repeat purchase, latent attrition, and spend given purchase. In contrast, CBCV models for contractual firms typically only need to account for heterogeneity in acquisition and (observable) attrition, because the purchase process is simpler and heterogeneity in spend is far smaller at non-contractual firms than at contractual firms. Most previous CBCV articles have noted

this distinction (Bonacchi et al. 2015b; Schulze et al. 2012), but not all (Gupta et al. 2004; Libai et al. 2009).

<u>Available data</u>

As alluded to in the previous section, another important contextual factor is whether internal or external data are available to the modeler. When internal data are available, it is easier to statistically identify theoretically reasonable effects such as cross-cohort dynamics (Gopalakrishnan et al. 2016), correlation between acquisition and retention propensities (Schweidel et al. 2008a), customer re-acquisition (Thomas et al. 2004), or "clumpiness"/"regularity" of customer purchasing (Platzer & Reutterer 2016; Zhang et al. 2014). The disaggregated data structure outlined in Section 2.3.1 would be able to identify these effects, if the disaggregated data were available for a large enough group of customers over a long enough period of time.

If the modeler only observes external data, it becomes difficult or impossible to identify these effects. Bodapati & Gupta (2004) warn us that when data are highly aggregated, even identifying heterogeneity (in their setting, using a finite mixture model) can be challenging. Model parsimony is important, particularly when the data are limited. We survey the relevant literature on estimating individual-level models with aggregated data and how our work relates to this literature in Section 2.6.2.

In addition to aggregation across customers and over time, publicly-disclosed external data frequently suffers from "missingness," which typically takes on one of two forms. First, there is the issue of left-censoring. For example, while satellite radio company Sirius XM (one of the companies we will consider in our empirical analysis for essay two) began commercial operations in 2001/2002,[10] it started disclosing paying customer data in Q3 2008. Second, there is the issue that some customer data

---

[10]The two companies that later merged to form Sirius XM started commercial operations in September 2001 and February 2002.

points (e.g., total active customers) are reported for some period of time before being complemented by other data points (e.g., total customers acquired). If only external data are available, the customer base model estimation procedure must be robust to these data challenges.

Another consideration is the source of the data. The data may come from "first-party" sources (i.e., data from the firm itself), "third-party" sources (e.g., data from a business intelligence firm), or a mixture of both. First-party disclosures and third-party disclosures based upon large, highly representative samples are observed with little or no measurement error. Third-party disclosures that are not based upon large, highly representative samples may suffer from error-in-variable bias if the parameter estimation procedure does not explicitly account for the existence of error in the data recording process. Modelers performing CBCV should assess whether their model specification requires a correction for measurement error based upon where the data are coming from. We assume throughout this dissertation that the data are observed without measurement error.

The difficulties which arise when the modeler only observes external data exist for both contractual or non-contractual firms. However, the challenges are more severe for non-contractual firms. While prior work has shown that two customer metrics are needed to perform CBCV for contractual firms – total customers acquired and total customers churned each period[11] (Gupta et al. 2004; Libai et al. 2009; Schulze et al. 2012; Wiesel et al. 2008) – it is unclear what metrics are needed and/or are most informative as inputs for non-contractual firms, for whom attrition is unobserved. The most popular and widely adopted/referenced workaround to this complication in non-contractual settings (Bain and Company 2000; Blattberg et al. 2001; Gupta

---

[11]Instead of disclosing the number of customers lost during the period, the firm may equivalently disclose the per-period churn rate, or the total number of customers that are active as of the end of each period.

et al. 2004; Libai et al. 2009; Seybold 2000) is to create some notion of an observable "retention rate" which, in truth, does not exist in a non-contractual setting, and proceed with the same framework that is used for contractual settings. This proxy may be defined, for instance, as the repeat rate, or proportion of customers who made a purchase last year who made another purchase this year (Farris et al. 2010). The repeat rate may be a useful proxy for customer longevity, but using it in a traditional CLV formula in a non-contractual setting will understate future purchase activity (and thus future profits) dramatically because customers who have not purchased in one year may still be alive[12]. We will study this topic in detail in essay three.

While it would be nice to incorporate other "bells and whistles" into the customer base model, it is most important that we allow for heterogeneity in the most relevant processes first (acquisition and retention for contractual firms; acquisition, latent attrition, repeat purchase, and spend for non-contractual firms). Ignoring heterogeneity, particularly in the attrition process, will tend to undervalue the firm considerably (Fader & Hardie 2010). What other effects we can identify over and above heterogeneity with external data will depend on what external data are available – in particular, the amount of left censoring, the quality and quantity of metrics available, the duration of time the metrics are available for, and the periodicity of the disclosures.

In this dissertation, we perform CBCV assuming external data are available in both contractual and non-contractual settings. Corporate valuation is performed most frequently using publicly available data, making this vantage point the most relevant one. Performing CBCV is also more challenging, and more interesting methodologically, when the data are limited than when data are abundant. Nevertheless, we will discuss other problem settings as part of our discussion of future work.

---

[12]For example, consumer products seller QVC notes that 6% of total sales in 2015 came from customers who had not purchased in over a year (QVC 2015).

2.6 Literature Review

Based on the unifying framework we developed in Section 2.3, and the importance of limited data to compute CBCV we discussed in Section 2.4, we provide a literature review of CBCV in Section 2.6.1, and modeling with missing and aggregated data summaries in Section 2.6.2.

2.6.1 Customer-based Corporate Valuation

The idea that maximizing shareholder value should be the objective when making decisions gained popularity in the 1980s, and the associated writings — especially Rappaport (1986) — brought the basic principles of firm valuation to a broader, non-finance audience.

In his review of the valuation literature, Damodaran et al. (2005, p. 1) writes: "Given the centrality of its role, you would think that the question of how best to value a business, private or public, would have been well researched. [...] [T]he research into valuation models and metrics in finance is surprisingly spotty, with some aspects of valuation, such as risk assessment, being deeply analyzed and others, such as how best to estimate cash flows ... not receiving the attention that they deserve."

Kim et al. (1995) were the first marketing academics to recognize the potential for using some of the models of customer behavior developed by marketing scientists to generate the key inputs for estimating cash flows. They used the logistic internal-influence model for the diffusion of an innovation (which is equivalent to Fisher & Pry (1971)'s model of technology substitution) to characterize (and then project) the market penetration of mobile phones (and therefore the associated revenues of a cellular communication company), resulting in an estimate of the market value of a

business explicitly based on a model of customer behavior.[13] Pioneering as it was, the biggest shortcoming in their analysis was that they did not consider the reality of customer churn (i.e., it assumed that once the customer has adopted the service, they remain as a customer forever).

Driven in part by the interest in moving from transaction-oriented/product-centric marketing strategies to relationship-oriented/customer-centric marketing strategies (with their emphasis on customer acquisition, retention, and development), the 1990s saw the notion of customer lifetime value (CLV) — defined as "the present value of the future cash flows attributed to the customer relationship" (Pfeifer et al. 2005, p. 17) — emerge from the confines of specialized direct/database marketing firms and become what is now a fundamental concept for most marketers. Blattberg & Deighton (1996) introduced the concept of "customer equity" (CE), which is the sum of the lifetime values of the firm's customers, both current and future. Kumar & Shah (2015) provide a comprehensive guide to the literature on customer equity.

The pioneering work of Gupta et al. (2004) (hereafter, GLS) was the first to explicitly link CLV and firm value. Underpinning their work was the logistic internal-influence model to characterize customer acquisitions and a simple model for the CLV of acquired customers. After calibrating the models using publicly available data (along with expert judgment), they arrived at estimates of market value for five listed companies. However, their treatment of the valuation problem suffers from two issues. First, their CLV calculations are performed assuming a constant retention rate. Second, their valuation framework does not incorporate key financial/accounting issues such as firm capital structure and non-operating assets.

A number of researchers have built on GLS's work. Most notably, Schulze et al. (2012) (hereafter SSW) provide a thorough treatment of how CE relates to firm

---

[13]It is important to note that they applied the model at the level of the industry, not the firm, in their empirical analysis.

value using financial valuation theory, addressing many of the financal/accounting issues associated with the valuation aspect of GLS's work. Several researchers have explored a number of technical issues associated with any valuation exercise. Other key papers include Kumar & Shah (2009), Libai et al. (2009), and Wiesel et al. (2008). Gupta (2009) provides a helpful summary of extant CBCV work and future research directions.

These ideas have been gaining attention and respect outside of marketing. Within the accounting literature, for instance, Bonacchi et al. (2015a) provide a systematic analysis across multiple companies linking the value of the firm's existing customers (current customer equity or CCE) to shareholder value. Other accounting-related work includes Andon et al. (2001), Bonacchi et al. (2008), Bonacchi & Perego (2012), Boyce (2000), Gourio & Rudanko (2014), and Hand (2015).

While many of the papers reviewed above discuss DCF methods for firm valuation (often anchoring on Rappaport (1986)'s expression for SHV), they do not explicitly make use of such a framework when generating an estimate of firm value. Rather, they take what Skiera & Schulze (2014) call a customer-based valuation approach. Skiera & Schulze (2014) first state that "[c]ustomer-based valuation first uses information about the customer base (for example, number of customers, contribution margin per customer, retention rate) to determine the value of the firm," and then describe an approach based on the (residual) lifetime value of existing customers and the lifetime value of as-yet-to-be-acquired customers, multiplying these two quantities by the number of current and expected future customers (respectively) and adjusting for various financial considerations. Their position is that DCF and "customer-based valuation" methods are fundamentally different.

We believe that this is more polarizing than it needs to be. Previous "customer-based valuation" methods are performing on a customer-by-customer basis what are

27

effectively NPV calculations over time and then summing up across customers. "DCF methods" (as outlined in Section 2.2.1) are summing up spends across customers in each time period and then performing an NPV calculation across the resulting total revenue figures. Given the same underlying matrix representing spends for each customer in each time period, and assuming all the various accounting issues are handled correctly, both approaches should yield the same estimate of firm value.

2.6.2 Modeling with Missing, Aggregated Data Summaries

Statistical issues arising from the need to estimate individual-level behaviors using aggregated data have not been addressed in the CBCV literature because to the best of our knowledge, all extant CBCV literature has assumed homogeneous models (i.e., they assume no variation in behaviors across customers). We cannot ignore these issues because the customer base models we build in essays two and three allow for heterogeneity in the propensity to be acquired, purchase, remain with the firm, and spend across customers and over time. Pfeifer (2011) explored the consequences of temporal aggregation of customer data disclosures by public firms. He showed that one must be careful with the timing of cash flows when estimating retention rates and CCE using publicly disclosed company data. These timing effects are due to the fact that while firms disclose financial metrics each quarter, the underlying cash flows are generated throughout the period. Pfeifer briefly alluded to the promise of allowing for heterogeneity in customer retention in a contractual setting, but did not pursue it further in the article.

Many marketing articles outside of the CBCV literature have studied the problem of inferring heterogeneous individual-level behaviors with aggregated data. Early literature focused upon the problem of estimating random coefficient or latent class discrete choice models with aggregate market share data. Historically, there had

been three common approaches to this problem. The first involved minimizing the difference between actual and expected market shares (Berry 1994; Boyd & Mellman 1980; Cardell & Dunbar 1980; Tardiff 1980). The second used simulation-based GMM procedures (Berry et al. 1995). The third obtained the marginalized likelihood function associated with the aggregated data, treating it as a convolution of the individual-level data, then maximized the marginalized likelihood (Bodapati & Gupta 2004; Kim 1995; Zenor & Srivastava 1993).

Since then, a number of marketing articles have extended both the breadth of the model used and the types of data that can be incorporated into the model. These extensions have tended to adopt two methodological approaches. The first approach uses more complex simulation-based GMM procedures. For example, Albuquerque & Bronnenberg (2009) combine different aggregated data sets – one involving market share data, and another involving other simple data summaries which marginalize across time such as brand penetration and purchase set size – using a more sophisticated GMM-based estimation technique. The second approach uses Bayesian methods, typically by treating the individual-level data as latent variables and sampling over the augmented posterior distribution. Chen & Yang (2007) estimate a more complex random coefficient discrete choice model which allows for purchase dynamics over time with market share data. Musalem et al. (2008) and Musalem et al. (2009) propose a similar Bayesian estimation technique which they use to estimate a discrete choice model on multiple aggregated data sources (market share and usage data). Jiang et al. (2009) approximate the likelihood of the aggregated data, then build a Bayesian model around it. Feit et al. (2013) incorporate both aggregate and disaggregate data into the same discrete choice model using a Bayesian model. More recently, Hui (2017) proposes a Bayesian model for both the adoption and repeat visit behavior of players in a mobile gaming setting, estimating this model with two com-

mon customer metrics, DAU (daily active users) and MAU (monthly active users). In essay three, we adopt an estimation approach more similar to the simulation-based GMM of Albuquerque & Bronnenberg (2009) than the Bayesian methods of Chen & Yang (2007), Musalem et al. (2008), and others.

Hui (2017) studies a problem setting that is most similar to the non-contractual CBCV setting we will study in essay three. Unlike previous articles, (1) it proposes a model for trial, repeat, and spend behaviors, (2) the aggregated data he estimates the model upon are closer to our own (e.g., a collection of common customer metrics, and not market share alone, or market share and simple summary metrics which marginalize across time), and (3) he does not use a rejection sampling-based method (i.e., his method can accept posterior samples whose individual-level parameters are not necessarily consistent with the observed aggregated data), which previous Bayesian aggregated data approaches have demanded. Hui uses this model to generate novel insights regarding a large cross-section of mobile games. However, there are some methodological concerns regarding Hui's approach, including (1) while the model can statistically identify mean effects, it has difficulty identifying heterogeneity effects, (2) difficulty identifying heterogeneity effects raises questions about the practical utility of using a computationally intensive Bayesian approach in lieu of a frequentist one, (3) the model requires that the error variance parameters be "hardcoded" in advance (in practice, Hui sets the error variance parameters equal to 5%).

We run extensive parameter identification tests benchmarked off of a commonly-used data set to show that we do not have parameter identification issues, and our proposed methodology does not require any parameters to be hardcoded. Our proposed customer base model specification greatly extends Hui's in that we allow for far richer models (e.g., heterogeneity across customers in the acquisition and spend processes, and time-varying covariates in the repeat purchase and spend processes).

Our available data are more limited (e.g., we observe customer metrics quarterly and not daily, and those metrics may be left censored with missingness) and more eclectic (e.g., in a non-contractual setting, we consider a set of six possible customer metrics which are an assortment of proportions, rates, and counts). We are answering different questions, including what the most informative collection of customer data summaries are, and how we can insert the resulting model into an overall valuation framework.

In summary, while the prior literature does not apply directly to our problem setting, it is very relevant and thought-provoking. It helps us derive the models we propose in essays two and three, and inspires research directions which modelers may pursue in the future.

2.7 Conclusions

The main objective of this essay is to motivate and clarify research issues related to customer-based corporate valuation. Understanding how models of customer behavior can be incorporated into the most common corporate valuation frameworks allows us to (1) improve the accuracy of those valuation models, (2) understand the risks associated with a company's valuation better, and (3) summarize the health of the customer base along meaningful dimensions (e.g., customer acquisition, customer retention, and purchase and spend propensities), and how these dimensions are changing over time.

We offer a unifying framework for CBCV, documenting the most common valuation methodologies, expanding upon the five processes which comprise the customer base model, and showing how the latter can be integrated into the former. We explain the limitations of the data which are available to perform CBCV, and how data avail-

ability and business type influence customer base model specification. We motivate and contextualize the problem settings that we will study in depth in essays two and three – namely, that of a passive external stakeholder performing a going concern valuation of contractual and non-contractual businesses using limited external data.

While we would not claim to present a complete listing of all possible CBCV issues and factors, we hope that it provides some useful guidance to future modelers attempting to perform CBCV. Let us turn to CBCV in a contractual setting next.

3

Valuing Contractual Firms with Publicly-Disclosed Customer Data

3.1 Introduction

The relevance and popularity of subscription-based businesses — businesses whose customers pay a periodically recurring fee for access to a product or service — has grown considerably in recent years. Previously dominated by newspapers, magazines, and telecommunications companies, the subscription-based business model has made strong inroads into consumer software (Microsoft 365), food preparation (Blue Apron), health and beauty products (Dollar Shave Club), and a large array of subscription-based software-as-a-service (SaaS) enterprises in the B2B space, as businesses look to increase the predictability of their revenue streams. Many experts have written in depth about this topic (Baxter 2015; Janzer 2015; Warrillow 2015).

The increased popularity of subscription-based businesses has brought with it an increase in the public disclosure of data on (but not limited to) customer churn, customer/subscriber acquisition costs, average revenue per user, and customer lifetime value (CLV). The price of a company's stock reflects and incorporates investors' beliefs regarding the future cash flows the company will generate. For subscription-based businesses, the primary source of future cash flows is customers. Therefore, customer

data are important to investors and are being used by analysts as they make their recommendations. For example, a class-action lawsuit was taken out again Netflix in response to changes in its reporting of such data (SCAC 2004). Analyst reports from Thomas Weisel Partners, Vintage Research, First Albany Capital and Delafield Hambrecht, made public as part of the litigation, all strongly emphasize customer data in general (and the size of the total subscriber base over time in particular) when justifying their investment recommendations.

The work of Gupta et al. (2004) was the first to explicitly link firm value to CLV for public subscription-based companies. However, their treatment of the valuation problem suffers from two major issues. First, their CLV calculations are performed assuming a constant retention rate, which can result in an undervaluing of existing customers (Fader & Hardie 2010). Second, their valuation framework does not incorporate key financial/accounting issues such as firm capital structure and non-operating assets. While other researchers, most notably Schulze et al. (2012), have built upon this work, the underlying models of customer behavior and the associated valuation frameworks are not up to the standards expected by marketers and financial professionals, respectively.

Our objective in this essay is to present a framework that is specifically designed for valuing subscription-based business. The parameters of the underlying model of customer behavior can be estimated using only publicly disclosed customer data, making it suitable for passive investors valuing a going concern. To do so, we explicitly account for the fact that publicly reported data are typically aggregated (temporally and across customers) and suffer from missingness (i.e., the reported data are not available for all periods). We present models of the firm's acquisition and retention processes that accommodate factors such as customer heterogeneity, duration dependence, seasonality, macroeconomic conditions, and changes in population size.

The essay is organized as follows. In the next section, we discuss the principles of customer-based corporate valuation in a subscription-based setting, exploring the nature of the customer data typically released by subscription-based businesses. We present our model of customer behavior, specifying models for customer acquisition, retention, and spend. We then provide an empirical analysis that explores how such a model can be fit to real public company data; the two firms considered in our analysis are DISH Network and Sirius XM. After demonstrating the validity of our model, we present our valuations of the firms, and explore other insights that can be derived using our model. We conclude with a discussion of the results and future work.

3.2 A Data Structure for Subscription-Based Businesses

As noted in the previous chapter, subscription-based businesses are a special case of non-subscription-based businesses. Because of this, we can simplify the general data structure that was outlined in Section 2.3.1 to one that is specifically suited to subscription-based firms when only external data are available. Let us first identify the data that are typically available inside subscription-based firms. Let us assume that the firm has a monthly internal reporting period. The key numbers of interest are monthly revenues, which we denote by $R(m)$ (where $m = 1$ corresponds to the firm's first month of commercial operations).

As we had noted in the previous chapter (Section 2.3), it makes sense to decompose the aggregate revenue numbers, separately model the constituent components, and then combine the forecasts of these components to arrive at the desired revenue forecasts. First, recognizing that revenue comes from customers, it would be a good start to decompose revenue into its "number of customers" and "average revenue

per customer" components. Second, as we think about the number of customers a subscription-based firm has in a month, it makes sense to decompose this quantity into the number of new customers acquired that month and the number of customers acquired in previous months who still have a relationship with the firm. Knowing the number of new customers acquired in each month is a critical input to any valuation exercise, especially for firms with high subscriber acquisition costs.

Let us think about what lies behind the "total number of customers" number. It is helpful to think of a "number of customers" matrix, $C(\cdot, \cdot)$, which tracks customer behavior by time of acquisition. With reference to Figure 3.1 (where the columns correspond to (calender) time since the start of the firm's commercial operations and the rows correspond to acquisition cohorts), let $C(m, m')$ be the number of customers acquired by the firm in month m who are still active in month m'. It follows that the total number of customers the firm has at the end of month m' is given by the column total $C(., m') = \sum_{m=1}^{m'} C(m, m')$. The number of customers in any cohort must be non-increasing over time (i.e., $C(m, m') \geq C(m, m'')$ for $m' < m''$).

The $C(\cdot, \cdot)$ matrix, along with $R(\cdot)$, lies that the heart of a number of customer metrics reported both internally and externally by subscription-based firms:

- A sophisticated subscription-based firm will report the $C(\cdot, \cdot)$ matrix internally, either in its raw form or as cohort-by-cohort survival percentages $(C(m, m')/C(m, m) \times 100\%)$ (e.g., Martinez-Jerez et al. 2011).

- The number of customers acquired each month by the firm is given by

$$A(m) = C(m, m).\tag{3.1}$$

Figure 3.1: The "Number of Customers" Matrix
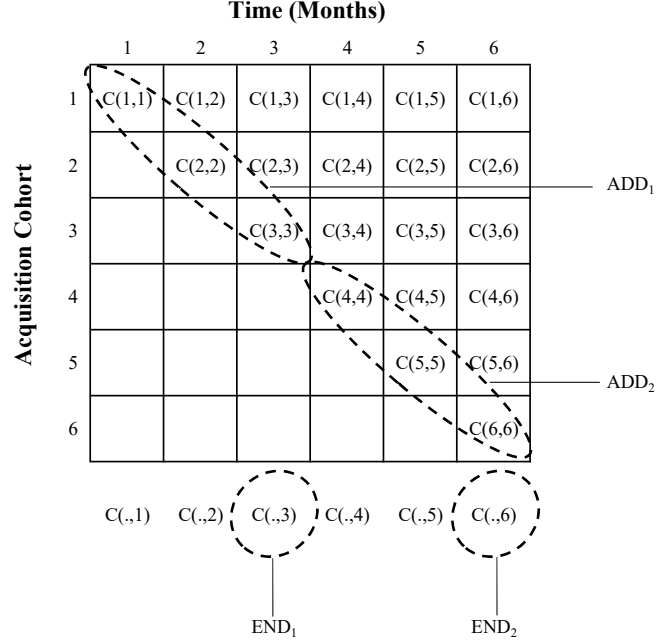
- The number of customers "lost" each month by the firm is given by

$$
\mathrm{L}(m) = \begin{cases} 0 & m = 1 \\ \mathrm{C}(.,m-1) - [\mathrm{C}(.,m) - \mathrm{C}(m,m)] & m = 2, 3, 4, \dots \end{cases} \tag{3.2}
$$

(It follows that an aggregate monthly churn rate can be computed as $\mathrm{L}(m)/\mathrm{C}(.,m-1)$.)

- For most firms with a subscription-based business model, the average revenue per subscriber is relatively constant across customers during a given period of time.[1] Let us denote this quantity by $\mathrm{ARPU}(m)$ and compute it in the following manner:

$$
\mathrm{ARPU}(m) = \mathrm{R}(m) \bigg/ \frac{\mathrm{C}(.,m-1) + \mathrm{C}(.,m)}{2} . \tag{3.3}
$$

---

[1]In contrast, average revenue per subscriber for firms with a non-subscription-based business model tends to vary considerably across customers in any given period of time.

(The denominator is the average number of customers the firm has during month m, assuming customers churn uniformly.)

Publicly disclosed customer data are typically reported quarterly with the associated unit of time being the quarter; as such, they represent a temporal aggregation of the true underlying process. Commonly reported measures include the number of customers active at the end of each quarter (END_q), and the number of customers added and lost each quarter (ADD_q and LOSS_q, respectively). Assuming the firm started operations at the beginning of a reporting quarter (i.e., $q = 1$ comprises $m = \{1, 2, 3\}$; equivalently, the first month of each quarter is either January, April, July or October),

$$\text{END}_q = C(., 3q) \tag{3.4}$$

$$\text{ADD}_q = A(3q - 2) + A(3q - 1) + A(3q) \tag{3.5}$$

$$\text{LOSS}_q = L(3q - 2) + L(3q - 1) + L(3q) \tag{3.6}$$

This mapping from the internal "number of customers" matrix to ADD and END is illustrated in Figure 3.1.

Quarterly revenues (REV_q) are given by

$$\text{REV}_q = R(3q - 2) + R(3q - 1) + R(3q). \tag{3.7}$$

The challenge we face is how to make projections of $R(m)$ and $A(m)$ far into the future (as required for calculating for the FCF numbers) using the publicly reported ADD, LOSS, END, and REV numbers. We pursue this important task in our Model Development section below but first we discuss how other researchers have utilized the valuation concepts and data structures discussed here.

3.3   Model Development

Our goal is to develop a model of customer behavior that can be used to generate long-run projections of R(m) and A(m), one whose parameters can be estimated using only publicly reported ADD, LOSS, END, and REV numbers. As we had noted in Section 2.5, it is important that our approach to parameter estimation accounts for the "missingness" and aggregation associated with the data reported by companies. With respect to aggregation, the publicly disclosed customer data are typically reported quarterly with the associated unit of time being the quarter, while for the purposes of our analysis, we assume that the firm is operating on a monthly basis.[2]

At the heart of this work are models for the customer acquisition and retention processes that allow us to project C($\cdot, \cdot$) into the future. Coupled with a model for ARPU(m), we can then generate our projections of R(m).

We start by describing our model for the retention process. This assumes we know how many customers the firm acquires each month. We then describe our model for the customer acquisition process, one that takes into consideration the possibility of reacquiring customers who have previously churned, and then examine how to jointly estimate the parameters of these two models. Finally, we present a simple model for the evolution of ARPU(m), and then outline how to bring together all of these submodels to generate the desired projections of R(m).

3.3.1   The Retention Process

Let the survival function $S_R(m' - m|m)$ denote the probability that a customer acquired in month m remains an customer for at least $m' - m$ months. Having acquired

---

[2]Our model can also be used in situations where customer data are only reported annually with minimal modification.

$A(m) = C(m, m)$ customers in month m, it follows that

$$C(m, m') = C(m, m) \times S_R(m' - m|m), m' \geq m.$$ (3.8)

Our objective is to specify an accurate yet parsimonious survival model for the duration of customers' relationships with the firm. In addition to capturing the effects of cross-sectional heterogeneity and duration dependence, we want to accommodate time-varying covariates to control for the effects of seasonality and macroeconomic conditions.

We use a proportional hazards model with a Weibull baseline, and capture cross-sectional heterogeneity in the baseline churn propensity using a gamma distribution. This is a well-accepted model for duration data and has been proven to be quite effective and robust in a number of different application settings (Morrison & Schmittlein 1980; Moe & Fader 2002; Schweidel et al. (2008b)).

Given a customer's individual-specific baseline propensity to churn (λ_R), a homogeneous retention process shape parameter (c_R), time-varying retention covariates $(\boldsymbol{X}_R(m, m') = [\mathbf{x}_R(m), \mathbf{x}_R(m+1), \ldots, \mathbf{x}_R(m')])$, and the coefficients associated with the retention covariates $(\boldsymbol{\beta}_R)$,

$$S_R(m' - m|m, \boldsymbol{X}_R(m + 1, m'); \lambda_R, c_R, \boldsymbol{\beta}_R) = \exp(-\lambda_R B_R(m, m')),$$

where
$$B_R(m, m') = \sum_{i=m+1}^{m'} [(i - m)^{c_R} - (i - m - 1)^{c_R}]e^{\boldsymbol{\beta}_R \mathbf{x}_R(i)}.$$ (3.9)

Following Schweidel et al. (2008b) and Jamal & Bucklin (2006), we expect $c_R \geq 1$. When $c_R = 1$, it reduces to an exponential baseline proportional hazards model.

Assuming λ_R is distributed gamma(r_R, α_R) across the population, the unconditional probability that a customer acquired in month m survives at least $m' - m$

40

months is

$$
\begin{aligned}
S_R(m' &- m | m, \boldsymbol{X}_R(m+1, m'); r_R, \alpha_R, c_R, \boldsymbol{\beta}_R) \\
&= \int_0^\infty S(m' - m | m, \boldsymbol{X}_R(m+1, m'); \lambda_R, c_R, \boldsymbol{\beta}_R) f(\lambda_R | r_R, \alpha_R) d\lambda_R \\
&= \left(\frac{\alpha_R}{\alpha_R + B_R(m, m' - m)} \right)^{r_R}.
\end{aligned}
\tag{3.10}
$$

Plugging this survival function into Equation 3.8 allows us to predict the number of active customers in future months for the month m cohort. Therefore, if we know $C(m, m)$ over all months, we can predict the remainder of the upper triangular matrix in Figure 3.1 as well as the number of customers lost each month (Equation 3.2). Estimates of LOSS_q and END_q follow from Equations 3.5 and 3.6 for all $q = 1, 2, \ldots$.

In theory, we could estimate the model parameters $(r_R, \alpha_R, c_R, \boldsymbol{\beta}_R)$ by minimizing the sum of squared differences between our model-based estimates of LOSS_q and the reported numbers (a non-likelihood-based approach in the spirit of Berry 1994). However this assumes we know the monthly customer acquisition numbers, $A(m)$, which is not the case. We only have quarterly customer additions ADD_q and some of these observations are probably missing. We therefore need to develop a model of the acquisition process whose parameters can be estimated using the reported ADD_q data. These two models will be estimated simultaneously to give us the required $A(m)$ and $L(m)$ numbers.

3.3.2 The Acquisition Process

At first glance, specifying a model for the acquisition of customers over time seems to be a relatively simple exercise. The Bass model (Bass 1969) or a simplified variant such as the logistic internal-influence model (as used by GLS and SSW) would appear to be the obvious choice. However, for the following four reasons, this is not the case:

(a) It assumes that all churning customers disappear forever — once an acquired customer has churned, he/she cannot re-enter the pool of 'potential adopters.' SSW attempt to overcome this problem by using the logistic internal-influence model to characterize the number of net total customers (i.e., the number of customers after churn).[3]

(b) It assumes the population size is fixed, when we know that the number of potential customers is typically increasing over time due to population growth.

(c) The Bass model and its simplified variants have a number of unfavorable properties, most notably the fact that the resulting adoption curve is symmetric about the period of peak acquisition. In real datasets, skewness about the peak is almost always present.

(d) It ignores the effects of seasonality and macro-economic events.

We therefore develop a model from first principles that addresses each of these issues.

Let $POP(m)$ denote the size of the population in the target market in month m, with $POP(0)$ being the population size when the firm first commences operations. We assume $POP(m)$ is non-decreasing over time.

Each month sees the formation of a new prospect pool of size $M(m)$. We set $M(0)$, the size of firm's prospect pool when it commences operations, to the size of the population at that time. The size of the prospect pool in the company's second month of operation is simply the increase in the size of the population over the preceding month. Thereafter, the size of the prospect pool is equal to the growth in the population during the preceding month, plus the number of customers who

---

[3]Libai et al. (2009) extend the basic Bass model to allow for lost customers re-entering the pool of potential adopters, but they assume a constant retention rate.

churned in the previous month:

$$M(m) = \begin{cases} \text{POP}(m) & m = 0 \\ \text{POP}(m) - \text{POP}(m-1) + L(m-1) & m = 1, 2, 3, ... \end{cases} \tag{3.11}$$

We assume that population growth is the only source of potential adopter growth over time, aside from previously churned customers.

Once a prospect pool has formed, some time will elapse before individuals within that pool are acquired as customers.[4] Let $F_A(m' - m|m)$ denote the probability that a member of prospect pool m is aquired by the end of month m'. It follows that the total number of new customers in month m is

$$A(m) = \sum_{i=0}^{m-1} M(i) \times \left[F_A(m - i|i) - F_A(m - i - 1|i) \right]. \tag{3.12}$$

We model the time it takes for a prospect to become a customer using a split-hazard model. A proportion π_{NA} of each prospect pool will never be acquired. For those that are potential customers, we characterize the time to acquisition using a proportional hazards model with a Weibull baseline, and capture cross-sectional heterogeneity in the baseline acquisition propensity using a gamma distribution.

Given a prospect's individual-specific baseline propensity to be acquired (λ_A), a homogeneous acquisition shape parameter (c_A), time-varying acquisition covariates ($\boldsymbol{X}_A(m+1, m') = [\mathbf{x}_A(m+1), \mathbf{x}_A(m+2), \ldots, \mathbf{x}_A(m')]$), and the coefficients associated with the acquisition covariates ($\boldsymbol{\beta}_A$),

$$F_A(m' - m|m, \boldsymbol{X}_A(m+1, m'); \lambda_A, c_A, \pi_{NA}, \boldsymbol{\beta}_A) = (1 - \pi_{NA})\left[1 - \exp(-\lambda_A B_A(m, m'))\right],$$
$$\tag{3.13}$$

---

[4]In line with most of the work on modeling the diffusion of innovations, we ignore the intermediate role of awareness as we do not have sufficient data to account for it.

where

$$B_A(m, m') = \sum_{i=m+1}^{m'} [(i-m)^{c_A} - (i-m-1)^{c_A}]e^{\boldsymbol{\beta}_A \mathbf{x}_A(i)}. \qquad (3.14)$$

Assuming λ_A is distributed gamma(r_A, α_A) across the population, the unconditional probability that a customer from prospect pool m will be acquired by the end of month m' is

$$F_A(m' - m|m, \boldsymbol{X}_A(m+1, m'); r_A, \alpha_A, c_A, \pi_{NA}, \boldsymbol{\beta}_A)$$
$$= \int_0^\infty F_A(m' - m|m, \boldsymbol{X}_A(m+1, m'); \lambda_A, c_A, \pi_{NA}, \boldsymbol{\beta}_A)f(\lambda_A|r_A, \alpha_A)d\lambda_A$$
$$= (1 - \pi_{NA})\left[1 - \left(\frac{\alpha_A}{\alpha_A + B_A(m, m')}\right)^{r_A}\right]. \qquad (3.15)$$

This acquisition model is flexible yet parsimonious. Parsimony is especially important in limited data settings (such as those considered here) because, as shown in Van den Bulte & Lilien (1997), ill-conditioning is a serious enough problem with small sample sizes that adding new predictors to alleviate model misspecification concerns may make the resulting model fit (and forecast) worse than it had been prior to the introduction of those covariates.

3.3.3 Parameter Estimation for the Acquisition and Retention Processes

We estimate the parameters of the acquisition and retention process models jointly using nonlinear least squares, minimizing the sum of squared differences between the actual and model-based estimates of quarterly acquisitions and losses. Let $\boldsymbol{\psi}$ denote the acquisition and retention process model parameters collectively, $\boldsymbol{\psi} \equiv (r_A, \alpha_A, c_A, \pi_{NA}, \boldsymbol{\beta}_A, r_R, \alpha_R, c_R, \boldsymbol{\beta}_R)$, and let Q be the number of quarters from the commencement of the firm's commercial operations to the end of the model calibra-

tion period.

If the firm reports the quarterly numbers from the very start of its operations, our parameters are those that minimize the following sum-of-squared errors:

$$\text{SSE}_{FULL} = \sum_{q=1}^{Q} \left\{ \left(\text{ADD}_q - \widehat{\text{ADD}}_q\right)^2 + \left(\text{LOSS}_q - \widehat{\text{LOSS}}_q\right)^2 \right\} + \left(\text{END}_Q - \widehat{\text{END}}_Q\right)^2, \quad (3.16)$$

where $\widehat{\text{ADD}}_q$, $\widehat{\text{LOSS}}_q$, and $\widehat{\text{END}}_q$ are the model-based estimates of these quantities computed using $\hat{\boldsymbol{\psi}}$.[5] Note that we optimize over all parameters jointly because of the dependence of the retention process on the acquisition process (i.e., customers cannot churn until they have been acquired) and vice versa (i.e., churning customers enter future prospect pools).

Equation 3.16 assumes that there is no missing data. However, this is rarely, if ever, the case. Most companies start disclosing ending customer count data (END_q) some number of quarters into the company's operations (call it q_A), then begin disclosing customer ADD and LOSS data in a later quarter (call it q_B, where $q_B \geq q_A$). In such cases, our parameters are those that minimize the following sum-of-squared errors:

$$
\begin{aligned}
\text{SSE}_{MISS} = {} & \left(\text{END}_{q_A} - \widehat{\text{END}}_{q_Q}\right)^2 \\
& + \sum_{q=q_A+1}^{q_B-1} \left[\left(\text{END}_q - \text{END}_{q-1}\right) - \left(\widehat{\text{END}}_q - \widehat{\text{END}}_{q-1}\right) \right]^2 \\
& + \sum_{q=q_B}^{Q} \left\{ \left(\text{ADD}_q - \widehat{\text{ADD}}_q\right)^2 + \left(\text{LOSS}_q - \widehat{\text{LOSS}}_q\right)^2 \right\} \\
& + \left(\text{END}_Q - \widehat{\text{END}}_Q\right)^2, \quad (3.17)
\end{aligned}
$$

This accounts for the shortened contiguous customer addition and loss data, and

---

[5]This assumes the firm started operation at the beginning of a reporting quarter, as discussed in Section 3.2. If this is not the case, minor modifications to Equations 3.16 and 3.17 are needed.

the missingness present at the beginning of the time series[6].

3.3.4   Average Revenue Per User

We make use of a simple time-series model to capture (and project) the evolution of ARPU(m). Assuming linear growth in ARPU,[7] we can use a simple time-trend regression:

$$\text{ARPU}(m) \;\; = \;\; \beta_0 + \beta_1 m + \epsilon(m), \quad \epsilon(m) \sim \mathcal{N}(0, \sigma^2). \tag{3.18}$$

The mean of many economic and financial time series is non-stationary (Zivot & Wang 2007). When this is so, the fitted residuals of the regression given in Equation 3.18 will fail tests for non-stationarity, the most popular of which is the Augmented Dickey-Fuller test (Dickey & Fuller 1979; Elliott et al. 1996). If this is the case, the parameter estimates from Equation 3.18 are invalid, and we should instead use an ARIMA(0,1,0) model:

$$\text{ARPU}(m) \;\; = \;\; \text{ARPU}(m-1) + \beta_0 + \epsilon(m), \quad \epsilon(m) \sim \mathcal{N}(0, \sigma^2). \tag{3.19}$$

The parameters of either model are estimated using maximum likelihood[8].

ARPU(m) is a standard internally reported measure for a subscription-based firm. Some firms do report quarterly ARPU publicly, but this data cannot be used in general because there are no well-accepted standards for calculating it. As DISH stated in

---

[6]We assume that the firm's decision to disclose is not strategic, so that data are missing at random (Little & Rubin 2014). As noted in the previous chapter, we performed a large scale disclosure analysis which suggests that this assumption is consistent with observed disclosure data.

[7]If we assume that ARPU grows at a constant growth rate over time, we would use $\log\big(\text{ARPU}(m)\big) = \beta_0 + \beta_1 m + \epsilon(m)$.

[8]As in GLS and SSW, we assume that spend and churn are uncorrelated. The data are too limited to identify such a correlation, and the lack of heterogeneity in spend limits the practical benefit of allowing for it.

its 2014 annual filing, "We are not aware of any uniform standards for calculating ARPU and believe presentations of ARPU may not be calculated consistently by other companies in the same or similar businesses." As there is no standard definition of ARPU, different firms may have different definitions for it, picking and choosing what sources of revenue to include in the numerator. As such, the reported ARPU numbers may not be representative of all revenue derived from the customer base.

Revenue numbers are more reliable. However, they are only provided quarterly, so we need to impute monthly revenues. For $m \in \{3q - 2, 3q - 1, 3q\}$, the revenue in month m is equal to the customer-weighted share of total revenue in quarter q:

$$\mathrm{R}(m) = \frac{\mathrm{C}(., m-1) + \mathrm{C}(., m)}{\mathrm{C}(., 3q-3) + 2\mathrm{C}(., 3q-2) + 2\mathrm{C}(., 3q-1) + \mathrm{C}(., 3q)} \times \mathrm{REV}_q. \quad (3.20)$$

(Strictly speaking, we are using $\widehat{\mathrm{C}}(\cdot, \cdot)$, computed using the estimated parameters of the acquisition and retention processes.) Having imputed $\mathrm{R}(m)$, our estimates of $\mathrm{ARPU}(m)$ follow from Equation 3.3.

3.3.5   Summary of Valuation Procedure

Taking a step back, recall that our goal since Section 3.2 has been to generate long-term projections of $\mathrm{R}(m)$ and $\mathrm{A}(m)$, from which we can compute estimates of period-by-period FCF and then the value of the firm. We now outline the process by which we compute these revenue numbers using the models described above.

(a) We estimate the parameters of the acquisition and retention processes (Section 3.3.3). Assuming the firm has been in operation for Q quarters, we then compute our estimate of the $3Q \times 3Q$ matrix $\mathrm{C}(\cdot, \cdot)$, the diagonal of which is our estimate of the number of customers acquired each month over this time period, and the rows of which are estimates of the number of customers in each

cohort that survive each of the subsequent months.

(b) As outlined in Section 3.3.4, we use $\widehat{C(\cdot, \cdot)}$ and the reported quarterly revenue numbers to impute the corresponding monthly revenue numbers, from which we estimate the parameters of our model for average revenue per user.

(c) In order to project $C(\cdot, \cdot)$ into the future, we need estimates of $POP(m)$ over the time horizon of interest. In some cases, such data may be available from a secondary source. In the absence of such a source, we can use a simple model for forecasting $POP(m)$. For example, we could use the long-run compound growth rate in $POP(m)$ and assume it holds going into the future.

(d) Having projected $C(\cdot, \cdot)$ far into the future (i.e., to a point in time where the present value of any associated cash flows is effectively zero), we compute the column totals $C(., m)$ to give us the total number of customers for each month.

(e) We compute expected $ARPU(m)$ across this time horizon using Equation 3.18 or 3.19. Rearranging Equation 3.3, it follows that

$$R(m) = ARPU(m) \times \frac{C(., m-1) + C(., m)}{2} .\qquad (3.21)$$

With the revenues estimated, the remainder of our valuation model is for all intents and purposes the same as what a financial professional would do when building a DCF valuation model. In the next section, we bring this valuation model to life from start to finish using data for two public companies.

3.4 Empirical Analyses

We first apply our model of customer behavior to data from DISH Network Corporation (Nasdaq: DISH), a large pay-TV service provider. We estimate the parameters of the model, evaluate its in-sample fit, evaluate the predictive validity of the model by performing rolling two-year-ahead forecasts over all possible calibration periods, and compare its performance to that of the models of customer behavior proposed by GLS and SSW. After demonstrating the validity of the model, we then use its revenue projections (along with the associated estimates of customer acquisition) to arrive at our estimate of the value of DISH's shareholder's equity. Next, to further establish the robustness of our proposed model, we apply it to a second publicly-traded company, Sirius XM Holdings (Nasdaq: SIRI), a satellite radio service provider. We conclude by exploring some other insights into customer behavior that can be derived using the model.

3.4.1 DISH Network

DISH commenced operations in March 1996 (DISH Network (2015)),[9] and end-of-period customer counts were first disclosed that quarter. However the gross customer acquisition data are left censored—the first time that gross customer additions were disclosed was seven quarters later, in Q1 1998 (i.e., with reference to Equation 3.17, $q_A = 0$ while $q_B = 7$). All historical customer data (ADD_q, LOSS_q, END_q and REV_q) come from DISH's quarterly and annual reports, Forms 10-Q and 10-K, respectively. We model this customer data up to and including Q1 2015 (i.e., $Q = 77$). The vast

---

[9]While DISH Network was technically incorporated in 1980, the relevant starting date for our analysis is when DISH actually commenced commercial operations and could thus begin acquiring customers.

majority of DISH's revenues come from its subscriber base.[10]

We use the same four time-varying covariates in our models of the acquisition and retention processes: three quarterly dummy variables to capture seasonal fluctuations in the propensity to sign-up to and churn from the service, and a "Great Recession" dummy variable to account for the diminished propensity to sign-up and the increased propensity to churn during that recession.[11] Given the nature of the DISH's service offering, our unit of population is the household. We use data on US household growth provided in the US Census Bureau's CPS/HVS data tables.

3.4.1.1 Parameter Estimates and Evaluation of Fit

We first estimate the parameters of the acquisition and retention models using all the available data. The parameters are reported in Table 3.1; the associated model SSE is 310,821. The story told by these parameters is consistent with what DISH has disclosed in its public filings. With reference to the coefficients of the quarterly dummies, consider DISH's comments on the seasonality of its operations in its 2015 annual report: "Historically, the first half of the year generally produces fewer gross new subscriber activations than the second half of the year, as is typical in the pay-TV industry. In addition, the first and fourth quarters generally produce a lower churn rate than the second and third quarters."

The negative effect of the 2008 recession on DISH's financials is unmistakeable; its effect upon acquisition and retention propensities was greater than all of the respective seasonal terms. The coefficient in the acquisition model is negative because customers have a lower propensity to acquire services during a recession, while the coefficient in

---

[10]In Q1 2015, 0.9% of DISH's revenue were derived from equipment sales, which are not core to the business and have not been growing over time. DISH has made investments in wireless spectrum over the past three years — wireless spectrum is a non-operating asset — but earns no revenue from it, and the core operations of the business do not depend upon it.

[11]The "Great Recession" started December 2007 and ended June 2009 (http://www.nber.org/cycles.html).

Table 3.1: Parameter Estimates: DISH Network

	Acquisition		Retention	
	Param.	Std. Err.	Param.	Std. Err.
r	11.440	5.123	1.648	0.232
α	29861.183	12406.027	81.196	4.897
c	2.001	.011	1.423	0.056
β_{Q1}	$-.052$	.008	$-.079$	.007
β_{Q2}	$-.057$	.007	.036	.009
β_{Q3}	.036	.008	.107	.008
β_{Rec}	$-.099$	.011	.129	.010
π_{NA}	.525	.006		

the retention model is positive because customers have a higher propensity to churn during a recession.

In Figure 3.2, we plot model estimates for gross customer additions, losses, and end-of-period total customer counts against what we actually observed. (The grey area indicates the duration of the Great Recession.) We must backcast gross customer additions and losses because DISH did not disclose ADD and LOSS data prior to Q1 1998. Backcasts are predictions of in-sample metrics during periods which we do not have observed data for. Our resulting fits are good; we see a clear seasonal pattern within acquisitions and losses, and lower acquisitions and higher losses during the recession of 2008. DISH appears to be past the point of peak adoption, a sentiment echoed by DISH CEO Charlie Ergen in DISH's Q1 2015 conference call: "My general sense is that the linear pay television business probably peaked a couple of years ago and that it's in a very slight decline."

Average revenue per user is modeled as per Section 3.3.4. We assume linear growth, which is consistent with comments made in DISH's annual financial reports. First fitting the simple time-trend regression given in Equation 3.18, we find that the model residuals fail the Augmented Dickey Fuller unit root test ($t = -2.6, p = .31$). We therefore model ARPU using the ARIMA(0,1,0) model specified in Equation 3.19,

Figure 3.2: DISH Network: Customer Additions, Losses and End-of-Period Customer Counts (Recession in Grey)
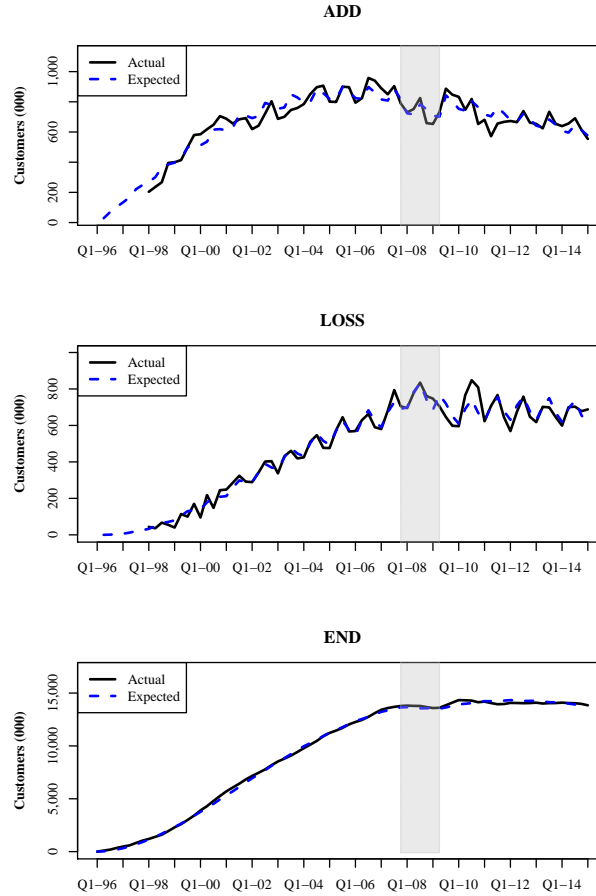
with $\hat{\beta}_0 = 0.246$ (s.e. .091) and an associated R^2 of 93%.

3.4.1.2   Predictive Validation and Comparison

While the analysis presented above shows that our in-sample fit is very good, it does not give us any real insight into the predictive validity of our model or how our model's predictions compare to those of alternative models (i.e., those presented in GLS and SSW). These are important questions, as the quality of our estimate of firm value is a direct function of the quality of the projections of revenue (and customer

acquisitions, or some combination thereof) coming from our model.

To shed light on these questions, we perform a rolling validation in which we vary the model calibration period and compare the model predictions of ADD, LOSS, and END with the actual numbers reported by DISH. Letting $Q = 10, 11, \ldots, 69$ (corresponding to all possible calibration periods ending from Q2 1998 to Q1 2013), we calibrate our model upon all data up to and including quarter Q, and then predict ADD, LOSS, and END for the next two years (i.e., ADD_{Q+q^*}, LOSS_{Q+q^*}, and END_{Q+q^*} for $q^* = 1, 2, \ldots, 8$). Because of missing data, only 3 quarters of ADD and LOSS data are available when $Q = 10$, making it a reasonable starting point to the rolling analysis. As a result, our evaluation of model performance is based on predictions made using 60 different calibration periods.

In Figure 3.3, we plot all resulting predictions over all calibration periods for ADD (first column), LOSS (second column), and END (third column) using GLS (first row), SSW (second row), and our proposed model (third row). While the general patterns of over- and under-estimation are similar for SSW and GLS, we see that the overall predictive validity of SSW is generally better than that of GLS. GLS underestimates future ADD, LOSS, and END figures, often severely so. This is primarily because the logistic internal-influence model for ADD and constant retention rate model for LOSS are unable to capture the underlying dynamics in customer behavior over time. Since SSW models END (rather than ADD) using the logistic internal-influence model, their resulting predictions for END are generally quite well-behaved and well-calibrated. Both methods have the most difficulty forecasting ADD, as evidenced by the large deviations between the predictions in grey and the actual data in black. This is important as ADD is an important input for these models' respective valuation models.

In contrast, our proposed model forecasts ADD, LOSS, and END very accurately,

Figure 3.3: DISH Network: Rolling Two-Year Predictive Validation Plots
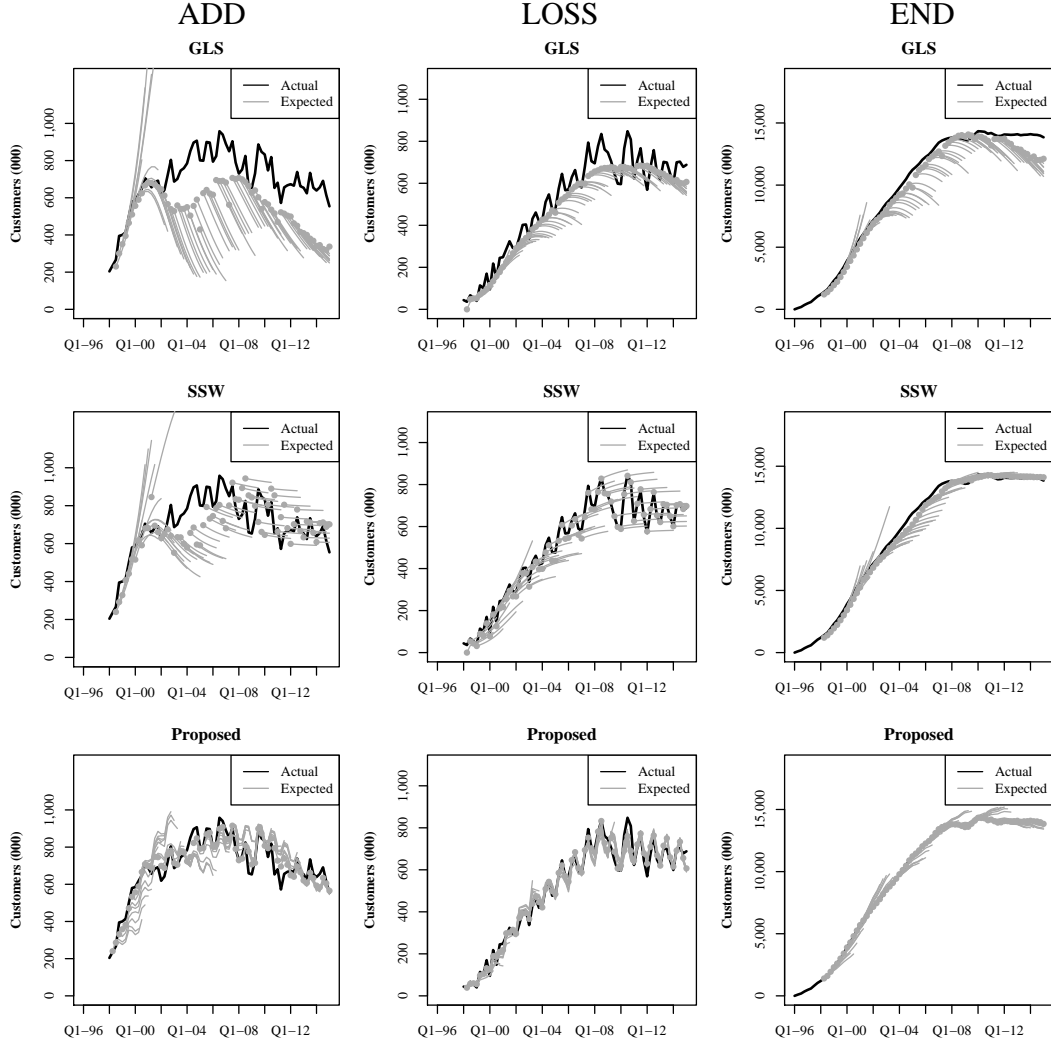
as evidenced by the tight correspondence between the grey and black lines in the bottom row of Figure 3.3. In contrast to the forecasts associated with the GLS and SSW models, this correspondence remains tight even for short calibration periods, which is further proof of the robustness of the model's predictions.

To summarize the relative performance of these three models, we compute the absolute percentage error in the ADD, LOSS, and END forecasts for each of the (rolling) eight holdout quarters and take the average across the 60 different calibration

periods. The resulting MAPE numbers are reported in Table 3.2. We see that the MAPE figures associated with the SSW model are generally half those of the GLS model, while the MAPE figures for our proposed method are generally one third smaller than those of SSW.

Table 3.2: DISH: MAPE of Predictions of ADD, LOSS, and END by Forecasting Horizon

Quantity	Horizon	GLS	SSW	Proposed
ADD	Q+1	26.0	14.7	8.2
	Q+2	29.2	16.4	9.5
	Q+3	32.5	16.4	10.4
	Q+4	36.5	16.3	11.3
	Q+5	41.1	17.7	13.2
	Q+6	45.5	19.8	14.5
	Q+7	50.6	21.5	15.6
	Q+8	55.8	22.9	16.0
LOSS	Q+1	26.0	14.7	8.2
	Q+2	29.2	16.4	9.5
	Q+3	32.5	16.4	10.4
	Q+4	36.5	16.3	11.3
	Q+5	41.1	17.7	13.2
	Q+6	45.5	19.8	14.5
	Q+7	50.6	21.5	15.6
	Q+8	55.8	22.9	16.0
END	Q+1	26.0	14.7	8.2
	Q+2	29.2	16.4	9.5
	Q+3	32.5	16.4	10.4
	Q+4	36.5	16.3	11.3
	Q+5	41.1	17.7	13.2
	Q+6	45.5	19.8	14.5
	Q+7	50.6	21.5	15.6
	Q+8	55.8	22.9	16.0

These conclusions are not affected by the fact that our model incorporates the effects of covariates while the model of GLS and SSW do not. We created variants of the GLS and SSW models that include the quarterly seasonality and Great Recession effects (through a logit formulation for retention, and a proportional hazards specifi-

cation for acquisition (GLS) and ending customers (SSW)), and do not observe any significant changes to our conclusions regarding the relative performance of the three models.

3.4.1.3 Valuation Results

Having demonstrated the performance of our proposed model, we now to turn to the primary reason why it was developed in the first place: computing an estimate of the value of the firm.

We first project revenues (Section 3.3.5) far enough into the future so that all subsequent profits/losses have no effect on our valuation; we choose 50 years. We forecast that POP will continue to grow at a per-month growth rate of 0.06% into the future; this is equal to the historical monthly US household growth rate over the period from March 1996 to March 2015.

Our revenue projections drive detailed financial projections that are used to estimate future free cash flows, the weighted average cost of capital, the value of non-operating assets, and net debt. We then add the value of the operating assets to the non-operating assets and subtract the net debt to arrive at our best estimate of shareholder value using Equation 2.1 — see Table 3.3.

We estimate a stock price of $64.62 based on Q1 2015 results, which were disclosed on May 11, 2015. The end-of-day stock price that day was $66.38, implying that we are within 3% of the then-current stock price. Holding all else constant, the DISH Network stock price estimates computed using the GLS and SSW models for customer acquisition and retention were $48.84 and $63.72, respectively.

Our valuation and the corresponding implied stock price are point estimates. So as to get a sense of the uncertainty in these estimates, we undertake the following sensitivity analysis. First, holding the parameters of the retention and ARPU pro-

Table 3.3: DISH Valuation Summary (End of Q1 2015)

Value of Operating Assets	$15.7B
Non-Operating Assets - Net Debt	$14.1B
Shareholder Value	$29.9B
Shares Outstanding	462.1MM
Implied Stock Price	$64.62
Actual Stock Price	$66.38
Over(under)-estimation	(2.7%)

cesses constant, we draw a new set of parameter values for the acquisition process model via bootstrap resampling of the model residuals (Efron and Tibshirani 1993, Chapter 9). Given this set of parameters, we compute the resulting revenue numbers, the corresponding estimate of the value of the firm and the implied stock price. We do this for 500 draws and compute a 95% interval for our implied stock price. We repeat this for the retention process (holding the parameters of the acquisition and ARPU processes constant) and the ARPU process (holding the parameters of the acquisition and retention processes constant). The interval associated with the acquisition process is [$64.48, $64.77] $(+/-0.2\%)$. The equivalent intervals for the retention and ARPU processes are [$62.47, $66.78] $(+/-3.4\%)$ and [$62.76, $66.49] $(+/-3.0\%)$, respectively. This suggests that it would be most beneficial to investors if DISH were to provide more or better data regarding customer retention (e.g., by disclosing LOSS figures monthly instead of quarterly).

3.4.2 Sirius XM

To test the robustness of our framework, we repeat our valuation exercise for a second company, Sirius XM, which is a broadcasting company that provides satellite radio services in the United States. Sirius XM is a good complementary example to that of DISH for a number of reasons:

(a) Sirius XM is a relatively high-growth business, while DISH is a mature business. ADD, LOSS, and END are all past their peak for DISH (Figure 3.2), while they are increasing for Sirius XM.

(b) Sirius XM suffers from more severe missingness than DISH. Sirius XM was formed by the merger of Sirius Satellite and XM Satellite, which began commercial operations in February 2002 and November 2001, respectively. Neither Sirius nor XM disclosed ADD, LOSS, or END data for paying customers. It was not until after the merger that these data were first publicly disclosed (Q3 2008). As a result, almost half of Sirius XM's customer data are missing.

(c) Sirius XM is a high fixed-cost business because its satellite radios are pre-installed in most new vehicles, while DISH Network is a high variable-cost business. Most of Sirius XM's operating expenses, net of subscriber acquisition costs (SAC), are fixed in nature, while most of DISH's operating expenses are variable. All else being equal, this substantially increases the marginal profitability of new Sirius XM users.

(d) Sirius XM has a very different customer base and customer profile than DISH. Sirius XM has a larger number of customers, each of whom generates less revenue but is much cheaper to acquire.

(e) Sirius XM sells almost entirely into cars, whereas DISH sells almost entirely into homes. All else being equal, this makes Sirius XM a more cyclical business than DISH.

See Table 3.4 for a comparison of the two companies on the basis of some basic measures.

Despite these differences, we proceed with virtually the same model. The main change is that the population unit for Sirius XM is cars (as opposed to households

58

Table 3.4: Comparison of DISH and Sirius XM (at point of valuation)

	DISH	Sirius XM
Total Paying Customers	13.8MM	22.9MM
Monthly ARPU	$88.72	$16.72
Weighted Average Cost of Capital	7.2%	6.9%
SAC / Customer	$716.46	$82.06
ARPU Growth Per Year	$2.95	$0.49

for DISH). The market size for Sirius XM is equal to the number of vehicles on the road, as provided by the Bureau of Transportation Statistics. Correspondingly, we use vehicle sales, as defined/provided by Federal Reserve Bank of St. Louis, as our macroeconomic covariate. We denote the coefficient associated with the vehicle sales covariate by β_{VS}.

The parameter estimates of the acquisition and retention process models are reported in Table 3.5; the associated model SSE is 146,799.[12] Once again, we assume linear growth when modeling average revenue per user. Fitting a simple time-trend regression (Equation 3.18) to the data, we find that the residuals do not fail the Augmented Dickey Fuller unit root test (test statistic: $t = -3.56, p = .04$); the associated parameter estimates are $\hat{\beta}_0 = 9.643$ (s.e. .218) and $\hat{\beta}_1 = 0.041$ (s.e. .002), with $R^2 = 88\%$.

In Figure 3.4, we plot model estimates for ADD, LOSS, and END against what we actually observed. We overlay a set of two-year rolling predictions corresponding to all possible calibration periods ending from Q3 2010 to Q1 2013, as we had done for DISH in Section 3.4.1.2. As was the case with DISH, the in-sample and out-of-sample fits for Sirius XM in terms of all three customer metrics are good.

As with DISH, we project revenues 50 years into the future. We project both future vehicles on the road and vehicle sales assuming monthly growth rates are equal to

---

[12]Unlike DISH, the Weibull-Gamma baseline retention process for Sirius XM is not significantly different from a Weibull baseline. We retain the more general formulation for consistency with DISH.

Table 3.5: Parameter Estimates: Sirius XM

	Acquisition		Retention	
	Param.	Std. Err.	Param.	Std. Err.
r	0.208	.239	153.206	35.129
α	7983.761	9213.648	2671.833	602.632
c	2.228	.154	1.066	.085
β_{Q1}	$-.106$	.012	.029	.014
β_{Q2}	$-.049$	.014	$-.039$	.016
β_{Q3}	$-.047$	.013	.003	.015
β_{VS}	.077	.003	$-.013$	.003
π_{NA}	.011	.075		

their respective historical cumulative average growth rates from 1980 until 2015. We perform a detailed margin and cash flow analysis to turn the revenue projections into monthly free cash flow projections. The resulting valuation is presented in Table 3.6. We estimate Sirius XM's operating assets to be worth $27.1B. After adding non-operating assets (Sirius XM has approximately $1.1B in net operating loss carry-forwards) and subtracting net debt, we estimate shareholder value to be $23.4B using Equation 2.1. This implies a stock price of $4.24 based on Sirius XM's Q1 2015 results, which were released on April 28, 2015. The end-of-day stock price that day was $3.90. Holding all else constant, the Sirius XM stock price estimates computed using GLS's and SSW's models for customer acquisition and retention were $0.41 and $6.55, respectively.

Table 3.6: Sirius XM Valuation Summary (End of Q1 2015)

Value of Operating Assets	$27.1B
Non-Operating Assets - Net Debt	$-$3.7B
Shareholder Value	$23.4B
Shares Outstanding	5513.7MM
Implied Stock Price	$4.24
Actual Stock Price	$3.90
Over(under)-estimation	7.4%

Figure 3.4: Sirius XM: Customer Additions, Losses and End-of-Period Customer Counts
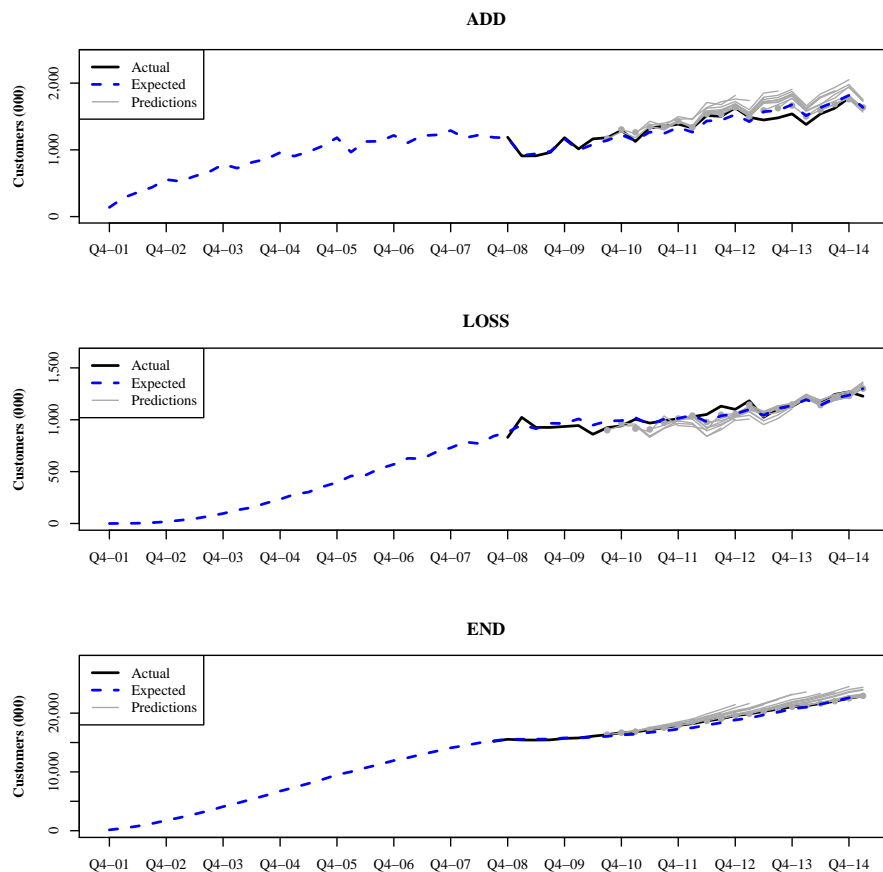
3.4.3 Additional Insights

Confident that our model provides accurate valuation estimates, we return to DISH to study other insights that we are able to draw from the model beyond stock price estimates. We look at the remaining/residual lifetime and lifetime value of DISH customers as a function of the length of their relationship (i.e., tenure) with the firm. We then decompose DISH's current customer equity (CCE) by tenure. While these seem like fairly ordinary applications of a customer-level model, it is important to keep in mind that we are doing these analyses with no customer-level data; all we

have are the aggregated summaries that companies disclose to the public.

3.4.3.1   Comparison of Residual Value by Tenure

Let us consider a DISH customer acquired at the end of Q1 2015 whom we call Recent Robin, and another DISH customer acquired 10 years earlier at the end of Q1 2005 whom we call Longtime Larry. One quantity of managerial interest is the expected remaining (or residual) lifetime of Recent Robin and how it compares to the expected residual lifetime of Longtime Larry. The GLS and SSW models both assume that all customers are equal and thus Recent Robin and Longtime Larry would be expected to share the same expected future lifetime. However, we intuitively expect that Longtime Larry is likely to remain a customer for a longer period of time because his long history with DISH thus far suggests that he has a lower churn propensity. By definition, the expected residual lifetime of a customer acquired in month m who is still a customer in month M equals

$$\sum_{i=0}^{\infty} \frac{\mathrm{S}(M + i - m|m, \boldsymbol{X}_R(m, M + i); \mathrm{r}_R, \alpha_R, \mathrm{c}_R, \boldsymbol{\beta}_R)}{\mathrm{S}(M - m|m, \boldsymbol{X}_R(m, M); \mathrm{r}_R, \alpha_R, \mathrm{c}_R, \boldsymbol{\beta}_R)} . \tag{3.22}$$

(See the Appendix for details of how we perform this calculation.) The expected residual lifetime of Recent Robin and Longtime Larry are 5.5 years and 9.4 years, respectively. This difference is in line with our intuition.

Investors should be interested in the expected lifetime of customers. Longer expected customer lifetimes imply more stable future cash flows, all else being equal, because future cash flows are less reliant on the acquisition of new customers. At DISH, we see that not only do longer tenured customers have longer residual lifetimes, but also that all customers live for a relatively long time, which should be heartening to investors. Reducing investors' perceived risk of future cash flows reduces the cost

of capital, raising firm valuation.

Another quantity of interest is the residual lifetime value (RLV) of customers.[13] Calculating this using nothing but the information provided in a firm's financial statements requires careful consideration of what expenses are fixed versus variable, and a proper handling of subscriber acquisition costs. (See the Appendix for details of how we perform this calculation.) We estimate that the (pre-tax) RLV associated with Recent Robin to be $1,426, excluding average initial acquisition costs of $854, while Longtime Larry is worth $1,932. While it is not possible to provide predictive validation of these customer insights because of the aggregated nature of the data, the predictive valuation analysis that we performed provides general validity for these results.

This information is useful to many stakeholders:

- Investors may track CLV relative to SAC per customer, viewing these metrics as financial barometers of customer health. Unfavorable trends in these figures (as has been evident at DISH, for example) could be indicative of decreasing customer (and thus firm) profitability.

- Competitors, comparable companies, and investors will be interested in the absolute level of CLV and RLV for benchmarking purposes. If a competitor estimates its own CLV to be less than DISH's, there may be opportunities to "close the gap," identifying what it could be that is causing the gap in average customer profitability. Investors may ask the same question and demand that changes be made to improve CLV and RLV.

While the preceding analysis has focused on *expected* residual lifetime and lifetime

---

[13]We make the distinction between CLV, which we reserve for as-yet-to-be acquired customers, and RLV, which applies to existing customers. Subject to minor accounting issues, these two quantities are equal when the we have constant retention rates (i.e., there is no heterogeneity across customers and/or duration dependence within customer). However, this is rarely the case and it is therefore important to make this distinction (Fader & Hardie 2010).

value, we can examine the distribution of these quantities across all possible Recent Robins and Longtime Larrys. In Figure 3.5 we provide a histogram representing 1MM samples from their respective RLV distributions. This provides us with additional information regarding the riskiness of future cash flows associated with new and existing customers. For example, we estimate that there is a 41% chance that the company will incur a loss on a Recent Robin (i.e., 41% of Recent Robin's RLV samples (Figure 3.5) lie to the left of $854, the SAC per customer for DISH). We infer a long right tail to Longtime Larry's pre-tax RLV — this drives up Longtime Larry's expected pre-tax RLV, but also implies a much higher variance about that expectation. Longtime Larry is more valuable but also has more variable cash flows (McCarthy et al. (2016)).

Figure 3.5: Histogram of 1MM Sampled RLV's — Recent Robin and Longtime Larry
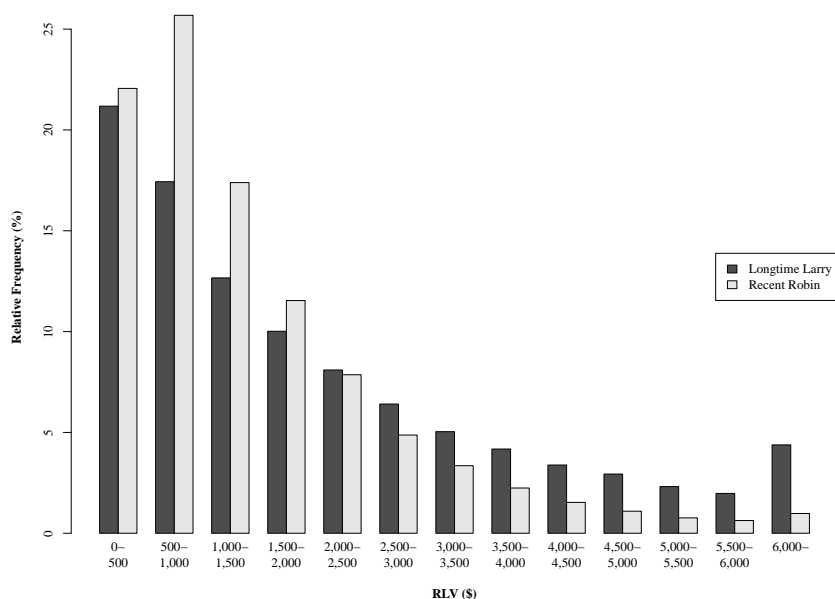
3.4.3.2 Customer-Base Decomposition

The raw data available from virtually any public source reveals nothing about the tenure of existing customers or how these "lifetimes" vary across the customer base. As suggested by the examples of Recent Robin and Longtime Larry, this can be important information to outside investors. Fortunately, as just demonstrated, our proposed model makes it easy for analysts to infer these lifetimes. We can go further and segment customers on this basis.

The proportion of currently active customers (i.e., active at the end quarter Q) who were born in month m is equal to

$$\frac{\widehat{C}(m, 3Q)}{\sum_{i=1}^{3Q} \widehat{C}(i, 3Q)} . \tag{3.23}$$

While knowing the count of customers within each segment is helpful, the value of those customers is probably of greater interest to investors and managers. Recall that the sum of RLV across all the firm's current customers is called current customer equity (CCE). It follows that the proportion of total CCE, as at the end of quarter Q, coming from customers who were born in month m is the RLV-weighted analog of Equation 3.23:

$$\frac{\widehat{C}(m, 3Q) E(\text{RLV}_{m,3Q})}{\sum_{i=1}^{3Q} \widehat{C}(i, 3Q) E(\text{RLV}_{i,3Q})} , \tag{3.24}$$

where $\text{RLV}_{m,3Q}$ is the residual lifetime value of a customer acquired in month m who is still active in month $3Q$. The resulting decomposition of DISH's customer base is presented in Table 3.7. We estimate, for example, that approximately one-eighth of DISH's customer base is comprised of highly loyal/inertial customers who have been DISH subscribers for 10+ years. We also infer that longer-lived segments comprise proportionally more of the total value of the customer base because they are inferred

to have higher residual lifetime values, as is evident from our comparison of Recent Robin and Longtime Larry above.

Table 3.7: Decomposition of Current Customer Equity (End of Q1 2015) by Tenure

Tenure (years)	% Customer base	% CCE
0–2	31	28
2–5	31	29
5–10	25	27
10+	12	16

This decomposition, and other granular inferences that can be drawn from our model, can provide useful insights for investors. In some sense, the overall corporate valuation shown earlier isn't necessarily very insightful by itself; it merely captures the "voice" of the financial market. It could be argued that the real value of our proposed approach is the ability to go beyond the macro valuation estimate to offer useful, operational diagnostics to better understand where that value is coming from, and what it might mean to the firm, its competitors, suppliers, investors, and possibly even public policy makers. In the case of DISH, the considerable amount of value coming from long-lived customers is indicative of a very mature business, and implies that the valuation of the business as a whole will be much more dependent on and sensitive to changes in the company's ability to retain existing companies, rather than acquire new ones. In contrast, Sirius XM's valuation is far more dependent on the the firm's ability to acquire new users and to earn a high rate of return on the firm's investment in those new users. This is important information for investors and managers alike.

3.5 Discussion

As noted at the outset of the essay, our objective has been to create an accurate model for customer acquisition and retention (which can be estimated using data publicly reported by firms with a subscription-based business model), and embed it within a standard financial framework for corporate valuation. Looking beyond the methods developed here, we hope this essay will serve as a call to action for firms, analysts, and investors to perform these kinds of analyses on a more regular and rigorous basis. We have provided several use cases for the insights that can be derived from our analyses, including but not limited to comparing CLV across comparable/competing firms, performing customer value segmentation, and providing investors with improved forward-looking sales visibility. All of this is possible because we have performed our valuation using a flexible, general-purpose model of customer behavior in contractual/subscription settings.

As noted in the previous chapter, while our particular model is suited to third parties analyzing publicly traded companies using their public disclosures, we contend that this same exercise can—and should — be undertaken internally as well. Firms can easily implement an equivalent version of this model using internal company data, enhancing its overall validity. While the estimation procedures differ slightly (given access to more granular data), the models for customer acquisition, retention, and spending, along with the proposed valuation framework, would remain essentially the same. Measuring and tracking CLV and RLV can improve the ROI of a company's acquisition and retention spending, and our valuation framework gives company executives the ability to estimate how much value this ROI improvement has created for the overall value of the firm. This provides executives with an important key performance indicator to hold themselves and their marketing managers accountable to.

While our valuation model for subscription-based firms is more flexible than previously published customer-based corporate valuation models (e.g., in terms of the dynamics that it can accommodate), it has nevertheless remained parametrically parsimonious because the available data are limited — and will likely stay that way for the foreseeable future. For example, it is highly unlikely that firms will begin to disclose the kinds of data required to properly account for other sources of customer value, such as the referral value of a customer (Kemper 2009; Kumar et al. 2010), the impact of social media (Luo et al. 2013; Yu et al. 2013), customer satisfaction (Anderson et al. 2004; Homburg et al. 2005; Luo & Bhattacharya 2006), or heterogeneity in the spend per customer (McCarthy et al. 2016). At the same time, indirect proxies for these factors may be obtainable in some cases through external data sources for a small subset of companies.

It may seem tempting to add in other "bells and whistles" to further enrich the model specification used here. This may be particularly appealing in this subscription-based setting, because the underlying customer model is simpler than in a non-subscription-based setting. We should be open to such possibilities but are nevertheless cautious about our ability to do so. For instance, it may be the case that individual-level acquisition and retention propensities are correlated (i.e., customers who take longer to acquire may have a lower propensity to churn once they have been acquired or vice versa — see Schweidel et al. 2008a — but our ability to empirically identify such a correlation is very limited, increasing the risk that we over-burden the limited data we have available. Many other theoretically reasonable extensions (e.g., allowing for cross-cohort effects (Gopalakrishnan et al. 2016), or specifying a more complex market potential model) will likely suffer from similar issues. When data are highly aggregated, even identifying heterogeneity can be challenging (Bodapati & Gupta (2004)). Model parsimony is a good thing.

Beyond the methodological improvements, our valuation framework could provide perspective to the ongoing discussion among marketing scholars regarding the accounting of customer equity and advertising spending. Consistent with Srinivasan (2015), the vast majority of DISH's SAC is expensed and not capitalized (82% in Q1 2015) — the primary component of SAC that is capitalized is spending to purchase satellite receivers, which are then owned by DISH and depreciated over a useful life of approximately 4 years. In contrast, just acquired customers have, on average, a longer "useful life" of 5.5 years (Section 3.4.3), and yet are not considered assets (Wiesel et al. 2008). As a result, subscriber acquisition activities create costs that are incurred immediately but whose benefits are received in the future; as such, the income statement is not reflective of the underlying economic condition of the business. It is no surprise, then, that DISH was generally unprofitable earlier in its history, and profitable in recent years.

As companies increasingly recognize the importance and merit of customer-centric business strategies (Fader 2012), and in turn disclose customer data on a more regular and thorough basis, there will be a growing opportunity for marketing scholars to study the behavior of large, publicly traded subscription-based companies through their customer data in conjunction with their financial statements. We hope that this essay lays a sound foundation for how future analyses will incorporate, and shed further light on, subscription-based company valuation. In the next essay, we turn our attention to non-subscription-based company valuation.

4 Valuing Non-Contractual Firms with Common Customer Metrics

4.1 Introduction

As we discussed in essay two, performing CBCV for "contractual" (i.e., subscription-based) firms entails forecasting three main quantities – (1) future customer acquisitions, (2) how long acquired customers will remain with the firm before they churn, and (3) the monetary value associated with customers, on average, while those customers are active. Prior work has shown that two customer metrics are needed to perform CBCV for contractual firms – total customers acquired and total customers churned each period[1] (Gupta et al. 2004; Libai et al. 2009; Schulze et al. 2012; Wiesel et al. 2008). However, it is unclear what valuation framework is most suitable for *non-contractual* firms (e.g., retail, travel/hospitality, media/entertainment, gaming, and more), what metrics external stakeholders need as inputs, and what estimation procedure external stakeholders can use to infer the parameters of the model given a particular collection of metrics. It is also unclear what collection of metrics is the most informative for external stakeholders if the firm is only willing to publicly disclose a

---

[1]Instead of disclosing the number of customers lost during the period, the firm may equivalently disclose the per-period churn rate, or the total number of customers that are active as of the end of each period.

limited number of them. The contribution of this essay is to address these issues, proposing a novel methodology to estimate an integrated individual-level model of customer acquisition, repeat purchase, and spend behavior using a small collection of commonly disclosed non-contractual customer metrics.

Recall from Section 2.5 that there are two main aspects of this problem which have limited progress to date. These aspects can be categorized broadly as model challenges and data challenges:

(a) *Model challenges*: Non-contractual firms are more difficult to value than contractual firms because attrition is unobserved in non-contractual settings, complicating the model required to probabilistically infer it. Even if a non-contractual customer were known to be alive, there is often substantial heterogeneity in both the propensity to purchase and the magnitude of spend given purchase across customers and over time. In contrast, there is no heterogeneity in the purchase rate and substantially less heterogeneity in the amount spent while contractual customers are alive. A principled non-contractual customer base model should account for heterogeneity in these processes.

(b) *Data challenges:* Complications related to the limited nature of available data are more acute in a non-contractual setting. If granular, individual-level data were available, there are a number of well-established, well-validated models specifically suited for non-contractual customer behavior (Fader et al. 2010; Platzer & Reutterer 2016; Schmittlein et al. 1987). In the absence of such data, customer behavior must be estimated using common customer metrics disclosed quarterly (e.g., active customers, or the repeat rate). While estimating individual-level behaviors using aggregate information is a well-studied problem (Albuquerque & Bronnenberg 2009; Chen & Yang 2007; Feit et al. (2013); Musalem et al. 2008; Musalem et al. 2009), our problem setting differs from

71

previous works in three ways. First, the data in our setting are a possibly incomplete, wide-ranging collection of cross-sectional summaries, and not cross-sectional market share data alone, or market share data augmented by customer purchase frequencies marginalized over time. Second, our goal is not only to estimate a model given a fixed set of disclosures, but also to recommend what set of metrics offers the highest predictive performance for the smallest number of disclosures. Third, our data often encompasses millions of prospects and customers over years or even decades.

In the customer-based valuation work of GLS, both contractual and non-contractual companies were valued with the same model. For non-contractual firms, a retention rate proxy was used as the retention rate in their CLV formula. Of the five companies valued in their empirical analysis, the only two non-contractual businesses, eBay and Amazon, were also the two most misvalued. These companies were undervalued by an average of 88% and 83%, respectively[2], even though these valuations were performed after the stock market had fallen sharply in the aftermath of the "tech bubble," and prominent Wall Street analysts were publicly questioning Amazon's solvency at the time (Arango 2001; Streitfeld 2001). Figure 4.1 reproduces the valuation estimates from GLS for reference. A model specifically suited to non-contractual businesses is needed.

The procedure we use to estimate the proposed valuation model relies on indirect inference (Gallant & Tauchen 1996; Gourieroux et al. 1993; Smith 1993, hereafter denoted as "II"). II is a versatile simulation-based estimation method which is particularly useful when fitting complex models using limited data (e.g., censored, missing, and/or aggregate data, or omitted covariates (Jiang & Turnbull 2004)), characteris-

---

[2]The other three businesses valued – Ameritrade, Capital One, and E*Trade – are all contractual as first defined by Schmittlein et al. 1987 because their churn is observable. Customers are required to maintain positive account balances to remain customers.

Figure 4.1: Market Value and Customer Value Over Time: Gupta, Lehmann, and Stuart (2004)
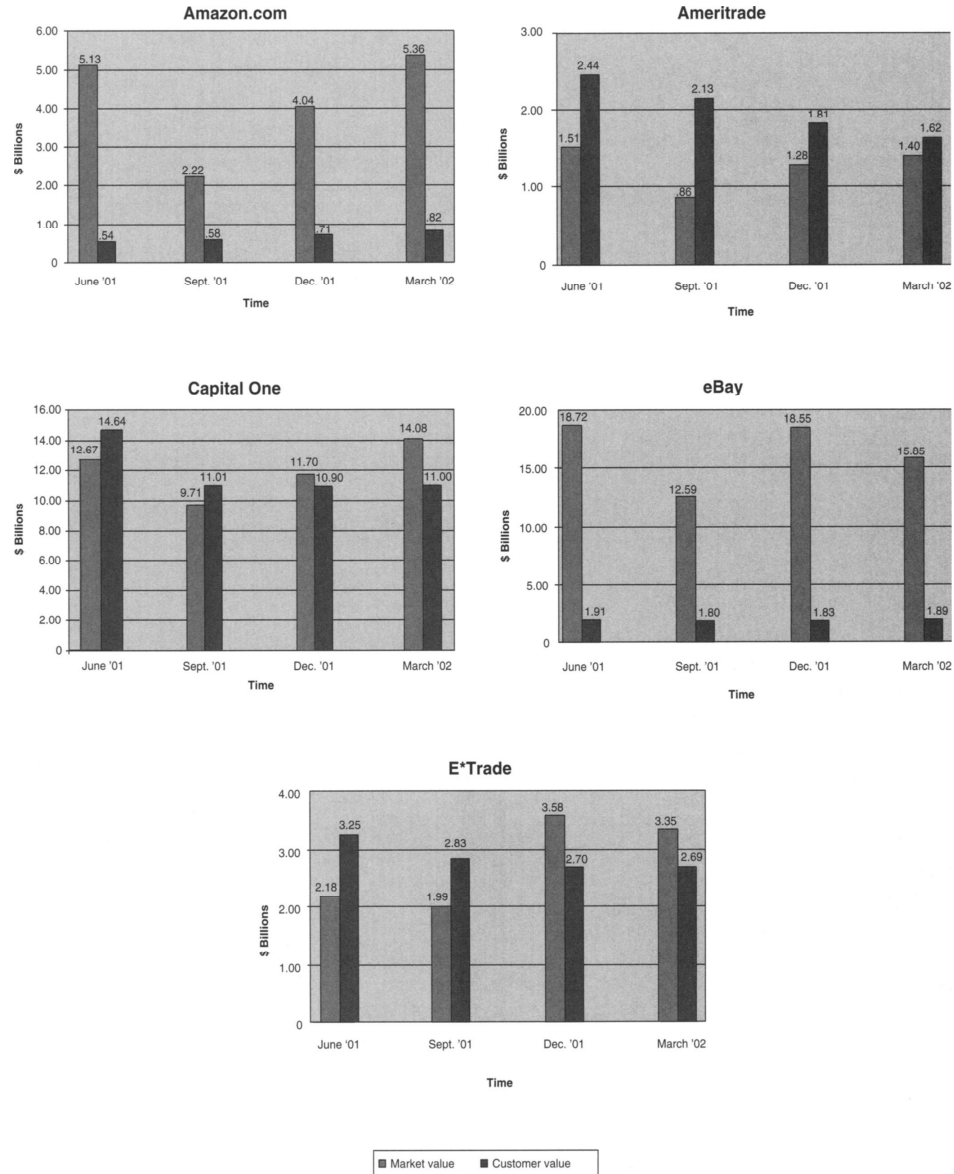
tics that describe our problem setting well. Generalized method of moments (GMM, Hansen 1982) can be thought of as a special example of II (Jiang & Turnbull 2004). II subsumes simulated method of moments (SMM, McFadden 1989) and conventional method of moments, which are popularly used parameter estimation techniques within

the marketing literature (Chintagunta 1992, Dutta et al. 1999, Erdem 1996; Gönül & Srinivasan 1993; Mehta et al. 2003; Schmittlein & Peterson 1994). SMM is not generally applicable in our setting because the set of metrics that we consider are usually not moments in a conventional sense (e.g., firms will not disclose metrics involving skewness or even variance). As with SMM, however, II is useful when it is difficult or impossible to analytically compute the likelihood function or even the moments of a proposed model, but easy to simulate from that model. We use II to find the set of true model parameters that minimize the expected distance between moments of convenient but misspecified "auxiliary models," which are derived from the observed disclosures, and what we expect those auxiliary moments should be.

In the next section, we present the model governing customers' acquisition, repeat purchasing, and spend, and how this model is used to drive an overall valuation for the firm. After providing the common customer metrics which will be used to estimate the customer model, we show how II is used to perform model estimation. We then analyze the predictive performance of all possible collections of these metrics through a large-scale simulation analysis. We apply this methodology to a 5.5-year transaction log data set from an e-commerce retailer business unit. We provide an overall valuation for the business unit, then decompose this valuation into existing versus yet-to-be-acquired customers, and analyze the unit economics of newly acquired customer cohorts. We conclude with a discussion of the results.

4.2   Model Development

In this section, we specify the individual-level model for the customer which we will use to forecast future customer activity and show how this model is embedded within an overall valuation framework for the firm.

4.2.1 Valuation Framework

We adopt the same firm valuation framework that we had used in the previous chapter, the discounted cash flow (DCF) model (for details, please refer to Section 2.2). We assume a weekly "clock" for disaggregate customer purchase activity, so that $w = 1$ represents the first week of the company's commercial operations. Week-by-week FCF projections are thus central to our estimation of OA and thus SHV in DCF models through Equation 2.2 in Section 2.2.1, analogous to how month-by-month FCF projections were central to OA and SHV for contractual firms in essay two. Using Equation 2.3 and assuming no cash flow-related adjustments (see Damodaran 2012), weekly FCF is equal to:

$$\text{FCF}(w) = \{\text{R}(w) \times [1 - \text{VC}(w)] - \text{FC}(w) - \text{CAC}(w) \times \text{A}(w)\}[1 - \text{TR}], \quad (4.1)$$

where as before, R, VC, FC, CAC, A and TR represent weekly revenue, variable cost ratio, fixed cost, customer acquisition cost per customer, gross acquisitions, and the tax rate, respectively.

Our goal, then, is to specify weekly processes for the acquisition of new customers, and the repeat purchase and spend of customers after they have been acquired. We estimate the parameters of these models so that the weekly behaviors implied by the model are consistent with the quarterly disclosures provided by the firm. We combine the projections from these processes to forecast R and A in Equation 4.1 into the future, then plug the resulting FCF projections into Equation 2.2 from chapter two to value the firm's operating assets.

4.2.2 Model Specification

Our proposed model for the timing of customer adoption for non-contractual busi-
nesses is identical to the model that we proposed in the last chapter for contractual
businesses, except for two small differences: (1) we model weekly and not monthly
acquisitions, because a weekly unit of time is more appropriate for non-contractual
behaviors than a monthly unit of time (Chintagunta 2002, Yoo et al. 2012); (2) we
assume that the market potential has been specified accurately, so that we may set
the proportion of each prospect pool who will never be acquired to zero (i.e., $\pi_{NA} = 0$
in Equation 3.13). In our empirical example, the company who supplied us with data
provided us with an internal-company estimate of the size of the target market. If we
so desired, we could easily reintroduce and estimate π_{NA} if this were not the case.

For completeness, we very briefly summarize the acquisition model. This model
consists of two parts: (1) the formation of pools of prospective customers over time,
and (2) the duration of time which elapses from the time a prospect is "born" to
when the prospect adopts the service. We drive the creation of prospect pools over
time off of the population size. At the beginning of the firm's commercial operations,
there is an initial prospect pool M(0) which is equal to the population size at the
time, POP(0). This prospect pool will eventually adopt in future weeks $w = 1, 2$, and
so on. The size of the prospect pool in a given week w is equal to population growth
during the week:

$$M(w) = \text{POP}(w) - \text{POP}(w - 1), \quad w = 1, 2, \ldots \tag{4.2}$$

After a prospect pool is formed, we model the duration of time until the customer
is acquired. Letting $F_A(w' - w|w)$ be the probability that a prospect from week w
will adopt by the end of week w', and letting A(w') be the number of new customers

acquired in week w', $\boldsymbol{\psi}_A$ be the collection of parameters underlying the acquisition model, and $\mathbf{x}_A(w')$ be a p-length vector containing values of our covariates in week w', then

$$A(w') = \sum_{w=0}^{w'-1} M(w)\{F_A[w'-w|w, \mathbf{x}_A(w'), \boldsymbol{\psi}_A] - F_A[w'-w-1|w, \mathbf{x}_A(w'-1), \boldsymbol{\psi}_A]\} \quad (4.3)$$

As before, the hazard model we use to model a prospect's time until adoption is a Weibull(λ_A) baseline with covariate effects incorporated through proportional hazards, with cross-sectional heterogeneity in the baseline propensity, λ_A, characterized by a gamma(r_A, α_A) distribution. Given a homogeneous acquisition shape parameter (c_A), homogeneous but possibly time-varying acquisition covariates up to week w' ($\boldsymbol{X}_A(w') = [\mathbf{x}_A(1), \mathbf{x}_A(2), \ldots, \mathbf{x}_A(w')]$), and coefficients for acquisition covariates ($\boldsymbol{\beta}_A$), the unconditional probability that a customer from prospect pool w will be acquired by the end of week w' is

$$
\begin{aligned}
F_A(w'-w|w, \boldsymbol{X}_A(w'), \boldsymbol{\psi}_A) &= \int_0^\infty F_A[w'-w|w, \lambda_A, c_A, \boldsymbol{X}_A(w'), \boldsymbol{\beta}_A] f(\lambda_A|r_A, \alpha_A) d\lambda_A \\
&= 1 - \left[\frac{\alpha_A}{\alpha_A + B_A(w, w'-w)}\right]^{r_A}, \quad \text{where} \quad (4.4) \\
B_A(w, w'-w) &= \sum_{i=w+1}^{w'} [(i-w)^{c_A} - (i-w-1)^{c_A}] e^{\boldsymbol{\beta}_A^T \mathbf{x}_A(i)} \quad (4.5)
\end{aligned}
$$

Our baseline model for the number of repeat purchases the customer makes is a generalization of the Beta-Geometric/Beta-Binomial (BG/BB) model for non-contractual customer base analysis (Fader et al. 2010). Immediately after the customer places her initial order, he/she is in an alive state. While alive, he/she makes a purchase in week w with probability $q(w)$. This probability may be higher or lower due to unobserved heterogeneity, or due to external factors such as the state of the macroeconomy or seasonality. We allow for both effects through a logit-normal

formulation, so that

$$q(w) = \frac{\exp[b_p + \boldsymbol{\beta}_p^T \mathbf{x}_p(w)]}{1 + \exp[b_p + \boldsymbol{\beta}_p^T \mathbf{x}_p(w)]}, \tag{4.6}$$

where the baseline purchase propensity b_p is distributed across the population according to a normal(μ_p, σ_p^2) distribution, $\mathbf{x}_p(w)$ are covariates associated with week w, and $\boldsymbol{\beta}_p$ are the coefficients associated with those covariates.

Each period, the customer may churn with probability θ. We assume that a customer who has churned has a 0% probability of making a future purchase. Because customer churn is not observed, this is a so-called "leaky always-a-share" model (Fader & Hardie 2014). We let θ vary across the population according to a gamma(γ, δ) distribution. Going forward, we will refer to this model as the Beta-Geometric/Mixed-Logit (BG/ML) model. Had we assumed that $q(w)$ varied across customers but not time and that $q(w)$ is distributed across the population according to a beta distribution, this model reduces to the BG/BB model.

The proposed repeat purchase model is similar in spirit to the latent attrition model in Braun et al. (2015) except that the proposed model operates in discrete time and not continuous time, and allows for covariate effects in the purchase process and not the latent attrition process. The purchase process is more empirically identified (and thus more able to accommodate covariate effects) than the latent attrition process is. While we could easily allow for time-varying covariates in both the purchase rate and attrition processes simultaneously, the parameters of the resulting model are not identified (Braun et al. (2015) observed the same phenomenon).

Finally, we model the underlying expected amount spent per purchase and how it varies across customers (Fader et al. 2005b and Schmittlein & Peterson 1994). The expected amount spent on a given purchase in week w, $\mathrm{E}[s(w)]$, is driven by a baseline propensity to spend, b_s, and a time trend term to allow expected spend to vary as a function of which quarter q the customer is in. A time trend in average spend

per purchase is often necessary – for example, it is evident in the empirical example which follows, and was evident in the two public company valuations from essay two. The baseline propensity to spend varies across customers according to a lognormal distribution:

$$\mathrm{E}[s(w)|b_s, \mu_q] = b_s + q\mu_q, \quad \text{where} \quad \log(b_s) \sim \mathcal{N}(\mu_0, \tau^2) \tag{4.7}$$

The spend formulation ensures that customers' expected spends are strictly positive with a heavy right tail, consistent with non-contractual transaction log data.

These models are flexible in their ability to capture variation across customers and over time while remaining parametrically parsimonious.

4.2.3 Valuation Procedure

Our goal is to use the fitted models from the previous section to forecast R and A in Equation 4.1 into the future, then use these projections to come up with an overall valuation estimate. In this section, we outline how to estimate the overall value of the firm's operating assets, after the parameters of the model have been estimated. We will discuss estimation (and the data used to do the estimation) over the next two sections.

First, we estimate weekly customer acquisitions far into the future, to a point in time when the present value of all future cash flows is negligible. Let the calibration period and forecasting horizon be W and W^* weeks long, respectively. To come up with weekly acquisition projections, we need estimates of $\mathrm{POP}(w)$ over the forecasting horizon. In the analyses which follow, we assume that projections of $\mathrm{POP}(w)$ over the forecasting horizon are provided to us. If projections of $\mathrm{POP}(w)$ are not available, we could use a simple model which estimates future growth as an extrapolation of

historical growth. We use Equation 4.2 to turn POP(w) projections into prospect pools M(w) over all periods $w = 1, 2, \ldots, W + W^*$. We then use the fitted acquisition parameters $(r_A, \alpha_A, c_A, \boldsymbol{\beta}_A)$, and the historical and projected future prospect pools [M(1),..., M($W + W^*$)] to obtain historical and future expected customer acquisitions [A(1),..., A($W + W^*$)] using Equation 4.3.

Second, we estimate weekly total (initial plus repeat) purchases, given the historical and future customer acquisitions [A(1),..., A($W + W^*$)] and the parameters of the repeat purchase process $(\mu_p, \sigma_p^2, \boldsymbol{\beta}_p)$. Because there is no closed-form expression for the unconditional expected number of repeat purchases by week, we use Monte Carlo simulation instead. Letting A* be the total number of customers that will eventually be acquired by the end of the forecasting horizon (i.e., A$^* = \sum_{w=1}^{W+W^*}$ A(w)), we repeatedly simulate a binary "total purchases matrix," TP, which records the purchases each eventually-acquired customer will make in each week. The first A(1) rows correspond to the customers acquired in week one, the following A(2) rows correspond to customers acquired in week two, and so on. Letting TP$^{(k)}$ be the kth simulated realization of this matrix,

$$
\text{TP}^{(k)} = \begin{bmatrix}
y_{1,1}^{(k)} & y_{1,2}^{(k)} & \cdots & y_{1,W+W^*}^{(k)} \\
y_{2,1}^{(k)} & y_{2,2}^{(k)} & \cdots & y_{2,W+W^*}^{(k)} \\
\vdots & \vdots & \ddots & \vdots \\
y_{A^*,1}^{(k)} & y_{A^*,2}^{(k)} & \cdots & y_{A^*,W+W^*}
\end{bmatrix},
\tag{4.8}
$$

where $y_{j,w}^{(k)}$ is a binary variable equal to one if the kth simulation of customer j made a purchase in week w, 0 otherwise. Assuming that customer j is acquired in week w, we simulate his/her individual-level parameters from their respective heterogeneity distributions, $\theta_j^{(k)} \sim \text{beta}(\gamma, \delta)$ and $b_{p,j}^{(k)} \sim \mathcal{N}(\mu_p, \sigma_p^2)$. The binary sequence associated with customer j in each week w' (i.e., the jth row of TP$^{(k)}$) is determined by the

80

following four scenarios:

$$
y_{j,w'} = \begin{cases} 1 \text{ with probability } 1 & w' = w \\ 0 \text{ with probability } 1 & w' < w \\ 1 \text{ with probability } (1 - \theta_j^{(k)})^{w'-w} \dfrac{\exp[b_p^{(k)} + \boldsymbol{\beta}_p^T \mathbf{x}_p(w')]}{1 + \exp[b_p^{(k)} + \boldsymbol{\beta}_p^T \mathbf{x}_p(w')]} & w < w' \\ 0 \text{ otherwise} \end{cases}
$$

The expected purchases made by all customers over all time periods, $\widehat{\text{TP}}$, is formed by taking the element-wise average of $\text{TP}^{(k)}$ over K replications (the appropriate value for K should be a function of A^*). The column sums of $\widehat{\text{TP}}$ are equal to the expected total number of purchases for the firm as a whole over the entire calibration period and forecasting horizon. We denote the $W + W^*$-length vector of expected historical and future purchases by $[\text{P}(1), \ldots, \text{P}(W + W^*)]$.

Third, we estimate weekly revenues, given the vector of expected weekly total purchases $[\text{P}(1), \ldots, \text{P}(W + W^*)]$ and the parameters of the spend process (μ_0, τ^2, μ_q). Let $\text{E}[s(w)]$ be the unconditional expected spend per purchase in week w. Assuming week w is in quarter q,

$$
\text{E}[s(w|\mu_0, \tau^2, \mu_q)] = \exp(\mu_0 + \tau^2/2) + q\mu_q.
$$

Expected firm revenues in week w, $\text{R}(w)$ (Equation 4.1), is equal to the product of expected total purchases in week w ($\text{P}(w)$) with expected spend per purchase in week w ($\text{E}[s(w)]$). We use future weekly revenues $[\text{R}(W+1), \ldots, \text{R}(W+W^*)]$ to estimate the valuation of the firm in week W.

As noted in the valuation framework, once we have estimated future revenues (R) and customer acquisitions (A), the remaining inputs needed to obtain FCF, OA, and SHV (Equations 2.1, 2.2, and 4.1) are obtained using the same procedure that

a financial professional would use. The operating unit we value in our empirical example is privately held, so all of these non-customer-driven line item projections were provided to us by the company's management team.

The valuation procedure assumes that we have already estimated the parameters of the model. In the next two sections, we discuss the data which is available to perform the estimation, and how we can estimate the model parameters with this information.

4.3   Candidate Customer Metrics

While no public non-contractual firms disclose granular data, some of these firms disclose aggregated customer metrics. Furthermore, these first-party disclosures released by the firms themselves are supplemented by third-party disclosures, mined by business intelligence and market intelligence firms such as 1010data, AppAnnie, Prosper Analytics, SecondMeasure, Slice Intelligence, and SurveyMonkey Intelligence[3]. What are the metrics that these firms do, and/or should, disclose? We discuss customer data summaries most relevant to each of the non-contractual customer model processes in turn. For all customer metrics, we follow firms' usual convention of providing this number on a trailing twelve month basis (e.g., the firm reports the number of customers acquired over the past year).

For the acquisition process, gross customers acquired is the most natural and most important customer data summary. It is the acquisition metric that was used in essay two for contractual firms. Dozens of contractual firms, and even some forward-thinking non-contractual firms (QVC 2015), regularly disclose this measure. From

---

[3]As noted in chapter two, we assume throughout this dissertation that the data are observed without measurement error so that there is no error-in-variables bias. This implies that our data come from first-party disclosures or third-party disclosures based upon large, highly representative samples.

a statistical relevance standpoint, gross customers acquired reliably identifies the parameters of our proposed acquisition model (see Appendix A.3 for a full factorial parameter recovery analysis). While we could easily consider other acquisition-related measures, these factors create little practical incentive to do so.

For the spend process, we obtain mean spend "for free," because it can be derived from total revenues (which must be disclosed by public firms) in conjunction with the purchase process. As noted earlier, however, spend per transaction is more variable than a routine subscription payment and thus it is more important for us to capture the shape of the distribution, rather than simply projecting the means. Median spend (i.e., the 50th percentile of all spend amounts) is a very natural companion measure to mean spend. Median spend is a communicable, intuitively appealing measure of basket size for external stakeholders. Disclosures of mean and median spend over time identify the parameters of the spend process well. As is the case with the acquisition process, while other disclosures could in theory be considered, it is not apparent what spend measures would be better, practically and statistically.

What is less clear is what should be recommended for the repeat purchase process. We focus on the following six common repeat purchase measures:

(a) Active users (AU): The number of customers who make 1+ purchases in the past year.

(b) Heavy users (HAU): The number of customers who make 2+ purchases in the past year.

(c) Repeat rate (RR): The proportion of customers who made a purchase last year, who purchase again this year.

(d) Repeat buyer proportion-customers (RBPC): The proportion of customers who made a purchase this year who also purchased before this year began.

(e) Repeat buyer proportion-orders (RBPO): The proportion of orders made this year by customers who also purchased before this year began.

(f) Average frequency (F): The average number of purchases made by all active customers over the past year.

A non-exhaustive list of first-party disclosures of these metrics (or very closely-related metrics) are provided in Table 4.1. Third-party disclosures are provided in parentheses. While the most commonly disclosed metric is active users, the other metrics are also frequently disclosed.

Table 4.1: Common Customer-Related Disclosures (Third-Party Disclosures in Parentheses)

Metric	Firms
Active Users	Amazon, Camping World, Caesars Acquisition Co., Evine, Facebook, Gamzio Mobile, Glu Mobile Inc., Go-Pro, HSN, International Game Technology, LinkedIn, MEDL Mobile Holdings, MeetMe, Inc., Quepasa Corp., QVC, Scientific Games Corp., Snap Interactive, Social Reality, Sohu Com Inc., Twitter, Uber, Wayfair, Zynga (Niantic)
Heavy Active Users	QVC (Aveda, Bare Essentials, ULTA)
Repeat Rate	QVC, Uber (Amazon)
Repeat Buyer Proportion	Amazon, Etsy, QVC, Wayfair (HSN, Instacart, Jet.com)
Average Frequency	Evine

We provide a simple numerical example in Table 4.2 to further illustrate how granular transaction log data is summarized along different dimensions by the customer metrics. For expositional purposes, we only show disclosures at the end of each *year* in this numerical example (i.e., at the end of years one, two, and three). In the analysis that follows, however, we assume that disclosures can be made each *quarter*. For example, AU as of the end of the fourth quarter of commercial operations represents the number of customers who made at least one purchase in quarters one through four, AU as of the end of the fifth quarter of commercial operations represents the number of customers who made at least one purchase in quarters two through five,

and so on. While our estimation procedure can be used for firms that only disclose once per year by treating the first three quarterly disclosures of each year as missing data, annual disclosure periodicity implies a very small number of data points to perform modeling with and thus may not be empirically identified.

Table 4.2: Numerical Example of Acquisition and Purchase Metrics

Number of Purchases by Customer/Year

	Y1	Y2	Y3
Customer 1	1	1	0
Customer 2		3	3
Customer 3			1
Gross Acquisitions	1	1	1
AU	1	2	2
HAU	0	1	1
RR	NA	100%	50%
RBPC	0%	50%	50%
RBPO	0%	25%	75%
F	1	2	2

An important question is how we use these customer metrics to estimate the latent variable model specified in the previous section. We answer this question next.

4.4   Estimation with Indirect Inference

The disclosures from the previous section are not conventional model-dependent moments. For example, the purchase disclosures are overlapping, cross-sectional, often highly non-linear summaries involving both the acquisition process and the repeat purchase process at once (i.e., these measures do not distinguish between initial and repeat purchases). In this section, we provide a brief overview of II and how it enables us to estimate the model parameters using these data summaries, when evaluation

of the likelihood function is analytically intractable. We start by providing a general framework which can be applied to any panel dataset. We then provide an illustration of how we use II to estimate the parameters of the acquisition process, and how the other two processes are estimated. We provide a detailed description of the auxiliary model specifications for the repeat purchase and spend processes in Appendix A.4.

4.4.1   General Parameter Estimation Framework

Let Z_{jt} be a random variable representing the behavior of customer j at time t, for $j = 1, 2, \ldots, J$ and $t = 1, 2, \ldots, T$. Assume that the distribution of Z_{jt} is a function of p model parameters $\boldsymbol{\psi}$ and time-varying covariates $\mathbf{x}_{jt}$, so that $z_{jt} \sim f(Z_{jt}|\boldsymbol{\psi}, \mathbf{x}_{jt})$. We estimate $\boldsymbol{\psi}$ through the following three-step procedure:

(a) Compute a q-dimensional "auxiliary statistic" $\hat{\boldsymbol{s}}$ which is a deterministic function of the empirical cumulative distribution function (ECDF) $\hat{F}_{JT}$ (i.e., $\hat{\boldsymbol{s}} = g(\hat{F}_{JT})$), where $q \geq p$. $\hat{\boldsymbol{s}}$ should represent key characteristics of the data. These auxiliary statistics are often moments of the data or deterministic functions of the true model parameters under a convenient but misspecified "auxiliary model." This allows us to match off of non-moments of the data under the true model, unlike the GMM estimator. In the next section we provide the auxiliary model we use for the acquisition process, and in Appendix A.4 we provide the auxiliary models for the repeat purchase and spend processes.

(b) Let the "auxiliary parameter" $\boldsymbol{s}$ be the limiting value of the auxiliary statistic $\hat{\boldsymbol{s}}$ as $(J, T) \to \infty$ (i.e., $\boldsymbol{s} = g(F_{\boldsymbol{\psi}})$). Define the "binding relationship," which relates the unknown model parameters $\boldsymbol{\psi}$ to the auxiliary parameter $\boldsymbol{s}$. Then $\boldsymbol{s} = \boldsymbol{s}(\boldsymbol{\psi})$.

(c) Use the binding relationship to find the model parameters $\boldsymbol{\psi}$ which minimize

the "distance" between the observed auxiliary statistic $\hat{s}$ and the auxiliary parameter under $\boldsymbol{\psi}$, $\boldsymbol{s}(\boldsymbol{\psi})$, under a suitable estimate of the distance or weight matrix W to be defined below[4]:

$$\hat{\boldsymbol{\psi}} = \text{argmin}_{\boldsymbol{\psi}}[\boldsymbol{s}(\boldsymbol{\psi}) - \hat{\boldsymbol{s}}]^T \hat{W}[\boldsymbol{s}(\boldsymbol{\psi}) - \hat{\boldsymbol{s}}]. \tag{4.9}$$

II is often used when no closed-form expression is available relating model parameters $\boldsymbol{\psi}$ to auxiliary parameters $\boldsymbol{s}(\boldsymbol{\psi})$. When there is no such closed-form expression, we simulate the binding relationship. For a given $\boldsymbol{\psi}$, we simulate K new datasets. Letting $\hat{F}_{JT}^{(k)}$ be the ECDF associated with simulation k, compute K auxiliary statistics under the K simulated datasets, $[\hat{\boldsymbol{s}}^{(1)}(\boldsymbol{\psi}), \hat{\boldsymbol{s}}^{(2)}(\boldsymbol{\psi}), \ldots, \hat{\boldsymbol{s}}^{(K)}(\boldsymbol{\psi})]$. Then for K large, $\boldsymbol{s}(\boldsymbol{\psi}) \approx \sum_{k=1}^{K} \hat{\boldsymbol{s}}^{(k)}(\boldsymbol{\psi})/K$. In the empirical analyses which follow, K is set so that stochastic variability in the average is acceptably low (in practice, it is set so that the average empirical coefficient of variation is less than .5%).

This procedure is consistent under mild regularity conditions as long as the weight matrix W is positive definite (for example, the identity matrix). However, the optimal weight matrix is the asymptotic inverse variance-covariance matrix of the estimator solution:

$$\widehat{W} = \lim_{K \to \infty} \frac{1}{K-1} \sum_{k=1}^{K} \{[\hat{s}(\boldsymbol{\psi}_A)^{(k)} - \boldsymbol{s}(\boldsymbol{\psi}_A)][\hat{s}(\boldsymbol{\psi}_A)^{(k)} - \boldsymbol{s}(\boldsymbol{\psi}_A)]^T\}^{-1}. \tag{4.10}$$

The optimal weight matrix is preferable to the identity matrix for three reasons: (1) well-identified moments have higher weights, (2) parameter estimation is invariant to changes in the scale of the auxiliary statistics, and (3) correlations between auxiliary statistics are accounted for (i.e., there is no "double counting" of highly

---

[4]This is the so-called "distance"-based II estimator of Gourieroux et al. 1993 and Smith 1993. There is also a "score"-based II estimator (Gallant & Tauchen 1996). Studies have shown that the finite-sample performance of the score-based estimator is not as strong (Duffee & Stanton 2008; Michaelides & Ng 2000)

correlated auxiliary statistics).

In contrast, standard nonlinear least squares (NLS) estimation gives equal weight to all observed data points, which diminishes its finite sample performance. This issue is particularly acute in our setting, where different disclosures operate on very different scales (e.g., we may observe RR of 50% and AU of 5MM), and firm disclosures may be strongly correlated with one another (e.g., AU and HAU). As with NLS, however, this procedure does not require calculation of the likelihood function. One must solve the objective function provided in Equation 4.9.

As an illustration, we show how we use this procedure to estimate the parameters of the acquisition process in the next section.

4.4.2 Acquisition Process Parameter Estimation

Assume that we observe an arbitrary collection of n gross acquisition disclosures,

$$\mathbf{ADD} \equiv [\mathrm{ADD}_{q_1}, \mathrm{ADD}_{q_2}, \cdots, \mathrm{ADD}_{q_n}],$$

whose ith element, ADD_{q_i}, is equal to the trailing year sum of gross customers acquired. We posit a non-parametric auxiliary model in which all of the prospects which exist by the end of the calibration period, $\mathrm{POP}(W)$, are independent and identically distributed, and can be acquired starting immediately after commercial operations begin. The acquisition time of each of the prospects under the auxiliary model is determined by a multinomial$(\pi_1^{(a)}, \pi_2^{(a)}, \cdots, \pi_Q^{(a)}, \pi_\emptyset^{(a)})$ draw. A non-parametric auxiliary model maximizes the amount of information we obtain from the observed data and over-identifies the true model. $\pi_q^{(a)}$ is the probability that the customer is acquired in quarter q, and $\pi_\emptyset^{(a)}$ is the probability that the customer is not acquired during the calibration period. $\mathbf{ADD}/\mathrm{POP}(W)$ are sums of empirical moments of the auxiliary

model under the observed data. For example, for $q_i \geq 4$,

$$\text{ADD}_{q_i}/\text{POP}(W) = \widehat{E}\left(\pi_{q_i-3}^{(a)} + \pi_{q_i-2}^{(a)} + \pi_{q_i-1}^{(a)} + \pi_{q_i}^{(a)}\right).$$

ADD/POP(W) is the auxiliary statistic for the acquisition process, $\widehat{\boldsymbol{s}}^{(a)}$.

For each set of acquisition parameters $\boldsymbol{\psi}_A$ that we consider, we obtain the corresponding auxiliary parameter $\boldsymbol{s}(\boldsymbol{\psi}_A)$. We do this by repeatedly simulating a "total acquisitions matrix", $\text{TA}(\boldsymbol{\psi}_A)$, which records the number of prospects within each prospect pool who are acquired in each week. The first row corresponds to the M(0) prospects from the week-0 prospect pool, the second row corresponds to the M(1) prospects from the week-1 prospect pool, and so on. Letting $\text{TA}^{(k)}$ be the kth simulated realization of this matrix,

$$\text{TA}(\boldsymbol{\psi}_A)^{(k)} = \begin{bmatrix} a_{0,1}^{(k)} & a_{0,2}^{(k)} & \cdots & a_{0,W}^{(k)} \\ a_{1,1}^{(k)} & a_{1,2}^{(k)} & \cdots & a_{1,W}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ a_{W-1,1}^{(k)} & a_{W-1,2}^{(k)} & \cdots & a_{W-1,W}^{(k)} \end{bmatrix}, \tag{4.11}$$

where $a_{w,w'}^{(k)}$ is equal to the kth realization of the number of prospects from M(w) prospect pool who are acquired in week w'. Consider the $(w+1)$st row of $\text{TA}(\boldsymbol{\psi}_A)^{(k)}$, corresponding to prospect pool M(w). Given $\boldsymbol{\psi}_A$, a simulated realization of this row is equal to

$$(a_{w,1}^{(k)}, \ldots, a_{w,w}^{(k)}) = 0 \text{ and } (a_{w,w+1}^{(k)}, \ldots, a_{w,W}^{(k)}) \sim \text{multinomial}[M(w); \phi_{w,w+1}, \cdots, \phi_{w,W}, \phi_{w,\emptyset}],$$

where $\phi_{w,w'}$ is defined via difference of CDF's using Equation 4.4:

$$\phi_{w,w'} \equiv F_A[w' - w | w, \boldsymbol{X}_A(w'), \boldsymbol{\psi}_A] - F_A[w' - w - 1 | w, \boldsymbol{X}_A(w' - 1), \boldsymbol{\psi}_A].$$

The kth realization of the number of acquisitions across all prospect pools is equal to the column sums of $\text{TA}^{(k)}$, so the corresponding auxiliary statistic $\widehat{s}(\boldsymbol{\psi}_A)^{(k)}$ is equal to

$$\widehat{s}(\boldsymbol{\psi}_A)^{(k)} = \left(\sum_{w'=52q_1-51}^{52q_1} \sum_{w=0}^{W-1} \text{a}_{w,w'}^{(k)}, \cdots, \sum_{w'=52q_n-51}^{52q_n} \sum_{w=0}^{W-1} \text{a}_{w,w'}^{(k)} \right) \Big/ \text{POP}(W)$$

The auxiliary parameter corresponding to a particular set of acquisition parameters $\boldsymbol{\psi}_A$ is equal to $\boldsymbol{s}(\boldsymbol{\psi}_A) = \sum_{k=1}^{K} \widehat{s}(\boldsymbol{\psi}_A)^{(k)}/K$.

To estimate the parameters of the acquisition process using indirect inference assuming this auxiliary model, we perform the following steps.

First, we initialize the distance metric to an n-dimensional identity matrix, $\widehat{W} = I_n$.

Second, we update our estimate of the acquisition parameters, $\widehat{\boldsymbol{\psi}}_A$. We find the set of parameters that minimizes the expected Wald distance of $\widehat{\boldsymbol{s}}^{(a)}$ from $\boldsymbol{s}(\boldsymbol{\psi}_A)$ under the distance metric $\widehat{W}$ using Equation 4.9.

Third, we update the distance metric $\widehat{W}$ given $\widehat{\boldsymbol{\psi}}_A$, using Equation 4.10.

Finally, we repeat steps two and three as many times as desired. Following Wooldridge 2002 (p. 193), we perform two iterations, and did not observe any noticeable change in the parameter estimates when we used more than two iterations.

The process is for all intents and purposes the same when we introduce the repeat purchase and spend processes into the estimation procedure. We posit auxiliary models corresponding to each process, then find the set of true model parameters whose corresponding auxiliary parameters are as close as possible to the observed auxiliary statistics which are derived from the observed disclosures using a two-step estimator for the optimal weight matrix. Auxiliary model specifications for the repeat purchase and spend processes are provided in Appendix A.4. We exploit the fact that under the proposed model, the true acquisition process is independent of the repeat purchase and spend processes, and the true repeat purchase process is independent

90

of the spend process, to improve the convergence speed of the algorithm[5]. This effectively allows us to optimize over no more than five parameters at a time.

4.5   Customer Metric Selection

Firms are unlikely to disclose all possible customer metrics within their quarterly and annual filings because of the perceived costs of disclosure (Lev 1992). Therefore, we must understand how the predictive validity of the model varies as a function of the size and composition of the collection of metrics used to train the model, and as a function of contextual factors. We generate many different data sets or "worlds" reflecting different patterns of customer acquisition, purchasing, and latent attrition by varying the values of the acquisition and repeat purchase processes. We vary six key inputs to these processes in a full factorial design to create $N_W = 64$ such data sets. We generate each dataset for an initial prospect pool of 10MM customers which grows at an annualized growth rate of 1.2% per year. We observe the behavior of this prospect pool over a six-year ($Q = 24$ quarter) period, which is used to train the model. To test the robustness of the metric collections to missing data, we left censor the data, deleting the first year of activity. This leaves us with 20 quarterly disclosures for each metric which is included in the metric collection, corresponding to the five years of uncensored quarterly data for the repeat purchase metrics and quarterly customers acquired. We estimate the parameters of the acquisition and repeat purchase processes with this data.

We consider all $N_C = 63$ possible combinations of the six repeat purchase metrics – all collections from size one (i.e., each of the metrics individually) to size six (i.e.,

---

[5]We estimate the parameters of the acquisition process using the procedure above, then estimate the repeat purchase process given the acquisition parameter estimates, then estimate the spend process given the acquisition and repeat purchase estimates, then estimate the processes jointly.

all of the available metrics at once). We assess metric pair performance within a particular data set based upon how well that metric pair predicts aggregate incremental purchases within a holdout period, which is a managerially relevant quantity linked to customer-based corporate valuation. We performed a similar analysis to assess parameter recovery and the results are qualitatively similar. The error measure that we choose is holdout mean absolute percentage error (MAPE) over the next five years ($Q^* = 20$ quarters). The results are robust to other error measures. Formally, letting $P_q^{(k)}$ denote the total observed number of purchases made within quarter q of data set k, and $\widehat{P}_q^{(k,c)}$ the expected number of repeat purchases within quarter q of data set k using metric collection c, the MAPE is equal to

$$
\text{MAPE}_{kc} = \frac{1}{Q^*} \sum_{q=Q+1}^{Q+Q^*} \left| \frac{P_q^{(k)} - \widehat{P}_q^{(k,c)}}{P_q^{(k)}} \right|, \quad c = 1, 2, \ldots, N_C, \quad k = 1, 2, \ldots, N_W. \quad (4.12)
$$

We obtain the expected MAPE corresponding to each metric collection c over each data set k. These estimates form the basis for our assessment of the performance of these metrics. For details regarding the simulation design, see Appendix A.5.

As a first overall assessment, we provide in Figure 4.2 a scatterplot of the average MAPE for each collection and each particular size, averaged across the 64 data sets. The average MAPE associated with the best size-one collection is 15.6%. The average MAPE of the best size-two collection is considerably lower, at .7%, and there is no further improvement when we move to larger collections (the best collections of size three to six all have average MAPE values of approximately .6%). This suggests that we need no more than two quarterly repeat purchase metrics to achieve adequate predictive performance. We also observe that there are size-two collections which have significantly smaller average MAPE values than other size-four and even size-five collections. This implies that the right pair of metrics has better predictive

performance than other collections which are far larger in size.

Figure 4.2: Average MAPE (%) of Each Collection by Collection Size (Count of Collections Exceeding 20% Average MAPE)
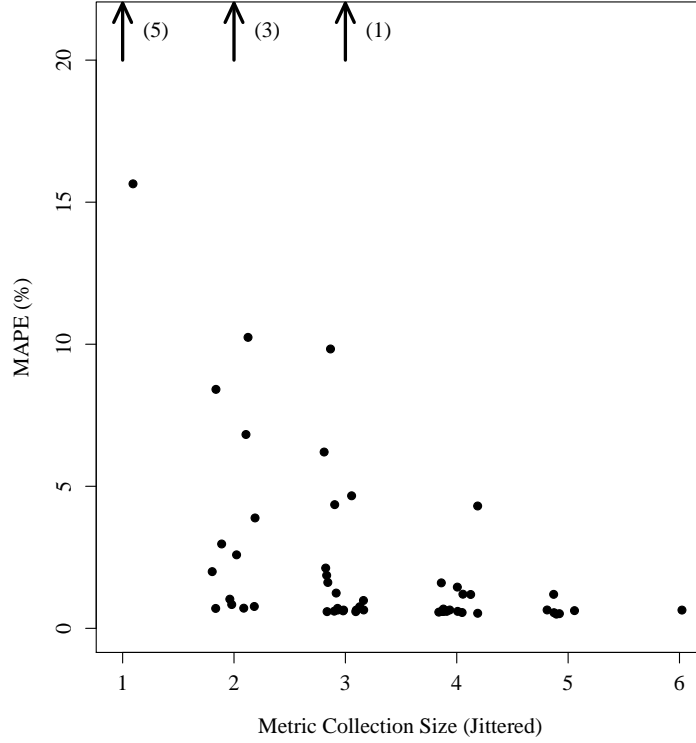

We explore the various size-two collections next. To better visualize which pairs have the best prediction accuracy, we plot in Figure 4.3 the MAPE of all 15 size-two collections, averaged across all 64 data sets. The results are shown in Figure 4.3[6], ordered from highest to lowest average MAPE, along with 95% standard error intervals.

Five size-two pairs have excellent overall predictive accuracy (average MAPE values less than or equal to 1%), and another four pairs have very good overall predictive

---

[6]The average MAPE values for AU+RR, RBPC+RR, and AU+RBPC are 45.6%, 47.4%, and 201.0%, respectively. The y-axis of the plot is truncated so that these pairs do not impede our ability to summarize the performance of the other 12 pairs.

Figure 4.3: MAPE (%) by Size-Two Metric Collection, Averaged Across Data Sets
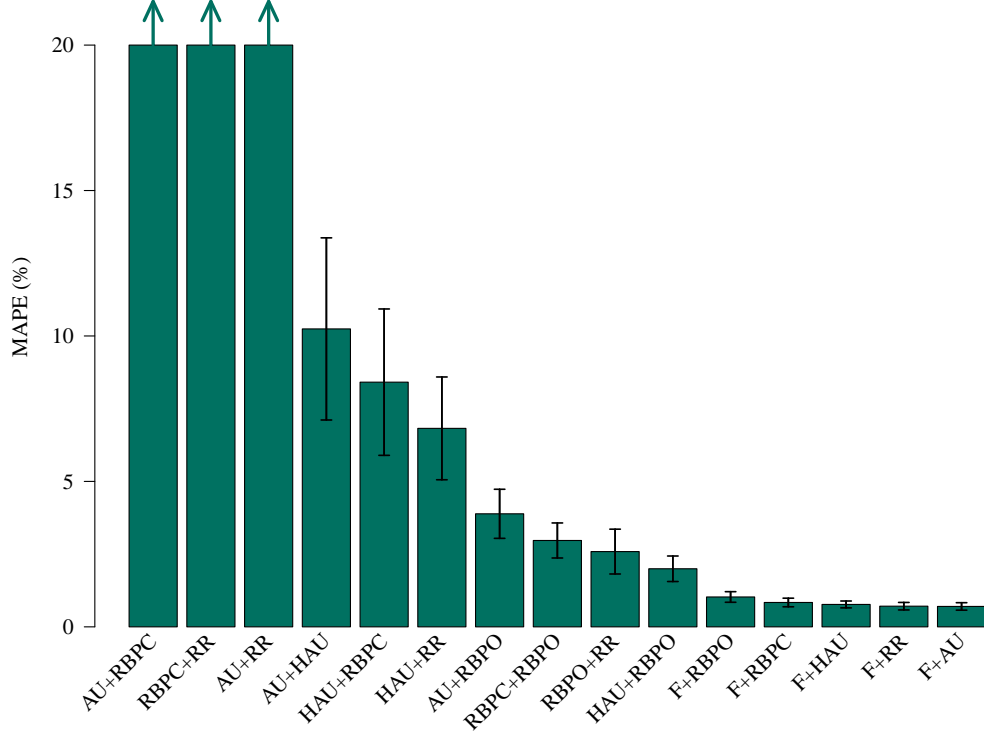
accuracy (MAPE values between 1% and 3%). All other pairs have average MAPE figures exceeding 5%, often significantly so.

The five best-performing pairs consist of F coupled with each of the other metrics, which suggests that F is the most informative metric in a marginal sense. The five F-pairs are also very robust – the maximum MAPE across all data sets is less than 4% for all F-pairs.

While the conceptual distinction between RBPO and RBPC may seem slight, there are significant differences in performance between these two measures. RBPO is included in the fifth through ninth best pairs by average MAPE. In contrast, RBPC is in three of the five worst pairs by average MAPE, including two pairs with average

MAPE values in excess of 45%. RBPO is a significantly more informative metric than RBPC in a marginal sense.

The quality of metrics within a particular collection is significantly more important for predictive accuracy than the quantity of metrics. For example, all but one of the size-two F-pairs have statistically significantly lower average MAPE values than the size-five collection which does not include F (AU+HAU+RBPC+RBPO+RR has an average MAPE of 1.2%, and the t-statistics of associated paired t-tests range from -2.72 to -3.74). Furthermore, some collections which are large in size have low predictive accuracy in an absolute sense (e.g., the average MAPE of AU+RBPC+RR is 46.3%).

This analysis suggests that F is the most informative of the proposed metrics to investors, in terms of its ability to improve overall purchasing (and thus revenue) forecasting accuracy. F is popularly used and studied by marketing scholars, often in conjunction with penetration/reach (Cheong et al. 2010; Danaher 2007; Sharp & Sharp 1997). Despite this, F is one of the lesser-disclosed common marketing metrics by public firms (see Table 4.1). We recommend that firms disclose F in conjunction with AU. While F is less popularly disclosed, AU is already widely adopted by firms and discussed by financial professionals. F+AU is closely related to so-called "means and zeroes" parameter estimation, a moment matching technique which is most commonly used to estimate the parameters of the Beta-Binomial and Negative Binomial distributions (Boylan 1997; Leckenby & Boyd 1984, Morrison 1969). This combination of metrics has the highest predictive accuracy in this simulation study, is small in number, and is an intuitive pair of key performance indicators for firms. AU summarizes purchase breadth (i.e., how well the firm is able to get a large number of buyers), while F summarizes purchase depth (i.e., how well the firm is able to get buyers to buy frequently). Furthermore, these metrics allow external stakeholders to

obtain total purchases over time when we observe that total purchases are the product of F with AU. This makes F+AU particularly compatible with investors' traditional financial models, which may decompose revenues into total purchases and average revenue per purchase, then further decompose total purchases into active customers and average purchases per active customer.

4.6   Empirical Analysis

We now apply the method developed over the previous sections to a data set of purchases from a large geographic business unit of a high-end online-only e-commerce retailer in the women's apparel industry. The data set consists of all disaggregate purchase data for the business unit since the beginning of commercial operations at the end of June 2010. The data set ends at the end of December 2015, so the observation period is 5.5 years (22 quarters) in length. Based upon the results from the large-scale simulation, we implement the estimation procedure using our recommended collection of customer metrics – gross customer acquisitions, active users, average purchase frequency, median spend, and total revenues. The results are virtually identical when we estimate the model upon all repeat purchase metrics. We split the 22-quarter data set into two periods – a $Q = 18$ quarter calibration period which is used to estimate the model, and a $Q^* = 4$ quarter holdout period which is used to assess the predictive validity of the model. The retailer disclosed the breakdown of their fixed and variable costs (FC and VC), their future expected subscriber acquisition costs (CAC), their weighted average cost of capital (WACC), and the size of the applicable market (POP). We summarize these projections when we value the business unit.

4.6.1 Model Assessment

Validation of acquisition and repeat purchase model fit focuses upon a series of six commonly used in-sample and out-of-sample goodness-of-fit diagnostic plots (Braun et al. 2015; Fader et al. 2005a; Fader et al. 2010; Feit et al. 2013). These plots summarize salient aspects of the data and test our ability to model and project them. Expected quantities of interest are formed by simulating K datasets using the estimated model hyperparameters, computing the desired quantity of interest each dataset, then averaging across datasets (we set $K = 1000$).

4.6.1.1 Parameter Estimates and Evaluation of Fit

We begin with the customer acquisition process. The fitted model rejected heterogeneity in the propensity to be acquired across customers ($\widehat{r}_A = 9.1E^9$). There was notable quarterly seasonality, with an uptick in calendar Q4 customer acquisitions. The estimated model parameters of the resulting Weibull with covariates model (and their associated standard errors) are $\widehat{\lambda}_A = 6.1E^{-5}$ ($1.7E^{-6}$), $\widehat{c}_A = 1.15$ (.008), and $\widehat{\beta}_A = .18$ (.006). Figure 4.4 shows expected and actual quarterly customer acquisitions on a quarter-by-quarter basis (left) and on a cumulative basis (right). The vertical dotted line denotes the beginning of the holdout period. While there is noise from quarter to quarter, we model and predict the flow of acquisitions well over time.

Turning to the latent attrition and repeat purchase process, seasonality was evident in the second calendar quarter for repeat purchases, so we include a seasonal covariate to account for it. The estimated BG/ML model parameters are $\widehat{\mu}_p = -3.92$ (.042), $\widehat{\sigma}_p^2 = 1.67$ (.036), $\widehat{\beta}_p = .06$ (.001), $\widehat{\gamma} = .53$ (.014), and $\widehat{\delta} = 3.65$ (.873). The actual data shown in the validation plots are constructed using the granular data. On the left within Figure 4.5, we show the expected and actual number of people making 0, 1, ..., 10+ repeat purchases during the calibration period. We cor-

rectly infer that the majority of customers (71%) are "one and done," making only their initial purchase and no subsequent repeat purchases. This suggests that the company's overall valuation is currently reliant upon a relatively small number of highly engaged customers. We will study this in more detail in the next section.

On the right within Figure 4.5, we plot the expected and actual number of purchases a customer will make in the holdout period, conditional upon the number of purchases the customer makes within the calibration period. We observe that the BG/ML model generates accurate predictions of expected behavior in the holdout period conditional upon what has occurred within the calibration period.

In Figure 4.6, we examine the flow of repeat purchases over time. The plot on the left corresponds to the expected and actual number of repeat purchases quarter-by-quarter, during the calibration period and the holdout period. The plot on the right shows the expected and actual number of repeat purchases cumulatively over time. The model generates good predictions of aggregate repeat purchase behavior longitudinally over time. Note that repeat purchases generally increase during the calibration period because the introduction of newly acquired customers during the calibration period (Figure 4.4) more than offsets the decline in purchases due to latent attrition. Repeat purchases decline during the holdout period because we only make projections in the holdout period for customers acquired during the calibration period.

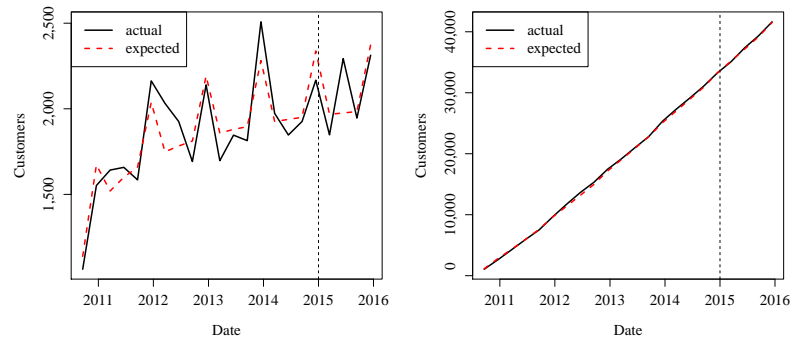Figure 4.4: Validating Incremental (left) and Cumulative (right) Gross Acquisitions

Figure 4.5: Predicted vs. Actual Frequency of Repeat Transactions (left), Conditional Expected Purchases (right)
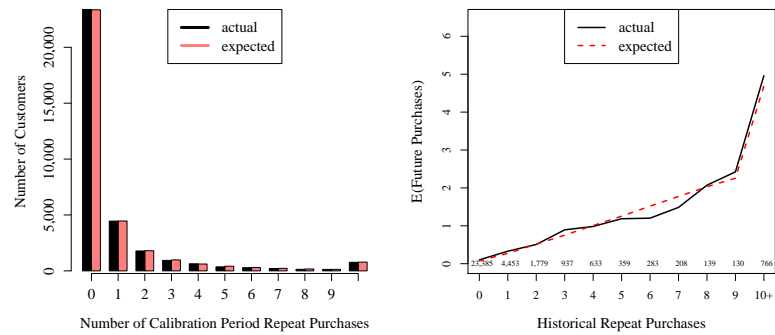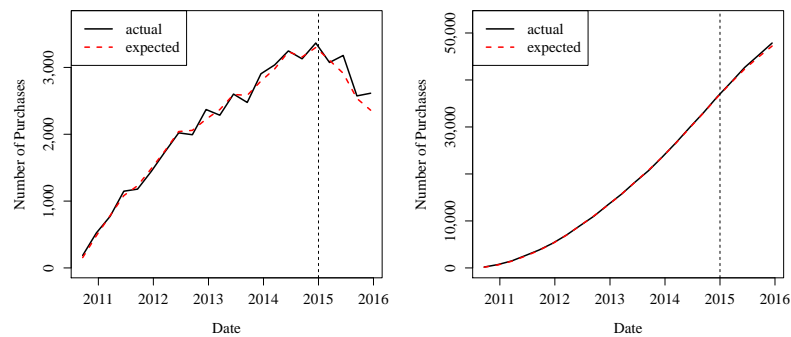

Figure 4.6: Validating Incremental (left) and Cumulative (right) Repeat Purchases

The final step is to validate the spend model. The estimated model parameters are $\widehat{\mu}_0 = 4.78$ (.018), $\widehat{\mu}_q = 1.63$ (.065), and $\widehat{\tau}^2 = .77$ (.009). To visualize the fit of the model, in Figure 4.7 we compute the expected marginal distribution of average spend across customers, and compare it to the empirical density of the observed average spends across customers. While the estimated modal spend amount is slightly less than the observed modal spend, we capture the overall shape of the distribution, including its long right tail. Figure 4.7 reinforces the need to model heterogeneity in spend across customers.

Figure 4.7: Marginal Distribution of Average Spend Across Customers
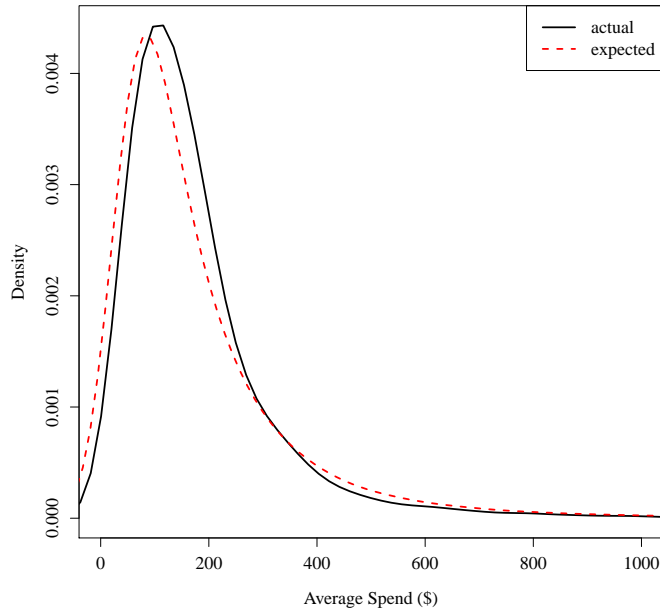
4.6.1.2 Model Comparison

Although the analysis thus far shows that the in-sample and out-of-sample fit of our proposed model is very good, it does not provide us with insight into how our model's performance compares to alternative models. In this section, we compare our

performance to two benchmark models.

We compare the proposed model to the methodology used in GLS. We also compare the proposed model to an extension of GLS (hereafter, GLS+) which allows for (1) seasonal fluctuation in the customer acquisition process, (2) time trend and seasonality covariates in the number of purchases made per active customer. In addition to validating in-sample and out-of-sample fit, we provide overall valuation estimates for the three models in the next section.

We summarize the relative performance of these three models in Table 4.3, which reports the mean absolute percentage error (MAPE) for (1) total quarterly customers acquired and (2) total quarterly purchases made by existing customers over the calibration and holdout periods.

	Customer Acquisitions			Total Purchases		
Model	Overall	OOS	IS	Overall	OOS	IS
Proposed	6.32	6.86	6.20	3.94	5.46	3.60
GLS+	26.20	53.41	19.80	13.56	15.25	13.16
GLS	29.12	58.84	22.13	41.59	25.08	45.48

Table 4.3: E-commerce Retailer: MAPE of Quarterly Customer Acquisitions and Total Purchases by Time Frame

The in-sample and out-of-sample fits of the proposed model are better than the benchmarks. We find that GLS underestimates future customer acquisitions. While allowing for seasonality in GLS's acquisition model reduces its error, the level of error remains high. This is because the Bass-like technological substitution model (TSM) for customer acquisition infers that customer acquisition has already hit its peak. Because the TSM must be symmetric about the period of peak acquisition (as with the Bass model), GLS predicts that future acquisitions will fall quickly in the holdout period.

GLS underestimates future total purchases more than it underestimates future

customer acquisitions, with MAPE values in excess of 25% in-sample and out-of-sample. GLS assumes a homogeneous retention rate which is equal to the then-current repeat rate. Because the vast majority of customers make only one purchase (Figure 4.5) the annual repeat rate is 25.9%, which quickly "kills off" active customers as we move forward into the holdout period. In reality, there is a considerable amount of heterogeneity in customers' propensity to remain with the firm, as evidenced by the repeat purchase parameter estimates. While most customers will churn quickly, some customers will remain with the firm for a long time. This underestimation is mitigated but not eliminated with the GLS+ model, which allows the number of purchases per active customer to trend upwards over time. For a more detailed analysis of these alternative specifications, see Appendix A.6.

4.6.2   Valuation

Following the process we outlined in the valuation procedure section, we move from a one-year forecast to 50-year ($2,600$-week) forecast, which is far enough into the future that the time value of money makes the incremental impact of all future profit or loss upon overall valuation negligible. This is the same forecasting horizon that we used to perform contractual CBCV in essay two. In Figure 4.8, we plot actual and expected quarterly customer acquisitions (top panel), total purchases (middle panel), and revenues (bottom panel). The acquisitions forecasts in Figure 4.8 are formed by "compressing" the observed acquisition data in Figure 4.4 to make room for a 50-year holdout period. These valuation results are not sensitive to the estimated total number of prospects, which was provided to us by the firm. Changes in the total number of prospects affect profits in later years the most, but those later years are of lesser importance to overall valuation because of the time value of money.

The weekly customer acquisition estimates over the 50-year forecasting horizon af-

Figure 4.8: E-commerce Retailer: Summary of Projections
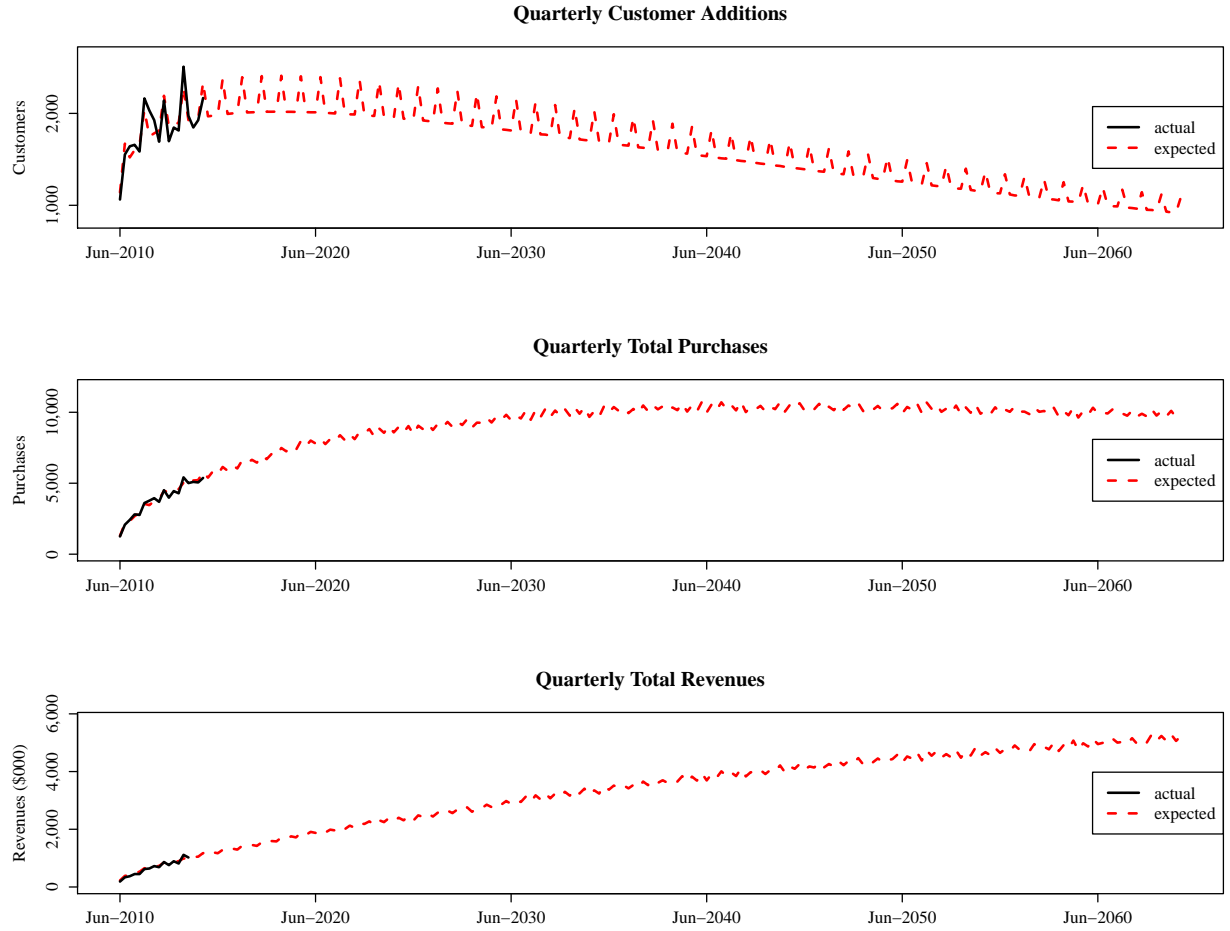
ter the 4.5-year calibration period (A(235),A(236),...,A(2834)), and the corresponding revenue estimates over the same period (R(235),R(236),...,R(2834)) are the primary customer model-driven inputs to future FCF using Equation 4.1. Projections of other inputs needed to estimate SHV in Equation 2.1 were provided to us by the company's management team. We use these figures to obtain weekly FCF projections over the forecasting horizon, the net present value of the resulting FCF stream, and finally, shareholder value (SHV). We estimate that SHV is equal to $22.8MM. The corresponding estimated valuations using GLS and GLS+ are $2.6MM and $2.8MM,

respectively. We provide a summary of the inputs driving the overall valuation in Table 4.4.

Table 4.4: Summary of Valuation

Total initial prospects (POP(0))	500,000
Annual growth rate of prospects	0%
Variable contribution margin (VC)	76.4%
Weekly fixed costs (FC)	$4,260
Cost per acquired customer (CAC)	$76
Tax rate (TR)	35%
Weighted average cost of capital (WACC)	6%
Non-operating assets - Net debt (NOA - ND)	$0
Shareholder value (SHV)	$22.8MM

4.6.3 Model Implications

In addition to estimating the overall valuation of the business unit, we may also use our model to make inferences into the value of newly acquired customers. The average customer lifetime value (CLV) of a newly acquired cohort of customers will be $76.2 per customer (a return of approximately 100% over the customer's acquisition cost), but that there will be substantial variability in profitability across these customers. The expected distribution of CLV across newly acquired customers is shown in Figure 4.9.

The overall valuation of the business unit is driven by a small number of customers. For example, we predict that 68% of newly-acquired customers will not be profitable and just 2.9% of newly-acquired customers will generate 80% of their collective value. These inferences are consistent with what we would have been able to ascertain if we had access to the granular data. For example, restricting our attention to customers acquired during the first year of the data set, 64% of customers were unprofitable and only 2.4% of customers generated 80% of their collective net profit by the end of the

Figure 4.9: Expected Distribution of CLV for Newly Acquired Customers
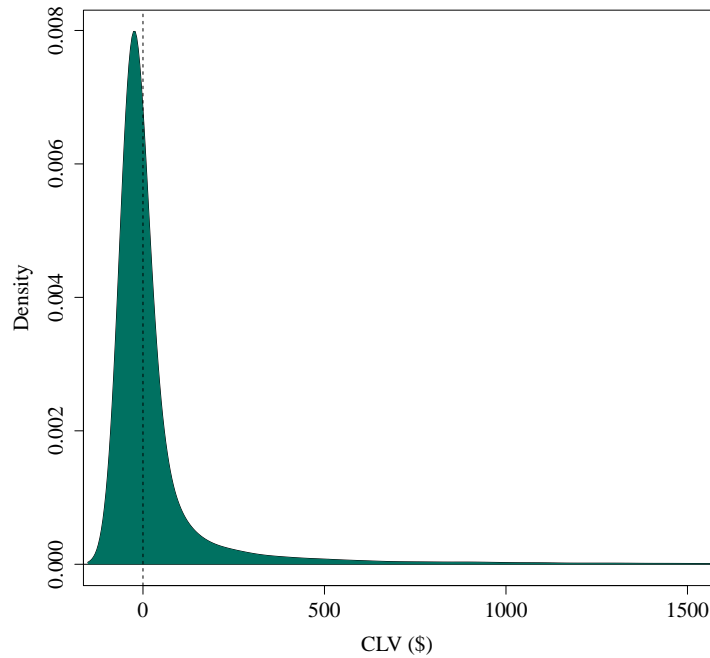
data set.

We may also infer how much of the overall valuation of the firm will come from existing customers versus not-yet-acquired customers. We estimate that the net present value of all future profits expected from the existing customer base (current customer equity or CCE) is \$3.5MM, or approximately 15% of the \$22.8MM overall valuation of the firm. This relatively low percentage is to be expected, given the relatively young age of the business unit. CCE is predicted to represent an increasing proportion of the overall valuation of the business unit in the future, as the existing customer base is increasingly comprised of customers with high retention and purchase propensities.

The business unit's valuation will be a "concentrated growth story." The chief valuation driver is whether the business unit can continue acquiring "die-hard" customers, whose CLV's exceed \$250 on average, because these highly engaged customers will generate virtually all of the business unit's profits and recoup the losses that the

firm will incur on the 68% of acquired customers who will be unprofitable.

These results also suggest that the business unit may increase its valuation if it is able to find a way to reduce the number of unprofitable customers and/or acquire fewer like them, because they currently represent a rather large proportion of newly-acquired customers. When we spoke with an executive at the retailer about this, he noted that the relatively large number of unprofitable customers are most likely due to international shipping issues (e.g., delays and errors) causing new customers to sour on the company after their initial purchases and not make repeat purchases. Fixing these shipping issues could reduce the large number of "one and done" customers in future acquisition cohorts, which could increase the overall valuation of the business if the incremental value created exceeds the cost of the fix.

4.7 Discussion

This essay proposes a novel methodology with which common customer metrics are used to estimate a latent variable model for non-contractual customer acquisition, repeat purchase, and spend. We not only show that some collections of customer metrics can reliably identify the customer model, but also which collections are the most predictive. The customer model is used to drive an overall valuation model for non-contractual firms, provide helpful color about where the value is coming from, and how much individual customers are worth.

The methodology has uses which extend beyond corporate valuation. These techniques may be useful for expert testimony in litigation cases where firms would like to provide enough information to confirm or deny specific points raised within the case, but no more than that. The methodology may also be useful in industries that would not usually be amenable to such an analysis. For example, it may be the case for

consumer packaged goods firms that customer acquisition figures are not available, but that a larger panel of repeat purchase metrics is. The right collection of repeat purchase metrics may identify the acquisition process as well. We leave these topics to future work.

More generally, the performance of the proposed indirect inference methodology in our large-scale simulation analysis and the empirical application highlights the promise of II for model estimation in large data settings. Marketing scholars working with datasets that are prohibitively large in size should consider (1) compressing the data to a smaller set of data summaries appropriate to the model being specified, (2) setting up the II procedure to estimate the specified model's parameters with the data summaries, and (3) validating the recoverability of the specified model's parameters through a simulation analysis using the same data summaries observed over the same duration of time. This procedure may yield solutions to many marketing problems that would otherwise be prohibitively difficult to answer.

As companies, business intelligence firms, and investors realize these other uses for customer metrics, we believe the demand for their disclosure will continue to grow. This would not be the first time – one of the most commonly disclosed and tracked retail metrics, same store sales (SSS), became popular because a Wall Street analyst showed just how useful it was at uncovering the true underlying financial condition of a fast-growing retailer in the 1970's (Blumenthal 2008). We believe the right set of customer metrics could allow investors to track the quality of existing customers much the same way that SSS allows investors to track the quality of existing stores. With physical stores ceding share to internet-based retail, the need for such a metric is more important than ever.

5

Areas of Future Research

We have developed general methodologies to perform customer-based corporate valuation with limited data in contractual and non-contractual settings. However, there are a number of extensions that we hope will be explored in the near future. Five extensions are discussed below:

(a) *Multiple company valuation analysis*: Throughout this dissertation, we focused our attention strictly upon conducting the valuation process for one company at a time, but our predictive accuracy may be improved if we were to study many companies at the same time through a hierarchical Bayesian model. This may alleviate some of the limited data issues by "borrowing strength" across firms. In addition, it may be informative to study competitive effects explicitly, modeling how customers allocate their limited discretionary income and time across companies within the same industry. However, both extensions require considerable methodological advancements to properly "share information" across firms and handle aggregated missing data.

(b) *Leveraging third party data sets*: As noted in chapter two, all data used to estimate our models are assumed to come from first-party data sources or third-party data sources which have negligible measurement error. The applicability of the methodology would be greatly extended if it could also be trained

upon third-party data sources which may have non-ignorable measurement error. Naively estimating the proposed model using data measured in error would result in error-in-variable bias (Fornell & Larcker 1981).

(c) *Active corporate value management with internal data*: All existing empirical CBCV literature has assumed that the modeler's objective is to passively measure the value of the firm using external data. If the modeling objective is passive measurement of company value, predicting future revenues and profits most accurately is of paramount importance. Correcting for endogeneity and including pricing or marketing mix data are of secondary importance in this setting, if at all. Incorporating endogenous variables into the customer base model would require additional data that may not be available, and an endogeneity-corrected valuation model may have lower holdout predictive validity than the same model which does not correct for endogeneity (Ebbes et al. 2011). However, there are a number of stakeholders who may be interested in using CBCV to both measure and manage the value of a firm over time. Private equity and venture capital firms may make an investment in a firm and then actively seek to increase the value of their investment, typically by changing high-level resource allocations (e.g., a private equity firm may invest in a capital-constrained company so that the company can increase its customer acquisition spending). The valuation of the firm after these allocations have been made will then be a very important consideration for whether the active investment firm will make the investment in the first place. Similarly, internal company management may be interested in using CBCV as an accountability mechanism for marketing campaigns, so that company management can estimate the value created through the campaign relative to the cost of the campaign itself (Day & Fahey 1988; Doyle 2009; Rappaport 1986; Srivastava et al. 1998). In this setting, it is more impor-

tant to understand the causal relationship between marketing actions/strategy and overall firm valuation, so the modeler may desire an endogeneity-corrected valuation model. An endogeneity-corrected valuation model will provide unbiased parameter estimates associated with important endogenous variables, even though this may diminish the model's prediction accuracy, raising interesting trade-offs and methodological challenges.

(d) *"Best in class" predictive CBCV modeling with internal data*: While internal data makes it feasible to consider adding strategically important endogenous variables to the CBCV model, internal data could also be used to simply make the CBCV model as accurate as possible. Company management could benefit from a highly accurate customer base model in a number of different ways. As a sales forecasting tool, it could serve as a reference point for (1) sales guidance it may provide to Wall Street analysts, (2) quotas it may set for the salesforce, because the sales forecasts are available at an account-by-account level, and (3) budgetary planning (e.g., expected inventory purchasing at the beginning of a selling season). As a valuation tool, it could serve as a reference point for corporate buybacks/share repurchases and equity issuance (Buckley 1989). If the value of the firm in the stock market is well above the internal data CBCV estimate, management may be inclined to issue new equity (or pay for employee compensation and/or corporate acquisitions through company stock instead of cash). It could also be included as a metric which factors into the level of executive compensation. While there are a wide variety of effects that could be incorporated into the model given the large quantity of granular data which is available, it is unclear what effects would be most helpful to include in such a model. The effects mentioned in Section 2.5 may be a helpful place for modelers to start.

(e) *The decision to disclose*: One of the limitations of the methodologies that we propose in essays two and three is that they assume the disclosure decision is not strategic. It could be that there is a forward-looking component to firms' decision to disclose (Mintz et al. (2016)), which may explain some of the residual variation in disclosure after controlling for business type and industry. The lack of non-contractual companies disclosing metrics makes it impossible to verify whether or not this is the case empirically for non-contractual firms. Furthermore, the large scale disclosure analysis we performed in Section 2.4 suggested that firms tend to disclose metrics relatively early on in their lives and do not tend to stop disclosing after they have started, which would suggest the decision to disclose would have little effect upon the resulting valuation estimates (similarly, Zantedeschi et al. (2016) use the lack of selectivity of the firm in their analysis to argue against endogeneity). However, the methodology we presented in essay three for non-contractual CBCV would allow us to perform a deeper and more systematic study.

We hope that this work encourages researchers to study this promising area in the future. We would be very happy to collaborate on these or other aspects of CBCV.

A

Appendices

A.1   Supplement for Chapter 3

Appendix

RLV can be expressed mathematically as

$$E(\text{RLV}) = \int_{t'}^{\infty} \text{E}[\text{V}(t)]\text{S}(t|t > t')\text{d}(t - t')\,\text{d}t\,,$$

where t' is the tenure of the customer at the point in time where their residual lifetime value is computed, $\text{E}[\text{V}(t)]$ is the expected net cash flow of the customer at time t (assuming they are alive at that time), $\text{S}(t|t > t'))$ is the probability that the customer has remained alive to at least time t (given they were alive at t', and $\text{d}(t - t')$ is a discount factor that reflects the present value of money received at time t (Fader and Hardie 2015).

This is acceptable as a mathematical representation of the definition of (expected) RLV but is of limited use in practice, as it ignores the accounting issues identified in Section 3.4.3.1. However, these accounting issues are considered when performing our valuation and an intermediate result from these calculations can be used to compute RLV. The intermediate result of interest is EBITDASAC, earnings before interest, taxes, depreciation, amortization, and subscriber acquisition costs.

Let $\tau_{m,M}^{(k)}$ be the kth sampled residual lifetime of a customer acquired in month m who is still active in month M, which is drawn from the distribution

$$\text{P}(\text{T}_{m,M} = \tau) = \frac{\text{S}_R(\tau + M - m - 1|m) - \text{S}_R(\tau + M - m|m)}{\text{S}_R(M - m|m)}, \quad \tau = 1, 2, \dots. \quad \text{(A1)}$$

The expected residual lifetime of a customer acquired in month m who is still active in month M can be computed as the average of many samples drawn from this

distribution:

$$
\mathrm{E}(\mathrm{T}_{m,M}) \;=\; \frac{1}{K}\sum_{k=1}^{K}\tau_{m,M}^{(k)}\,. \tag{A2}
$$

We set $K = 1{,}000{,}000$ in our analysis.

Given monthly EBITDASAC numbers, the value of an average customer in month $M + m^*$ is

$$
\frac{\mathrm{EBITDASAC}(M + m^*)}{\left[\widehat{\mathrm{C}}(.,M+m^*-1)+\widehat{\mathrm{C}}(.,M+m^*)\right]/2}\,.
$$

Therefore, the pre-tax RLV of a customer with sampled residual lifetime $\tau_{m,M}^{(k)}$, is

$$
\mathrm{rlv}_{m,M}^{(k)} \;=\; \sum_{m^*=1}^{\infty} \frac{\mathrm{EBITDASAC}(M + m^*)}{\left[\widehat{\mathrm{C}}(.,M+m^*-1)+\widehat{\mathrm{C}}(.,M+m^*)\right]/2}\frac{\mathbb{1}(m^* \le \tau_{m,M}^{(k)})}{(1+\mathrm{WACC})^{m^*}}, \tag{A3}
$$

(The empirical distribution of these draws is plotted in Figure 3.5.) Averaging over these sampled realizations of residual lifetime gives us the expected RLV of a customer acquired in month m who is still active in month M,

$$
E(\mathrm{RLV}_{m,M}) \;=\; \frac{1}{K}\sum_{k=1}^{K}\mathrm{rlv}_{m,M}^{(k)}\,. \tag{A4}
$$

A.2 Supplement for Chapter 4

A.3 Acquisition Parameter Recovery Analysis

We test our ability to recover the parameters of the acquisition model across a wide variety of data generating processes. We consider $3^3 = 27$ different scenarios: $r_A \in (0.5, 1.0, 1.5)$, median time to acquisition of 10, 15, or 20 years; and

$c_A \in (1.0, 1.25, 1.50)$. We simulate five years of weekly data, aggregate the weekly data to report acquisition data every quarter, and estimate the parameters by MLE. We assume an initial population of 1MM households and 1.2% annual growth in the population. For each scenario considered, we report the mean absolute percent error (MAPE) for each parameter, as well as the overall MAPE averaged across the parameters. Results are averaged over 100 replications and are shown in Table A.1.

Table A.1: True Parameters, Estimated Parameters, and MAPE

True Parameters			MAPE (%)			
r_A	α_A	c_A	r_A	α_A	c_A	Overall
0.5	173.3	1.00	.40	.27	.09	.25
1.0	520.0	1.00	.83	.60	.09	.51
1.5	885.3	1.00	1.30	1.04	.09	.81
.5	260.0	1.00	.56	.38	.10	.35
1.0	780.0	1.00	1.28	1.01	.10	.80
1.5	1327.9	1.00	2.13	1.82	.11	1.35
.5	346.7	1.00	.73	.50	.11	.45
1.0	1040.0	1.00	1.86	1.54	.12	1.17
1.5	1770.5	1.00	2.97	2.65	.12	1.91
.5	827.7	1.25	.42	.29	.08	.27
1.0	2483.2	1.25	.95	.63	.09	.56
1.5	4227.4	1.25	1.49	1.14	.09	.91
.5	1374.0	1.25	.66	.41	.10	.39
1.0	4122.1	1.25	1.70	1.30	.11	1.04
1.5	7017.5	1.25	2.80	2.35	.12	1.75
.5	1968.7	1.25	.94	.60	.11	.55
1.0	5906.0	1.25	2.71	2.23	.13	1.69
1.5	10054.4	1.25	4.76	4.18	.14	3.02
.5	3952.6	1.50	.47	.35	.09	.30
1.0	11857.8	1.50	1.14	.72	.10	.65
1.5	20186.9	1.50	1.85	1.37	.10	1.11
.5	7261.4	1.50	.88	.50	.11	.50
1.0	21784.2	1.50	2.45	1.78	.13	1.45
1.5	37085.8	1.50	4.34	3.57	.14	2.68
.5	11179.7	1.50	1.47	.90	.14	.84
1.0	33539.0	1.50	4.56	3.72	.16	2.81
1.5	57097.3	1.50	8.11	7.18	.17	5.15

A.4 Auxiliary Model Specifications

In this section, we specify the auxiliary models for all processes except the acquisition process, which was specified in the section "Acquisition Process Parameter Estimation."

There are two non-parametric auxiliary models associated with the repeat purchase process. One auxiliary model is associated with AU, HAU, and F while the other auxiliary model is associated with RR, RBPC, and RBPO. As with the acquisition process, these non-parametric auxiliary models were chosen to minimize loss of information.

We posit an auxiliary model associated with AU, HAU, and F in which all customers who will eventually be acquired by the end of the calibration period, denoted by A^* (i.e., $A^* = \sum_{i=1}^{Q} \text{ADD}_q$), are exchangeable with one another and may make purchases in the first quarter of the data. The auxiliary model does not distinguish between initial purchases and repeat purchases. The number of purchases made in quarter q by all customers who will eventually be acquired by the end of the calibration period is determined by a multinomial$[\pi_0^{(p)}(q), \ldots, \pi_{12}^{(p)}(q)]$ draw. $\pi_x^{(p)}(q)$ is the probability that the customer makes x purchases within quarter q, and the multinomial distribution associated with each quarter is independent of other quarters. Assume that we observe an arbitrary collection of n AU disclosures, $\mathbf{AU}$, where

$$\mathbf{AU} = [\text{AU}_{q_1}, \text{AU}_{q_2}, \cdots, \text{AU}_{q_n}],$$

whose ith element, AU_{q_i}, is equal to the number of customers who made at least one purchase in the four quarters preceding quarter q_i. Then AU_{q_i}/A^* is equal to the complement of the probability of no purchases taking place over the prior four

116

quarters under the auxiliary model:

$$\mathrm{AU}_{q_i}/\mathrm{A}^* = 1 - \widehat{E}\left[\prod_{t=0}^{3} \pi_0^{(p)}(q_i - t)\right].$$

Assume that we observe an arbitrary collection of n HAU disclosures, **HAU**, where

$$\mathbf{HAU} = [\mathrm{HAU}_{q_1}, \mathrm{HAU}_{q_2}, \cdots, \mathrm{HAU}_{q_n}],$$

whose ith element, HAU_{q_i}, is equal to the number of customers who made at least two purchases in the four quarters preceding quarter q_i. Then $\mathrm{HAU}_{q_i}/\mathrm{A}^*$ is equal to the complement of the probability of either zero or one purchases taking place over the prior four quarters under the auxiliary model:

$$\mathrm{HAU}_{q_i}/\mathrm{A}^* = 1 - \widehat{E}\left[\prod_{t=0}^{3} \pi_0^{(p)}(q_i - t)\right] - \widehat{E}\left[\sum_{s=0}^{3}\prod_{t=0}^{3} \pi_{\mathbb{1}(s=t)}^{(p)}(q_i - t)\right],$$

where $\mathbb{1}(s = t)$ is a binary indicator variable equal to one if $s = t$ and zero otherwise.

Finally, assume that we observe an arbitrary collection of n F disclosures, **F**, where

$$\mathbf{F} = [\mathrm{F}_{q_1}, \mathrm{F}_{q_2}, \cdots, \mathrm{F}_{q_n}],$$

whose ith element, F_{q_i}, is equal to the average number of purchases made over the past 12 months by a customer, given the customer was active over this time period. Using Bayes' theorem, it follows that

$$\mathrm{F}_{q_i} = \widehat{E}\left[\frac{\sum_{t=0}^{3}\sum_{s=1}^{13} s\pi_s^{(p)}(q_i - t)}{1 - \prod_{t=0}^{3}\pi_0^{(p)}(q_i - t)}\right]$$

We posit a different auxiliary model associated with RR, RBPC, and RBPO. All A^* customers who will eventually be acquired by the end of the calibration period

are exchangeable with one another and may make purchases in the first quarter of the data, as with the auxiliary model associated with AU, HAU, and F. However, the number of purchases made each quarter for each customer depends upon whether the customer made at least one purchase in any prior quarter or not. The number of purchases made in quarter q by all customers who have not yet made their first purchase is determined by a multinomial$[\pi_{A,0}^{(p)}(q), \pi_{A,1}^{(p)}(q), \cdots, \pi_{A,13}^{(p)}(q)]$ draw, while the number of purchases made in quarter q by all customers who have made at least one purchase in any prior quarter is determined by a multinomial$[\pi_{B,0}^{(p)}(q), \pi_{B,1}^{(p)}(q), \cdots, \pi_{B,13}^{(p)}(q)]$ draw. Conditional upon the prior purchase history of the customer, the multinomial distributions associated with each quarter and each regime are independent of other quarters.

Assume that we observe an arbitrary collection of n RR disclosures, **RR**, where

$$\mathbf{RR} = [\mathrm{RR}_{q_1}, \mathrm{RR}_{q_2}, \cdots, \mathrm{RR}_{q_n}],$$

whose ith element, RR_{q_i}, is equal to the proportion of customers who made one or more purchases within the past four quarters, given that they made one or more purchases one year ago (i.e., five to eight quarters ago). This is the complement of making no purchases over the past four quarters, given that the customer made a purchase prior to the beginning of the year:

$$\mathrm{RR}_{q_i} = 1 - \widehat{E}\left[\prod_{t=0}^{3} \pi_{B,0}^{(p)}(q_i - t)\right]$$

Similarly, assume that we observe an arbitrary collection of n RBPC disclosures,

$$\mathbf{RBPC} = [\mathrm{RBPC}_{q_1}, \mathrm{RBPC}_{q_2}, \cdots, \mathrm{RBPC}_{q_n}],$$

whose ith element, RBPC_{q_i}, is equal to the proportion of customers who made one or more purchases before the year began, given that they made one or more purchases within the past year. By Bayes' Theorem, this is equal to the probability that the customer made one or more purchases in both periods divided by the probability that the customer made one or more purchases this year, where the probability that the customer made one or more purchases this year depends upon whether or not the customer made one or more purchases before the year began:

$$
\begin{aligned}
\text{RBPC}_{q_i} &= \widehat{E}\left(\frac{A_1}{A_1 + B_1}\right), \quad \text{where} \\
A_1 &= \left[1 - \prod_{t=0}^{3} \pi_{B,0}^{(p)}(q_i - t)\right] \left[1 - \prod_{t=4}^{q_i-1} \pi_{A,0}^{(p)}(q_i - t)\right] \quad \text{and} \\
B_1 &= \left[1 - \prod_{t=0}^{3} \pi_{A,0}^{(p)}(q_i - t)\right] \left[\prod_{t=4}^{q_i-1} \pi_{A,0}^{(p)}(q_i - t)\right]
\end{aligned}
$$

Finally, assume that we observe an arbitrary collection of n RBPO disclosures,

$$
\mathbf{RBPO} = [\text{RBPO}_{q_1}, \text{RBPO}_{q_2}, \cdots, \text{RBPO}_{q_n}],
$$

whose ith element, RBPO_{q_i}, is equal to the proportion of orders within the past year which were made by customers who made one or more purchases before the year began. This is equal to the expected number of orders made this year by customers who purchased one or more times before the year began, divided by the number of orders made this year:

$$
\begin{aligned}
\text{RBPO}_{q_i} &= \widehat{E}\left(\frac{A_2}{A_2 + B_2}\right), \quad \text{where} \\
A_2 &= \left[\sum_{t=0}^{3}\sum_{s=1}^{13} s\pi_{B,s}^{(p)}(q_i - t)\right] \left[1 - \prod_{t=4}^{q_i-1} \pi_{A,0}^{(p)}(q_i - t)\right] \quad \text{and}
\end{aligned}
$$

$$B_2 = \left[\sum_{t=0}^{3}\sum_{s=1}^{13} s\pi_{A,s}^{(p)}(q_i - t)\right]\left[\prod_{t=4}^{q_i-1} \pi_{A,0}^{(p)}(q_i - t)\right]$$

In summary, **AU/A***, **HAU/A***, **F**, **RR**, **RBPC**, and **RBPO** are the auxiliary statistics associated with non-parametric auxiliary models, and have been selected so that they converge asymptotically to (finite-valued) true population-level figures. We obtain the repeat purchase auxiliary parameters corresponding to these auxiliary statistics by repeatedly simulating from the true acquisition and repeat purchase processes.

The spend auxiliary model is the true model itself. The parameters of the spend process are identified by the empirical quarterly mean and median of spend amounts over time. The empirical median of spend amounts each quarter is directly disclosed by the firm. The empirical mean of spend amounts each quarter is known, conditional upon the parameters of the acquisition, latent attrition and repeat purchase processes, in conjunction with total quarterly revenues, which is directly disclosed by the firm.

A.5   Large-Scale Simulation Details

In this section, we specify the 64 "worlds" or data sets which we generate data from to evaluate the predictive validity of the proposed repeat purchase customer metric pairs. We obtain the parameter values corresponding to each world through a full factorial design, perturbing the parameter values estimated from a canonical data set which has been studied and benchmarked extensively in the marketing literature, the CDNOW dataset (Abe 2009; Fader et al. 2005b; Zhang et al. 2014).

This dataset consists of 2,357 customers who were acquired in the first quarter of 1997 and observed over a period of 1.5 years. We discretize the data at the weekly level, modeling customers' decision to make at least one purchase each week, and

summing all spend amounts which occur during purchase-weeks. There is no seasonal pattern in the data, so we do not include a seasonal covariate in the model.

While it is possible to estimate the parameters of the BG/ML model via maximum marginal likelihood, this is computationally expensive because the marginal likelihood expression is not available in closed-form. Instead, we use latent method of moments estimation. We estimate the parameters of a model which only differs in terms of the parametric distribution used to characterize unobserved heterogeneity in the repeat purchase process but whose marginal likelihood is available in closed-form, the Beta-Geometric/Beta-Binomial model (Fader et al. 2010). We then match the moments of the purchase and death heterogeneity distributions of the BG/ML with those of the BG/BB[1].

The estimated hyperparameters of the weekly BG/BB model applied to the CD-NOW dataset are $\alpha = .58, \beta = 13.13, \gamma = .39$, and $\delta = 9.72$. We match the mean and variance of the heterogeneity distributions with respect to the individual-level purchase and death hyperparameters, q and θ respectively. Because the heterogeneity distributions with respect to θ are the same for the BG/BB model and the BG/ML model, the estimated death hyperparameters for the BG/ML are the same as the estimated death hyperparameters for the BG/BB.

For the purchase process, we simulate a large number (1MM) of samples of q from the beta(α, β) distribution for the BG/BB model. We then apply the logit function to these values of q, computing $\text{logit}(q^{(k)}) = \log[q^{(k)}/(1 - q^{(k)})]$ for each sample k. The empirical mean and variance of these 1MM $\text{logit}(q)$ samples are -4.15 and 3.91, respectively. Therefore, the latent method of moments estimator for the BG/ML

---

[1]The BG/ML model without covariates assumes that the per-period probability of purchase while alive, q, is distributed across the population according to a logit-normal(μ_p, σ_p^2) distribution, while the BG/BB model assumes that heterogeneity in q is beta(α, β)-distributed. Both models share the same heterogeneity distribution with respect to the death process.

Figure A.1: Predicted vs. Actual Frequency of Repeat Transactions
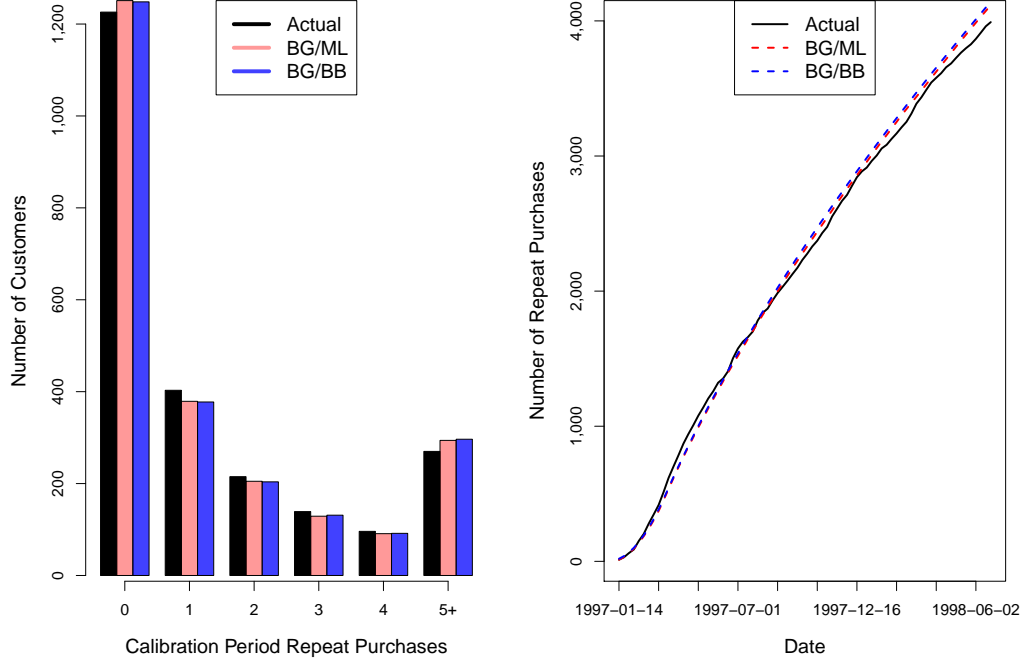
model hyperparameters is

$$\mu_p = -4.15, \quad \sigma_p^2 = 3.92, \quad \gamma = .388, \quad \delta = 9.724 \tag{A5}$$

The corresponding model fits and forecasts for the BG/ML are virtually identical to those of the BG/BB, as shown in Figure A.1.

We assume the following two parameter sets for the acquisition process:

$$r_A = .5, \quad \alpha_A \in (5523.9, \ 15624.1), \quad c_A = 1.5, \quad \beta_A = \log(2) \tag{A6}$$

β_A is set such that the baseline hazard function is doubled during the fourth calendar quarter. These scenarios correspond to a median baseline time until acquisition of 12.5 and 25 years, respectively. By the end of the calibration period, 32.4% and 16.2%

of the total prospect pool will have been acquired, respectively.

The 64 data sets are created by varying the parameters of the acquisition, and repeat purchase models along six dimensions relative to the baseline parameters in Equations A5 and A6. We consider high purchase rate and purchase homogeneity scenarios by doubling μ_p and σ_p^2. We also consider high death rate and death homogeneity scenarios by doubling the mean and polarization indices of the beta(γ, δ) distribution, $\gamma/(\gamma + \delta)$ and $1/(1 + \gamma + \delta)$, respectively (Sabavala & Morrison (1977)). Finally, we consider a high seasonality scenario in which the baseline purchase rate is doubled during the fourth calendar quarter (i.e., $\beta_p = |\mu_p|$).

A.6   GLS Model Valuation and Comparison

In this section, we elaborate upon the model forecasts and valuations provided in the empirical analysis of Section 4.6. GLS estimates the parameters of the acquisition process using the technological substitution model (TSM). However, seasonality is evident in the fourth calendar quarter. Therefore, we create a variant of the TSM which incorporates covariates using proportional hazards. Using GLS notation (Equation 9 in GLS), denoting the time t vector of seasonal covariates by $\mathbf{x}_A(t)$, and the corresponding vector of parameter estimates by $\boldsymbol{\delta}_A$, the extended expression for the cumulative number of customers N_t at any time t is given by

$$N_t = \alpha \times \left(1 - \exp \left(\sum_{i=1}^{t} -\gamma + \log \left(\frac{1 + \exp(-\beta - \gamma(i-1))}{1 + \exp(-\beta - \gamma i)} \right) \exp[\boldsymbol{\delta}_A^T \mathbf{x}_A(i)] \right) \right)$$

Parameters are estimated via non-linear least squares, fitting the acquisition parameters to N_t, the cumulative customer additions data. We provide the resulting incremental and cumulative quarterly estimates of customer acquisitions for our pro-

Figure A.2: Incremental (left) and Cumulative (right) Customers Acquired
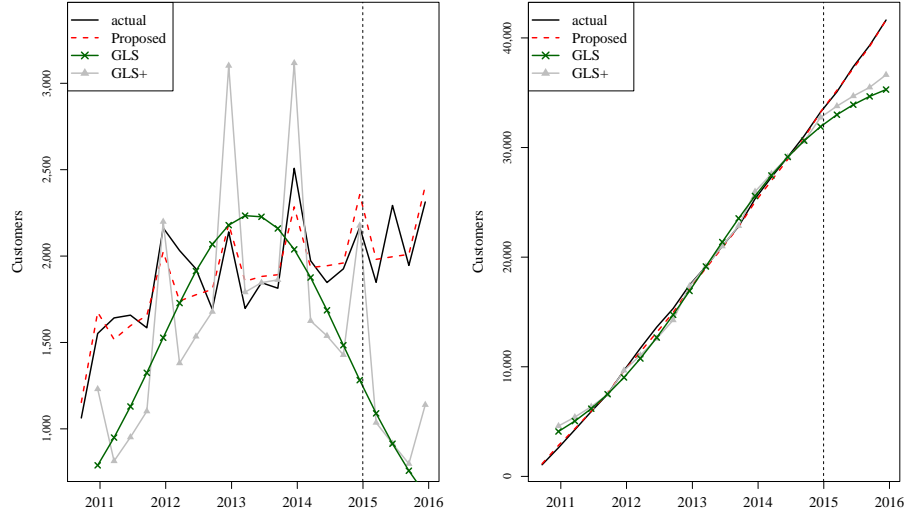
posed method, GLS, and GLS+ in Figure A.2, alongside the observed data.

The estimated acquisition parameters for GLS and GLS+ are shown below, using notation from GLS and denoting by β_{Q_4} the quarterly seasonal dummy:

	α	β	γ	β_{Q_4}
GLS	37.823	-2.345	.237	.000
GLS+	39.645	-2.377	.208	.507

The annual repeat rate for the e-commerce retailer is 25.9% as of the end of the calibration period. Therefore, the retention rate for GLS is 25.9%. It is unclear how this process can be easily modified. If customer churn were observed, we could allow the retention rate to vary over time as a function of covariates through a logit formulation. Because churn is not observed, we assume a 25.9% retention rate for GLS+ but counteract this assumption through the repeat purchase process next.

While GLS does not explicitly model future purchases, it does model the dollar margin per active customer, which is assumed to be equal to the trailing four quarter average. Dollar margin per active customer can be decomposed into margin per

purchase and the number of purchases per active customer. Assuming that these latter two quantities are constant because their product is constant, future purchases per active customer-quarter under GLS are equal to .495.

The historical evolution of inferred repeat purchases per active customer is upward-sloping over time, with a seasonal increase in the second calendar quarter. Therefore, we create a variant of the purchase process which allows for both a time trend and seasonality. We estimate the following equation for GLS+ purchases per active customer, denoting by q the number of quarters after the beginning of commercial operations:

$$\frac{\widehat{\text{Purchases}}}{\text{Active Customer}} = .254 + .0116 \times q + .0372 \times \mathbb{1}[Q(q) = 2],$$

where $\mathbb{1}[Q(q) = 2]$ is an indicator variable equal to one if quarter q is within the second calendar quarter and zero otherwise. The incremental and cumulative quarterly estimates of purchases for our proposed method, GLS, and GLS+ are shown in Figure A.3, alongside the observed data.

While allowing purchases per active customer to trend over time improves the GLS model's fit and forecast dramatically, it continues to underestimate future purchases.

Using the assumptions made above, dollar sales per purchase is equal to the trailing four quarter average of $221.80. However, sales per purchase is rising linearly during the calibration period. Therefore, we extend GLS to allow sales per purchase to have a time trend which is estimated with a regression. We estimate the following equation for GLS+ spend per purchase:

$$\frac{\widehat{\text{Spend}}}{\text{Purchase}} = 145.58 + 4.621 \times q.$$

Using the fitted parameter estimates for the GLS and GLS+ customer acquisition, retention, purchases per active customer and spend per purchase, we make future

Figure A.3: Incremental (left) and Cumulative (right) Repeat Purchases
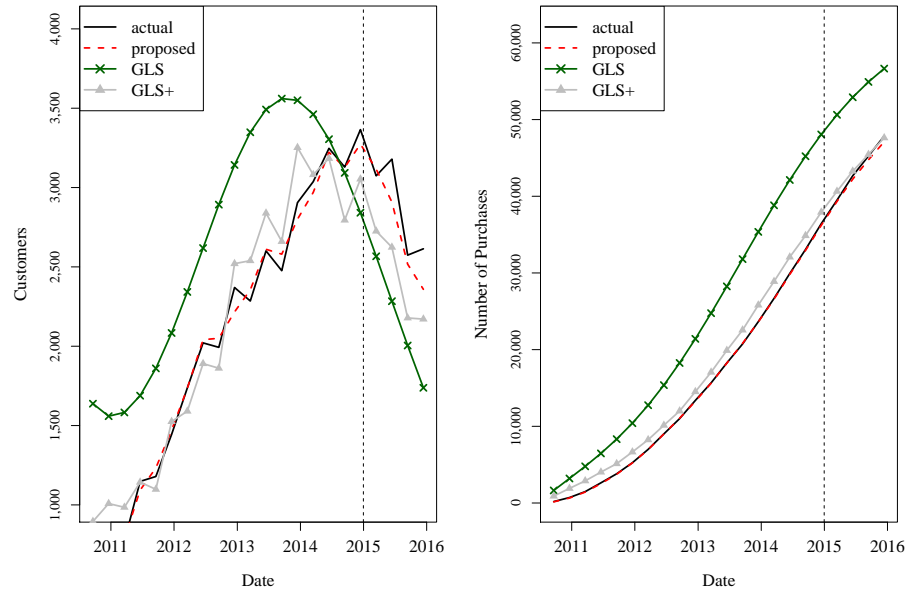
revenue projections. We then convert these revenue projections into overall valuation estimates using the same assumptions that we had made to convert the revenue forecasts of our proposed model into an overall valuation estimate.

Bibliography

Abe, M. (2009). Counting your customers one by one: A hierarchical Bayes extension to the Pareto/NBD model. *Marketing Science*, 28(3), 541–553.

Albuquerque, P. & Bronnenberg, B. J. (2009). Estimating demand heterogeneity using aggregated data: An application to the frozen pizza category. *Marketing Science*, 28(2), 356–372.

Alford, A. W. (1992). The effect of the set of comparable firms on the accuracy of the price-earnings valuation method. *Journal of Accounting Research*, 30(1), 94–108.

Anderson, E. W., Fornell, C., & Mazvancheryl, S. K. (2004). Customer satisfaction and shareholder value. *Journal of Marketing*, 68(4), 172–185.

Andon, P., Baxter, J., & Bradley, G. (2001). Calculating the economic value of customers to an organisation. *Australian Accounting Review*, 11(23), 62–72.

Arango, T. (2001). Exec says publisher declined to extend Amazon's credit terms. TheStreet.com. URL: https://news.thestreet.com/story/1340697/1/exec-says-publisher-declined-to-extend-amazons-credit-terms.html (visited on 2017-03-14).

Ascarza, E., Netzer, O., & Hardie, B. (2017). Some customers would rather leave without saying goodbye. *Marketing Science (forthcoming)*.

Bain and Company (2000). The value of online customer loyalty. Working Paper. URL: http://www.bain.com/Images/Value_online_customer_loyalty_you_capture.pdf (visited on 2017-03-14).

Bass, F. (1969). A new product growth for model consumer durables. *Management Science*, 15(5), 215–227.

Baxter, R. K. (2015). *The Membership Economy: Find Your Super Users, Master the Forever Transaction, and Build Recurring Revenue*. New York, NY: McGraw Hill Professional.

Bayer, E., Tuli, K. R., & Skiera, B. (2017). Do disclosures of customer metrics lower investors' and analysts' uncertainty, but hurt firm performance? *Journal of Marketing Research (forthcoming)*.

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 63(4), 841–890.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 25(2), 242–262.

Blattberg, R. C. & Deighton, J. (1996). Manage marketing by the customer equity test. *Harvard Business Review*, 74(4), 136.

Blattberg, R. C., Getz, G., & Thomas, J. S. (2001). *Customer equity: Building and managing relationships as valuable assets.* Boston, MA: Harvard Business Press.

Blumenthal, K. (2008). *Grande Expectations: A Year in the Life of Starbucks' Stock.* New York, NY: Crown Publishing.

Bodapati, A. V. & Gupta, S. (2004). The recoverability of segmentation structure from store-level aggregate data. *Journal of Marketing Research*, 41(3), 351–364.

Bonacchi, M., Ferrari, M., & Pellegrini, M. (2008). The lifetime value scorecard: From e-metrics to internet customer value. *Managerial and Financial Accounting*, 18, 193–226.

Bonacchi, M., Kolev, K., & Lev, B. (2015a). Customer franchise: A hidden, yet crucial asset. *Contemporary Accounting Research*, 32, 1024–1049.

Bonacchi, M., Kolev, K., & Lev, B. (2015b). Customer franchisea hidden, yet crucial, asset. *Contemporary Accounting Research*, 32(3), 1024–1049.

Bonacchi, M. & Perego, P. (2012). Measuring and managing customer lifetime value: A CLV scorecard and cohort analysis in a subscription-based enterprise. *Management Accounting Quarterly*, 14(1), 27–39.

Boyce, G. (2000). Valuing customers and loyalty: The rhetoric of customer focus versus the reality of alienation and exclusion of (de-valued) customers. *Critical Perspectives on Accounting*, 11(6), 649–689.

Boyd, J. H. & Mellman, R. E. (1980). The effect of fuel economy standards on the us automotive market: an hedonic demand analysis. *Transportation Research*, 14(5-6), 367–378.

Boylan, J. E. (1997). *The centralisation of inventory and the modelling of demand.* PhD thesis, University of Warwick.

Braun, M., Schweidel, D. A., & Stein, E. (2015). Transaction attributes and customer valuation. *Journal of Marketing Research*, 52(6), 848–864.

Buckley, F. H. (1989). When the medium is the message: Corporate buybacks as signals. *Indiana Law Journal*, 65, 493.

Campbell, J. Y. & Shiller, R. J. (1988a). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1(3), 195–228.

Campbell, J. Y. & Shiller, R. J. (1988b). Stock prices, earnings, and expected dividends. *The Journal of Finance*, 43(3), 661–676.

Cardell, N. S. & Dunbar, F. C. (1980). Measuring the societal impacts of automobile downsizing. *Transportation Research*, 14(5-6), 423–434.

Chen, Y. & Yang, S. (2007). Estimating disaggregate models using aggregate data through augmentation of individual choice. *Journal of Marketing Research*, 44(4), 613–621.

Cheong, Y., De Gregorio, F., & Kim, K. (2010). The power of reach and frequency in the age of digital advertising. *Journal of Advertising Research*, 50(4), 403–415.

Chintagunta, P. K. (1992). Estimating a multinomial probit model of brand choice using the method of simulated moments. *Marketing Science*, 11(4), 386–407.

Chintagunta, P. K. (2002). Investigating category pricing behavior at a retail chain. *Journal of Marketing Research*, 39(2), 141–154.

Damodaran, A. (2010). It is all relative multiples, comparables and value! Working Paper. URL: http://pages.stern.nyu.edu/ adamodar/pdfiles/country/relvalAIMR.pdf (visited on 2017-03-14).

Damodaran, A. (2012). *Investment valuation: Tools and Techniques for Determining the Value of Any Asset*. Hoboken, N.J.: John Wiley & Sons.

Damodaran, A. et al. (2005). Valuation approaches and metrics: a survey of the theory and evidence. *Foundations and Trends in Finance*, 1(8), 693–784.

Danaher, P. J. (2007). Modeling page views across multiple websites with an application to internet reach and frequency prediction. *Marketing Science*, 26(3), 422–437.

Dawson, G. (2017). At sweetgreen, a suitcase full of cash wont buy you lunch. Forbes.com. URL: https://www.nytimes.com/2016/07/31/business/where-a-suitcase-full-of-cash-wont-buy-you-lunch.html (visited on 2017-03-18).

Day, G. & Fahey, L. (1988). Valuing market strategies. *Journal of Marketing*, 52(3), 45–57.

Dickey, D. A. & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a), 427–431.

DISH Network (2015). Form 10-K (Annual Report for the Fiscal Year Ended December 31, 2014).

Doyle, P. (2009). *Value-based marketing: marketing strategies for corporate growth and shareholder value*. West Sussex, England: John Wiley & Sons.

Duffee, G. R. & Stanton, R. H. (2008). Evidence on simulation inference for near unit-root processes with implications for term structure estimation. *Journal of Financial Econometrics*, 6(1), 108–142.

Dutta, S., Narasimhan, O., & Rajiv, S. (1999). Success in high-technology markets: Is marketing capability critical? *Marketing Science*, 18(4), 547–568.

Ebbes, P., Papies, D., & Van Heerde, H. J. (2011). The sense and non-sense of holdout sample validation in the presence of endogeneity. *Marketing Science*, 30(6), 1115–1122.

Elliott, G., Rothenberg, T. J., & Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica*, 64(4), 813–836.

Erdem, T. (1996). A dynamic analysis of market structure based on panel data. *Marketing Science*, 15(4), 359–378.

Eskin, G. J. (1973). Dynamic forecasts of new product demand using a depth of repeat model. *Journal of Marketing Research*, 10(2), 115–129.

Fader, P. (2012). *Customer Centricity: Focus on the Right Customers for Strategic Advantage*. Philadelphia, PA: Wharton Digital Press.

Fader, P. S. & Hardie, B. G. (2010). Customer-base valuation in a contractual setting: The perils of ignoring heterogeneity. *Marketing Science*, 29(1), 85–93.

Fader, P. S. & Hardie, B. G. (2014). The Pareto/NBD is not a lost-for-good model. Working paper.

Fader, P. S., Hardie, B. G., & Huang, C.-Y. (2004). A dynamic changepoint model for new product sales forecasting. *Marketing Science*, 23(1), 50–65.

Fader, P. S., Hardie, B. G., & Lee, K. L. (2005a). "Counting your customers" the easy way: An alternative to the Pareto/NBD model. *Marketing Science*, 24(2), 275–284.

Fader, P. S., Hardie, B. G., & Lee, K. L. (2005b). RFM and CLV: Using iso-value curves for customer base analysis. *Journal of Marketing Research*, 42(4), 415–430.

Fader, P. S., Hardie, B. G., & Shang, J. (2010). Customer-base analysis in a discrete-time noncontractual setting. *Marketing Science*, 29(6), 1086–1108.

Fama, E. F. & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1), 3–56.

Farris, P. W., Bendle, N. T., Pfeifer, P. E., & Reibstein, D. J. (2010). *Marketing metrics: The definitive guide to measuring marketing performance.* Upper Saddle River, NJ: Pearson Education.

Feit, E. M., Wang, P., Bradlow, E. T., & Fader, P. S. (2013). Fusing aggregate and disaggregate data with an application to multiplatform media consumption. *Journal of Marketing Research*, 50(3), 348–364.

Fisher, J. C. & Pry, R. H. (1971). A simple substitution model of technological change. *Technological Forecasting and Social Change*, 3, 75–88.

Fornell, C. & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39–50.

Fourt, L. A. & Woodlock, J. W. (1960). Early prediction of market success for new grocery products. *Journal of Marketing*, 25(2), 31–38.

Gallant, A. R. & Tauchen, G. (1996). Which moments to match? *Econometric Theory*, 12, 657–681.

Gönül, F. & Srinivasan, K. (1993). Modeling multiple sources of heterogeneity in multinomial logit models: Methodological and managerial issues. *Marketing Science*, 12(3), 213–229.

Gopalakrishnan, A., Bradlow, E., & Fader, P. (2016). A cross-cohort changepoint model for customer-base analysis. *Marketing Science*, (pp. 1–20).

Gourieroux, C., Monfort, A., & Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, 8(1), S85–S118.

Gourio, F. & Rudanko, L. (2014). Customer capital. *The Review of Economic Studies*, 81(3), 1102–1136.

Greenwald, B., Kahn, J., Sonkin, P. D., & Van Biema, M. (2004). *Value Investing: From Graham to Buffett and Beyond.* Hoboken, N.J.: John Wiley & Sons.

Gupta, S. (2009). Customer-based valuation. *Journal of Interactive Marketing*, 23(2), 169–178.

Gupta, S., Lehmann, D. R., & Stuart, J. A. (2004). Valuing customers. *Journal of Marketing Research*, 41(1), 7–18.

Hand, J. R. (2015). Discussion of "customer franchise – a hidden, yet crucial, asset". *Contemporary Accounting Research*, 32, 1050–1052.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, 50(4), 1029–1054.

Holthausen, R. W. & Zmijewski, M. E. (2013). *Corporate Valuation: Theory, Evidence and Practice.* Cambridge, MA: Cambridge Business Publishers.

Homburg, C., Koschate, N., & Hoyer, W. D. (2005). Do satisfied customers really pay more? A study of the relationship between customer satisfaction and willingness to pay. *Journal of Marketing*, 69(2), 84–96.

Hui, S. K. (2017). Understanding repeat playing behavior in casual games using a bayesian data augmentation approach. *Quantitative Marketing and Economics*, 15(1), 29–55.

Jamal, Z. & Bucklin, R. E. (2006). Improving the diagnosis and prediction of customer churn: A heterogeneous hazard modeling approach. *Journal of Interactive Marketing*, 20(3-4), 16–29.

Janzer, A. H. (2015). *Subscription Marketing.* Cuesta Park Consulting.

Jiang, R., Manchanda, P., & Rossi, P. E. (2009). Bayesian analysis of random coefficient logit models using aggregate data. *Journal of Econometrics*, 149(2), 136–148.

Jiang, W. & Turnbull, B. (2004). The indirect method: inference based on intermediate statistics—a synthesis and examples. *Statistical Science*, 19(2), 239–263.

Kamakura, W. A. & Russell, G. J. (1993). Measuring brand value with scanner data. *International Journal of Research in Marketing*, 10(1), 9–22.

Kemper, A. (2009). *Valuation of network effects in software markets: a complex networks approach.* Heidelberg: Physica-Verlag.

Kim, B.-D. (1995). Incorporating heterogeneity with store-level aggregate data. *Marketing Letters*, 6(2), 159–169.

Kim, N., Mahajan, V., & Srivastava, R. K. (1995). Determining the going market value of a business in an emerging information technology industry: The case of the cellular communications industry. *Technological Forecasting and Social Change*, 49(3), 257–279.

Koller, T., Goedhart, M., & Wessels, D. (2010). *Valuation: Measuring and Managing the Value of Companies.* Hoboken, N.J.: John Wiley & Sons.

Kumar, V., Aksoy, L., Donkers, B., Venkatesan, R., Wiesel, T., & Tillmanns, S. (2010). Undervalued or overvalued customers: Capturing total customer engagement value. *Journal of Service Research*, 13(3), 297–310.

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

Kumar, V. & Shah, D. (2015). *Handbook of Research on Customer Equity in Marketing.* Edward Elgar Pub.

Leckenby, J. D. & Boyd, M. M. (1984). An improved beta binomial reach/frequency model for magazines. *Current Issues and Research in Advertising*, 7(1), 1–24.

Lev, B. (1992). Information disclosure strategy. *California Management Review*, 34(4), 9–32.

Libai, B., Muller, E., & Peres, R. (2009). The diffusion of services. *Journal of Marketing Research*, 46(2), 163–175.

Little, R. J. & Rubin, D. B. (2014). *Statistical analysis with missing data*. Hoboken NJ: John Wiley & Sons.

Luo, X. & Bhattacharya, C. (2006). Corporate social responsibility, customer satisfaction, and market value. *Journal of Marketing*, 70(4), 1–18.

Luo, X., Zhang, J., & Duan, W. (2013). Social media and firm equity value. *Information Systems Research*, 24(1), 146–163.

Martinez-Jerez, F. d. A., Steenburgh, T. J., Avery, J., & Brem, L. (2011). HubSpot: Lower churn though greater CHI. Harvard Business School Case 9-110-052. Boston, MA: Harvard Business School Publishing.

McCarthy, D., Fader, P., & Hardie, B. (2016). V(CLV): Examining variance in models of customer lifetime value. Available at SSRN: http://ssrn.com/abstract=2739475.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica: Journal of the Econometric Society*, 57(5), 995–1026.

Mehta, N., Rajiv, S., & Srinivasan, K. (2003). Price uncertainty and consumer search: A structural model of consideration set formation. *Marketing science*, 22(1), 58–84.

Michaelides, A. & Ng, S. (2000). Estimating the rational expectations model of speculative storage: A monte carlo comparison of three simulation estimators. *Journal of Econometrics*, 96(2), 231–266.

Mintz, O., Gilbride, T. J., Currim, I. S., & Lenk, P. (2016). Metric effectiveness and use in marketing-mix decisions: Correcting for endogenous selection effects and ex-ante expectations. Working paper.

Moe, W. W. & Fader, P. S. (2002). Using advance purchase orders to forecast new product sales. *Marketing Science*, 21(3), 347–364.

Morrison, D. G. (1969). Conditional trend analysis: A model that allows for nonusers. *Journal of Marketing Research*, 6(3), 342–346.

Morrison, D. G. & Schmittlein, D. C. (1980). Jobs, strikes, and wars: Probability models for duration. *Organizational Behavior and Human Performance*, 25(2), 224–251.

Musalem, A., Bradlow, E. T., & Raju, J. S. (2008). Who's got the coupon? Estimating consumer preferences and coupon usage from aggregate information. *Journal of Marketing Research*, 45(6), 715–730.

Musalem, A., Bradlow, E. T., & Raju, J. S. (2009). Bayesian estimation of random-coefficients choice models using aggregate data. *Journal of Applied Econometrics*, 24(3), 490–516.

Pfeifer, P. E. (2011). On estimating current-customer equity using company summary data. *Journal of Interactive Marketing*, 25(1), 1–14.

Pfeifer, P. E., Haskins, M. E., & Conroy, R. M. (2005). Customer lifetime value, customer profitability, and the treatment of acquisition spending. *Journal of Managerial Issues*, 17(1), 11–25.

Platzer, M. & Reutterer, T. (2016). Ticking away the moments: Timing regularity helps to better predict customer activity. *Marketing Science*, 35(5), 779–799.

QVC (2015). 10-K Report. SEC.
URL: http://www.sec.gov/Archives/edgar/data/1254699/000125469916000050/qvc_10kx12312015.htm
(visited on 2017-03-14).

Rappaport, A. (1986). *Creating shareholder value: the new standard for business performance.* Free press.

Reinartz, W., Thomas, J. S., & Kumar, V. (2005). Balancing acquisition and retention resources to maximize customer profitability. *Journal of Marketing*, 69(1), 63–79.

Sabavala, D. J. & Morrison, D. G. (1977). A model of TV show loyalty. *Journal of Advertising Research*, 17(6), 35–43.

SCAC (2004). Case Summary: Netflix, Inc. Securities Litigation,"Securities Class Action Clearinghouse, Stanford Law School. URL: http://securities.stanford.edu/filings-case.html?id=103168 (visited on 2015-10-15).

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

Schmittlein, D. C. & Peterson, R. A. (1994). Customer base analysis: An industrial purchase process application. *Marketing Science*, 13(1), 41–67.

Schulze, C., Skiera, B., & Wiesel, T. (2012). Linking customer and financial metrics to shareholder value: The leverage effect in customer-based valuation. *Journal of Marketing*, 76(2), 17–32.

Schweidel, D. A., Bradlow, E. T., & Fader, P. S. (2011). Portfolio dynamics for customers of a multiservice provider. *Management Science*, 57(3), 471–486.

Schweidel, D. A., Fader, P. S., & Bradlow, E. T. (2008a). A bivariate timing model of customer acquisition and retention. *Marketing Science*, 27(5), 829–843.

Schweidel, D. A., Fader, P. S., & Bradlow, E. T. (2008b). Understanding service retention within and across cohorts using limited information. *Journal of Marketing*, 72(1), 82–94.

Seybold, P. (2000). Don't count on Amazon. Business 2.0.
URL: http://boards.fool.com/business-20-dont-count-out-amazon-13436734.aspx (visited on 2017-03-14).

Sharp, B. & Sharp, A. (1997). Loyalty programs and their impact on repeat-purchase loyalty patterns. *International Journal of Research in Marketing*, 14(5), 473–486.

Skiera, B. & Schulze, C. (2014). *Handbook of service marketing research*, chapter Customer-based vluation: similarities and differences to traditinoal discounted cash flow models, (pp. 123–134). Edward Elgar Publishing: Cheltenham, UK.

Smith, A. A. (1993). Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics*, 8(1), S63–S84.

Srinivasan, S. (2015). Capitalizing advertising spending. Presentation to the Marketing Accountability Standards Board; Chicago, IL.

Srinivasan, S. & Hanssens, D. M. (2009). Marketing and firm value: Metrics, methods, findings, and future directions. *Journal of Marketing Research*, 46(3), 293–312.

Srivastava, R. K., Shervani, T. A., & Fahey, L. (1998). Market-based assets and shareholder value: a framework for analysis. *Journal of Marketing*, 62(1), 2–18.

Streitfeld, D. (2001). Analysts, vendors increasingly wary about Amazon. The Washington Post. URL: https://www.washingtonpost.com/archive/business/2001/02/21/analysts-vendors-increasingly-wary-about-amazon/82aee135-065e-4ce9-a293-edbbc83a6552 (visited on 2017-03-14).

Tardiff, T. J. (1980). Vehicle choice models: review of previous studies and directions for further research. *Transportation research Part A: General*, 14(5-6), 327–336.

Thomas, J. S., Blattberg, R. C., & Fox, E. J. (2004). Recapturing lost customers. *Journal of Marketing Research*, 41(1), 31–45.

Van den Bulte, C. & Lilien, G. L. (1997). Bias and systematic change in the parameter estimates of macro-level diffusion models. *Marketing Science*, 16(4), 338–353.

Warrillow, J. (2015). *The Automatic Customer: Creating a Subscription Business in Any Industry*. Westminster, U.K.: Penguin.

Wiesel, T., Skiera, B., & Villanueva, J. (2008). Customer equity: An integral part of financial reporting. *Journal of Marketing*, 72(2), 1–14.

Wooldridge, J. (2002). *Introductory Econometrics Analysis of Cross Section and Panel Data*. Cambridge, MA: The MIT Press.

Yoo, S., Hanssens, D. M., & Kim, H. (2012). Marketing and the evolution of customer equity of frequently purchased brands. *Working Paper*.

Yu, Y., Duan, W., & Cao, Q. (2013). The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decision Support Systems*, 55(4), 919–926.

Zantedeschi, D., Feit, E. M., & Bradlow, E. T. (2016). Measuring multichannel advertising response. *Management Science*.

Zenor, M. J. & Srivastava, R. K. (1993). Inferring market structure with aggregate data: A latent segment logit approach. *Journal of Marketing Research*, (pp. 369–379).

Zhang, Y., Bradlow, E. T., & Small, D. S. (2014). Predicting customer value using clumpiness: From RFM to RFMC. *Marketing Science*, 34(2), 195–208.

Zivot, E. & Wang, J. (2007). *Modeling Financial Time Series with S-Plus®, 2nd edn.* New York: Springer Science & Business Media.