Beyond Customer Lifetime Valuation: Measuring the Value of Acquisition and Retention for Subscription Services

Hamidreza Badri*
Netflix Inc.
Los Angeles, USA
hbadri@netflix.com

Allen Tran*
Netflix Inc.
Los Angeles, USA
allent@netflix.com

ABSTRACT

Understanding the value of acquiring or retaining subscribers is crucial for subscription-based businesses. While customer lifetime value (LTV) is commonly used to do so, we demonstrate that LTV likely over-states the true value of acquisition or retention. We establish a methodology to estimate the monetary value of acquired or retained subscribers based on estimating both on and off-service LTV. To overcome the lack of data on off-service households, we use an approach based on Markov chains that recovers off-service LTV from minimal data on non-subscriber transitions. Furthermore, we demonstrate how the methodology can be used to (i) forecast aggregate subscriber numbers that respect both aggregate market constraints and account-level dynamics, (ii) estimate the impact of price changes on revenue and subscription growth and (iii) provide optimal policies, such as price discounting, that maximize expected lifetime revenue.

CCS CONCEPTS

• **Computing methodologies → Markov decision processes**; • **Applied computing → Economics**; • **General and reference** → *Measurement*.

KEYWORDS

Customer Lifetime Valuation, Observational Causal Inference, Markov Decision Processes

ACM Reference Format:
Hamidreza Badri and Allen Tran. 2022. Beyond Customer Lifetime Valuation: Measuring the Value of Acquisition and Retention for Subscription Services. In *Proceedings of the ACM Web Conference 2022 (WWW '22), April 25–29, 2022, Virtual Event, Lyon, France.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3485447.3512058

1 INTRODUCTION

Understanding the incremental value of subscribers is essential to subscription services. Marketing or product investments generally aim to increase acquisition of new subscribers or retention of existing subscribers. The evaluation of these core investments is difficult because the costs of such investments are easily measured, yet the *monetary* benefits are not obvious. Without the ability to evaluate investments, businesses cannot optimize towards profitable future investments leading to sub optimal business outcomes. This paper develops a methodology to accurately quantify the monetary value of acquired or retained subscribers.

By definition, subscription-based services grow by acquiring and retaining subscribers. To do so, they may launch new content or features [2], run marketing campaigns [20], or offer pricing packages that better suit subscriber needs [6]. Measuring the causal impact on acquisition or retention (in units of subscribers) from these interventions can be difficult, but approaches based on randomized control trials [22] or observational studies [16] exist.

However, it is not clear what monetary value to assign to an acquired or retained subscriber. Although it is common practice, we demonstrate that using LTV (or remaining LTV in the case of retention) will tend to overstate the value of acquisition or retention for a subscription business. Instead, we demonstrate that the difference between on and off-service LTV, *incremental LTV*, is the more appropriate quantity of interest and develop a Markov chain based estimation approach.

Conceptually, acquiring a new subscriber or retaining an existing subscriber transitions (or saves from the perspective of the business) a customer from an off-service state to an on-service state. Naturally then, the value of that acquisition or retention to the business is the difference between the expected cumulative revenue from that individual in the on-service state and from that same individual if they were in the off-service state. Note crucially that the latter is positive if there is a positive probability of an individual becoming a subscriber when off-service. By using LTV to value acquisition or retention, one is implicitly assuming that the value of off-service states are zero. However, if there is positive probability of a non-subscriber joining, or rejoining, then LTV is likely upwards biased since it fails to subtract the baseline value associated with the off-service state.

As an example, suppose prices are increased and we have some method to determine the subscribers who churned as a result. For each of these former subscribers, calculating the difference in expected lifetime value if they had stayed on service from the revenue expected in their churned state provides an estimate of the value of churn for each former subscriber which in sum is the aggregate cost of the price increase. The aggregate cost of the price increase is a crucial quantity in assessing the success of the price increase. Since it is denominated in dollar units, the difference between the revenue

*Both authors contributed equally to this research.

WWW '22, April 25–29, 2022, Virtual Event, Lyon, France

ACM ISBN 978-1-4503-9096-5/22/04…$15.00
https://doi.org/10.1145/3485447.3512058

gains of the price increase and the total cost is itself in dollar units, providing a dollar valuation of the price increase intervention.

This paper introduces a model that estimates the incremental customer lifetime value in a subscription service in an event of acquisition or retention. Our proposed methodology models the customer lifetime through a finite Markov chain, allowing for multiple subscriptions and transitions between states such as on-service/off-service, tenure, subscription plans, or more generally, any other attributes that impact acquisition and churn probabilities. By chaining together the many probabilistic transitions from state to subsequent states, our methodology can model the expected lifetime and cumulative revenue from any given initial state. We can simply find the appropriate counterfactual for various events such as acquiring a new subscriber or retaining a current subscriber, and use the difference between the expected revenue starting from a state and its counterfactual as the incremental lifetime value. More specifically, the methodology can answer several important questions related to subscription services such as forecasting aggregate subscriber figures or estimating the impact of price changes on subscriber growth and revenue.

The contributions of this paper are: (1) present a scalable framework to estimate the incremental lifetime value of an acquired or retained subscriber that is more accurate than LTV and extends to multiple subscriptions, (2), extends the proposed methodology to forecast subscribers in a manner satisfying natural constraints such as the remaining size of the market and (3), demonstrate how the approach can be used to set policies, such as pricing, optimally to maximize lifetime revenue.

The paper is organized as follows. Section 2 shows how we can use a finite Markov chain to estimate the incremental lifetime value of a subscriber in an acquisition or retention event. Section 3 extends our framework to three applications common to subscription services (i) forecasting aggregate subscriber counts, (ii) estimating the incremental revenue or subscription impact due to price interventions, and (iii) optimizing sequences of prices to maximize lifetime revenue. Finally Section 4 concludes.

1.1 Related Literature

The majority of prior research on customer lifetime valuation focuses on predicting LTV, as the definition is taken as given: the discounted sum of expected revenue from a customer [3, 10]. There exist a vast literature in machine learning and economics on understanding customer lifetime value. Gupta et al. [8] review modeling advances in the customer lifetime valuation literature and provide empirical insights obtained from various models. Most of the proposed models focus on machine learning techniques that use consumer level features [1, 4, 21] such as recency, frequency and monetary characteristics to predict individual LTV outcomes. The most similar approach to ours is an approach that sets up a discrete choice model resulting in Markov probabilistic switching between brands [19]. In this paper, we aggregate LTV over cohorts of subscribers where effects of fine-grained features would wash out in aggregate. Moreover, our focus is on incremental LTV where the estimand of interest is the difference between LTV at different states.

Compartmental models used in epidemiology focus on modeling transitions of people between compartments (such as susceptible, infected and recovered states in SIR models), with transition probabilities governing population dynamics [14]. By modeling the population in a closed system, it is straightforward to forecast the spread of diseases under various assumptions underlying the disease. We take a similar closed system approach by explicitly modeling never or former subscribers, which is key to estimating the incremental value of an acquired or retained subscriber.

Various methodologies have been undertaken to forecast demand of subscription services [7, 18]. The majority of these studies use traditional time series methods that aim to minimize shorter-term forecast errors [12]. However, in the case of subscription services, such models will perform poorly when longer forecast horizons are required as they don't account for market size constraints. In other words, they are unaware of structural constraints that bind over longer time horizons. By modeling the subscription dynamics in a closed system which naturally incorporates a market size constraint, our forecasting methodology provides more accurate longer term forecasts.

In subscription-based industries, one of the key levers to enhance profitability is price. Many mathematical models have been proposed to find the optimal pricing strategy in the literature. In [11], authors propose a methodology to set the optimal pricing for a magazine publishing firm facing stochastic demand over multiple periods. They model the dynamics of customer subscription and retention/attrition as a function of production quantity, subscription price and newsstand price. They propose a dynamic programming formulation to find the optimal policy. In this paper, we adopt a similar methodology and show that if an estimates of the elasticity of the Markov Chain transition probabilities with respect to prices are available, our framework can be used to find the optimal policy that maximizes the long-term revenue given the distribution of customers across the states.

2 METHODOLOGY

We're interested in the LTV we can causally attribute to the acquisition or retention of new/existing members. To do so, we first outline assumptions that enable a causal estimate from observational data. Next, we set up a finite Markov chain to model the customer lifetime in a subscription service. Finally, we illustrate how to use the proposed Markov chain to estimate the incremental LTV of an acquired or retained subscriber.

To start, denote V as the remaining cumulative discounted revenue for a household in state s. More concretely, dropping the s for notational convenience:

$$V = \sum_{k=0}^{\infty} \beta^k m_k \cdot c_k \tag{2.1}$$

where m_k is an indicator equal to 1 if the household is a member k periods in the future and c_k the price likewise.

Given the notation, the question we seek to answer is: what fraction of V is due to the intervention of shifting a household into a membership state?

2.1 A Causal Interpretation

Borrowing from the potential outcomes framework [13], define two potential LTV outcomes starting from an initial state s: m if the household is a member in the next state and $\neg m$ if they are not a member in the next state. Correspondingly, V_m is potential LTV if the household is a member and $V_{\neg m}$ is potential LTV if the household is not a member.

The realized observation, V, is one of these potential outcomes depending on whether the household becomes a customer, $m = 1$, or not. $m = 0$.

$$V = m \cdot V_m + (1 - m) \cdot V_{\neg m} \tag{2.2}$$

Our object of interest is incremental LTV, the additional LTV we expect to receive from the household being in a subscriber versus a non-subscriber state or more concretely, the difference in potential LTV outcomes. To estimate this, we average across households and seek an estimate of the average treatment effect where the treatment is an intervention that shifts the household into the membership state.[1]

$$\mathbb{E}\left[\Delta V|s\right] \equiv \mathbb{E}_{V_m|s}\left[V_m|s\right] - \mathbb{E}_{V_{\neg m}|s}\left[V_{\neg m}|s\right] \tag{2.3}$$

Unfortunately, for each household we only observe one of the potential outcomes. To be able to estimate incremental LTV from data, we need to ensure the standard assumptions from causal inference hold: overlap and unconfoundedness [13]. Assuming overlap holds (i.e ensuring that we observe transitions into the subscriber and non-subscriber state from each initial state), unconfoundedness is the assumption warrants further discussion. It requires that the potential LTV outcomes are independent with respect to the realized membership indicator m, conditional on the state s.

$$(V_m, V_{\neg m}) \perp\!\!\!\perp m \mid s \tag{2.4}$$

Another way to say this is that we need enough variables in the states to ensure that conditional independence holds. For example, if households from a particular geo-region have higher potential LTV as members, V_m, and are more likely to join as non members, then unconfoundedness does not hold unless geo-region is part of the state-space.

Assuming overlap and unconfoundedness hold, we can estimate incremental LTV from observational data, where a similar derivation can be made for expectation of the non-subscriber potential outcome in equation (2.3).[2]

$$\mathbb{E}_{V_m|s}\left[V_m|s\right] = \mathbb{E}_{V|s,m=1}\left[V|s, m = 1\right] \tag{2.5}$$

Combining equations (2.3) and (2.5) show that incremental LTV from state s can be estimated as the difference in observed LTV in the subsequent subscriber state and observed LTV in the subsequent non-subscriber state.

$$\mathbb{E}\left[\Delta V|s\right] = \mathbb{E}\left[V|s, m = 1\right] - \mathbb{E}\left[V|s, m = 0\right] \tag{2.6}$$

[1] We're explicitly quantifying the causal effect on LTV from acquiring/retaining the member which is different from the causal effect on acquisition/retention from some intervention. In the former, the intervention is the acquisition/retention event and in the latter the intervention is some investment designed to acquire or retain members.

[2] Crucially, note that V is observed whereas V_m and $V_{\neg m}$ are not.

2.2 The Markov Chain Model

The primary difficulty in estimating equation (2.6) is that the second part of the difference represents LTV of a household in a non-subscriber state. Even for web based businesses, data on non-subscribers is at best limited and commonly unavailable.[3] To overcome this problem, we build a Markov chain where we are able to estimate non-subscriber LTV from limited data on non-subscriber acquisitions. By chaining together transitions in a probabilistic manner, we implicitly recover the distribution over the possible paths across the state space and hence the expected LTV from a non-subscriber state.

To begin, consider a subscription service with two different plans (A and B), each with different prices. Subscribers pay a fee at a predetermined cadence (e.g. weekly or monthly) to have access to the firm's products or services for a billing cycle. At the end of each billing cycle, subscribers have the option to renew, upgrade, downgrade, or cancel their subscription. We also assume that the firm can identify rejoiners, subscribers who have a prior subscription, cancelled and resubscribe.

Based on their subscription status (on- or off-service), billing cycle, and plan type, a subscriber is in one the following states:

$$\mathbb{S} = \{0, A_1, B_1, \ldots, A_N, B_N, A_{-1}, B_{-1}, \ldots, A_{-M}, B_{-M}\}$$

where A_i denotes a subscriber with the firm for i consecutive billing cycles and currently enrolled in Plan A, A_{-j} denotes a former subscriber off-service for j consecutive billing cycles most recently enrolled in plan A during her last billing period on the service. Finally a customer is in state 0 if she hasn't been a subscriber of the firm in the past (never-subscriber). The states defined in $\mathbb{S}$ do not overlap and every subscriber is in one of the states.[4] We also assume N and M are two integers which are large enough that subscriber behaviour in terms of acquisition and retention would stay constant, for example:

$$p\left(A_{i+1}|A_i\right) \approx p\left(A_{j+1}|A_j\right), \forall i, j \geq M.$$

Figure 1 depicts the Markov chain for the subscription service. The Markov chain is characterized by an $(2N+2M+1)\times(2N+2M+1)$ transition probability matrix, where $p(s'|s)$ shows the transition probability that a subscriber will move to state s' in the next billing cycle given that she is currently in state s.

Given these transition probabilities, we are in a position to derive both finite and infinite horizon LTV, and their incremental equivalents.

Define $V(s)$ as the value function at state s, the expected discounted revenue generated by a customer starting from state s over an infinite horizon. Then value function is described by the following recursive equation

$$V(s) = c(s) + \beta \times \sum_{s'} p(s'|s) \times V(s') \tag{2.7}$$

where $c(s)$ is the subscription price of state s and β is the discount factor.

[3] Obvious exceptions are advertising or tracking companies where significant engineering effort is spent on tracking users across the web.

[4] For ease of exposition, we consider only a minimal number of states. In practice, the state space should include all states that satisfy the Markov property and to eliminate unobserved heterogeneity that affects the transition probabilities.

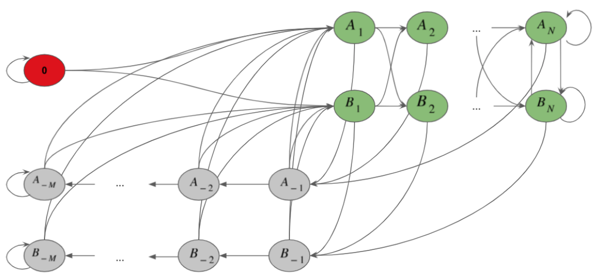


Figure 1: Markov chain modeling a customer's journey in a subscription service.

Solving for $V(s)$ is straightforward since one can rewrite (2.7) in vector notation as

$$(\mathbf{I} - \beta\mathbf{P})\mathbf{V} = \mathbf{c} \tag{2.8}$$

where **c** and **V** are the stacked vector equivalents of $c(s)$ and $V(s)$ and **P** is a square matrix filled with transition probabilities. Since **P** is a transition matrix with transition probabilities and $\beta < 1$, the matrix $(\mathbf{I} - \beta\mathbf{P})$ has eigenvalues bounded below by $1 - \beta$ which guarantees $(\mathbf{I} - \beta\mathbf{P})$ is invertible and hence there exists a unique and non-degenerate solution to **V**.

Alternatively, if one is interested in a finite horizon value function, $V_T(s)$ one can simply simulate the Markov chain starting from each initial state and accumulate revenue (or subscribers) until the terminal period.

At time 0, suppose one is interested in LTV over T periods. Denote the k-step ahead transition probabilities recursively as

$$p(s''|s;k) = \Pi_{s'} p(s'|s;k-1)p(s''|s';1) \quad \forall k > 1 \tag{2.9}$$

and

$$p(s'|s;1) = p(s'|s) \tag{2.10}$$

Given these k-step transition probabilities, the finite horizon LTV is simply

$$V_T(s) = \sum_{t=0}^{T} \beta^t \sum_{s'} p(s'|s;t)c(s'). \tag{2.11}$$

Similarly, given an initial distribution of subscribers across the states, $n_0(.)$, the number of subscribers at state s at T is

$$n_T(s) = \sum_{s'} n_0(s')p(s|s';T). \tag{2.12}$$

A direct application of this is the number of subscribers at T

$$S_T = \sum_{s} n_T(s) 1_{on-service(s)}. \tag{2.13}$$

where $1_{on-service(s)}$ is an indicator variable denoting where state s is a on-service state.

2.3 Value of Acquired or Retained Subscriber

In the previous section, we showed how to model LTV with a finite Markov chain. The payoff in modeling LTV with a value function is that we can decompose LTV in the service of estimating the incremental value of an acquired or retained subscriber.

Suppose the subscription service runs a marketing campaign aiming to acquire (or re-acquire) and retain subscribers and we can causally attribute acquired or retained subscribers to that campaign, what is the monetary benefit of the marketing campaign?

For the Markov chain above, there are 3 types of channels we have to consider: new acquisitions, acquisitions of former subscribers and retention of existing subscribers.[5] At a high level, we will value each of these channels similarly. First, we calculate their forward-looking LTV at the time of the marketing campaign. This is the revenue we expect to receive from the subscriber going forward, accounting for churn risk and the possibility of rejoins. Crucially however, we also calculate the revenue we expect to receive if the subscriber had not joined/rejoined or churned. The difference between the two is the additional revenue we expect to gain, incremental LTV, due to the marketing campaign causally shifting the subscriber from the off-service state to the on-service state.

Figure 2 depicts these three channels. More concretely, the value of acquiring a new subscriber is:

$$\Delta V_{acq}(i) \equiv V(i_1) - V(0) \quad i \in \{A, B\}$$

where i denotes the plan the subscriber chooses upon acquisition.

Similarly, reacquiring a subscriber who was previously on plan j off-service for k periods

$$\Delta V_{reacq}(i,j,k) \equiv V(i_1) - V\left(j_{\min\{-M,-k-1\}}\right)$$

$$i, j \in \{A, B\}, 1 \leq k \leq M$$

where i shows what types of plan the subscriber signs up into, j shows the latest plan that subscriber were enrolled in her last billing cycle, and k shows for how many billing cycles the subscriber has been off-service.

Finally, retaining a subscriber who chooses plan i and was previously on plan j with tenure k:

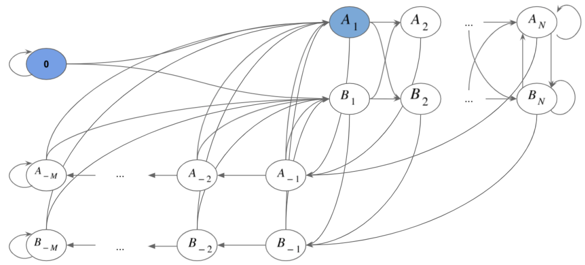
$$\Delta V_{ret}(i,j,k) \equiv V(i_{\max\{N,k+1\}}) - V(j_{-1})$$

$$i \in \{A, B\}, 1 \leq j \leq M.$$

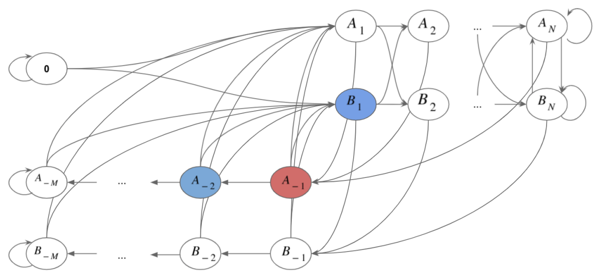
Intuitively, each of these channels is lower if the baseline level of joining or rejoining is higher. In other words, there is little value to a marketing campaign that acquired new subscribers if those never-subscribers would have likely joined tomorrow absent the marketing campaign. Hence using LTV alone, without subtracting the appropriate baseline, will tend to overvalue acquisition and retention and consequently, overvalue investments aiming to increase acquisition and retention.

One important caveat is that because these estimates of incremental LTV are based on static transition probabilities, they can only be used to estimate the value of acquisition and retention coming from interventions that have only transitory effects on transition probabilities. For example, launches of new content or marketing campaigns typically have effects that decay rapidly over time and hence are good candidates for incremental LTV whereas new product features that fundamentally alter a service are not.

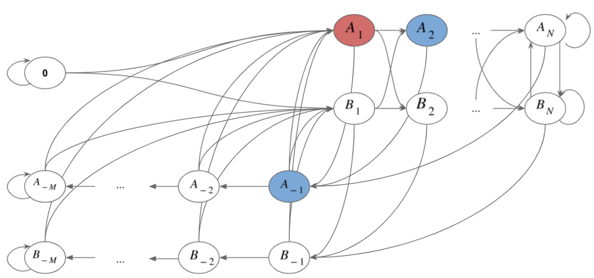
[5]In practice, there are many other behavioral responses that occur such as changes in engagement, account sharing, etc. These are all possible to value if the appropriate states are included in the state-space.



(a)



(b)



(c)

Figure 2: (a) Incremental value of acquiring a never-subscriber as a type-A subscriber can be estimated as $V(A_1) - V(0)$ (b) Incremental value of re-acquiring a subscriber being in state A_{-1} as a type-B subscriber can be estimated as $V(B_1) - V(A_{-2})$ (c) Incremental value of retaining a subscriber being in state A_1 can be estimated as $V(A_2) - V(A_{-1})$

2.4 Limitations

One possibility is that the underlying stochastic process may be non-Markov: either the relevant history cannot be reduced to the preceding state and/or we are missing states in the state space. This prevents the causal interpretation of incremental LTV and potentially biases estimates of the value of acquisition or retention. For example, if higher income households retain with higher probability and we don't observe income, the value of retention in a given state is biased upwards since estimates of LTV for members implicitly sub-sample wealthier households. Ideally, the state-space would would include as many covariates needed to condition away these hidden states but the curse of dimensionality limits a naive approach. The two solutions we propose enable conditioning on a high dimensional set of covariates without significantly increasing the state space.

Similar to [10] who model a binary retention decision, one can construct transition matrices on the original state space while conditioning on a set of covariates, and in a manner that accounts for unobserved heterogeneity. For any state, there exist a sample of households each characterized by covariates. One can model the outcome of households decisions at that state as a Dirichlet distribution, where the concentration parameters of the Dirichlet distribution depend on covariates and the current state.[6] If the state-space is of dimension d, one estimates d separate Dirichlet models each modeling the transition to states within a subspace of the full state-space. Note that since each model includes the state as part of the input along with the covariates, we avoid estimating a separate model for each state which would be computationally prohibitive.

As an example, suppose that the two states are member or not, and plan type A or B. By fitting two Dirichlet models, one for membership and another for plan choice, one can construct a transition matrix over the original state space that is conditioned on covariates, assuming that choices for each state are independent. Moreover, by sampling from the Dirichlet models for a given set of covariates, one recovers the full distribution of "personalized" transition matrices which explicitly accounts for unobserved heterogeneity that remain after conditioning. Hence this approach deals with the existence of hidden states in two ways: first, conditioning away the hidden state and secondly, characterizing the distribution of transition matrices instead of a single point estimate based transition matrix which accounts for any remaining unobserved heterogeneity.

An alternative direction is to collapse the potentially high dimensional set of covariates into a scalar propensity score (probability of transitioning to a member state conditional on covariates) and to include the propensity score as a state. If the covariates are sufficient to ensure unconfoundedness holds, it is sufficient to condition on the propensity score alone as the proof in [17] holds.[7] This approach summarizes the covariates with the propensity score but other latent variable based approaches are also possible [15].

For instance, to calculate the value of acquisition at state s, first calculate propensity scores with a model of the probability of transitioning into membership conditional on covariates. Then calculate the value of acquisition at each state and binned propensity score, which requires constructing transition matrices for the original state space with the addition of the propensity score states. Finally, integrate out the propensity scores by taking a weighted average of those estimates across the propensity score states to get the value of acquisition at s. As with the previous approach, a rich set of covariates can be used in calculating the propensity score while limiting the expansion of the state space.

[6] Note that the baseline method proposed already does this, where a constant is the only covariate, and hence every household in that state gets the same transition probabilities.

[7] For continuous valued states, one should use the generalized propensity score described in [9].

3 APPLICATIONS

In this section, to illustrate the wide ranging applicability of the methodology, we use our proposed methodology to answer some common yet challenging questions that subscription services face. In particular, we demonstrate how the methodology can be used to forecast future subscribers, estimate the impact of price change on business metrics and finally, to set optimal prices that trade-off increased revenue now against increased revenue in the future via subscriber growth.

Transition probabilities between states are the heart of the methodology. They govern the path and distribution of subscribers across the states and hence revenue and subscriber aggregates (see (2.11), (2.12), (2.13)). As such, our methodology for forecasting and estimating counterfactual outcomes relies on forecasting and estimating counterfactual transition probabilities.

The advantage of narrowing the problem to transition probabilities is that each state-to-state transition selects a small cohort of subscribers, minimizing unobserved heterogeneity that may mask the effect of the interventions. For example, the typical time series or event based model of aggregate subscribers embeds a host of dynamics that cloud the effect of an intervention.

3.1 Forecasting Transition Probabilities

The simplest model of transition probabilities is to assume a constant trend from a number of periods before the intervention (or announcement thereof) or forecast date. However, estimates of the counterfactual or forecast become noisy if the impact horizon is large. For example, a large known content launch in the future may higher the likelihood of acquisition and hence shift the true underlying counterfactual or forecast transition probability. Figure 3 shows scenarios under which (a) the constant model is sufficient and (b) where a more complex model is required.

We can improve on the constant model and deal with such concerns by using time series forecasting methods that incorporate covariates. Most forecasting methods are tuned to minimize k-step ahead forecast errors and hence optimized for the impact horizon one is interested in [12].

For example, suppose one uses a vector autogregressive model with covariates such as

$$\mathbf{p}_t(\cdot|s;t) = f_{s'|s}\left(\mathbf{x_t}, \mathbf{p_{t-1}}(\cdot|s), \ldots, \mathbf{p_{t-q}}(\cdot|s)\right) \tag{3.14}$$

where q is the lag length and $\mathbf{x_t}$ is a vector of covariates exogenous to the price intervention. Chaining these probabilities using equations (2.9) and (2.10) gives an estimate of the k-step ahead transition probability, $\hat{p}_k(s'|s)$.

3.2 Subscriber Forecasting

An obvious yet impactful application of the methodology is forecasting market penetration. Subscriber aggregates are by definition the result of disaggregate acquisition and churn decisions, which is precisely what our Markov chain measures. One of the challenges in forecasting market-wide aggregates is that classical time-series forecasting methods often lack structure and cannot account for natural constraints such as slowing growth due to market saturation. We explicitly account for these natural constraints in a parsimonious way by modeling customer dynamics within a closed system. Growth naturally slows as markets grow, as the pool of non-subscribers shrinks.

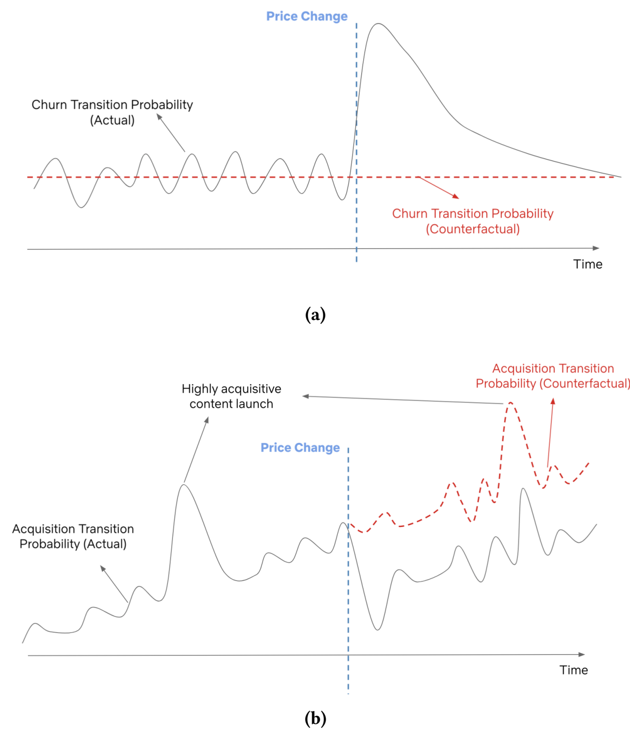


Figure 3: (a) Estimating counterfactual probability for a stationary transition (b) Estimating counterfactual probability for a non-stationary transition that depends on external factors such as subscription service content offering.

One of the keys to forecasting accurately, is choosing a large enough set of states to sufficiently characterize the transition probabilities over the forecast horizon. For example, suppose that churn rates fall as subscriber tenure increases. If tenure is not part of the set of states, the model's forecasts of growth will likely be biased downwards in growing markets as forecasts are based on historical transition probabilities which come from relatively low tenure subscribers with relatively high churn rates. The advantage of modeling transition probabilities is that even in growing markets, there typically is enough variation in the cross-section of existing subscribers such that forecasts that extrapolate beyond the current market size are accurate.

However, the benefits of more states are often offset by increased sparsity in the data. Broadly speaking, we propose two solutions.

First, we can use dimension reduction techniques that group states together that have similar transition probabilities and are nearby logically (e.g binning time-based data). Techniques also exist that are specific to Markov chains that cluster states with similar empirical distributions of transition paths [23].

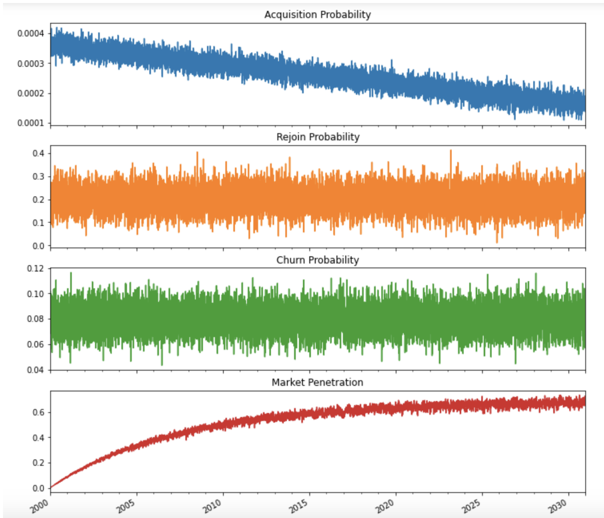


Figure 4: Simulated transition probabilities and number of subscribers (normalized by market size) for a streaming service over time

Another approach is to explicitly account for uncertainty in the transition matrices.[8] For example, we may observe only a small number of transitions out of a particular initial state which we are not confident of. Instead of using the point estimate, we can think of modeling the distribution of those transition probabilities themselves. For example, the transition probabilities out of an initial state all need to sum to 1 so we can model the distribution of transition probabilities as a Dirichlet distribution, $\mathbf{D}(\alpha)$. The concentration parameters, α, dictate how sharp the distribution is around the observed fractions and hence we can set each of them to the observed number of transitions. By repeating this for all the states, we have a distribution of transition matrices from which we can sample from.

Given a state space and historical transition matrices, we can use the techniques described in Section 3.1 to forecast transition probabilities. From there, forecasting aggregates at forecast horizon T is a straight-forward application of equation (2.13).

3.2.1 Example: Simulated Data. To demonstrate the advantages of our Markov chain closed system approach, we compare it against a baseline model which is estimated on time series aggregates. Consider an example with six states: never-subscriber, previous-subscriber, current-subscriber where each subscriber state can be either low or high income. We simulate data based on a random transition matrix across those states to generate time series data, shown in Figure 4, where low income households probability of transitioning into a member state is lower than high income households. Note the randomness in the time series which comes from the random nature of a Markov chain and slowing acquisition probabilities over time as market share grows.

We build three models, each forecasting the future number of subscribers at some terminal date T. A baseline ARIMA model trained on market penetration directly, second, a set of ARIMA models that forecast transition probabilities for the full set of states and subsequently predict top-line market penetration based on those forecast transition probabilities using equation (2.13) and finally, a set of ARIMA models that forecast transition probabilities treating income as a unknown state. This last model mimics the scenario whereby a hidden unobserved state is part of the stochastic process. For each set of models, we choose hyperparameters such as lag length of the AR and MA components, to minimize out-of-sample error.

[8]This is similar to [10] who estimate the parameters of a Beta distribution, where the outcome is a distribution over churn probabilities in the context of LTV models.

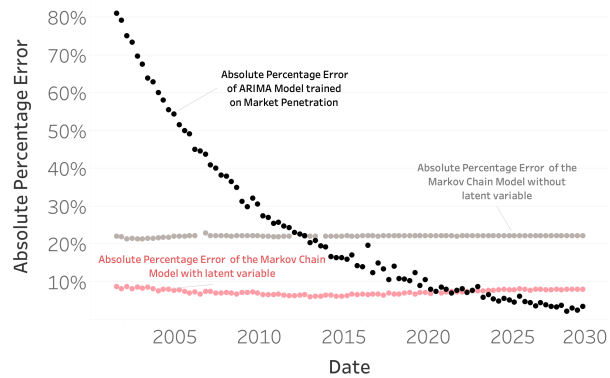


Figure 5: Absolute percentage error of the baseline and the full-state and hidden-state Markov chain based models over time.

Figure 5 shows the absolute percent error of each model over time, where forecasts at t only use data before t. The Markov chain based approach with the knowledge of the relevant state-space has relatively low error over a long horizon, even as market penetration grows from 0 to around 70 percent. On the other hand, the forecast errors of the baseline model are particularly high when market penetration is rapidly changing early on, precisely because it is unable to predict the slowdown in acquisition that comes from a reduced pool of non-subscribers. This result is consistent with results from the forecasting literature which show that forecasts made in non-stationary environments are difficult for models that implicitly rely on stationary data [5]. As growth slows and dynamics move closer to the stationary distribution, the baseline model eventually has lower error than the full-state Markov approach. As expected, the Markov based approach's main advantage is forecasting over longer horizons.

However, Figure 5 also shows that the existence of hidden states increase forecasting errors of the Markov based approach. Without knowledge of the relevant state-space, forecast errors are consistently higher than a Markov based approach incorporating the full set of states. Moreover, the period during which the Markov based model is better than the naive ARIMA model is shorter. In other words, the Markov-based approach's benefit of understanding long-run dynamics can be offset by errors induced by unobserved hidden states, as suggested in Section 2.4. Although not depicted, the size of the errors induced by the hidden state(s) depend on the extent to which transition probabilities for the state one cares about, in this case membership status, depend on the hidden state.

3.3 Pricing Application

In a subscription-based service, the two broad levers available to increase revenue are prices increases and/or subscriber growth. Focusing on pricing, prices have to be set optimally to balance short term first-order revenue changes against longer term effects on revenue via subscriber dynamics. Moreover, price elasticities evolve over the tenure of a subscriber. For example, longer tenure or higher engagement subscribers may for instance, have lower price elasticities. Consequently, the effect of a price change depends on the distribution of subscribers across states, which in turn governs the distribution of price elasticities. Therefore setting prices optimally requires accurate estimates of price elasticities.

Generally speaking, it is difficult to estimate price elasticities via experimentation at subscription businesses as the limited number of plans and repeated nature of the subscriptions means prices are highly visible making causal inference difficult due to SUTVA violations. Alternatively, we use observational data and the Markov chain in section 2.2 to estimate the incremental effect of price changes on aggregate revenue and subscriber growth, as well as more disaggregate cohorts of subscribers.

Assume one is interested in the impact of a price intervention after k periods. Since we already have realized revenue and subscriber figures, all we need to estimate the impact is an estimate of revenue and subscribers under the counterfactual where prices did not change.

Section 3.1 describes how we can construct counterfactual transition probabilities, using a model trained on data before the price intervention and applied post-intervention using covariates that are exogenous to the intervention.

With these, the counterfactual discounted revenue for a cohort described by states $s \in \Omega$ over the k periods is

$$\sum_{s\in\Omega} n_0(s)\hat{V_k}(s) = \sum_{t=0}^{k}\sum_{s\in\Omega} n_0(s)\beta^t \sum_{s'} \hat{p}_t(s'|s)c(s').$$

Again, counterfactual subscriber estimates are similar

$$\hat{S_k}(\Omega) = \sum_{s\in\Omega} n_0(s) \sum_{s'} \hat{p}_k(s'|s) 1_{on-service(s')}.$$

3.4 Optimal Policies

Implicit in the previous section was a method for obtaining the response of transition probabilities to price interventions. These "price elasticities" show the complexity of responses to price changes, examples which include increased churn, decreased acquisition, plan changes etc. With enough observed price elasticities, it is possible to model transition probabilities as a function of prices.[9] In other words, going from estimates of the price elasticity, $\frac{\partial dp(\cdot|s)}{\partial \mathbf{c}}$, and observations on baseline transition probabilities, $p(\cdot \mid s)$, to a model of transition probabilities as a function of prices $p(\cdot \mid s; \mathbf{c})$.

Let $n \in \mathbb{R}_+^{|S|}$ denote the distribution of households across the states (i.e $n'\mathbf{1} = \bar{n}$ where $\bar{n}$ is the total number of households). Having a model of transition probabilities is powerful and enable one to write the value function, the optimal discounted revenue given an initial distribution n across the states, by the following recursive function:

$$V^*(n) = \max_{\mathbf{c}\geq 0} n'\mathbf{c} + \beta \sum_{n'} V^*(n')p(n'|n;c) \tag{3.15}$$

Note that prices are written as state-contingent prices but in practice the dimension of prices is much smaller (e.g the number of plans available). As an approximation, assume $\bar{n}$ is large enough that the transition to next period's distribution n' is certain. In that case, we can write the value function as

$$V^*(n) = \max_{\mathbf{c}\geq 0} n'\mathbf{c} + \beta V^*(\mathbf{P}(\mathbf{c})n) \tag{3.16}$$

where $\mathbf{P}(\mathbf{c})$ is the transition matrix as a function of prices c where the i, j element is the probability of transitioning from j to i.

With some assumptions, Blackwell's conditions trivially apply here so equation (3.16) describes a contraction and hence a unique fixed point exists and can be found by computational methods. In particular, care needs to be taken such that prices cannot be raised indefinitely so discounted revenue is unbounded. For example, the space of functions is bounded above if there is a maximum price $\bar{c}$ such that $\forall c >= \bar{c}$, $\mathbf{P}(\mathbf{c})$ is such that all members churn to the non-member state and all non-members remain non-members. A lower bound is guaranteed from the positive restriction on prices.

Given a fixed point of (3.16), the optimal price at n is simply the policy function

$$c^*(n) = \arg\max_{\mathbf{c}\geq 0} n'\mathbf{c} + \beta V^*(\mathbf{P}(\mathbf{c})n) \tag{3.17}$$

4 CONCLUSION

While the existing literature on customer lifetime value (LTV) is rich, we demonstrated that incremental LTV is the correct quantity to estimate to value acquired or retained subscribers whereas LTV tends to overvalue acquired or retained subscribers.

This paper establishes a methodology to estimate incremental LTV and as such, the monetary value of an acquired or retained subscriber. We leverage a finite state Markov chain to model the on and off-service behaviour of subscribers in a closed system. Unlike LTV models, subscribers do not disappear at churn and instead have positive forward looking LTV as they may rejoin in the future. Hence we can compare the forward looking LTV when a subscriber is acquired or retained, against that if they had remained a non-subscriber or churned. The difference represents the additional revenue the business can expect to receive with the customer in a subscriber state versus as non-subscribers. This allows for monetary valuation for a host of investments that are common in subscription businesses, yet typically measured in units of subscribers.

Finally, we demonstrated additional applications of the model. We used the Markov chain methodology to answer three common yet challenging problems that subscription services face, (i) forecasting the future number of subscribers, (ii) estimating the impact of subscription price increase on revenue and subscription growth, and (iii) optimizing policies, such as price discounting, that maximize expected lifetime revenue.

[9] Doing so is beyond the scope of this paper but the key is overcoming sparsity with sensible assumptions that incorporate domain expertise. For example, prices beyond the support of data leading to sensible changes in transition probabilities.

REFERENCES

[1] Josef Bauer and Dietmar Jannach. 2021. Improved Customer Lifetime Value Prediction With Sequence-To-Sequence Learning and Feature-Based Models. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 15, 5 (2021), 1–37.

[2] Tarek Ben Rhouma and Georges Zaccour. 2018. Optimal marketing strategies for the acquisition and retention of service subscribers. *Management Science* 64, 6 (2018), 2609–2627.

[3] Paul D Berger and Nada I Nasr. 1998. Customer lifetime value: Marketing models and applications. *Journal of Interactive Marketing* 12, 1 (1998), 17–30.

[4] Benjamin Paul Chamberlain, Angelo Cardoso, CH Bryan Liu, Roberto Pagliari, and Marc Peter Deisenroth. 2017. Customer lifetime value prediction using embeddings. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1753–1762.

[5] Ching-Hsue Cheng, You-Shyang Chen, and Ya-Ling Wu. 2009. Forecasting innovation diffusion of products using trend-weighted fuzzy time-series model. *Expert Systems with Applications* 36, 2 (2009), 1826–1832.

[6] Peter J Danaher. 2002. Optimal pricing of new subscription services: Analysis of a market experiment. *Marketing Science* 21, 2 (2002), 119–138.

[7] Ömer Fahrettin Demirel, Selim Zaim, Ahmet Çalişkan, and Pinar Özuyar. 2012. Forecasting natural gas consumption in Istanbul using neural networks and multivariate time series methods. *Turkish Journal of Electrical Engineering & Computer Sciences* 20, 5 (2012), 695–711.

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

[9] Keisuke Hirano and Guido W. Imbens. 2004. *The Propensity Score with Continuous Treatments*. John Wiley Sons, Ltd, Chapter 7, 73–84. https://doi.org/10.1002/0470090456.ch7 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/0470090456.ch7

[10] David Hubbard, Benoit Rostykus, Yves Raimond, and Tony Jebara. 2021. Beta Survival Models. In *Proceedings of AAAI Spring Symposium on Survival Prediction - Algorithms, Challenges, and Applications 2021 (Proceedings of Machine Learning Research)*, Russell Greiner, Neeraj Kumar, Thomas Alexander Gerds, and Mihaela van der Schaar (Eds.), Vol. 146. PMLR, 22–39. https://proceedings.mlr.press/v146/hubbard21a.html

[11] Woonghee Tim Huh, Soulaymane Kachani, and Ali Sadighian. 2010. Optimal pricing and production planning for subscription-based products. *Production and Operations Management* 19, 1 (2010), 19–39.

[12] Robin John Hyndman and George Athanasopoulos. 2018. *Forecasting: Principles and Practice* (2nd ed.). OTexts, Australia.

[13] Guido W. Imbens and Donald B. Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press. https://doi.org/10.1017/CBO9781139025751

[14] William Ogilvy Kermack and Anderson G McKendrick. 1927. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 115, 772 (1927), 700–721.

[15] Christos Louizos, Uri Shalit, Joris Mooij, David Sontag, Richard Zemel, and Max Welling. 2017. Causal Effect Inference with Deep Latent-Variable Models. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*. Curran Associates Inc., Red Hook, NY, USA, 6449–6459.

[16] Sungjoon Nam, Puneet Manchanda, and Pradeep K Chintagunta. 2007. The effects of service quality and word of mouth on customer acquisition, retention and usage. *Retention and Usage* (2007).

[17] Paul R. Rosenbaum and Donald B. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70, 1 (04 1983), 41–55. https://doi.org/10.1093/biomet/70.1.41

[18] Matthew Rowe. 2011. Forecasting audience increase on Youtube. *In Workshop on User Profile Data on the Social Semantic Web* (2011).

[19] Roland Rust, Katherine Lemon, and Valarie Zeithaml. 2004. Return on Marketing: Using Customer Equity To Focus Marketing Strategy. *Journal of Marketing* 68 (01 2004), 109–127. https://doi.org/10.1509/jmkg.68.1.109.24030

[20] John Thøgersen. 2009. Promoting public transport as a subscription service: Effects of a free month travel card. *Transport Policy* 16, 6 (2009), 335–343.

[21] Ali Vanderveld, Addhyan Pandey, Angela Han, and Rajesh Parekh. 2016. An engagement-based customer lifetime value system for e-commerce. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 293–302.

[22] Huizhi Xie and Juliette Aurisset. 2016. Improving the sensitivity of online controlled experiments: Case studies at Netflix. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 645–654.

[23] Jie Xiong, Väinö Jääskinen, and Jukka Corander. 2016. Recursive learning for sparse Markov models. *Bayesian analysis* 11, 1 (2016), 247–263.