Digital Marketing Twins

Samuel Levy

June 3, 2024

Abstract

This research provides a novel methodology, Digital Marketing Twins, that automatically extracts latent features from individual-level brand survey responses to inform a statistically-principled, deep generative model of customer-side brand affinity and firm-side performance factors. The proposed model enables marketers to find drivers of individual-level brand affinity, as opposed to traditionally observed metrics that must be analyzed in aggregation. The framework serves a counterfactual purpose at the customer level. The generative part of the model *completes* the distribution of survey responses over time, and across firms – thereby addressing the archetypal missing data problem – by imputing customer responses in counterfactual regimes. This research applies Digital Marketing Twins methodology to the competitive landscape of the U.S. wireless telecommunications retail market, leveraging a unique dataset of large-scale quarterly brand surveys from all three major carriers (AT&T, T-Mobile, and Verizon) from 2020 to 2022. Empirically, this approach reveals latent asymmetries in competition in terms of brand affinity, together with a nonlinear increase in brand affinity for certain types of drivers, such as satisfaction with network speed, but a nonlinear decrease in brand affinity for customers who report greater likelihoods of changing plans, providers, or devices, relative to their current wireless services.

Keywords:
Customer Satisfaction, Competitive Environment, Survey Research, Deep Generative Models.

Customer surveys are ubiquitous tools. Marketers leverage them to fuel brands and boost corporate growth, as well as determine the causes of customer satisfaction and of customer churn. Marketing researchers adopt them to learn customer preferences, gauge customer satisfaction, identify competitive offers, improve existing products and services, tailor marketing strategies, and innovate personalized services. Due to the ever growing complexity, and frequency of customer surveys and survey touchpoints though, firms today mostly rely on third-party platforms (e.g., Salesforce, HubSpot) or survey companies (e.g., Ipsos and Kantar) to execute large-scale customer surveys, known commercially as *brand surveys*. The increasing scale and scope of such brand surveys also is part of a broader trend of marketers embracing data-driven approaches, with an emphasis on the use of behavioral metrics to compute customer lifetime value (CLV) (e.g., Venkatesan and Kumar 2004; Fader, Hardie, and Lee 2005). Behavioral metrics help identify which customers are at risk of churning, though recent research also calls for efforts to distinguish the causes from the predictors of churn (Braun and Schweidel 2011; Ascarza et al. 2018; Ascarza 2018). That is, predictors of churn include demographics and behavioral patterns that statistically indicate a high likelihood of discontinuing products or services, but the causes of churn might include poor customer service, high prices, product quality issues, or more attractive offerings from competitors. Customer satisfaction targets the root issues that lead to customer attrition (Gustafsson, Johnson, and Roos 2005). Moreover, customer satisfaction and dissatisfaction drive companies' stock prices asymmetrically: dissatisfaction harms returns far more than a one-unit increase enhances them (Malshe, Colicev, and Mittal 2020). Accordingly, a new methodology is needed to carve out the "paths of least resistance" to individual-level customer satisfaction; this research proposes a prescriptive framework that pioneers such policy optimization by relying on customer surveys.

Despite the numerous benefits of customer surveys for marketers, there are two issues – one theoretical and the other practical – that plague most large-scale surveys carried out for customer relationship management (CRM) purposes. First, the surveys are difficult to integrate into prescriptive frameworks. Linear functional forms, strong parametric assumptions, and limited consideration of customer heterogeneity as a result of limited or incomplete individual-level data across brands makes it difficult for marketers to understand customer churn and retention in mature, competitive environments. Which marketing action should be recommended when customer satisfaction declines? Although one obvious lever is promotional offers, there are others; for example, in the wireless telecommunications industry, managers can increase customer satisfaction by improving network quality and network speeds, providing better data plans, strengthening brand perception, and solving problems that customers encounter when using providers' devices and services. However, the question remains: which aspects should be prioritized at the customer level? Second, in practice, customer surveys often are repeated cross-sectional. Because repeated cross-sectional surveys represent *different* sets of customers at various points in time, they cannot track individual changes. Unlike longitudinal surveys, they take "snapshots" and thereby provide less depth of information about individual respondents. If the sampled population changes significantly over time, comparisons between different cross-sectional surveys become challenging, if not impossible. For this reason, the nature of data variations in the repeated cross-sectional format of brand surveys is described as *pseudo-longitudinal*.

In this paper, we propose a novel methodology – Digital Marketing Twins – that leverages large-scale brand surveys conducted by a focal firm and its competitors in the U.S. wireless telecommunications retail market. This methodology finds paths of least resistance to individual-level customer satisfaction, in a statistically principled way. It uses a unique dataset from a representative sample of customers of AT&T, T-Mobile, and Verizon, the three major players in the U.S. telecommunications market. Using quarterly cross-sectional survey responses that span ten quarters – from 2020 to 2022 – the methodology overcomes both the substantive and technical limitations previously mentioned. The framework builds a generative model of customer preferences by flexibly mapping individual-level surveyed characteristics to various dimensions of customer satisfaction. Generative models capture the joint probability distribution between observed and latent variables of interest. In practice, they provide the steps that explain how the data are assumed to be generated, allowing marketers and researchers to incorporate domain expertise into their models. The generative aspect not only supports the forecasting of customers' responses in the next quarter, but also provides counterfactual responses according to different scenarios, such as customer responses as if they were using a different wireless carrier, all else being equal.

The digital twins approach, already an established method for counterfactual simulations in the realm of manufacturing, presents a novel and previously unexplored avenue for application within the marketing field. Until now, this innovative approach has, to the best of our knowledge, remained unapplied to this context. Digital twins integrate data from different sources to mimic the behavior of physical objects or systems; they can be used to test hypotheses, simulate scenarios, and optimize the performance of the systems. The use of generative models as a basis for digital twins is not novel; for example, generative adversarial networks (GANs) and conditional GANs can learn distributions of interest in structures with material nonlinearities and uncertainties (Tsialiamanis et al. 2021). In marketing contexts, digital twins serve a counterfactual purpose at the customer level. The generative part of the model *completes* the distribution of survey responses over time, and across firms, such that it can address the archetypal missing data problem. The proposed Digital Marketing Twins methodology offers a solution to the missing data problem that arises from the pseudo-longitudinality of brand surveys – in which individual respondents can be only observed in one time period and at one company – by imputing customer responses in the next time period and in a counterfactual regime in which all individually observed characteristics remain constant. The goal is to infer customer satisfaction under counterfactual regimes of their experiential, engagement, and usage characteristics, identifying the potential causes of satisfaction.

Table 1 provides a summary of the conceptual benefits of the Digital Marketing Twins framework. Quasi-experimental methods can be particularly useful for understanding the effect and strategic value of an intervention on an outcome of interest from recorded CRM data. A combination of propensity score matching and a flexible Bayesian parametric or non-parametric model such as a GAM, a Gaussian process (GP), or a mixture-of-normals can be useful for making counterfactual predictions. However, these methods are seldom scalable out-of-the box; they require careful selection of kernel functions and/or hyperparameters, which in turn requires expert knowledge. The quasi-experimental methods' inference (especially of natural experiments) is necessarily to quantifying what *has* happened, rather than what *could* happen. Although state-of-the-art predictive CRM methods provide forward-looking analytical insights, often with the help of rich machine learning techniques, they lack the ability to provide counterfactuals by focusing on predictors of customer churn, neglecting complex structures such as competitive effects, and relying on flexible statistics of customer behavior.

	Predictions / Tactical	**Counterfactuals / Strategic**
Retroactive	Basic CRM Reports	Quasi-Experimental Methods
Proactive	Predictive CRM Models	Digital Twins

Table 1: Digital Marketing Twins as a proactive and strategic framework for customer analytics.

From a technical standpoint, the goal of this research is to develop a novel deep generative and probabilistic latent factor model, as well as to leverage grid search to find the best marketing actions to recommend at the individual level, from a large-scale survey. The Digital Marketing Twins model captures customer-side brand affinity at the individual level, for each brand, in each time period, controlling for observed heterogeneity and firm-side factors. The inference model mapping from data to the latent space is parameterized by a neural network, for high flexibility. Latent, customer-side brand affinity provides an interpretable layer that maps to a latent utility model that in turn yields an ordinal logit structure for brand survey questions pertaining to customer satisfaction. To generate counterfactual responses and missing quarters, it relies on amortized inference, learning a set of parameters that can map any data point to the latent space. We train the model using stochastic variational inference with mini-batching, for high scalability and uncertainty quantification. After training, a grid search is used for the individual-level, latent, customer-side brand affinity for customers of the focal firm, to discover the marketing actions most likely to increase customer satisfaction. This search leads to a path of least resistance at the individual level, enabling marketers to use surveys to identify causes of satisfaction.

The remainder of the paper is organized as follows: we first present a literature review. The subsequent section contains a description and exploration of the data, using simple descriptive techniques. Then, the methodology is provided. Fit metrics, benchmarks for the Digital Marketing Twins model against nested baseline models, analysis of the probabilistic latent factors, and counterfactual results are in the next section. The last section concludes.

RELATED LITERATURE

This work contributes to academic literature on digital twins using generative modeling, as well as to literature on machine learning methods in marketing for competitive environments; it also shows that early models for customer satisfaction are special cases of the proposed framework.

Digital Twins and Generative Modeling

Tsialiamanis et al. (2021) suggest how to advance system simulation by creating digital twins for specific systems, referring to fields such as manufacturing, control systems, the Internet of Things (IoT), smart cities, social networks, and management. Digital twins can help predict the behavior of structures in different situations, thus maximizing the operational lives of the structures and minimizing costs. However, the construction of digital twins is inherently complex and uncertain. Aleatory uncertainty, related to random events, and epistemic uncertainty, related to a lack of knowledge, are key considerations. To address these issues, Tsialiamanis et al. (2021) propose the use of generative models as the foundation for a digital twin, providing estimations of aleatory and epistemic uncertainty. They study two types of generative models: the Stochastic Finite Element (SFE) method – a physics-based, white-box model – and the Conditional Generative Adversarial Network (cGAN) – a data-driven, black-box model. Each has strengths and limitations. For example, SFE models excel in predefined conditions but struggle with unknown scenarios, whereas cGANs can perform across a wide range of conditions but cannot extrapolate beyond available data. With a hybrid, grey-box approach, incorporating both models to overcome these limitations, generative models might better accommodate uncertainty in digital twins. By combining a generative white box (SFE) and a generative black box (cGAN), they propose a fully generative grey box that they assess in relation to other existing models, such as variational auto-encoder (VAE) and Gaussian processes (GP). In this paper, on the other hand, we use a deep generative model building on a variational auto-encoding neural network to mirror the competitive environment, and use a grid search to optimize the latent customer-side brand affinity.

Kapteyn, Pretorius, and Willcox (2021) propose a new mathematical foundation for digital twins, that is, computational models that mirror structures, behaviors, and contexts of physical assets. Because digital twin applications usually require extensive resources and expertise for implementation, the authors propose a unifying mathematical model that uses dynamical systems theory and probabilistic graphical models, with the digital twin and the physical asset modeled as coupled dynamical systems that evolve over time, and the digital twin constantly updating its internal models according to observational data. By dmonstrating this approach with a digital twin of an unmanned aerial vehicle (UAV), they show how the model aids in calibration by updating internal models, and facilitating decision-making. They conclude with the presentation of an abstract state-space formulation for digital twins, describing a realized dynamic decision network based on this mathematical model and illustrating its application to a UAV's structural digital twin.

Finally, Yu et al. (2021) propose a health monitoring solution for complex systems in smart manufacturing, applying a digital twin approach with a nonparametric Bayesian network model. With advancements in sensor technology and artificial intelligence, modern manufacturing systems need to be intelligent, visual, and capable of self-assessing their health throughout their life cycle. The Prognostic and Health Management (PHM) process is crucial in this context, and the proposed model offers an innovative solution for tracking the health states of such systems. The model collects sensor data from the physical world, updating its simulated physical model in real time and providing optimization and decision support. Their nonparametric Bayesian network model can adapt in real-time too, thus reducing model uncertainty. Yu et al. (2021) also include model validation experiments on electro-optical systems, and provide more accurate health monitoring than a traditional data-driven Convolution Neural Network (CNN) approach.

Machine Learning Methods in Marketing for Competitive Environments

Among the proposals for machine learning techniques to study market structures and competitive landscapes, Netzer et al. (2012) systematically analyze online user-generated content to "listen" to what customers write about a focal firm's and competitors' products; they use text mining to overcome the difficulties involved in extracting and quantifying the wealth of online data that customers generate and network analysis tools to convert the mined relationships into co-occurrences among brands or between brands and terms. (Tirunillai and Tellis 2014) extract latent dimensions of customer satisfaction with quality, using an unsupervised latent Dirichlet allocation model, and Lee and Bradlow (2011) automatically elicit product attributes and extract brands' relative positions from online customer reviews, providing both predictive and descriptive support for managerial decision making.

Brand competition occurs not only in single markets, but also in different sub-markets and structured markets. The hypothesis of multiple structured markets (Kannan and Wright 1991) helps us understand how brands compete by including marketing mix variables. In type-primary markets, "switchers" are highly responsive to changes in marketing mix variables whereas in brand-primary markets, the "loyal" segment remains relatively unresponsive to marketing programs (e.g., in contexts of ground coffee purchases or store panel records). Ringel (2023) recently proposed the visualization of brand competition in a multimarket membership product (MMP) context, in which products that compete in multiple submarkets that are each characterized by distinct competitors and customer preferences, with competitive relationships inferred from customers' online searches using bootstrapped neural network product embeddings in the digital camera market.

Customer Satisfaction and Survey Research

This research contributes to the rich literature on customer satisfaction in marketing. Customer satisfaction is one of the most prominent metric of interest for marketers; for example, improvements in customer satisfaction are linked to reduced risk in both systematic and idiosyncratic stock returns, suggesting that customer satisfaction not only enhances financial performance but also stabilizes it, making a compelling case for firms to report these metrics in their annual financial disclosures (Tuli, Kohli, and Bharadwaj 2007). Marketing models have also focused on understanding the drivers of customer satisfaction. For example, using an ordered probit model, Kekre, Krishnan, and Srinivasan (1995) study determinants of customer satisfaction for software products and service support for mainframes and workstations. Their main dependent variable is an overall satisfaction score, measured on an ordered categorical scale. The authors can explain how certain features of the software, such as reliability, capability and usability, affected overall satisfaction. They consider other explanatory variables, such as the type of product and the type of user, allowing for interaction effects. If an ordered logit were substituted for their ordered probit, their model can be nested within the proposed framework, by replacing amortized neural networks with linear functions, assuming that business key performance indicators (KPI) have no impact on customer satisfaction and considering only overall satisfaction as a unique target variable. Importantly, any specific survey question on satisfaction can only solicit one of the many facets of a customer's overall brand affinity. This is indeed how survey questions ought to be designed. Recognizing this need for a holistic metric, this paper extends the understanding of customer satisfaction by operationalizing an overall metric of brand affinity, as a latent variable learned across the multitude of satisfaction questions.

EXPLORING THE DATA

The data for this study consist of repeated cross-sectional responses from a brand survey for all major U.S. telecom carriers (AT&T, T-Mobile, and Verizon) between the third quarter of 2019 and the third quarter of 2022. Responses are recorded quarterly. For each carrier – not necessarily in the order previously mentioned, for confidentiality – we observe responses from a sample of 8770 customers, 7129 customers, and 4370 customers, in every quarter. The number of customers is the same across quarters for a given carrier, but the sample of customers for each carrier differs between quarters (i.e., repeated cross-sectional data).

According to managerial sources from one of the three major U.S. telecom carriers, the objectives of this survey were threefold: (1) gain an understanding of how wireless, internet, and pay TV customers view and rate customer experience with their carrier or provider, (2) determine the driving factors of customer satisfaction, and (3) determine what the focal firm does well and where it falls behind competitors, according to not only customer satisfaction (i.e., net promoter score) but also specific drivers and attributes.

Inputs and Outputs

The questions in the survey data fall into three categories, representing three different goals. The first group of questions provides customer characteristics; their characteristics have a predictive function, because they cannot be manipulated by managers (e.g., age and ethnicity cannot be influenced by any marketing action). The second group of perceptual questions offer immediate strategic value to managers, in that they ask customers about their feeling toward competing carriers. The third group of questions relates directly to customer satisfaction and form the basis for the proposed digital twin approach, because we assume that managers aim to maximize customer satisfaction. Therefore, the survey questions reflect three categories:

- Predictive Variables: during the inference phase, and at test time, these variables are fixed. In the grid search phase, they remain fixed. A key assumption of the model is that invariant predictors completely characterize customer heterogeneity. For the empirical application, we use large numbers of socio-demographic and usage questions, including age, gender, race or ethnicity, annual household income before taxes, devices at home, name of the wireless service provider, type of plan, tenure with provider, dollar amount paid per month for the plan, data usage, 5G usage, and rewards program.
- Strategic Variables: at training time and test time, these variables are fixed. At grid search time, these variables are the arguments of the optimization problem, and are assumed to be manipulated by the marketing analyst. They include:
  - Importance of any of the following according to likelihood to recommend: network in rating, price / value; billing; customer service; general feeling; plans; rewards and benefits; other factors.
  - Satisfaction with network speed; network reliability; data plans that meet my needs; value of price paid; accuracy of billing; rewards and recognition; ease of doing business; solving problems for the first time; "brand for me"; total cost of wireless service; device selection.
- Target Variables: these variables are reconstructed at inference time, and predicted at test time, and optimized at grid search time. They include[1]:
  - Likelihood to recommend (LTR) (0-10);
  - Likelihood to recommend current provider's phone to a friend or a colleague (Phone LTR) (0-10);

[1] The complete list of strategic and target variables is available in the appendix; because there are more than 300 one-hot encoded predictors, they are not listed here. The complete list remains available upon request.

- Likelihood to switch wireless service providers within the next 12 months (Intention to Switch) (0-4);
- Overall satisfaction with current provider (0-9);
- Overall feeling about current provider (0-4);
- Overall feeling about competitions' providers[2] (0-4).

The model also controls for different aspects of firm performances, using Generally Accepted Accounting Principles (GAAP) and non-GAAP measures published quarterly by AT&T, T-Mobile, and Verizon between the second quarter of 2020 and the fourth quarter of 2022. For brevity, we denote these variables as business key performance indicators (KPIs). They include total revenue, operating revenue, cost of revenue, gross profit, operating expense, churn, and average revenue per user (ARPU). The measures are standardized. Tables 4 and 5 (in the appendix) lists all questions included as target variables and strategic variables. Figure 8 includes summary statistics at the question level, per carrier.

Multivariate Analysis

Before providing a generative model of the target variables (LTR, Phone LTR, Satisfaction, Overall Feeling about Carrier $\{1, 2, 3\}$, Intention to Switch), it is helpful to understand how they are associated with one another. Therefore, we undertake a correlation analysis of the target variables, at the carrier level.

First, we recode the Intention to Switch as Retention Likelihood, applying the formula $f(x) = 4 - x$. Intuitively, Satisfaction, Likelihood to Recommend and Overall Feeling about Own Provider should correlate positively with Retention Likelihood; an empirical analysis verifies these correlations (Figure 1). For each carrier, the correlation between retention likelihood and satisfaction ranges from .28 to .31. The correlation between Retention Likelihood and Feeling about Current Provider also is positive (respectively, .50, .43, and .46 for Car-

[2]For example, an AT&T customer is asked about their overall feeling about T-Mobile and Verizon, as two separate questions.

riers 1, 2, and 3). Phone LTR is also positively correlated with Retention Likelihood. The strong positive correlations between LTR and Feeling about Own Provider and Satisfaction suggest that marketers at least indirectly capture a measure of satisfaction when they record the popular Net Promoter Scores[3] (Reichheld 2003).

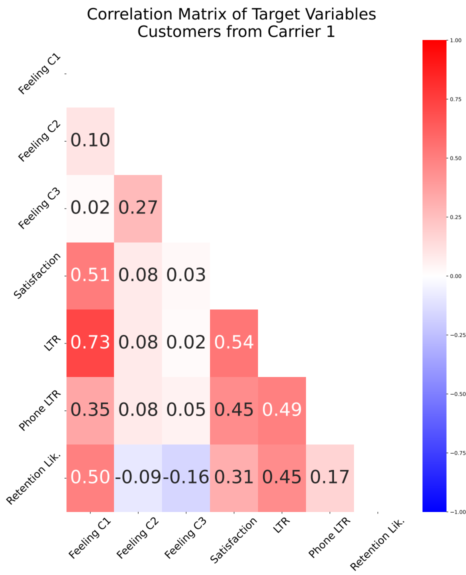
It is more challenging to understand the relationship between Feeling about Competitions' Providers and other target variables. More positive feelings are associated with a lower Retention Likelihood (correlations from -.08 to -.18, Figure 1). However, more positive feelings are also associated with higher LTR, Phone LTR, and Satisfaction, suggesting that customers may simply be "happier" about the telecommunications industry in general. This suggestion is corroborated by the slightly positive correlation between all Feeling measures about Own and Competitor's providers.

Finally, it is interesting to notice symmetries and asymmetries between the three carriers in terms of correlations across target variables. In terms of symmetries, customers from all carriers tend to have more positive or more negative feeling about all competitions' carriers simultaneously; moreover, the variable that correlates most with Retention Likelihood is Feeling about Own carrier, whereas Overall Satisfaction comes second. In terms of asymmetries, customers from Carrier 3 do not express positive or negative associations with their Feeling about Own Carriers and Competitions' Carriers (correlations of .01 and .03, in Figure 1) whereas customers from Carrier 1 and especially Carrier 2 tend to have a stronger associations (correlations of .14 and .16 for Carrier 2, Figure 1).

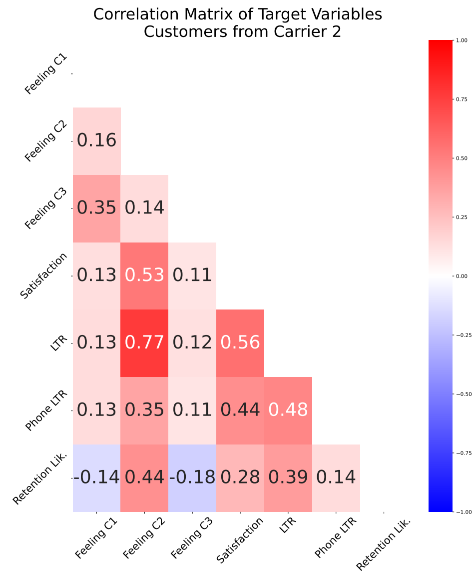
Linear Modeling is Limited for Analyzing Brand Surveys

The first step in gauging whether a nonlinear model is needed to investigate the relationship between explanatory variables (invariant predictors) is to compare two simple discriminative machine learning models. For simplicity, we use multiple output linear regression as a benchmark for the discriminative linear model, and multiple output random forest as a

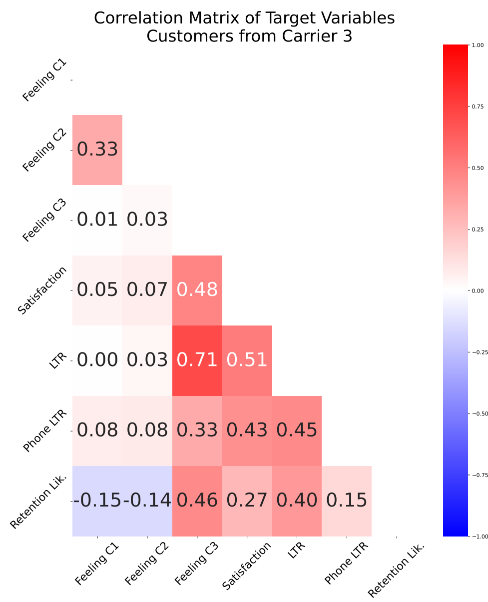
[3]The Net Promoter Score is the measure that transforms LTR by assigning a -1 to respondents who indicate a LTR of 0 to 6, 0 for LTR between 7 and 8, and +1 for LTR of 9 and 10, then averaging across all respondents.



(a)



(b)



(c)

Figure 1: Correlation matrices for target variables in the survey, for each carrier. Notes: Feeling C1, Feeling C2, and Feeling C3 respectively refer to Overall Feeling about Carrier 1 (a), Carrier 2 (b), and Carrier 3 (c), respectively.

benchmark for the discriminative nonlinear model.

Multiple output linear regression is a generalization of the simple linear regression model when more than one output variables is considered. The model learns a linear relationship between the input variables and each of the output variables. Each output variable is modeled as a linear combination of the input variables plus an error term. In contrast, random forests are ensemble learning methods that operate by constructing a multitude of decision trees and outputting the mean prediction for a regression task. A multiple output random forest is an extension of this technique to handle multiple output variables. This method is useful when the output variables are not independent of each other and share some correlation, as indicated in Subsection 3.2.

Fitting the multiple-output linear regression and the multiple-output random forest on the pooled data shows that the coefficient of determination score R^2 is higher for all target variables in the random forest analysis (Table 2). This higher goodness-of-fit metric indicates that the nonlinear model better captures the relationships between input (invariant and strategic) variables and target variables.

Although linear regression and random forest are useful models, they are fundamentally predictive and do not necessarily provide meaningful interpretations of the relationships between inputs and outputs. For counterfactual reasoning, other techniques may be more desirable, such as structural equation modeling (SEM). Yet, even if SEM may be more interpretable, it also yields questionable assumptions, because it is difficult to know *a priori* how the different input variables relate to latent constructs of interest that explain the various target variables that managers care about and try to optimize.

MODELING FRAMEWORK

The first subsection details the model architecture and training, based on mapping customer-side and firm-side input variables to latent variables using amortized neural networks. After documenting the latent variables and their marketing interpretations, we present the model

Data Source	Multiple Output Discriminative Models	
	Linear Regression (Test R^2)	**Random Forest (Test R^2)**
Q12 (Likelihood-to-recommend, LTR)	.78	**.82**
Q27 (Phone LTR)	.30	**.42**
Q18 (Satisfaction)	.58	**.67**
Q20Ar1 (Overall Feeling Carrier 1)	.38	**.58**
Q20Ar2 (Overall Feeling Carrier 2)	.40	**.57**
Q20Ar3 (Overall Feeling Carrier 3)	.39	**.55**
Q20 (Intention to Switch)	.52	**.60**

Table 2: Goodness-of-fit of two discriminative models on a test sample of the data. Notes: Multi-output Linear regression is a standard benchmark, while multi-output random forest is a nonlinear benchmark. A nonlinear model better explains the variation in the data.

layer for ordered categorical variables. Next, we detail the inference procedure, with implementation details. We further outline the predictions tasks, and finally, we offer a description of the grid search phase.

Mapping Customer-Side and Firm-Side Input Variables to Latent Parameters through Amortized Neural Networks

The survey data refer to N respondents and K carriers. In a single quarter t, for a given firm k, a subset N_{kt} of these N respondents are surveyed, such that $\sum_{k=1}^{K}\sum_{t=1}^{T} N_{kt} = N$. The number of respondents within a firm remains constant over time, but individuals are not surveyed more than once, such that $N_{k1} = \cdots = N_{kT}$ for all $k = 1, \ldots, K$. Throughout this paper, $K = 3$, referring to the three largest carriers: AT&T, T-Mobile, and Verizon[4]. The survey data has J questions. We label the J^{pred} predictive variables $\mathbf{x}_{it}^{pred}$, the J^{str} strategic variables $\mathbf{x}_{it}^{str}$ and the $J^{targets}$ targets $\mathbf{y}_{ijtk}$, where $i = 1, \ldots, N$ customers, $j = 1, \ldots, J$ questions, $k = 1, 2, 3$ firms. In summary, $J^{str} + J^{pred} + J^{targets} = J$. The training phase does not make a conceptual distinction between invariant predictors and strategic variables, because they enter the same neural network. Therefore, $\mathbf{x}_{it} = [\mathbf{x}_{it}^{str}, \mathbf{x}_{it}^{pred}]^T$. For a given quarter t and a given firm k, business KPI are $\mathbf{x}_{kt}^{KPI}$, which is a vector of size H.

Amortized inference refers to inference over variational parameters that are parameterized

[4]Sprint was also a major carrier but merged with T-Mobile U.S. on April 1, 2020. It was the fourth-largest telecommunications carrier in the United States before the merger. Since the data starts in Q2 2020, Sprint customers are omitted here.

by a function of the data, instead of approximating separate variables for each data point (Zhang et al. 2018). For this research, the parameterized function is the neural network $f(.)$ that represents the parameters of the variational distribution across all data points from the J questions in the survey. An alternative would be to separately learn a set of parameters for each data point, rather than learning a set of mean and location parameters for each customer, at each time period, and each firm. The word "amortized" herein means that the cost of learning the variational parameters is "amortized" over all the data points.

Amortized inference is a powerful way to infer the posterior over customer-level and firm-level latent variables according to $\mathbf{x}^{KPI}$, $\mathbf{x}^{str}$ and $\mathbf{x}^{pred}$. Using variational inference to approximate the posterior distribution of customer-side and firm-side latent variables implies replacing the locational variational parameters with a function of the data where parameters – weights and biases of neural networks – are shared across all data points, for all firms and at all quarters. The neural network parameters automatically learn a complex representation of the inputs across firms and over time, and this representation is mapped to latent variables that are the building blocks of the target variables. As a major methodological contribution, the current study proposes amortized inference as a way to augment repeated cross-sectional data.

The feed-forward neural networks $f(.)$ and $g(.)$ map customer-side and firm-side, respectively, input variables to a set of latent location and scale parameters that generatively model the target variables. In such feed-forward neural networks, hidden layers are dense and sequentially connected. Consider the feed-forward neural network function $f(\mathbf{x}; \theta_f)$ with D hidden layers; is detailed as follows. The input layer is $d = 0$, hidden layers are $d = 1, 2...$, and the output layer is D. The weights connecting layer d and layer $d + 1$ can be referred to as $W^{(d)}$, and the biases in layer $d + 1$ are indicated by $b^{(d)}$. The pre-activation at layer $d + 1$ can be denoted as $a^{(d+1)}$, and the post-activation is $h^{(d+1)}$. The activation function is the hyperbolic tangent ($tanh$), such that:

$$
\begin{aligned}
\mathbf{a}^{(1)} &= W^{(0)}\mathbf{x}_{it} + \mathbf{b}^{(0)} && (1)\\
\mathbf{h}^{(1)} &= \tanh(\mathbf{a}^{(1)}) && (2)\\
\mathbf{a}^{(2)} &= W^{(1)}\mathbf{h}^{(1)} + \mathbf{b}^{(1)} && (3)\\
&\vdots \\
\mathbf{a}^{(D)} &= W^{(D-1)}\mathbf{h}^{(D-1)} + \mathbf{b}^{(D-1)} && (4)\\
\mathbf{h}^{(D)} &= \tanh(\mathbf{a}^{(D)}) && (5)\\
\mu_{ikt} &= W_{\mu}^{(D)}\mathbf{h}^{(D)} + \mathbf{b}_{\mu}^{(D)} && (6)\\
\nu_{ikt} &= \exp\left(W_{\nu}^{(D)}\mathbf{h}^{(D)} + \mathbf{b}_{\nu}^{(D)}\right) && (7)
\end{aligned}
$$

Here, $\mathbf{x}_{it}$ is the batch input to the network. Because ν_{ikt} is a variance and must be non-negative, we apply an exponential function to obtain it from $\mathbf{a}^{(D)}$. The weights and biases (collectively referred to as $\boldsymbol{\theta}_f$) are learned by training the network. These weights and biases are parameters of amortized neural networks.

For the feed-forward neural network $g(\mathbf{x}^{KPI}; \boldsymbol{\theta}_g)$ with D' hidden layers, the inputs are the KPI for the three major carriers in the U.S. market (AT&T, T-Mobile, and Verizon), published quarterly over the corresponding 10 quarters of survey data. The neural network's output at a batch level is a concentration parameter γ_{ktl} and a rate parameter ω_{ktl}. The function g also relies on amortization to learn a shared representation across all quarters and firms instead of learning individual γ_{ktl} and ω_{ktl}. The dimension l refers to a set of L latent dimensions summarizing the various aspects of performance across firms and over time. These L latent dimensions provide dimensionality reduction, similar to principal components in principal component analysis.

Finally, the use of the $tanh$ activation function introduces non-linearities between layers, allowing the network to learn complex mappings from inputs to outputs. The $tanh$ function is particularly well-suited to the empirical application, due to its differentiability and its output range of -1 to 1, which helps with the normalization of the outputs.

Interpreting the Probabilistic Latent Factors Generating Digital Twins

For the latent parameter layer of the digital twin architecture, which includes the latent variables and their prior distributions, recall that i indexes customer identifiers from 1 to N; k indexes firms from 1 to K; t indexes time from 1 to T. Customer-side factors include the following latent variables:

- $z_{ikt} \sim \mathcal{N}(\mu_{ikt}, \nu_{ikt})$: this latent factor has a Normal prior distribution. Because μ_{ikt} and ν_{ikt} are functions of an amortized neural network, this prior is highly flexible and encodes a wide range of customer characteristics, automatically accounting for interactions and nonlinearities. This parameter is interpreted as the *customer-side brand affinity factor*; it represents, for a given customer at a given time, their affinity with brand k.

- $\alpha_{jkt} \sim \mathcal{N}(0, 1)$: The parameters α_{jkt} represent the *baseline* for question j for firm k at time t. It has a standard Normal prior distribution for simplicity.

- $\beta_{jl} \sim \mathcal{N}^{+}(0, 1)$ The parameter β_j is interpreted as the *polarization* of question j in the l-th dimension of service quality, that is, how much question j elicits a response on the l-th service characteristic.

The firm-side factors include the following latent variables:

- $\phi_{ktl} \sim \mathcal{G}(\gamma_{ktl}, \omega_{ktl})$: The parameterization of a prior on ϕ_{ktl}, a firm-side latent factor on dimension l for firm k at time t, has a Gamma prior distribution. Because γ_{ktl} and ν_{ktl} are functions of an amortized neural network, this prior also is highly flexible; it encodes a wide range of firm characteristics, automatically accounting for interactions and nonlinearities.

Support for both ϕ_{ktl} and β_{jt} is the real positive line, for identification. The sign of z_{ikt} becomes then immediately interpretable, as explained in Subsection 4.3.

Model Layer for Ordered Categorical Outcomes

Because the target variables are all ordered categorical variables, we use an ordered logit specification. Let y^*_{ijkt} denote the latent response of respondent i to the entire set of J questions. Questions have different numbers of scale points: some questions have five scale points (0-4) whereas others have 10. For $M+1$ common and ordered cut points $\{c_m : c_{m-1} \leq c_m, m = 1, \ldots, M\}$ where $c_0 = -\infty$ and $c_M = +\infty$, latent utility values y^*_{ijkt} depend linearly on α_{jkt}, which are baseline values for question j at firm k; the customer-side brand affinity z_{ikt}; the polarization of question j in firm-side latent dimension l, β_{jl}; and the firm-side factors ϕ_{ktl}:

$$y^*_{ijkt} = \alpha_{jkt} + z_{ikt} \sum_{l=1}^{L} (\beta_j \phi_{kt})_l + \varepsilon_{ijkt} \qquad \text{where } \varepsilon_{ijkt} \underset{i.i.d.}{\sim} EV(0,1) \tag{8}$$

The individual responses y_{ijkt} for a customer $i = 1, \ldots, N$ of firm $k = 1, \ldots, K$ at question $j = 1, \ldots, J_{targets}$, at time $t = 1, \ldots, T$ take the following values: $m = 1, 2, \ldots, M$ where M is the maximum number of scale points for question j. Because questions differ in their total number of scale points, m and M should have a subscript j, but we omit it for simplicity. The following holds:

$$y_{ijkt} = m \qquad \text{if } c_{j,m-1} \leq y^*_{ijkt} \leq c_{j,m} \tag{9}$$

where a Dirichlet prior model applies to ordinal probabilities, which serves to induce cut points indirectly. This approach enables a proper, principled prior on the cut points, which is useful when some categories are not strongly separated due to their data sparsity in some categories (Betancourt 2020).

By marginalizing out the latent utilities y^*_{ijkt}, it is possible to write the probability of observing category m for question j in customer i of firm k at time t:

$$p(y_{ijkt}|c_{j,1},\ldots,c_{j,M}) = \begin{cases} \Pi(c_{j,1} - \alpha_{jkt} - z_{ikt}\sum_{l=1}^{L}(\beta_j\phi_{kt})_l) & \text{if } m = 1 \\ \Pi(c_{j,m} - \alpha_{jkt} - z_{ikt}\sum_{l=1}^{L}(\beta_j\phi_{kt})_l) & \\ -\Pi(c_{j,m-1} - \alpha_{jkt} - z_{ikt}\sum_{l=1}^{L}(\beta_j\phi_{kt})_l)) & \text{if } 1 < m < M \\ 1 - \Pi(c_{j,m-1} - \alpha_{jkt} - z_{ikt}\sum_{l=1}^{L}(\beta_j\phi_{kt})_l)) & \text{if } m = M \end{cases} \tag{10}$$

where $\Pi(.)$ is the cumulative distribution function of the Type I extreme value distribution, that is, the logistic function.

Inference and Implementation

The set of latent variables to infer is $\tilde{\mathbf{z}} = [\boldsymbol{z}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\mu}, \boldsymbol{\nu}, \boldsymbol{\gamma}, \boldsymbol{\omega}]$. The set of parameters to learn is $\boldsymbol{\theta} = [\boldsymbol{\theta}_f, \boldsymbol{\theta}_g]$. By writing $\mathbf{y}$ as the set of all observations (survey data and KPI), it is possible to approximate the posterior distribution $p_{\boldsymbol{\theta}}(\tilde{\mathbf{z}}|\mathbf{y})$.

Because of the size of the data, and the use of neural networks to parameterize the latent variables, exact inference (e.g., Markov chain Monte Carlo algorithms) is not feasible. Therefore, it is necessary to rely on approximate Bayesian inference; stochastic variational inference (SVI) aims at determining a variational distribution $q_{\boldsymbol{\lambda}}(\tilde{\mathbf{z}})$ that is as close as possible to the posterior $p_{\boldsymbol{\theta}}(\tilde{\mathbf{z}}|\mathbf{y})$ as measured by Kullback-Leibler (KL) divergence. Minimizing the KL divergence is equivalent to maximizing the evidence lower bound (ELBO) on the log marginal probability of the data $\log p_\theta(\mathbf{y})$, with $\log p_\theta(\mathbf{y}) \geq$ ELBO and $\log p_\theta(\mathbf{y}) -$ ELBO $=$ KL $(q_{\boldsymbol{\lambda}}(\tilde{\mathbf{z}}))\|p_{\boldsymbol{\theta}}(\tilde{\mathbf{z}}|\mathbf{y}))$.

The evidence lower bound (ELBO) is:

$$\mathcal{L}(\boldsymbol{\lambda}) = \mathbb{E}_{q_{\boldsymbol{\lambda}}(\tilde{\mathbf{z}})}\left[\log p_{\boldsymbol{\theta}}(\mathbf{y}, \tilde{\mathbf{z}})\right] - \mathbb{E}_{q_{\boldsymbol{\lambda}}(\tilde{\mathbf{z}})}\left[\log q_{\boldsymbol{\lambda}}(\tilde{\mathbf{z}})\right] \tag{11}$$

The ELBO creates two expectations with respect to the variational distribution. The first expectation, $\mathbb{E}_{q_{\boldsymbol{\lambda}}(\tilde{\mathbf{z}})}[\log p_{\boldsymbol{\theta}}(\mathbf{y}, \tilde{\mathbf{z}})]$, represents the expected log-likelihood of the data given the model parameters, which encourages densities that place their mass on configurations of the latent variables that explain the observed data (Blei, Kucukelbir, and McAuliffe 2017). The second expectation, $\mathbb{E}_{q_{\boldsymbol{\lambda}}(\tilde{\mathbf{z}})}[\log q_{\boldsymbol{\lambda}}(\tilde{\mathbf{z}})]$, is the negative divergence between the variational density and the prior. Maximizing the ELBO is akin to finding a balance between encouraging the model to fit the data well (maximizing the first term) and encouraging densities close to the prior (maximizing the second term).

In line with standard practice, this study uses mean-field variational approximation. The model implementations relies on the machine learning framework Google JAX for fast computation on Graphics Processing Unit (GPU), and the NumPyro probabilistic programming language (Phan, Pradhan, and Jankowiak 2019). With the Adam optimization algorithm (Kingma and Ba 2014), a Monte Carlo version of the loss function is optimized in Equation (11) and a test set is used to determine all model hyperparameters, namely, the number of hidden layers per neural network, number of hidden units per neural network, and number L of latent firm-side dimensions.

Digital Twin Generative Process

To summarize, the specification is such that for all $m = 1, \ldots, M$, $i = 1, \ldots, N$, $j = 1, \ldots, J$, $k = 1, \ldots, K$ and $t = 1, \ldots, T$:

$$
\begin{aligned}
\begin{bmatrix} \mu_{ikt} \\ \nu_{ikt} \end{bmatrix} &= f(\mathbf{x}_{it}, \boldsymbol{\theta}_f) && \text{where } f \text{ is a feed-forward neural network} \\
\begin{bmatrix} \gamma_{kt} \\ \omega_{kt} \end{bmatrix} &= g(\mathbf{x}_{kt}^{(\mathrm{KPI})}, \boldsymbol{\theta}_g) && \text{where } g \text{ is a feed-forward neural network} \\
\beta_j &\sim \mathcal{N}^+(0,1) \\
z_{ikt} &\sim \mathcal{N}(\mu_{ikt}, \nu_{ikt}) \\
\phi_{ktl} &\sim \mathcal{G}amma(\gamma_{ktl}, \omega_{ktl}) \\
y^*_{ijkt} &= \alpha_{jkt} + z_{ikt} \sum_{l=1}^{L} (\beta_j \phi_{kt})_l + \varepsilon_{ijkt} && \text{where } \varepsilon_{ijkt} \underset{i.i.d.}{\sim} EV(0,1) \\
\pi(\mathbf{c}_j | \kappa, \lambda) &= \mathcal{D}(p(\mathbf{c}_j, \lambda)|\kappa).|J(\mathbf{c}_j, \lambda)| \\
y_{ijkt} &= m && \text{if } c_{j,m-1} \leq y^*_{ijkt} \leq c_{j,m}
\end{aligned}
$$

where $\mathcal{D}$ is the Dirichlet probability density function, and J is the Jacobian matrix. A uniform Dirichlet prior $\kappa = (1, 1 \ldots, 1)$ and the anchor point $\lambda = 0$ identify the model, without loss of generality. Figure 2 provides an illustration of the full digital twin architecture. The hybrid deep learning and probabilistic generative framework allow the best of both worlds: high flexibility and representation power of the amortized neural network on the left hand side of the illustration, in orange, and high interpretability and theory-based for counterfactual reasoning on the right hand side, in green.

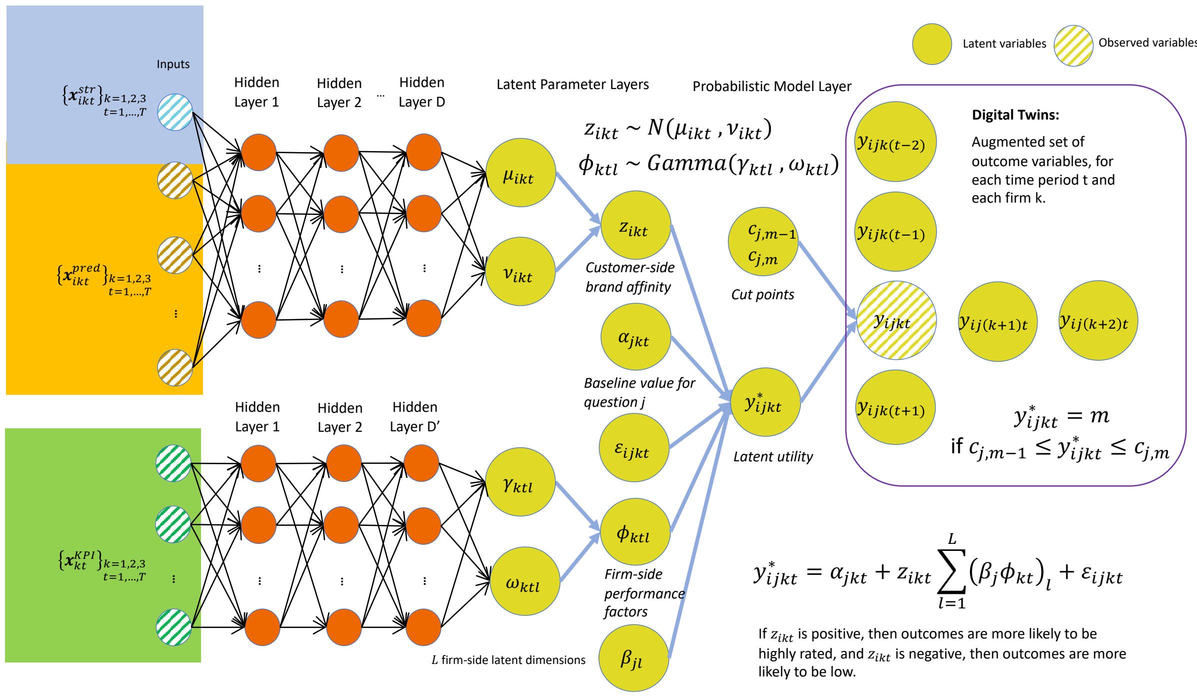


Figure 2: A deep probabilistic architecture of the modeling framework. An amortized inference neural networks (in orange, top) take survey data question as inputs and parameterize latent customer-side brand affinity z_{ikt}. Another amortized inference neural network (in orange, bottom) takes KPI data as input and parameterize firm-side performance factors ϕ_{ktl}. These two types of latent variables are then combined in the probabilistic model layer according to the Equation (8) as a latent customer-level utility. This latent utility is then evaluated against latent question-level cut points to present ordered categorical variables for target customer satisfaction questions.

Identification

Two core empirical challenges prevent marketing analysts from drawing respondent-level counterfactual inferences from the observed outcomes of brand surveys, or even when utilizing discriminative models. Among the T repeated cross-sectional surveys for a carrier k, it is highly unlikely that any respondent would repeatedly manifest across surveys. This data regime is not only common in commercial brand surveys, but across typical survey designs in the social sciences (Groves et al. 2011). Hence, the first challenge is that such pseudo-longitudinal setting disallows the application of standard longitudinal panel models to analyze these types of data. Second, except in rare cases, the presence of any customer i of carrier k in the U.S. wireless telecom sector precludes the possibility of their simultaneous presence as a customer of competing carrier k'. These two identification challenges motivate the development of Digital Marketing Twins.

In this section, we establish the mechanism of Digital Marketing Twins' generative framework in identifying respondent-level counterfactual outcomes across K carriers and T periods, while assuming a data generating process of the survey outcomes whereby any respondent i is only ever observed for a single carrier k, in a single period t.

The counterfactual identification strategy described below extends missing data approaches found in marketing and statistics (Rubin 1976; Little and Rubin 2019; Rossi and Allenby 2003) to formalize a set of antecedent modeling assumptions and how they translate into the consequent posterior predictive distribution. Where appropriate, testable vs. untestable assumptions necessary for internal validity are delineated, along with their implications on external validity (i.e., managerial actionability). Finally, we draw parallels to the closely related missing data paradigm of the Rubin Potential Outcomes framework (Rubin 1974), as well as contrast the counterfactuals under Digital Marketing Twins versus causal estimands based on potential outcomes – namely, in the analysis of brand surveys, there does not exist the notion of treatment interventions, which is the focal design of any causal inference undertaking.

Target counterfactual. The Digital Marketing Twins framework enables the identification of the posterior predictive distribution of:

$$p_\theta(\boldsymbol{z}_{ik'}|\mathbf{x}_{ikt}) \tag{12}$$

where $k' \neq k$ such that $\boldsymbol{x}_{ikt}$ is the J-length vector of observed survey outcomes for customer i of firm k who responded in period t, $\mathbf{z}_{ik'}$ is a T-by-D matrix of customer-side brand-affinity if i were – counterfactually – a customer of firm k' instead.

Let $\mathcal{D}$ denote the observed data generating process, which face the two aforementioned repeat cross-sectional empirical limitations; and $\mathcal{D}^*$ denote an oracle data generating process whereby all N respondents appear in all T periods and for all K carriers. Samples drawn from the above target distribution (Eq. 12) are defined as counterfactual samples if the corresponding stationary posterior given $\mathcal{D}$ is identical to the stationary posterior given $\mathcal{D}^*$.

If given oracle data with sufficiently large N, $\{\boldsymbol{x}\}_{i=1}^{N} \subseteq \mathcal{D}$, an arbitrarily flexible generative model can robustly and consistently infer the above target distribution (Eq. 12) by simply learning a bijective mapping of a respondent's outcomes from any period t, under any carrier k, to respondent's own outcomes for any other period $t' \in \{1, \ldots, T\}$ and carrier $k' \in \{1, \ldots, K\}$. However, oracle data are infeasible to collect, both due to cost of carrying out large-scale longitudinal brand surveys as well as the market reality that the vast majority of U.S. wireless consumers procure service from a single carrier at any time. Therefore, the identification of the counterfactual posterior predictive distribution from observed data relies on (1) desired empirical regularities in $\mathcal{D}^*$ that have equivalents in $\mathcal{D}$, and (2) undesired empirical regularities in $\mathcal{D}$ that must be controlled for via model specifications.

Formally, the posterior predictive distribution, conditional on observed data, that produces the desired counterfactuals must meet the following criterion on the KL-divergence, a measurement of the difference between distributions:

$$D_{KL}\left\{p_\theta(\mathbf{z}_{ik'}|\mathbf{x}_{ikt} \subset \mathcal{D}^*) \| p_\theta(\mathbf{z}_{ik'}|\mathbf{x}_{ikt} \subset \mathcal{D})\right\} = 0 \tag{13}$$

- **Assumption 1: Ignorability in y.** Extending the classic econometric age-period-cohort (APC) approach (e.g., Mason et al. 1973; Yang 2006) to modeling repeat cross-sectional panels via partial pooling, here ignorability posits that idiosyncratic differences across time and carriers can be deconfounded (i.e., ignorable) via the latent variables α_{jkt} and ϕ_{kt}. Whereas the APC framework assumes all cohort differences (i.e., selection artifacts and other unobservables) are captured by the additive parameter α_{jkt}, in Digital Marketing Twins, this deconfounding mechanism is extended to also include the multiplicative term ϕ_{kt}. Together, as amortized parameters, α_{jkt} and ϕ_{kt}

serve to flexibly control for confounding arising from unobservable factors that would bias the counterfactual inference of $\mathbf{z}_{ik'}$ in repeat cross-sectional settings.

- **Assumption 2: Comparability in x.** As shown in the model-free evidence, the brand surveys exhibit strong overlapping empirical support in input features x across periods and carriers. Having this overlap in the observed data generating process $\mathcal{D}$ signifies that – despite any respondent i is only ever observed for a single carrier k, in a single period t – the distributions of x are comparable across any other period $t' \in \{1, \ldots, T\}$ and carrier $k \in \{1, \ldots, K\}$. Under comparability, any sample from the posterior, when conditioned on identical values of x but varying in period and/or carrier, can be considered as interpolations within the empirical support – i.e., robust and consistent to the equivalent posterior under $\mathcal{D}^*$.

- **Assumption 3: Exchangeability in z.** Given assumptions 1 and 2, it follows that $p_\theta(\mathbf{z}_{ik'}|\mathbf{x}_{ikt})$ (Eq. 12) is robust and consistent to any permutation in the indexing of z and x. Should the indexing entail $p_\theta(\mathbf{z}_{ik'}|\mathbf{x}_{ikt})$, then we can interpret this posterior predictive distribution as the counterfactual distribution of the customer-side brand-affinity of customer i if they were – exchangeably – a customer of firm k' instead.

Lastly, while Eq. 12 has a canonical form of a conditional distribution, its validity is asserted through exchangeability, which is a weaker assumption than conditional independence. Whereas the latter can be assessed through empirical hypothesis testing, the former arises in counterfactual and missing data contexts where the validity of the inference on the unobserved outcome(s) must arise from assumptions on the data generating process, as done above. In summary, recognizing the "chasm" between the observed data generating process of surveys, $\mathcal{D}$, versus the ideal data generating process $\mathcal{D}^*$, the Digital Marketing Twins framework utilizes flexibly parameterizations to control for observed and unobserved confounders (Assumption 1), as well as exploits essential empirical regularities in $\mathcal{D}$ that mimics $\mathcal{D}^*$ (Assumption 2), to establish that the counterfactual inferences capable of being drawn from Eq. 12 are *exchangeably* valid across time and firms (Assumption 3) – despite the absence of the ideal, but unrealistic, longitudinal survey data.

Relation to Prior Literature

Variational Autoencoders

The model is novel in its use of customer-level predictors and strategic variables that parameterize an amortized neural network for high flexibility, that output structured latent variables that can be subsequently interpreted by marketers, and generate a coherent model of customer satisfaction. However, neither the use of amortized neural networks nor the use of grid search in marketing is novel.

The model resembles a Variational Autoencoder (VAE) (Kingma and Welling 2013), which is also a generative model that also uses variational inference for learning. The differences lie primarily in the specific structure of the model and the form of the decoder. In the VAE, the encoder is a neural network that takes the observed data as inputs, then outputs parameters of a distribution over the latent variables. The proposed model has two such "encoders", f and g, each of which produces parameters for different distributions over subsets of the latent variables, such that f encodes μ_{ikt} and ν_{ikt} for customer-side brand affinity z_{ikt}, and g encodes γ_{kt} and ω_{kt} for firm-side factors ϕ_{kt}.

In a VAE, the latent variables capture unobserved factors of variation in the data, whereas z_{ikt} and ϕ_{kt} capture observed factors of variation in the data, because they are parameterized by survey and KPI inputs. The only unobserved factors of variations come through the type I extreme value that affects latent utilities.

The decoder in a VAE takes the latent variables and generates parameters for the distribution over the observed data. The equations involving y_{ijkt} and y^*_{ijkt} can be interpreted as part of a kind of decoder that uses the latent variables, together with an ordered logit model, gives a distribution over the observed variable y_{ijkt}. However, unlike a VAE, this decoder does not involve a neural network but is determined by an ordered logit model and a latent factor model that decomposes firm-side and customer-side effects.

MODEL RESULTS

Fit and Benchmarks

The comparison of the proposed model with three benchmarks confirms its validity, as the goodness-of-fit metrics in Table 3 reveal. The three competing specifications are as follows: Model (1) is a linear version of the proposed model, in which the neural networks have been replaced by a linear layer. It is akin to a traditional SEM, with observed inputs and outputs, but has latent variables parameterizing the relationship between inputs and outputs. A traditional SEM is therefore a special case. Model (2) omits individual predictive and strategic variables from the full specification, but retains individual level customer-side brand affinity z_{ikt} for all $i = 1, .., N$, $k = 1, ..., K$, $t = 1, ..., T$ as "random effects". Model (3) omits KPI, but retains firm level performance factors ϕ_{ktl} for all $l = 1, ..., L$, $k = 1, ..., K$, $t = 1, ..., T$ as "random effects".

Figure 3 shows evidence of convergence, in that the average training and test loss (i.e., negative evidence lower bound – ELBO) decrease rapidly before stabilizing. The average test loss is close to the training loss, suggesting good generalizability. The greater variance of the test loss results from the test sample being smaller than the training sample.

In terms of goodness-of-fit, the proposed model (Model (4)) consistently performs better than other models across all carriers and metrics. The average training and test losses are lowest in this model, indicating that it offers the best fit to the data. For example, the Average Training Loss for Model (4) is 11.54, lower than the corresponding values for other models. The same trend is apparent in the Average Test Loss, in which Model (4) outperforms other models with a loss of 12.81. Furthermore, with regard to the Mean Absolute Error (MAE) for each carrier across different tasks, Model (4) generally exhibits the smallest error, suggesting it the most capable of accurately reconstructing the data. Some exceptions involve the Phone LTR and Carrier Satisfaction for Carriers 1 and 2, and the Retention Likelihood for Carrier 3, for which Model (4) does not perform best. However, the overall performance of Model (4) remains superior.

Thus, Table 3 suggests that a neural network model that includes individual predictors and KPI performs best among the presented models across a variety of reconstruction tasks. It also indicates the complexity of the relationships in the data and the ability of neural networks to better capture these complex relationships and extract predictive latent features.

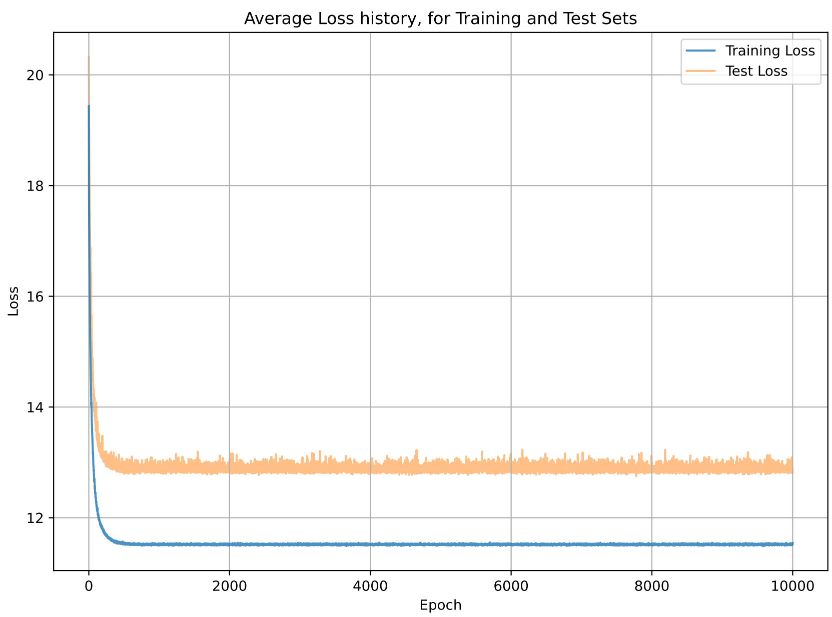


Figure 3: Average training and test loss (ELBO), over all individuals in training and test sets, respectively.
Notes: The plot suggests that convergence is reached after about 1000 epochs, but the model was trained for 10,000 epochs in total. Test loss is slightly greater than training loss, but remains constant after convergence, as expected.

Analyzing and Interpreting the Probabilistic Latent Factors

A crucial aspect of the proposed framework is its ability to analyze and interpret the estimated probabilistic latent factors while relaxing the functional form between various predictors and these factors for maximum flexibility.

Figure 4 plots the counterfactual customer-side brand affinities z^*_{ikt} using posterior means

Models		(1)	(2)	(3)	(4)
ϕ		Linear	NN	w/o KPI	NN
z		Linear	w/o indiv. predictors	NN	NN
No. epochs		10000	10000	10000	10000
Avg. Training Loss		12.22	14.06	11.57	**11.54**
Avg. Test Loss		13.49	15.81	12.94	**12.81**
Test Mean Absolute Error (MAE)					
Carrier 1	LTR	.84	1.59	67	.64
	Phone LTR	1.39	1.57	1.24	1.27
	Carrier Satisfaction	1.10	1.51	.95	1.00
	Overall Feeling Carrier 1	.53	.71	.49	.42
	Overall Feeling Carrier 2	.68	.61	.68	.64
	Overall Feeling Carrier 3	.55	.47	.56	.50
	Retention Likelihood	.74	.90	.69	.71
Carrier 2	LTR	.88	1.75	.69	.67
	Phone LTR	1.31	1.48	1.15	1.17
	Carrier Satisfaction	1.11	1.53	.97	1.01
	Overall Feeling Carrier 1	.70	.70	.69	.68
	Overall Feeling Carrier 2	.51	.75	.47	.43
	Overall Feeling Carrier 3	.61	.57	.60	.57
	Retention Likelihood	.85	1.01	.81	.82
Carrier 3	LTR	.84	1.54	.63	.62
	Phone LTR	1.35	1.60	1.20	1.21
	Carrier Satisfaction	1.11	1.54	.95	1.00
	Overall Feeling Carrier 1	.75	.69	.75	.70
	Overall Feeling Carrier 2	.76	.70	.77	.71
	Overall Feeling Carrier 3	.49	.68	.47	.38
	Retention Likelihood	.74	.89	.68	.69

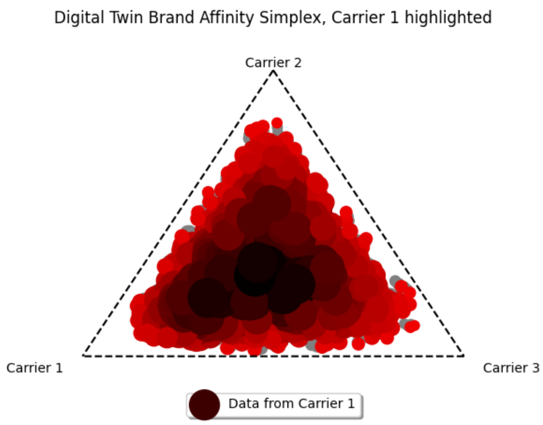
Table 3: Goodness-of-fit metrics. Training and Testing Loss, and Mean Absolute Error for Reconstruction Tasks. Model (4) is benchmarked against nested versions (1,2,3). Model (1) assumes a linear link between inputs (predictive variables, strategic variables, key performance indicators) and latent variables. Model (2) omits individual predictors. Model (3) omits key performance indicators.

across all customers. A darker color signifies higher density. Each dot represents the triplet $(z_{i1t}, z_{i2t}, z_{i3t})$ where $t = 0$, after transformation through a softmax function, which is then converted in a barycentric coordinate system to obtain values that live in a three-dimensional simplex. For example, a customer in Carrier 2 with a dot near the "Carrier 1" vertex reflects that the brand affinity level that customer would obtain if they were assigned to Carrier 1, i.e., the *digital twin* of that customer under the counterfactual regime that this customer's carrier is now Carrier 1.

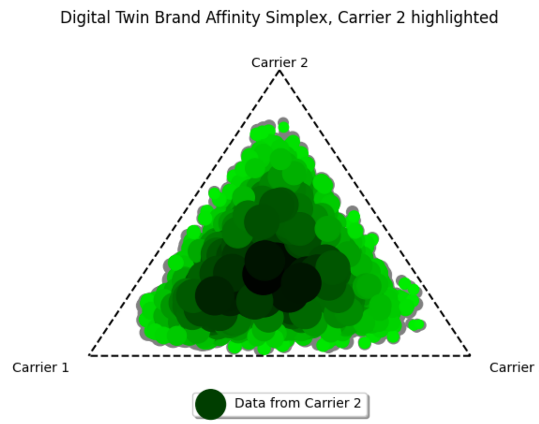
One interesting phenomenon to notice is the higher density of customers toward Carrier 1, for all carriers' customer bases. This digital twin representation suggests a large group of customers who would have a high brand affinity with Carrier 1, if they were ever assigned to be its customers. This latent asymmetry in brand affinity could not be identified without a rigorous counterfactual analytical framework. Carrier 1 likely should target this group of prospective customers to steal them from Carrier 2's and Carrier'3 customer bases.

Figure 5 plots baseline values α_{jkt} for each target variable over time, showing posterior mean and 95% credible intervals. These values represent the base utility for a given question, at a given time, and for a given carrier, after accounting for nonparametric variations in individual predictors or KPI. Baseline values for LTR seem to be lower for Carrier 2, though they seem higher in terms of recommending their carrier's phone. Unsurprisingly, baseline values for Overall Feeling about Carriers 1, 2, and 3 are higher for corresponding customer bases. Finally, baseline values for Retention Likelihood for Carrier 1 and 3 are higher than for Carrier 2; customers of Carrier 2 are more likely to switch to the competition, on average.

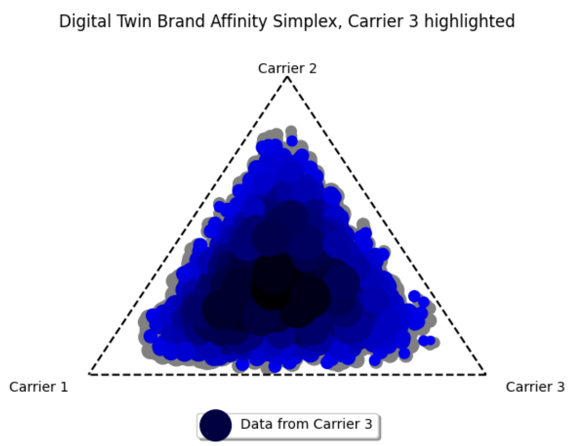
Figure 6 plots the carrier performance loading on each target variable over time. The evidence is more mixed, because reflecting considerable uncertainty, as also indicated by the credible interval. This uncertainty is propagated into the model's predictive performance, explaining why the model with KPI performs only marginally better than the model without them. An interesting aspect to notice is that carrier performance loading on Overall Feeling about Own Carrier and Competition's Carrier are well informed, as is also reflected in the



(a) Carrier 1



(b) Carrier 2



(c) Carrier 3

Figure 4: Plotting counterfactual customer-side brand affinities z^*_{ikt}, using posterior means across all customers.

Notes: A higher density is signified by darker colors. Brand affinities have been transformed using a softmax to fit into a simplex. Each dot represents a customer from a given carrier, and can be projected onto the edges of the triangle to reveal the manifest *digital twins*, i.e., counterfactual brand affinities summarizing target variables.

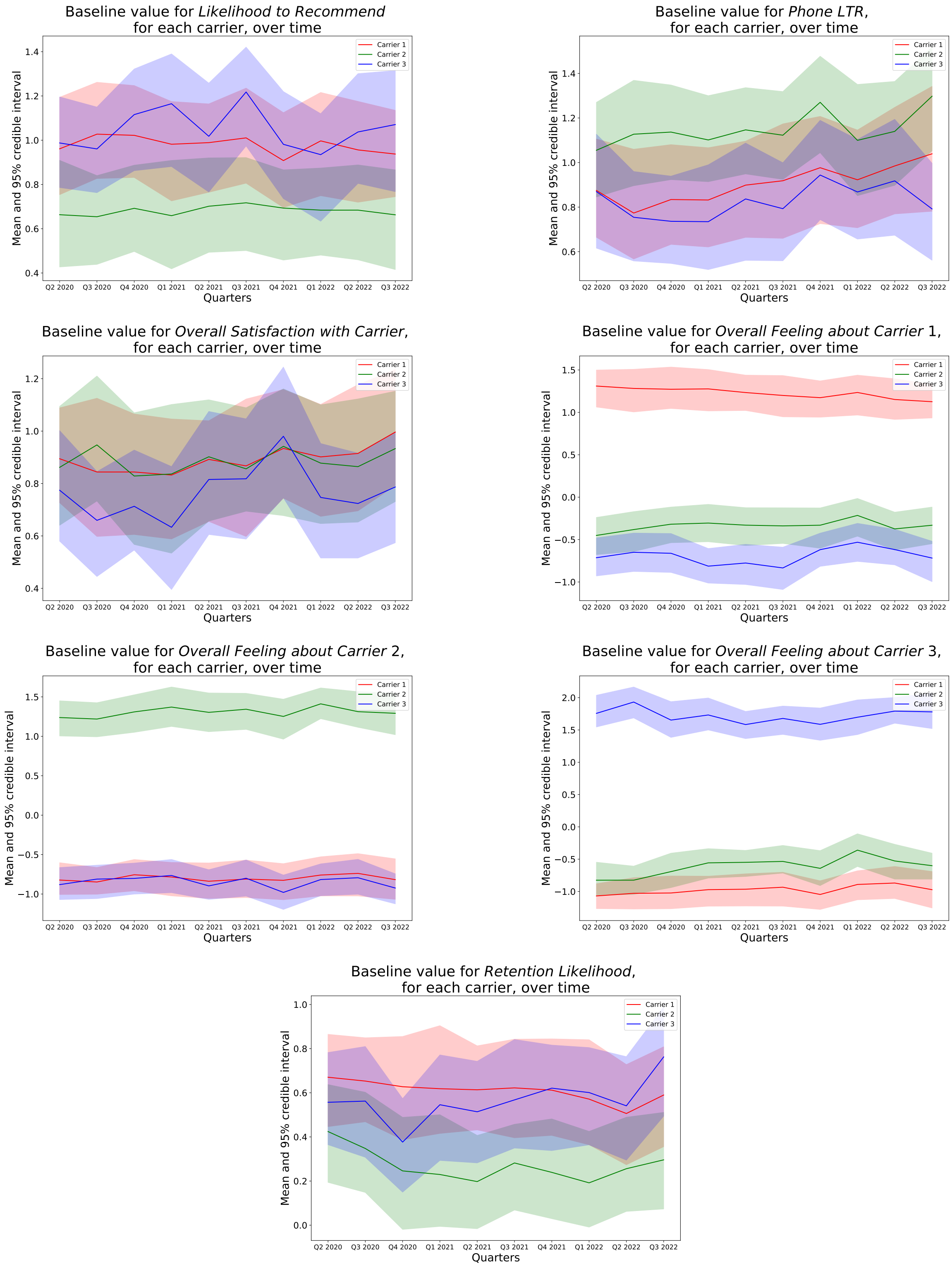


Figure 5: Plot of the baseline values α_{jkt} for each target variable over time.
Notes: These values represent the base utility for a given question, at a given time, and for a given carrier, abstracting away from individual predictors or key performance indicators.

lower test MAE shown above in Model (4) compared with Model (3).

Application: Personalized Marketing Communication Campaigns using Grid Search on Strategic Variables

After the model is trained, digital twins can be used to optimize personalized marketing communications campaigns. Certain features of telecommunications services are complex and not well understood or known by customers. The proposed framework allows marketers to automatically rearrange current communication strategies automatically, to focus on the most critical aspects of customer satisfaction, at the individual level.

With a grid search on strategic variables for all customers in the test sample of the dataset, we decrease each customer's current value by 2 points for each strategic variable with at least five point scales, then gradually increase each current value by .1 increments, until it reaches +2 points. These changes are simultaneously implemented across all customers in 2020 Q2. Figure 7 plots individual-level responses in brand affinity after changing strategic variables from -2 to 2, using the posterior mean for z_{ikt} as a summary statistic. Carrier 2 seems to have a group of customers with lower brand affinity. Carriers 1 and 3 are remarkably close in terms of brand affinity level. The plots also offer strong evidence of customer-level nonlinearities for two strategic variables. A more positive response by customers to a change in wireless service within the next six months would induce a lower brand affinity, up to a certain point. However, a more positive response by customers to a satisfaction question on their providers' network speed would induce a rapid increase in brand affinity for most customers.

CONCLUSION

The novel Digital Marketing Twins methodology promises to address the challenges of analyzing large-scale customer surveys, as demonstrated in the context of the U.S. wireless telecommunications retail market. This study thus addresses two major issues: (1) the

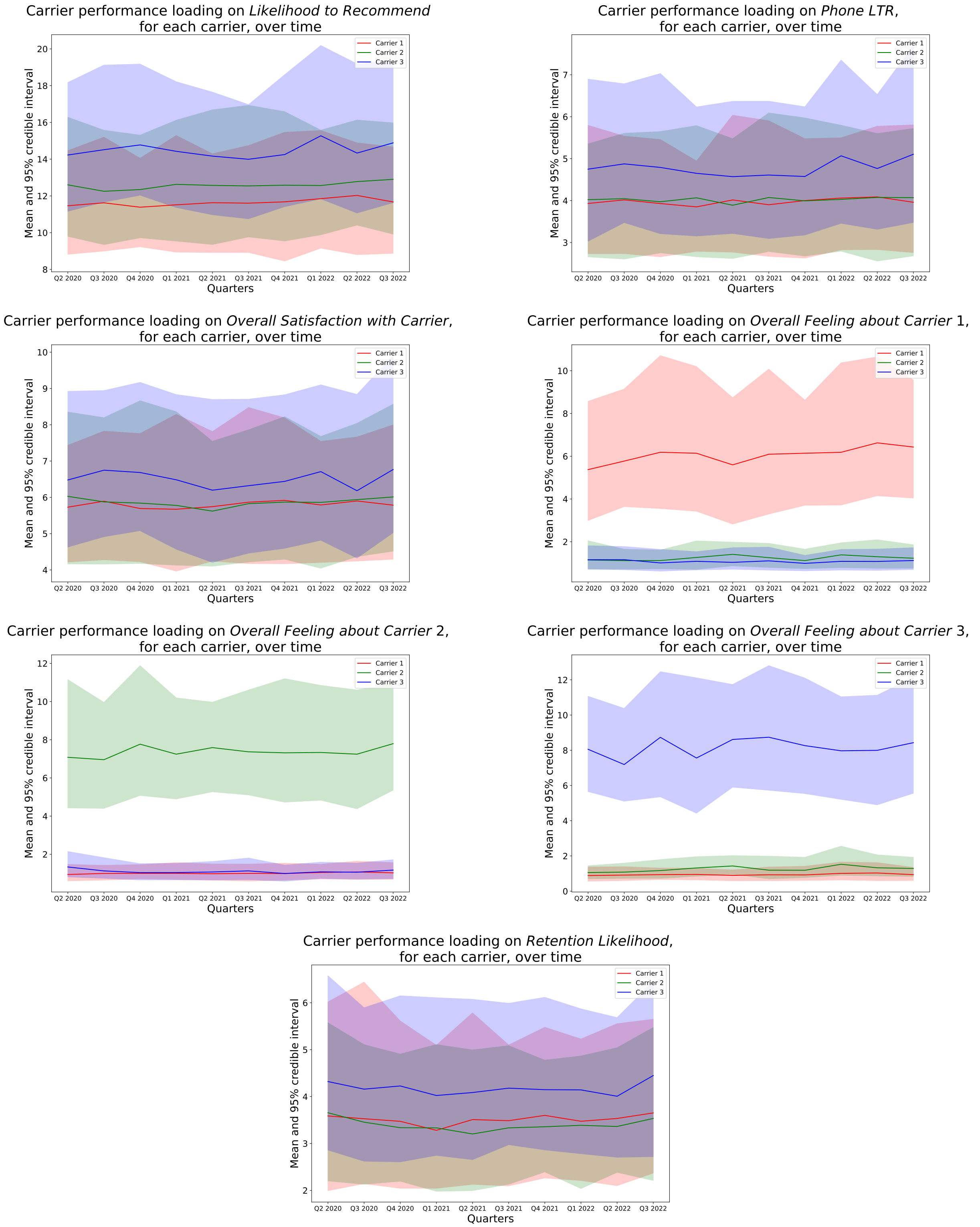
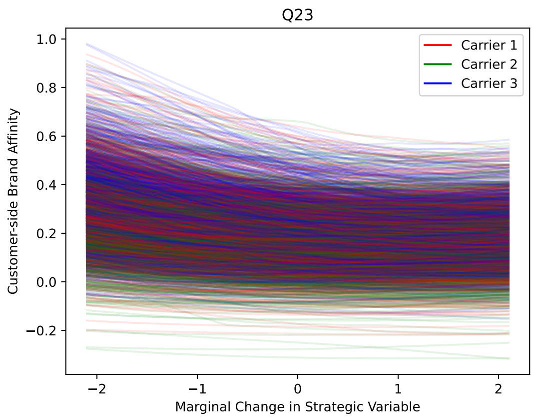
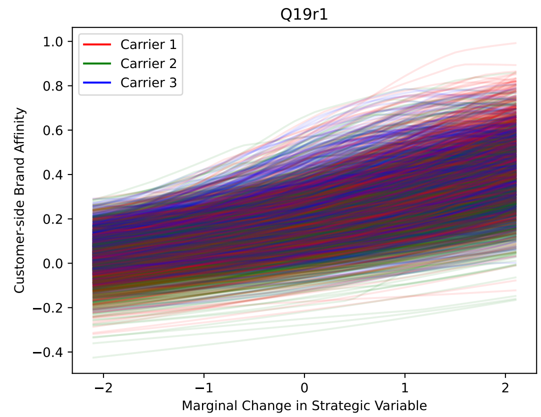


Figure 6: Plot of the carrier k performance loading on each target variable j $(\sum_{l=1}^{L} \beta_{jl}\phi_{lkt})$ over time t.

Notes: These values represent the contribution from KPI to target questions. These values are multiplied to brand affinity.



(a) "How likely are you to change anything about your current wireless service in the next 6 months?"



(b) How satisfied are you with your carrier's performance on the following aspect of your overall wireless service experience: Network Speed

Figure 7: Plotting variations in counterfactual customer-side brand affinities z^*_{ikt} in response to marginal changes in strategic variables, between -2 and 2. Each line represents an individual customer's response from a given carrier, in Q2 2020. These plots show evidence of individual-level nonlinearities in brand affinity responses to marginal changes in strategic variables.

theoretical difficulty of integrating customer surveys into a prescriptive framework and (2) the practical problem posed by repeated cross-sectional surveys, such that it contributes to the literature on digital twins, machine learning methods for competitive environments, and customer satisfaction.

The proposed methodology provides counterfactual responses under different scenarios, which can serve as a powerful tool in the realm of customer analytics. The technique also addresses the missing data problem that is typical of repeated cross-sectional surveys, thereby presenting a comprehensive approach to understanding and leveraging customer survey data at scale.

The implementation of the methodology involves the development of a deep generative and probabilistic latent factor model that captures customer-side brand affinity at the individual level, for each brand and each time period, while controlling for observed heterogeneity and firm-side factors. The methodology provides a grid searche to maximize individual-level, latent, customer-side brand affinity, thereby leading to a "path of least resistance" at the individual level.

The findings have implications for marketers who seek to improve customer satisfaction by understanding the causes of satisfaction from surveys. Furthermore, because the methodology appears generalizable to other sectors and contexts, and therefore, it suggests new avenues for research and applications in the field of marketing and customer analytics.

REFERENCES

Ascarza, Eva (2018), "Retention futility: Targeting high-risk customers might be ineffective," *Journal of Marketing Research*, 55 (1), 80–98.

Ascarza, Eva, Scott A Neslin, Oded Netzer, Zachery Anderson, Peter S Fader, Sunil Gupta, Bruce GS Hardie, Aurélie Lemmens, Barak Libai, David Neal et al. (2018), "In pursuit of enhanced customer retention management: Review, key issues, and future directions," *Customer Needs and Solutions*, 5, 65–81.

Betancourt, Michael "Ordinal Regression Case Study, section 2.2," (2020) `https://betanalpha.github.io/assets/case_studies/ordinal_regression.html`, [Online; accessed 30-June-2023].

Blei, David M, Alp Kucukelbir, and Jon D McAuliffe (2017), "Variational inference: A review for statisticians," *Journal of the American statistical Association*, 112 (518), 859–877.

Braun, Michael and David A Schweidel (2011), "Modeling customer lifetimes with multiple causes of churn," *Marketing Science*, 30 (5), 881–902.

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

Groves, Robert M, Floyd J Fowler Jr, Mick P Couper, James M Lepkowski, Eleanor Singer, and Roger Tourangeau (2011), *Survey methodology* John Wiley & Sons.

Gustafsson, Anders, Michael D Johnson, and Inger Roos (2005), "The effects of customer satisfaction, relationship commitment dimensions, and triggers on customer retention," *Journal of Marketing*, 69 (4), 210–218.

Kannan, PK and Gordon P Wright (1991), "Modeling and testing structured markets: A nested logit approach," *Marketing Science*, 10 (1), 58–82.

Kapteyn, Michael G, Jacob VR Pretorius, and Karen E Willcox (2021), "A probabilistic graphical model foundation for enabling predictive digital twins at scale," *Nature Computational Science*, 1 (5), 337–347.

Kekre, Sunder, Mayuram S Krishnan, and Kannan Srinivasan (1995), "Drivers of customer satisfaction for software products: implications for design and service support," *Management Science*, 41 (9), 1456–1470.

Kingma, Diederik P and Jimmy Ba (2014), "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*.

Kingma, Diederik P and Max Welling (2013), "Auto-encoding variational Bayes," *arXiv preprint arXiv:1312.6114*.

Lee, Thomas Y and Eric T Bradlow (2011), "Automated marketing research using online customer reviews," *Journal of Marketing Research*, 48 (5), 881–894.

Little, Roderick JA and Donald B Rubin (2019), *Statistical analysis with missing data*, Vol. 793. John Wiley & Sons.

Malshe, Ashwin, Anatoli Colicev, and Vikas Mittal (2020), "How main street drives wall street: Customer (dis) satisfaction, short sellers, and abnormal returns," *Journal of Marketing Research*, 57 (6), 1055–1075.

Mason, Karen Oppenheim, William M Mason, Halliman H Winsborough, and W Kenneth Poole (1973), "Some methodological issues in cohort analysis of archival data," *American sociological review*, pages 242–258.

Netzer, Oded, Ronen Feldman, Jacob Goldenberg, and Moshe Fresko (2012), "Mine your own business: Market-structure surveillance through text mining," *Marketing Science*, 31 (3), 521–543.

Phan, Du, Neeraj Pradhan, and Martin Jankowiak (2019), "Composable effects for flexible and accelerated probabilistic programming in NumPyro," *arXiv preprint arXiv:1912.11554*.

Reichheld, Frederick F (2003), "The one number you need to grow," *Harvard Business Review*, 81 (12), 46–55.

Ringel, Daniel M (2023), "Multimarket membership mapping," *Journal of Marketing Research*, 60 (2), 237–262.

Rossi, Peter E and Greg M Allenby (2003), "Bayesian statistics and marketing," *Marketing Science*, 22 (3), 304–328.

Rubin, Donald B (1974), "Estimating causal effects of treatments in randomized and nonrandomized studies.," *Journal of Educational Psychology*, 66 (5), 688.

Rubin, Donald B (1976), "Inference and missing data," *Biometrika*, 63 (3), 581–592.

Tirunillai, Seshadri and Gerard J Tellis (2014), "Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent Dirichlet allocation," *Journal of Marketing Research*, 51 (4), 463–479.

Tsialiamanis, George, David J Wagg, Nikolaos Dervilis, and Keith Worden (2021), "On generative models as the basis for digital twins," *Data-Centric Engineering*, 2, e11.

Tuli, Kapil R, Ajay K Kohli, and Sundar G Bharadwaj (2007), "Rethinking customer solutions: From product bundles to relational processes," *Journal of marketing*, 71 (3), 1–17.

Venkatesan, Rajkumar and Vita Kumar (2004), "A customer lifetime value framework for customer selection and resource allocation strategy," *Journal of Marketing*, 68 (4), 106–125.

Yang, Yang (2006), "Bayesian inference for hierarchical age-period-cohort models of repeated cross-section survey data," *Sociological Methodology*, 36 (1), 39–74.

Yu, Jinsong, Yue Song, Diyin Tang, and Jing Dai (2021), "A Digital Twin approach based on nonparametric Bayesian network for complex system health monitoring," *Journal of Manufacturing Systems*, 58, 293–304.

Zhang, Cheng, Judith Bütepage, Hedvig Kjellström, and Stephan Mandt (2018), "Advances in variational inference," *IEEE transactions on pattern analysis and machine intelligence*, 41 (8), 2008–2026.

APPENDIX

Variable Name	Description
Q14R1	[pipe:hCurrentProvider]'s Network - Which of the following directly contributed to you giving [pipe:hCurrentProvider] a rating of ?
Q14R2	[pipe:hCurrentProvider]'s Price / Value - Which of the following directly contributed to you giving [pipe:hCurrentProvider] a rating of ?
Q14R8	[pipe:hCurrentProvider]'s Billing process - Which of the following directly contributed to you giving [pipe:hCurrentProvider] a rating of ?
Q14R3	[pipe:hCurrentProvider]'s Customer Service - Which of the following directly contributed to you giving [pipe:hCurrentProvider] a rating of ?
Q14R4	General feelings about [pipe:hCurrentProvider] - Which of the following directly contributed to you giving [pipe:hCurrentProvider] a rating of ?
Q14R5	[pipe:hCurrentProvider]'s Plans - Which of the following directly contributed to you giving [pipe:hCurrentProvider] a rating of ?
Q14R9	[pipe:hCurrentProvider]'s Rewards and benefits - Which of the following directly contributed to you giving [pipe:hCurrentProvider] a rating of ?
Q14R6	[pipe:hCurrentProvider]'s Devices - Which of the following directly contributed to you giving [pipe:hCurrentProvider] a rating of ?
Q14R7	Other (please specify) - Which of the following directly contributed to you giving [pipe:hCurrentProvider] a rating of ?
Q19R1	Network speed - How satisfied are you with [pipe:hCurrentProvider]'s performance on the following aspects of your overall wireless service experience?
Q19R2	Network reliability - How satisfied are you with [pipe:hCurrentProvider]'s performance on the following aspects of your overall wireless service experience?
Q19R3	Data plans that meet my needs - How satisfied are you with [pipe:hCurrentProvider]'s performance on the following aspects of your overall wireless service experience?
Q19R4	Value for the price paid - How satisfied are you with [pipe:hCurrentProvider]'s performance on the following aspects of your overall wireless service experience?
Q19R5	Accuracy of billing - How satisfied are you with [pipe:hCurrentProvider]'s performance on the following aspects of your overall wireless service experience?
Q19R6	Rewards and recognition - How satisfied are you with [pipe:hCurrentProvider]'s performance on the following aspects of your overall wireless service experience?
Q19R7	Easy to do business with - How satisfied are you with [pipe:hCurrentProvider]'s performance on the following aspects of your overall wireless service experience?
Q19R8	Solves problems the first time you contact them - How satisfied are you with [pipe:hCurrentProvider]'s performance on the following aspects of your overall wireless service experience?
Q19R9	Is a brand for me - How satisfied are you with [pipe:hCurrentProvider]'s performance on the following aspects of your overall wireless service experience?
Q19R10	Total cost of wireless service - How satisfied are you with [pipe:hCurrentProvider]'s performance on the following aspects of your overall wireless service experience?
Q19R11	Device Selection - How satisfied are you with [pipe:hCurrentProvider]'s performance on the following aspects of your overall wireless service experience?
Q19R1aux (resp. Q19R2-Q19R11)	Aware of [pipe:hCurrentProvider]'s performance on the aspects mentioned in Q19R1? (resp. Q19R2-Q19R11)
Q23	How likely are you to change anything (plan, provider, device) about your current wireless service in the next 6 months?
Q35	On a scale of 1 to 5, how well do you feel you understand the details of your wireless plan with [pipe:hCurrentProvider]?

Table 4: List of Strategic Variables

Variable Name	Description
Q12	Thinking about your overall experience with your wireless service provider, on a scale of 0 to 10, how likely are you to recommend [pipe:hCurrentProvider] to a friend or family member?
Q18	Q18: Overall, how satisfied are you with [pipe:hCurrentProvider]?
Q20	How likely are you to switch wireless service providers within the next 12 months?
Q20AR1	Carrier 1 - What best describes your overall feeling about each wireless service provider?
Q20AR2	Carrier 2 - What best describes your overall feeling about each wireless service provider?
Q20AR3	Carrier 3 - What best describes your overall feeling about each wireless service provider?
Q27	How likely are you to recommend your [pipe:Q26] phone to a friend or a colleague?

Table 5: List of Target Variables

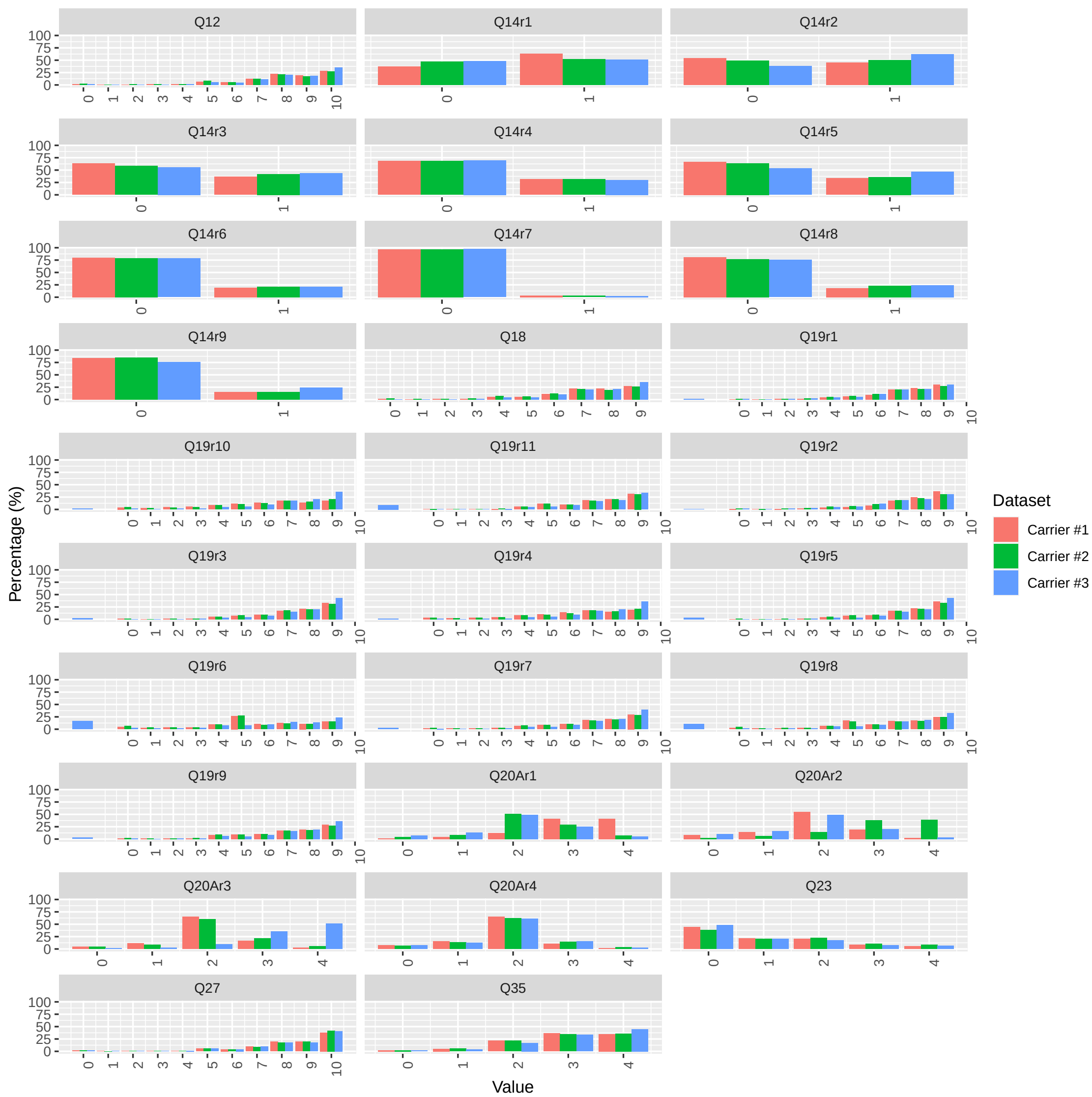


Figure 8: Summary Statistics for Target Variables and Strategic Variables, Per Carrier