Detecting Routines in Ride-sharing: Implications for Customer Management

Ryan Dew,[1*] Eva Ascarza,[2] Oded Netzer,[3] Nachum Sicherman[3]

[1] The Wharton School, University of Pennsylvania
[2] Harvard Business School, Harvard University
[3] Columbia Business School, Columbia University

December 10, 2021

Abstract

Routines often shape many aspects of day-to-day consumption, including transportation choice, use of mobile apps, or visits to a gym. While prior work has established the importance of habits in consumer behavior, little work has been done to understand the implications of routines, which we define as repeated behaviors with recurring temporal structures, for customer management. One possible reason for this lack of research is the difficulty of statistically modeling routines with customer-level transaction data, particularly when routines may vary substantially across customers. In this paper, we propose a new approach to measuring routine consumption, which we apply in the context of ride-sharing. We model customer-level routines with a hierarchical, Bayesian nonparametric Gaussian process, leveraging a novel kernel structure that allows for flexible yet precise estimation of routine behavior. We then nest this Gaussian process in an individual-level inhomogeneous Poisson point process, which allows us to estimate individual-level routines from transaction data, and decompose a customer's overall usage into routine and non-routine components. We show that more routine users tend to be more valuable customers, with higher individual-level "routineness" being associated with higher future usage, lower churn rates, and more resilience to service failures.

Keywords: routines, customer management, customer relationship management, Bayesian nonparametrics, Gaussian processes, machine learning, ride-sharing

*Ryan Dew (ryandew@wharton.upenn.edu) is the corresponding author. He is an Assistant Professor of Marketing at the Wharton School, University of Pennsylvania. Eva Ascarza is the Jakurski Family Associate Professor of Business Administration at Harvard Business School. Oded Netzer is the Arthur J. Samberg Professor of Business at Columbia Business School. Nachum Sicherman is the Carson Family Professor of Business, Columbia Business School. The authors thank Via for providing the data, and Saar Golde and Coleman Humphrey in particular for their help and input. The authors thank Zhangyi (David) Fan, Xuanming (Donny) Gu, and Nikhil Kona for excellent research assistance. Finally, the authors thank seminar participants at UT Austin, CU Boulder, CityU, Temple, and the CBS "Quant Lab", as well as participants at the AI/ML 2021, TPM 2021, and Marketing Science 2020 conferences for valuable feedback.

1. Introduction

Routines are an integral feature of consumers' daily lives: for many people, from the time they wake up in the morning, to the moment they go to sleep at night, their time is structured around routines. Such routines often involve consumption, like picking up coffee from their favorite coffee chain each morning, checking their weather app before leaving home in the morning and before leaving work in evening, or choosing the mode of transportation to take to and from work. Moreover, consumers are often different in their routines: while one may drink their coffee only in the mornings, seven days per week, another may prefer to have their coffee after lunch, and only on weekdays. Marketers can greatly benefit from understanding consumer routines. Yet, while routines are intuitively important drivers of consumer behavior, prior research has not explored the presence of such routines in consumers' behavior and their implications for customer management. Accordingly, the objectives of this research are first, to build a statistical model that can capture customer routines at the individual customer-level, and second, to explore the relationship between such routines and customer-level outcomes like purchase frequency and churn.

We define routine behavior as a behavior with a defined, recurring, temporal structure, such that the same behavior occurs at roughly the same time, period after period. We focus specifically on the period of a week, as weekly routines capture many common routines, including, for instance, weekday commutes, weekday lunches, weekend brunches, and weekly grocery shopping.[1] Routines are related to habits, which have been studied more extensively in marketing (e.g. Drolet and Wood, 2017). It is the emphasis on temporal structure that differentiates a routine behavior from habitual (or repeat) behavior. For example, a consumer who always shops at the same store may do so out of habit. A customer who always shops at that same store every Thursday evening exhibits a routine.

Little research has been done on capturing routines and understanding their impact on consumer behavior and firm profitability. There are many reasons why routines may matter,

[1] While weekly routines capture much of the richness of recurring consumption, there are also routines that exist over longer periods, like getting a haircut the first Friday of a month, or biweekly Sunday dinners at your parents' house, which will not be captured by focusing on weekly routines. Our approach could be easily extended to cyclicalities other than a week

and why firms may wish to understand their existence, antecedents, and consequences. For instance, from a demand forecasting perspective, knowing that a consumer is in a routine can aid firms in making more precise forecasts of demand, not only at the aggregate level (e.g., weekly), but at specific points in time. This is particularly important for firms that need to plan around peak demand periods, like the ride-sharing service we study in our empirical application. Furthermore, if the strength of a routine moderates customer responses to firms' interventions, knowing that a customer is in a routine may be particularly helpful for marketers, who can better tailor or anticipate the impact of their marketing actions. In this work, we focus on the implication of routines for customer value. In particular, we hypothesize that customers who use a product as part of their routine may be higher value customers, insofar as they may (1) consume the product more often, and (2) be less likely to churn than non-routine customers. They may also be better customers in other ways, including having lower price sensitivities, and being more resilient to disruptions in service. We hypothesize that the effect of routines exists over and above the effect of mere habitual, but non-routine usage, as is already captured in many existing customer relationship management (CRM) models such as the recency-frequency-monetary value (RFM) model (e.g., Blattberg et al., 2008; Neslin et al., 2013).

To measure customer "routineness," we develop a statistical model that allows us to identify the routine pattern of each individual customer, and isolate the share of the customers' consumption that can be attributed to a routine behavior. Through this model, we are able to differentiate between, for instance, one customer who typically rides during weekday commute hours, for a total of 10 routine trips per week, and another customer who rides twice a week to her Yoga class on Tuesday and Thursday afternoons. Specifically, our model is an individual-level, inhomogeneous Poisson process that captures individual-specific patterns in consumption across periods, with a unique Bayesian nonparametric specification of its rate. The individual-specific rate of consumption is decomposed into both a random component that varies across periods, capturing changing levels of idiosyncratic consumption, and a routine component that varies within periods, which is modeled using a Gaussian process prior with a unique kernel structure. This kernel structure incorporates intuitive aspects of consumption over time — specifically, that certain days exhibit similarities in consumption (e.g., a Tuesday might be more similar to a Thursday than to a Sunday) and that consumption within a day exhibits 24-hour cycle (e.g., 12:05am is similar to 11:55pm) — to precisely estimate individual-specific variation in recurring behavior. Using the routine component of the Poisson rate parameter, we construct an individual-specific "routineness" metric that measures to what degree an individual's behavior is structured around a routine. In addition to the "routineness" metric, the model infers the form or temporal "shape" of the routine for each consumer (e.g., whether a consumer has a Monday through Thursday AM routine, or a Tuesdays evening routine).

We apply our model and routineness metric to data from Via, a leading New York City-based ride-sharing company, to estimate consumer routines in requesting rides. Ride-sharing is a particularly rich setting for studying routines, as travel is often an integral part of many day and week-level routines. We identify various patterns of routines in using the ride-sharing service across users, including both predictable commuting routines, as well as more complex, idiosyncratic routines. More importantly, we show that, as hypothesized, users who are more routine in their behavior, are also more valuable to the firm, in terms of both higher future usage rates and lower propensity to churn, even after controlling for past usage patterns. The effect is robust to numerous specifications. Having established the value of routineness in customer value, we then show that routines also play a role in driving and moderating other aspects of the customer-firm relationship, including price sensitivity and customer response to service failures. Finally, we show that our proposed routineness measure is a stronger predictor of future customer value than other (related) metrics. Specifically, we investigate how our routineness metric compares with (1) routines in terms of trip locations (i.e., *what* someone buys, not *when* they buy), and (2) an extant metric of regularity of inter-transaction times (Platzer and Reutterer, 2016). We find that our temporal routineness metric is a stronger predictor of customer behavior than either, suggesting its relevance over and above these other constructs.

The rest of the paper is organized as follows: we start by discussing the prior literature on habits, routines, and the possible relationship between routines and CRM. We then present our model for capturing and measuring customer customer-specific routines. Moving next to our empirical application, we describe the ride-sharing data we use to test our model. We then we describe the results of applying our model: we first apply the model on synthetic data that mimics the real data, validating the model's ability to recover different types of routines, and then apply the model on the the ride-sharing data, characterizing the types of routines exhibited by riders, and validating the model's fit. Finally, we explore the idea of routineness more deeply, by highlighting the relevance of routines for customer management, and by comparing routineness to other constructs. We conclude with discussion and directions for future research.

2. Routines, Habits and Customer Management

While research on routines is relatively scant, the closely related topics of habits and repeat behaviors have been studied extensively, both in marketing and in related disciplines. Early work in marketing used the term repeat buying habit to simply indicate repeatedly buying the same product or repeatedly buying from the same company, without considering the more psychological construct of a habit or habit formation (Ehrenberg and Goodhardt, 1968). Capturing the regularities of repeat purchasing has subsequently been the focus of many models in customer base analysis, including popular buy-till-you-die models (e.g., Schmittlein et al., 1987) and more general RFM-based specifications (e.g., Dew and Ansari, 2018). Repeat buying is also central to other important marketing constructs, including brand and store loyalty and brand inertia (e.g., Guadagni and Little, 1983), all of which can also be viewed as forms of habitual behavior. Moving beyond studying simply repeat purchasing, Shah et al. (2014) generalized the idea of habits to extend to recurring behaviors like returning products, purchasing on promotion, and purchasing low-margin items. They showed that these repeat behaviors are linked to firm profitability, and that, moreover, firm actions can influence the formation of habitual behaviors.

Habit formation has also been studied in economics, often in the context of consumption and expenditure more generally, where it is typically defined as current expenditures depending on lagged expenditures through a "habit stock." In this literature, habits have been used to explain the smoothness of consumption over time, even in the presence of shocks to income, although evidence for the existence of habit formation in aggregate consumption is mixed (Dynan, 2000; Fuhrer, 2000).

Much of the theory behind why habits matter, how they develop, and how they can be changed has come from the psychology and consumer behavior literatures. Habits have been studied in psychology since as early as the 19th century (James, 1890). In this literature, habits are often defined as tendencies to repeat behaviors, typically automatically or subconsciously (Ouellette and Wood, 1998; Wood et al., 2002), and often in goal-directed manner (Aarts and Dijksterhuis, 2000), or triggered by contextual cues (Neal et al., 2012). Especially relevant for our empirical application of ride-sharing, habits have recently been identified as a primary driver of travel mode choice (e.g., Verplanken et al., 2008), which is of particular interest for developing more sustainable consumer choices (White et al., 2019). A noteworthy finding in this literature is the habit discontinuity hypothesis, which states that context changes that disrupt individuals' habits can lead to deliberate choice consideration, and thus to consumers' breaking their habits (Verplanken et al., 2008). This phenomenon has also been observed in the customer management literature, particularly in Ascarza et al. (2016), who show that customers who continue to transact with a focal firm out of habit may be driven to churn by company retention efforts, even when those retention efforts are intended to save the customer money, simply by means of disrupting their inertia.

In this research, we move beyond studying repeated behavior or habits, to studying specifically routines. We show that beyond just repeatedly purchasing, the *temporal structure* of when customers interact with the firm also matters for customer management. In a sense, routines are a specific type of habit, where a habitual behavior is built into a temporal structure. Thus, many of the predictions made elsewhere in the literature about habitual behavior and customer loyalty (e.g., Ascarza et al., 2018) carry over to routines: we postulate, for instance, that routines can lead to nearly automatic choices, and will thus be more difficult to break, resulting in stickier long-run behavior, and lower likelihood to react negatively to price increases or service failures. However, we argue that routines are a special kind of habit, that deserve special attention, because they imply not only repeated behavior, but also behavior that is also temporally consistent. Thus, a customer who is *routinely* consuming a focal product or service may be even more valuable than one who is merely habitually (i.e., repeatedly) consuming the product, but not in a temporally routine manner.

Our work is also related to the growing literature on extending traditional CRM frameworks like RFM to incorporate individual-specific data about usage and purchase timing. Notable contributions in this stream include the inclusion of clumpiness of transactions in RFM models by Zhang et al. (2015), and the modeling of regularity of transactions by Platzer and Reutterer (2016). The concept of regularity is particularly closely related to the concept of routineness. In Platzer and Reutterer (2016), the regularity of transactions is modeled by relaxing the standard Poisson process transaction model common to many customer base models, allowing for customer-specific gamma-distributed intertransaction times. They find that regularity is associated with higher-value customers, and incorporating it can improve customer-level predictions. Although conceptually related, routineness and regularity are distinct: regularity focuses exclusively on the distribution of interpurchase times, while routineness is based on the specific temporal structure of transactions. This means that some routines may be regular, while others are not. We conceptually and empirically contrast our approach with regularity. We find that, while the metrics are correlated, our temporal routineness metric is a stronger predictor of customer behavior, both in terms of number of requests and inactivity, thus highlighting the importance of considering routineness over and above regularity and habits. As its own unique construct, we suggest that routineness is a valuable metric for marketers looking to build interpretable yet accurate CRM models, thus addressing an on-going need for new advances in this space (Neslin et al., 2006).

Methodologically, our model merges an inhomogeneous Poisson process with a Bayesian nonparametric Gaussian process. While the basis of many customer base analysis models is a homogeneous Poisson process (Schmittlein et al., 1987), inhomogeneous Poisson process transaction models have been employed to capture more complex dynamics in usage or transaction behavior (e.g., Ho et al., 2006; Ascarza and Hardie, 2013). In our model, the rate parameter of the Poisson process is modeled partly using a Gaussian process (Rasmussen and Williams, 2006), a specification closely related to the log-Gaussian Cox process (Møller et al., 1998). In marketing, Gaussian processes have seen increased use in recent years, in both aggregate-level and individual-level CRM and brand choice contexts (Dew and Ansari, 2018; Dew et al., 2020; Tian and Feinberg, 2021). In the case of routines, Gaussian processes offer an ideal solution to flexibly modeling customer-level rates of usage, as they are flexible, but also allow us to encode prior knowledge about the structure of time which is difficult to encode in other flexible functional estimation methods. We elaborate more on this point as we describe our model below. In the broader literature, our model aims to capture time-varying purchasing or usage patterns, and is thus related to a long line of dynamic models in marketing (e.g., Kim et al., 2005; Du and Kamakura, 2012). Finally, our novel approach of identifying and isolating routines using transaction data and relating them to the customer value is consistent with Du et al. (2021)'s call to move toward a richer characterization of behavior, and relating such behaviors to firm growth through customer value.

3. Model

In this section, we specify a model of usage that yields a natural metric for how routine a customer's behavior is, and what pattern of weekly routine behavior the customer exhibits. By "usage," we mean the consumer interacting with the firm in some way, and by "weekly routine," we mean the structure of usage within a given week, which is the main focus of this research. In later sections, where we apply this model to ride-sharing data, the dependent variable will be requesting rides. However, our model is fully general, and can be applied using timing data from any context, and for myriad customer behaviors of interest (e.g., using a mobile app, making purchases with the firm, visiting the firm's website).

At a high level, the goal of the model is to capture the degree of temporal similarity in usage over time, at the customer level. To that end, we divide the relevant unit of time (a week) into sub-units (hours of days), such that the model learns, for each individual, what are the within-week patterns that are consistent across weeks. More formally, we define time as $t = (w, d, h)$, with w indexing weeks (i.e., $w = 5$ is the 5th week since the start of the data), d indexing days of the week starting with $d = 1 =$ Sunday, and $h = 0, \ldots, 23$ indexing 24 hours. Our dependent variable, y_{it}, is defined as the number of times user i interacts with the firm during time period t. To simplify notation, we will use the unit of "day-hours," which we denote as $j = (d-1) \times 24 + h$, such that $j = 1, \ldots, 168$, captures all the hours in a week. The dependent variable of interest will then be $y_{it} = y_{iwj}$, or the number of interactions customer i has with the firm in week w at day-hour j.[2,3]

3.1. Usage Model

To model usage, we consider an individual-level, discretized inhomogeneous Poisson process, such that:

$$y_{it} \sim \text{Poisson}(\lambda_{it}). \tag{1}$$

To capture the part of an individual's usage that can be attributed to random needs versus a routine, we decompose λ_{it} as:

$$\lambda_{it} = \exp(\alpha_{iw} + \mu_j) + \exp(\gamma_{iw} + \eta_{ij}). \tag{2}$$

In each term of this decomposition, there are two parameters: one that varies over weeks (w), and one that varies over day-hours j. Our substantive focus is primarily on variation over day-hours. Thus, noting that λ_{it} could be rewritten as $\lambda_{it} = \exp(\alpha_{iw})\exp(\mu_j) + \exp(\gamma_{iw})\exp(\eta_{ij})$, we will refer to the terms with j subscripts (i.e., μ_j and η_{ij}) as *rate* terms, reflecting the rate of usage across day-hours. We will refer to the terms with w subscripts (i.e., α_{iw} and γ_{iw}) as *scaling* terms, as these terms scale up or down each of the rate terms over customers and weeks.

In the first term, $\exp(\alpha_{iw} + \mu_j)$, the day-hour variation term, μ_j, does not have an i subscript: it thus captures general patterns of when the service is commonly used across customers. For example, in our later ride-sharing application, it will capture that customers tend to take rides during the day, but not in the middle of the night. Said differently, if any given user were to randomly have need of the service, μ_j captures when we might expect that random need to arise, and how the distribution of random needs may deviate from a uniform distribution over day-hours. We thus refer μ_j as the rate of *random* usage. Then, intuitively, the term α_{iw} captures the scale of random usage for customer i in week w. It controls how many rides we expect to see in a given week arising from random needs. In contrast, in the second term,

[2]For the majority of time periods, $y_{it} = 0$ or 1.

[3]Note that that this approach is also generalizable to any other specification of temporal units, such as monthly or yearly routines.

$\exp(\gamma_{iw} + \eta_{ij})$, the day-hour variation term, η_{ij}, *does* have an i subscript. Mathematically, η_{ij} captures an individual's *routine* usage — that is, the part of usage that is specific to that user, specified over day-hours, consistent over weeks — and γ_{iw} captures the scale of routine usage for customer i in week w. We will later define additional restrictions on γ_{iw} to ensure that routines are consistent over weeks, which allow us identify these quantities from the data.

By splitting λ_{it} into two terms, $\exp(\alpha_{iw} + \mu_j)$ and $\exp(\gamma_{iw} + \eta_{ij})$, we allow our model to be equivalently expressed as the sum of two count processes (Kingman, 1992), $y_{it} = y_{it}^{\text{Random}} + y_{it}^{\text{Routine}}$, such that:

$$y_{it}^{\text{Random}} \sim \text{Poisson}(\exp(\alpha_{iw} + \mu_j)), \tag{3}$$

$$y_{it}^{\text{Routine}} \sim \text{Poisson}(\exp(\gamma_{iw} + \eta_{ij})). \tag{4}$$

This decomposition allows for a natural definition of the levels of random usage and routine usage, through the expectation of Poisson random variables. Specifically, we define two metrics, E_{iw}^{Random} and E_{iw}^{Routine}, which are the expected number of random and routine interactions (respectively), within a single week w, for customer i, such that:

$$E_{iw}^{\text{Random}} = \sum_j \exp(\alpha_{iw} + \mu_j), \tag{5}$$

$$E_{iw}^{\text{Routine}} = \sum_j \exp(\gamma_{iw} + \eta_{ij}). \tag{6}$$

In plain English, these two terms capture how often a user is expected to interact with the firm in a given week, decomposing the total number of interactions into the expected number of interactions happening at random, and the number of interactions stemming from the user's routine. These metrics allow us to identify how routine customers' behaviors are, and are at the heart of paper's focus and intended contribution. In particular, we call E_{iw}^{Routine} the *routineness* of customer i in week w, and will use this metric and terminology throughout our analysis.

3.2. Specifying the Components of the Usage Rates

Our usage model specifies individual-level, time-varying usage through two count processes, the rates of which each have two parts: scaling terms, α_{iw} and γ_{iw}, and day-hour rates, μ_j, and η_{ij}. The term α_{iw} governs the time-varying random usage level. We therefore consider a straightforward Markovian state space model, such that:

$$\alpha_{iw} \sim N(\alpha_{iw-1}, \tau).$$

This specification captures the fact that usage in week w is likely related to usage the previous week, but imposes no further assumptions. To model the parameters μ_j, γ_{iw}, and η_{ij}, we first recast the problem as estimating latent *functions*, $\mu(j)$, $\gamma_i(w)$, and $\eta_i(j)$. This switch from subscript notation to functional notation is merely a conceptual pivot: by recasting the problem of estimating rates as a problem of estimating unknown functions, we can capture uncertainty over those rates using Gaussian processes. As we will show below, this in turn allows us to encode prior knowledge and assumptions about these parts of the model, beyond those that could be incorporated in a simple state space specification.[4]

Gaussian processes (GPs) provide a way of specifying prior distributions over spaces of functions. With this prior, we can encode structural information about the functions in that space, like the smoothness and differentiability of the functions, or other a-priori knowledge about the shape of the functions, that is difficult to capture with other techniques. In this way, GPs allow for flexible estimation of functions, like the individual-level rates in our model, while optimally leveraging both information sharing and a-priori knowledge, in order to improve the efficiency of those estimates. GPs have been popular for some time in geostatistics and machine learning as a means of placing structure over unknown functions (Rasmussen and Williams, 2006), and have also seen recent applications in marketing (Dew and Ansari, 2018; Dew et al., 2020; Tian and Feinberg, 2021). A Gaussian process is a distribution over functions, $f(x) : \mathbb{R}^d \rightarrow \mathbb{R}$, defined by two other functions: a mean function, $m(x)$, which captures the

[4]We could also model α using a GP: we've found the two specifications yield similar results, but the specification with the simple autoregressive formulation is faster to estimate. Moreover, since we do not want to impose any smoothness assumptions on α, the Markovian model is natural. We also note that there are many links between GPs and state space models, a discussion of which is beyond the scope of this paper (see, e.g., Loper et al., 2020).

a-priori expected function value at inputs x, and a kernel function $k(x, x')$, which captures a-priori how similar we expect the function values $f(x)$ and $f(x')$ to be, for two inputs x and x'. Modeling $f(x)$ using a GP is denoted $f(x) \sim GP(m(x), k(x, x'))$. For a finite, fixed set of inputs, $x = (x_1, \ldots, x_N)$, $f(x) \sim GP(m(x), k(x, x'))$ is equivalent to:

$$f(x_1, \ldots, x_N) \sim N(m(x_1, \ldots, x_N), K), \text{ such that } K_{ij} = k(x_i, x_j). \tag{7}$$

Mathematically, the matrix K is the the kernel $k(x, x')$, evaluated pairwise over all inputs, and is called the kernel matrix. Intuitively, a GP specifies a multivariate Gaussian prior over the outputs corresponding to any combination of inputs, by means of its mean function and kernel. Thus, these two objects are the primary source of model specification in GP-based models. In practice, it is common to set the mean function $m(x)$ to be zero or a constant, and let the dependencies between the outputs be solely captured by the kernel (Rasmussen and Williams, 2006). The mean function and kernel themselves are typically parameterized through an additional set of parameters referred to as hyperparameters. From this relationship, it can be seen that the primary restriction in specifying the kernel is that the corresponding kernel matrix be positive definite.

Returning to our specification, we model:

$$\gamma_i(w) \sim GP(\gamma_0, k_{\text{SE}}(w, w'; \phi_\gamma)), \tag{8}$$

$$\mu(j) \sim GP(0, k_{\text{DH}}(j, j'; \phi_\mu)), \tag{9}$$

$$\eta_i(j) \sim GP(0, k_{\text{DH}}(j, j'; \phi_\eta)), \tag{10}$$

where ϕ_γ, ϕ_μ, and ϕ_η are kernel hyperparameters and $k_{\text{SE}}(w, w'; \phi)$ is the standard squared exponential kernel (e.g. Dew and Ansari, 2018), given by:

$$k_{\text{SE}}(w, w'; \phi = \{\sigma, \rho\}) = \sigma^2 \exp\left[-\frac{(w - w')^2}{2\rho^2}\right]. \tag{11}$$

The other kernel, k_{DH}, is a novel kernel which we term the *day-hour kernel*, and which we describe subsequently.

The term $\gamma_i(w)$ captures the scaling of routine usage. As such, and by the definition of routines capturing behavior that is repeated over time, we do not want to allow this usage term to widely fluctuate between weeks. Accordingly, we use the SE kernel, which captures functions that are assumed to be relatively smooth, with a smoothness parameter ρ (also called the lengthscale) and an amplitude parameter σ. By imposing additional assumptions on ρ, as we detail in the next section, we can ensure that $\gamma_i(w)$ evolves slowly.

For the other components of the model, we use our novel day-hour kernel, $k_{\mathrm{DH}}(j, j'; \phi)$, which has a functional form designed to capture the a priori structure we know exists within a week, specifically that hours follow a 24-hour cycle, and that certain days are more similar to other days (e.g., weekends versus weekdays, or work days versus days off). To capture these properties, we fuse a periodic kernel (Rasmussen and Williams, 2006, chapter 4) with an unstructured estimate of the correlation between different days of the week. Specifically, we define:

$$k_{\mathrm{DH}}(j, j'; \phi = \{\sigma, \rho, \Omega\}) = \sigma^2 \Omega_{d,d'} \times \exp\left\{ \frac{1}{2\rho^2} \sin^2\left(\frac{\pi |h - h'|}{24} \right) \right\}, \tag{12}$$

where the matrix Ω is a correlation matrix over days of the week. The right-hand side of this product is the periodic variant of the squared exponential kernel, defined with a 24-hour cycle.[5] The kernel matrix implied by our DH kernel is given by,

$$K_{\mathrm{DH}} = \sigma^2 \Omega \otimes K_{\mathrm{Per}}. \tag{13}$$

Ω is constrained to be a correlation matrix, and is thus positive definite. K_{Per} is guaranteed to be positive definite, since $k_{\mathrm{Per}}(h, h'; \rho)$ is a valid kernel. Thus, since the Kronecker product of two positive definite matrices is also positive definite, we see that $k_{\mathrm{DH}}(j, j')$ is a valid kernel. To estimate the correlation matrix Ω, we use a Lewandowski-Kurowicka-Joe (i.e., LKJ) prior for correlation matrices (Lewandowski et al., 2009), such that $\Omega \sim \mathrm{LKJ}(2)$, which puts a weak prior toward the identity matrix (Barnard et al., 2000).

[5] This kernel can be easily adjusted to other types of routines. For example, if the model was aimed at capturing yearly routines, one could set up a single kernel with 52 sub-units (i.e., weeks), or alternatively, specify two kernels, one capturing the correlations across months and another one capturing days (or weeks) within a month, which could be multiplied together.

Intuitively, this day-hour kernel allows us to place a prior over functions that exhibit two natural properties when dealing with weekly usage data: we allow for arbitrary relatedness of days through the unstructured correlation matrix Ω, and for a natural 24-hour cycle through $k_{\text{Per}}(h, h')$, which accounts for the fact that usage at $h = 0$ (12 AM) will be similar to usage at $h = 23$ (11 PM). Finally, through its multiplicative structure, it assumes these two forces operate together: if day d is similar to day d', as captured by Ω, and hour h is similar to hour h', a GP modeled with this kernel is expected to have similar function values at (d, h) and (d', h'). By encoding this natural prior information into our model structure, we facilitate the efficient inference of the mean and individual-level rate functions, $\mu(j)$ and $\eta_i(j)$.

3.3. Identifying Assumptions

To identify routines in usage, we make two additional assumptions. First, and without loss of generality, since the terms γ_{iw} and η_{ij} both vary at the individual-level, and only their sum determines the level of routine usage, their scales are not separately identified. We thus fix $\eta_{i0} = 0$. Second, we assume that the baseline term for the routine part of usage, $\gamma_i(w)$, evolves slowly. In our functional parlance, we assume that $\gamma_i(w)$ is a smooth function of w. This assumption is a key part of how our model identifies routines from the data, and what separates the random and routine components of usage: routine usage is defined with a rate term that is specific to each individual, thus enabling the model to capture a wide range of routines present in the data, and a baseline that is assumed to evolve slowly, capturing the consistent or "sticky" nature of routines. In contrast, random usage is defined with a relatively unrestricted baseline term, that may change period after period, but where the day-hour variation is restricted to follow the general pattern of usage of the population. Thus, random needs may arise with differing frequencies week after week, and the day-hours in which they arise are expected to mirror the general pattern at which all users, on average, use the service.

To enforce the stickiness of routines over weeks in our model, we assume that the length-scale parameter, ρ, of $\gamma_i(w)$'s squared exponential kernel is fixed to a value that yields smooth draws. In other words, while each individual's degree of routineness is allowed to vary over time, the week to week change is assumed to be relatively small. Specifically, we fix $\rho = 3$. We

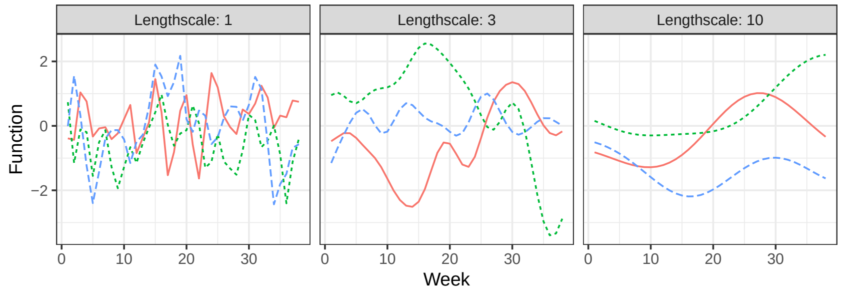

Figure 1: Draws from a GP prior.
The effect of the lengthscale parameter on draws from a GP with the squared exponential kernel. Each panel represents a different lengthscale (ρ). Each line is an independent draw from a GP prior with an SE kernel with that lengthscale. In our application, we fix the lengthscale $\rho = 3$, which ensures relatively smooth variation over weeks, consistent with the idea of routines being somewhat (but not absolutely) sticky.

illustrate this choice in Figure 1, where we plot draws from a GP with a squared exponential kernel, across several values of ρ. The choice boils down to a trade-off: smaller values allow for more flexibility in the second term of the rate, by assuming less consistency in routines over time.[6]

3.4. Inference

We estimate the model parameters in a fully Bayesian fashion using NUTS, a gradient-based MCMC sampler, implemented in Stan (Carpenter et al., 2017). For all previously unspecified parameters, we use weakly informative priors. However, in its simplest form, the above model is computationally intractable: by discretizing the arrival times into hourly buckets, we force the model to do likelihood computations over many time periods in which nothing happened. That is, customers often interact with the firm sparsely (e.g., at most, once or twice per day); yet, our likelihood function is specified as a count variable over all time periods $t = (w, d, h)$, which forces us to consider all of the zeroes. To help facilitate inference in this set up, we draw on a property of Poisson variables described in Gopalan et al. (2015). Specifically, the log-likelihood

[6] We tested multiple values of ρ. We found that the substantive results are generally not sensitive to these values, provided ρ induces enough smoothness to be consistent with the idea of a routine, although overly smooth values $\rho > 7$ negatively impact predictive performance.

of our model for all observations from a single customer i can be decomposed into two terms:

$$\log p(y_i \mid \lambda_i) = \sum_{y_{it} \neq 0} y_{it} \log(\lambda_{it}) - \sum_t \lambda_{it} + C, \tag{14}$$

where C is a constant with respect to λ_{it}. The first term in this expression depends only on the non-zero values of y_{it}, while the second term is a simple sum over all λ_{it}. In this way, the likelihood can operate only on the non-zero values of y_{it}, circumventing the problematic sparsity.[7]

4. Application: Ride-sharing

We apply our model to data from Via, a popular NYC-based ride-sharing company. The data contain detailed records on a subsample of customers who we observe from the day they joined the ride-sharing platform. Specifically, in our application, we focus on a subsample of 2,000 customers, with data spanning 48 weeks. For each customer, we discard their first three weeks of data after acquisition. Of the 48 weeks, we use the first 38 weeks for calibration, and reserve the final 10 weeks for holdout validation.

Like most ride-sharing platforms, Via uses a proposal system for matching riders with rides. Specifically, when a customer uses the company's app to request a ride, their request generates a proposal, assuming a match can be found. The rider can then accept or reject that proposal. Unlike Uber or Lyft, however, Via operates primarily as a ride-*sharing* service, where customers typically share their ride with other customers, and often need to walk short distances from their request locations to their pick-up locations, and from their drop-off locations to their requested destinations. Thus, the proposal on Via includes standard information like the cost of the ride and how long the driver will take to get there, and also information about how far the user will have to walk to meet the driver. Occasionally, a rider requests a ride and then rejects it, possibly multiple times, looking for a better proposal. To handle situations like this, the company uses a unit of analysis called a session, which is a grouping of back-to-back requests. Following the company's lead, the dependent variable we focus on in our analyses is

[7] Our Stan code is available upon request.

Table 1: Summary statistics.
Summary statistics for our ride-sharing data, summarized over the training data, unless otherwise noted.

Total Customers	2,000
Total Weeks (Training)	38
Total Weeks (Holdout)	10
Number of Sessions	86,952
Sessions / Customer	43.48
Sessions / Customer / Week	3.10
Weeks in Data / Customer	14.02

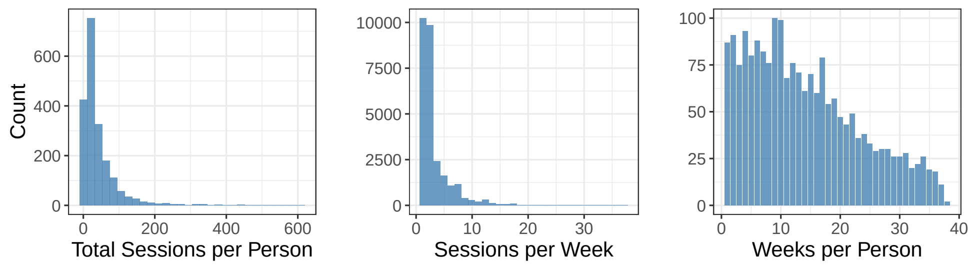

Figure 2: Distributions of summary statistics.
Distribution of three summary statistics in our training data: (1) the total number of sessions per customer; (2) the number of sessions per customer per week; (3) the number of weeks per customer.

the number of sessions a given user has in a given hour. [8]. Summary statistics for our data are presented in Table 1 and in Figure 2. The vast majority of riders have either zero or one session per hour, and most users have less than 10 sessions per week.

Importantly, a user can have a session without actually taking a ride, if the user declines all of the proposals. We focus on requests, and not whether the ride was eventually accepted or completed, because it is the most granular level of engagement with the company. A request means the rider was interested in using the service at that time. That said, we further leverage the information about acceptance and rejection of proposals when we subsequently investigate the implications of routines for customer behavior and customer management.

[8]throughout the paper, we will use the terms "request" and "session" interchangeably, always referring to sessions

5. Results

5.1. Simulations

To show that the model is indeed able to capture relevant routines, and to illustrate the different components in the model, we now turn to a quasi-simulation that combines real and synthetic customers. Specifically, we simulated the usage of 15 hypothetical customers, with rates of usage typical of customers in our data, and whose usage follows pre-specificied, managerially meaningful patterns. These patterns include different types of routines and different patterns of overall usage, including switching between random and routine usage, and churning from the platform.[9] Then, we merged the data from these 15 simulated customers with a sample of 500 real customers, and estimated the model on this partly synthetic dataset. By combining the simulated case studies with a much larger set of real customer data, we ensure that the population-level parameters are consistent with reality. For the sake of brevity, the remainder of this section presents the individual-level model results from two of these simulated customers — one exploring the model's ability to detect routines separately from random usage, and the other illustrating how the model captures churn in the data. The results for the remaining 13 simulated customers are reported in the appendix.

5.1.1. Simulated Case 1: Random then Routine

In Figure 3, we plot the key model estimates for a simulated individual for whom routine behavior emerges over time. Specifically, this individual was simulated by drawing day-hour request times in two ways: for the first half of the data (i.e., before week 20), each week, we drew five day-hours completely at random, and assumed the individual makes one request at each of these five day-hours. Since the five day-hours are drawn anew each week, there is no pattern to this user's usage, and thus the model should capture this as random activity. Then, at week 19, we simulate this user suddenly adopting a routine. To simulate routine usage, we first drew a set of five random day-hours (e.g., Sunday at 2pm, Tuesday and Wednesday at 8pm, and Thursday at 8am and 6pm), and then assumed the user requests a ride at these

[9] Note that these patterns were *not* generated from the model itself.

same five day-hours each week. Since the user is making requests at the same times, week over week, the model should detect that a routine has emerged around week 20.

There are five panels in Figure 3: at the top left, we plot the posterior median estimates of E_{iw}^{Random} (black/solid) and E_{iw}^{Routine} (red/dashed). To the right of the decomposition, we show the posterior median estimates of the random scale parameter, α_{iw}, and the routine scale parameter, γ_{iw}. Finally, below those, we show the posterior median estimate of the routine rate η_{ij}, plotted as an intensity over day-hours, and below that, the model's expectations for when this user will request rides during the last week of the training data ($w = 38$).

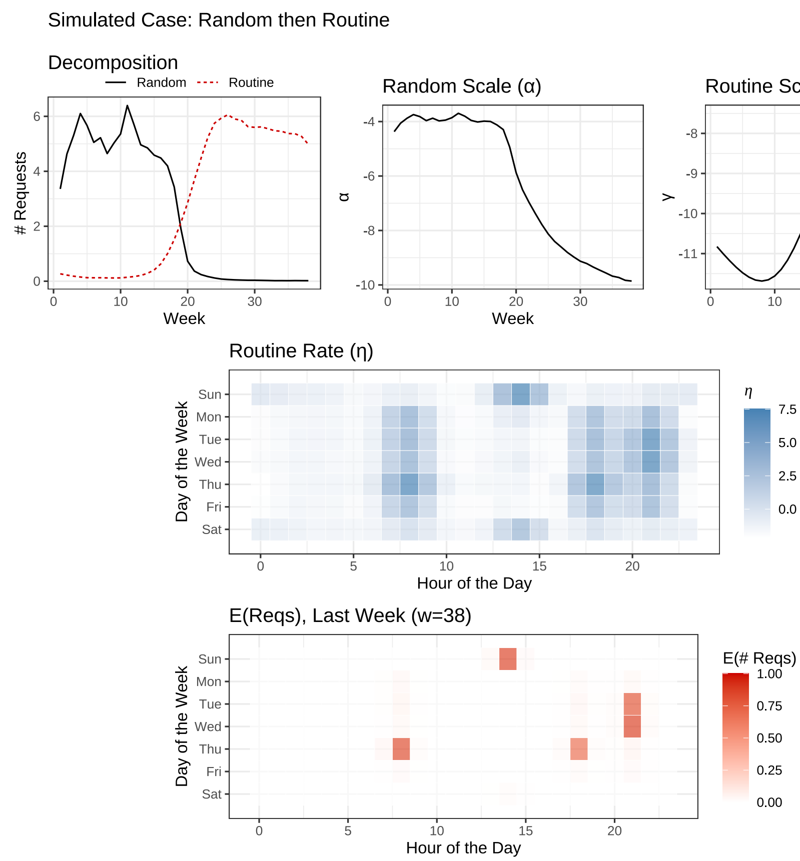

Figure 3: Simulated Case: Random then Routine.
Model estimates for a simulated individual who first uses the service randomly, then adopts a routine.

From Figure 3, we see the model is correctly able to parse this user's behavior: in the Decomposition panel, we see the random component E_{iw}^{Random} is high at the start, capturing around 5 rides per week. We can also see this reflected in the relatively high value of the random scale, α_{iw}. Then, at the middle, we see a sudden shift, with E_{iw}^{Random} falling to zero, and E_{iw}^{Routine} rising to around 5, corroborating the model's ability to detect routines. The times that the user is expected to request a ride are captured in the user's routine rate, $\eta_i(j)$, for which we can see there are five peaks in usage, and these peaks, when combined with the routine scale $\gamma_i(w)$ in week 38, translate exactly to five expected requests at exactly the hours simulated: Sunday at 2pm, Tuesday and Wednesday at 8pm, and Thursday at 8am and 6pm.

5.1.2. Simulated Case 2: Routine then Churn

In Figure 4, we see the results for a different simulated user, who first exhibited routine behavior (generated analogously to the last 19 weeks of the simulated customer in case study 1), but then stopped using the service altogether. Although our model is not explicitly designed to detect churn, churn can be captured in our framework when both scaling terms become very negative, essentially implying zero expected requests. Indeed, we see that this is exactly how the model behaves: we see in the first panel that the decomposition correctly detects, at first, a high level of routine usage, which then dips to zero at the midpoint, when the user churns. Looking at the model components, we see this pattern of routine requests is driven by the routine scale parameter, γ, which starts out relatively high (when the user is active), but then plummets and stays low until the end of the data. Again, the routine rate, η, is able to recover the correct routine for this user, with five peaks (Fri at 1pm, Sat at 3pm, and so forth). However, as reflected in the bottom figure, when that routine rate is combined with a very negative routine scale, we see that the model essentially predicts no requests at all for the last week of the data, when the customer has churned. [10,11]

[10]While the scale of α appears similar to the scale of γ in terms of magnitude of the parameter's value, this is misleading: recall that the number of requests coming from random needs and routine needs are $\exp(\alpha + \mu)$ and $\exp(\gamma + \eta)$ respectively (ignoring the subscripts). Thus, the scales of α and γ can only be interpreted relative to μ and η, which themselves may have different scales and variances.

[11]One may also notice that there is a decreasing pattern in the random scale: this decrease is an artifact of several things. First, when combined with the estimate for μ and exponentiated, all of these very negative numbers still suggest zero requests. Moreover, in the first period, there are two other things going on that lead to this observed behavior: first, the prior for the first period, α_{i1}, is shared across individuals, leading to some pooling in the first period. Second, since the model has no data to the left with which to smooth, we often find less reliable estimates at the very start of the data, than towards the end.

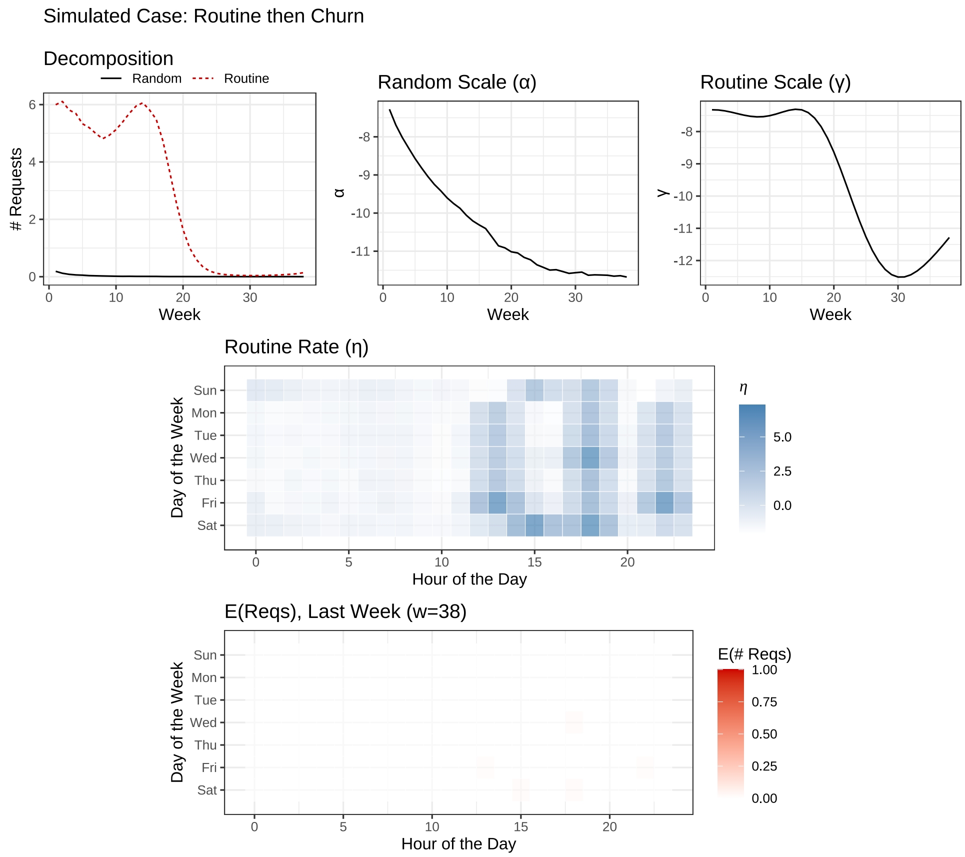

Figure 4: Simulated Case: Routine then Churn
Model estimates for a simulated individual who uses the service in a routine pattern, but then churns.

5.2. Model Estimates

Having established the model's ability to separate routine behavior from random behavior, we now turn to describing the results from the real data, estimated on the full sample of 2,000 customers over the period of 38 weeks used for model calibration. We first describe some of the population-level parameter estimates which characterize usage patterns broadly; for example, what days and times exhibit the highest level of usage across customers, and how often users exhibit random vs. routine behavior. We then describe some individual case studies, exploring the degree of routineness and the specific routine patterns for individual consumers.

5.2.1. Population Parameters

There are two main population-level parameters of interest: the common population-level rate parameter, $\mu(j)$, which governs when users tend to take rides (randomly), and the correlation matrix Ω from the day-hour kernel, which describes how different days are related to one another. We plot the posterior mean of $\mu(j)$ in Figure 5 and visualize the posterior mean of Ω as a correlation plot in Figure 6.

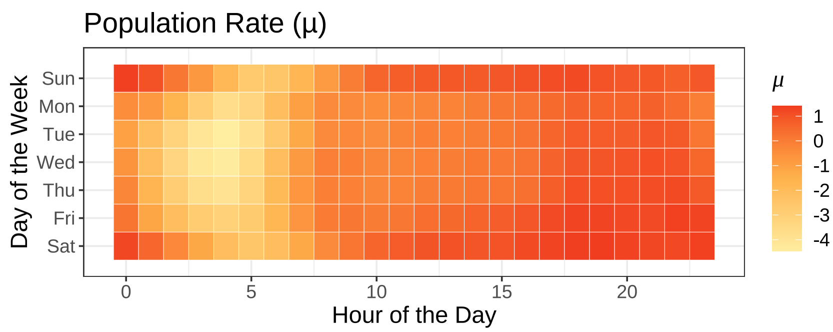

Figure 5: Posterior mean of $\mu(j)$.
Posterior mean of $\mu(j)$, the common rate of usage across individuals at the day-hour level.

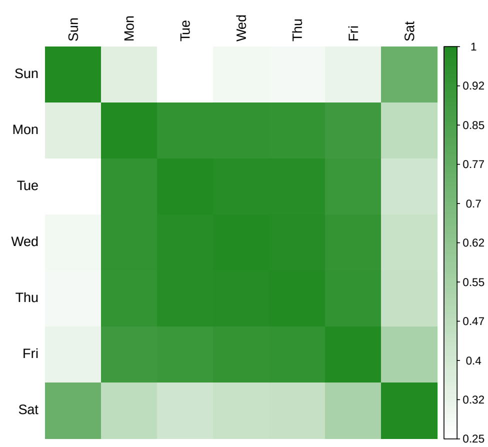

Figure 6: Posterior mean of Ω.
Visualization of the posterior mean of Ω, the correlation matrix across days for routines. Darker colors indicate higher correlation.

Some intuitive patterns emerge: first, from Figure 5, we see that random needs tend to arise during all times, except in the middle of the night (i.e., hours 2-5, or 2 AM to 5 AM). This pattern is moderated somewhat on the weekends, when travel times shift a bit later, and when there's a noticeable drop in usage at 4 AM, corresponding to the closing time of many bars in New York City. On weekdays, we also also observe a slight increase in usage of the service in the evenings, but the daytime variation is much less stark than the variation between day and night. Similarly intuitive, the correlation matrix in Figure 6 captures the expected pattern that weekdays tend to be more similar to one another than weekends. Saturday and Sunday are somewhat correlated, as are Friday and Saturday.

5.2.2. Usage Decomposition

A key output of our model is the decomposition of usage into routine and random requests. Essentially, the model infers whether each individual request is random or routine, depending on the user's temporal consistency, week over week. Before examining the distribution of random and routine requests in the population, we first corroborate that overall usage captured by the model resembles the true data patterns. Figure 7 compares expected usage (i.e., $E(y_{iw})$) and actual number of requests made in the training data. We see a strong correlation between our model's expectation and reality, reflecting very good fit (correlation $r = 0.904$, $p < 0.001$).

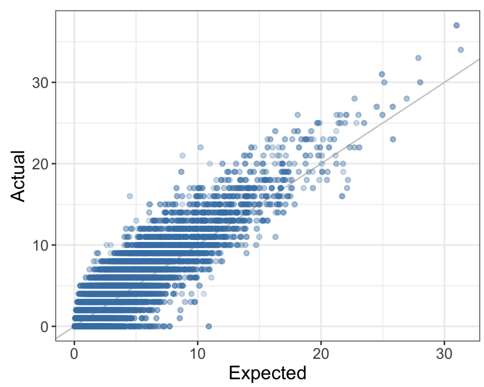

Figure 7: In-sample fit.
In-sample fit, where expected is $E(y_{iw}|\lambda_{iw}) = E_{iw}^{\text{Routine}} + E_{iw}^{\text{Random}}$, and actual is the actual number of requests a customer made.

We now explore the decomposition $E(y_{iw}|\lambda_{iw}) = E_{iw}^{\text{Routine}} + E_{iw}^{\text{Random}}$ for all users in our sample. Figure 8 shows the joint distribution of the two parts E_{iw}^{Routine} and E_{iw}^{Random} in the last week of our data. We find a somewhat L-shaped distribution, suggesting that heavy usage customers are either primarily routine or primarily random but rarely both. The vast majority of customers fall in the lower left part of the figure, with few requests per week, balanced between random and routine. Although Figure 8 shows the decomposition pattern for the last week of the data, we also find similar weekly decompositions throughout the data period.

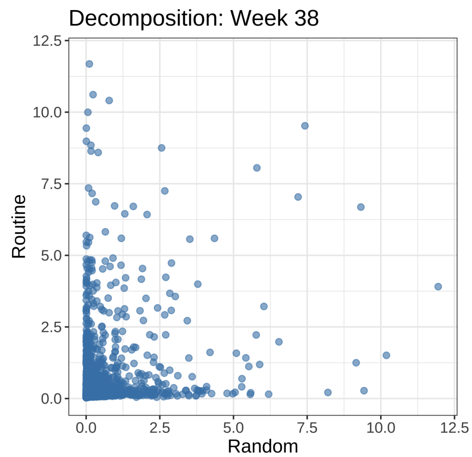

Figure 8: Joint distribution, E_{iW}^{Routine} and E_{iW}^{Random}.
The joint distribution of the posterior medians of E_{iW}^{Routine} and E_{iW}^{Random}, where $W = 38$, the last week of the data.

5.2.3. Individual Customers' Routine Patterns

We now zero in on the individual-level parameters, to illustrate the insights provided by the model. Relative to the simulated examples, the results on real users are less clean cut in their interpretation, but still offer valuable customer-level insights. In Figures 9 – 10, we show the same posterior estimates and decompositions for two real customers, as we did for the simulated case studies in Figures 3 – 4.

In Figure 9, we show an example of a very common type of routine: commuting. As shown in the decomposition, this customer is a fairly heavy user, making roughly 14 ride requests per week, with a high-level of routine usage. This routine usage tends to cluster around commuting hours, 8am and 5pm, as can be seen both in the routine rate and in the expected numbers of requests. In contrast, in Figure 10, we show a customer with an increasing and random pattern of usage, who we could characterize this user as a "casual" rider. In the decomposition, we see the random component trending upward, driven by the increase in this customer's random

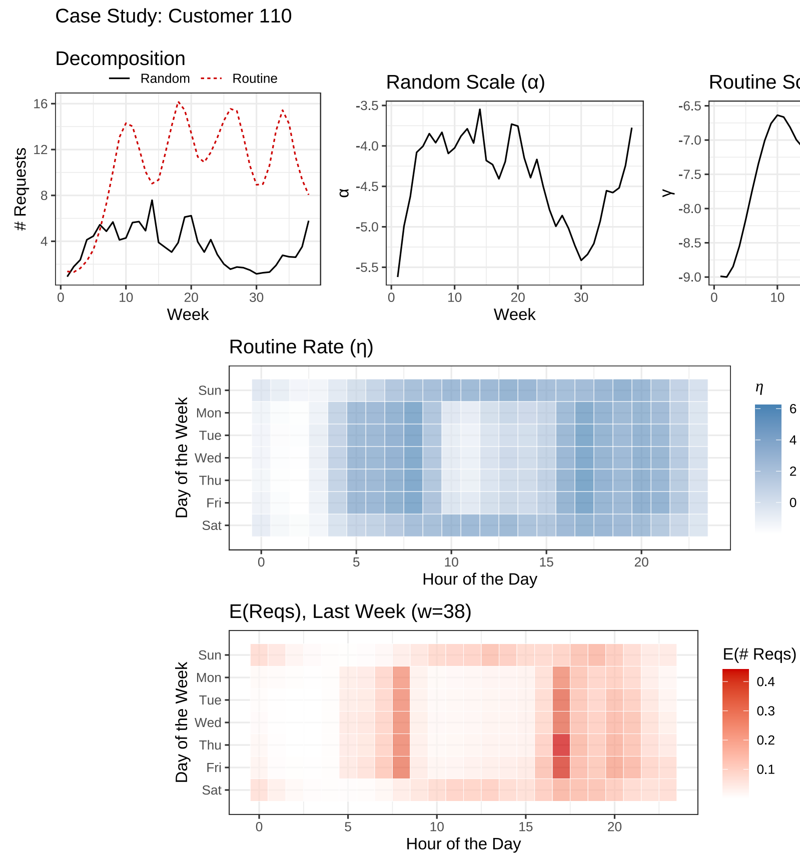

Figure 9: Real User: Commuting Routine.
Model estimates for an individual who uses the service in a routine, typically in commuting hours.

scale. While the model does still learn a routine rate $\eta_i(j)$ for this customer, when combined with the very low routine scale $\gamma_i(w)$, we see that the customer's expected pattern of usage is very diffuse, very much resembling the population pattern shown in Figure 5.

While the case study in Figure 9 captures a fairly intuitive and common routine, there are, in fact, many different routines present in our data. In Figure 11, we show a set of six different routine rates ($\eta_i(j)$) of customers with a relatively high level of routine usage. Starting from the top-left, we see:

- Customer 44 - "Night Owl": a routine clustered in the very early mornings, especially on Saturday and Sunday nights, perhaps suggesting a person with an active nightlife, or a person who works a night shift. Note that this user's usage, especially during the weekdays, is essentially the opposite of the population pattern.

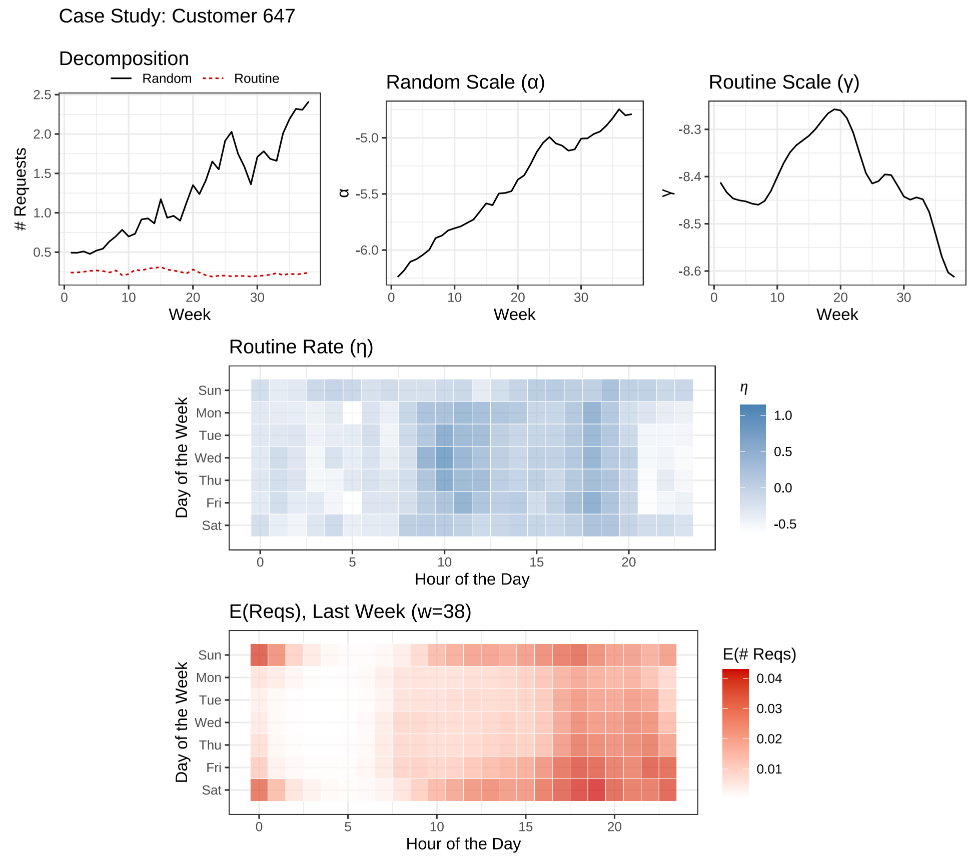

Figure 10: Real User: Random Usage.
Model estimates for an individual who mostly uses the service at random.

- Customer 235 - "Early Bird": a strong weekday morning routine, but with less of an evening routine, and a strong weekend afternoon routine.
- Customer 734 - "Evenings Out": a strong evening routine, but with less of a morning routine, and a strong weekend evening routine.
- Customer 826 - "Workaholic": a very stark 8am and 11pm routine, even on the weekends, perhaps suggesting someone working long days and weekends.[12]
- Customer 937 - "Nightlife": consistently traveling between 2-4am, especially on weekends, suggestive of a person with a healthy nightlife.
- Customer 945 - "Work Hard, Play Hard": a three-part routine, with many trips in the morning (around 6-7am), afternoon (around 3pm), and evening (around 11pm).

[12] A doctoral student, perhaps?

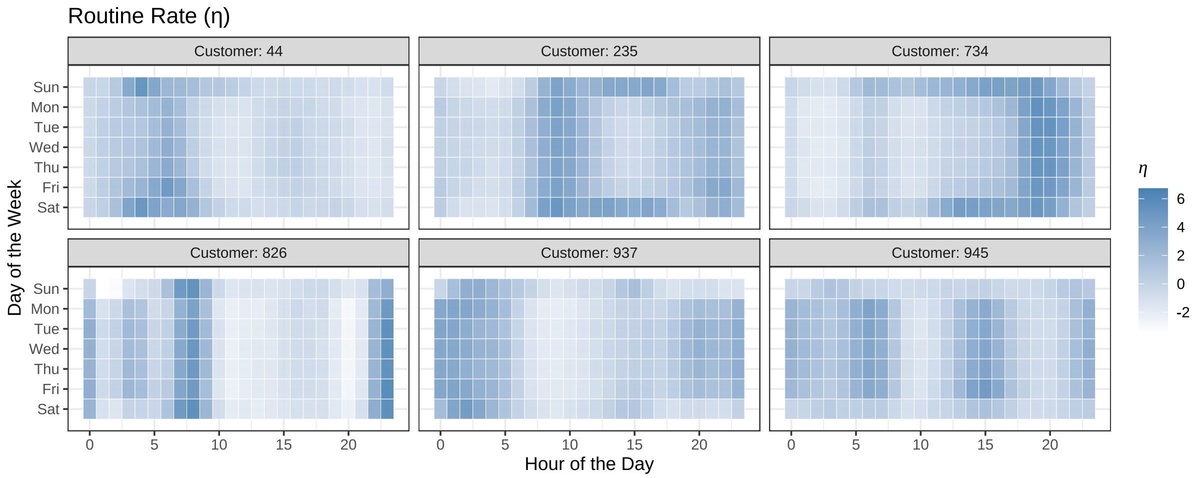

Figure 11: Examples of other routines.
Six other routines found in the real data. Shown are the posterior median estimates of $\eta_i(j)$.

Taken together, these case studies illustrate the flexibility of the model to capture customer-level routines and isolate interesting and often complex usage patterns. Identifying such patterns at the customer level can be useful for the firm both for targeting purposes, suggesting to customers promotions at times that are relevant to them, and from a supply perspective, making sure that the supply of driver matches demand.

5.3. Model Validation

Before exploring the implications of routineness for customer relationship management, we first validate the model, by examining its ability to both explain and predict the data. Note that the primary goals of our model are decomposition (between random and routine consumption) and the illumination of individual-level routines — not prediction. That being said, our model is, at its core, a probabilistic model of customer behavior, and can thus be used to predict individual-level behavior. We use predictive checks as a means of establishing that our model is doing a reasonable job at explaining the data.

Mechanically, making predictions from the model is straightforward. Since $\mu(j)$ and $\eta_i(j)$ do not vary over weeks, the only two components of the model that need to be projected forward are α_{iw} and $\gamma_i(w)$. For α_{iw}, we simply simulate forward the Markovian process, generating future weeks according to $\alpha_{iw+1} \sim N(\alpha_{iw}, \tau^2)$. For the GP term, $\gamma_i(w)$, we utilize the fact that GPs are marginally Gaussian to derive the posterior predictive values, given what we previously observed. Let w^* indicate a new week that we want to make a prediction for, having observed weeks $\mathbf{w} = (w_1, \ldots, w_W)$, with estimated scale values $\gamma_i(\mathbf{w})$. Then $\gamma_i(w^*) \sim N(m^*, s^*)$, where,

$$m^* = K(w^*, \mathbf{w})K(\mathbf{w}, \mathbf{w})^{-1}\gamma_i(\mathbf{w}), \tag{15}$$

$$s^* = k(w^*, w^*) - K(w^*, \mathbf{w})K(\mathbf{w}, \mathbf{w})^{-1}K(\mathbf{w}, w^*), \tag{16}$$

and where $K(w^*, \mathbf{w})$ is the vector formed by evaluating the kernel $k(w^*, w)$ for all $w \in \mathbf{w}$, and likewise for $K(\mathbf{w}, \mathbf{w})$ and $K(\mathbf{w}, w^*)$.

Using this machinery, we can predict different aspects of future behavior. For instance, we can compute how many rides someone will take in the future, similar to what we showed for in-sample fit in Figure 7. Indeed, we find a strong positive correlation between predicted and actual out-of-sample request volume ($r = 0.55$, $p < 0.001$). However, more interesting and unique for the present model is predicting *when* someone will request rides, in terms of day-hours. This will be our focus for the remainder of the section.

We compare the predictive performance of the model to a nested benchmark in which we zero-out $\gamma_i(w)$ and $\eta_i(j)$, which we refer as the "No Routine" model. In terms of capturing week-over-week fluctuations in usage, this model is very flexible: there are no smoothness assumptions in weekly variability, and the weekly scale term is estimated at the individual-level (as in the full model). However, the only component of this benchmark that can predict day-hour variation in when users will request rides is $\mu(j)$. Hence, the model essentially predicts that all users will follow the same pattern of requests, making requests at the times when the average customer tends to take rides. In that sense, this model is a good "no information" baseline: by comparing fit and prediction to this benchmark, we get a sense of whether the data exhibits systematic patterns of day-hour variation in requests at the individual-level, and more importantly, whether the focal model is able to capture those interesting patterns.

5.3.1. Metrics for Request Time Validity

We use four metrics to measure how well the model is able to predict request times: two extant metrics — mean average precision (MAP) and conditional precision (CP) — and two novel metrics — routine conditional precision (RCP) and an extended version of the routine conditional precision (RCP+x). The basis of all of these measures is the *top-k precision*, denoted $p(k)$, which is computed as follows: Let T denote the set of day-hours that a customer requested rides in a given week, and let R_k denote a ranking of the k most likely day-hours for that customer to request a ride, as predicted by the model. Then, ignoring the i and w subscripts, $p(k) := |R_k \in T|/k$, which captures the fraction of those day-hours in R_k in which the customer actually requested a ride. As a running example, let us consider a person who took four rides, i.e., $y = |T| = 4$. Suppose those rides happened at day-hours $T = \{11, 22, 33, 44\}$. Suppose then that the model's top-6 predicted ride times were $R_6 = \{11, 22, 32, 33, 45, 44\}$. Then $p(1) = 1$ (since the top-ranked ride actually happened), $p(2) = 1$ (since both of the top-2 rides happened), but $p(3) = 2/3$ (since the rank-3 ride did not happen), and so on.[13] Having defined $p(k)$, we can now define our four key metrics used to validate the model.

Mean Average Precision (MAP) The mean average precision is often used to evaluate the success of rankings, particularly in recommender systems, where the goal is to return a ranked list of items a customer may like.[14] In our case, the relevant ranking is the ranking of day-hours when a customer is most likely to request a ride in a given week. To define MAP, we first define the average precision (AP) for a given customer (again omitting the i and w subscripts),

$$AP = \frac{1}{y} \sum_{k=1}^{168} p(k)\, \mathbb{I}(\text{Request @ k}). \tag{17}$$

where $\mathbb{I}(\text{Request @ k}) = 1$ if the user made a request at the day-hour ranked k and 0 otherwise. To illustrate this, let's return to our example with true ride times $T = \{11, 22, 33, 44\}$, and

[13]Note that this metric does not separate adjacent mispredictions (e.g., if the true day-hour of a ride were 33, a prediction of day-hour 34 or 14 would both count as mispredictions). However, note that our model is likely to rank similarly adjacent day-hours due to the correlation between days and hours induced by the day-hour kernel.

[14]For an excellent exposition of MAP and recommender systems, see: http://sdsawtelle.github.io/blog/output/mean-average-precision-MAP-for-recommender-systems.html

$R_6 = \{11, 22, 32, 33, 45, 44\}$. The average precision for this user would be $\frac{1}{4}(1 + 1 + 0 + \frac{3}{4} + 0 + \frac{4}{6} + 0 + \ldots) = 0.854$. The AP is always between 0 and 1, with higher values indicating that the model is producing better rankings of the day-hours for that customer. The AP will be 1 if all of the user's rides happened during the highest ranked hours. Finally, the MAP is just the average AP across users.

Conditional Precision (CP) The conditional precision is similar to the classic precision metric. Suppose that we know a given user rode y times in a given week; the conditional precision captures which of those y day-hours the model predicts correctly. Mathematically, CP is equal to $p(y)$.

Routine Conditional Precision (RCP) In our model, we do not necessarily expect the average CP to be high: rather, we only expect to be able to accurately predict the ride times of customers who are actually in routines. For those customers who have a high level of random usage, by definition, we do not expect to be able to predict their ride times. Thus, we also explore what we call the *Routine Conditional Precision* (RCP): instead of conditioning on y, we condition on the rounded expected number of routine requests in a week. Using the shorthand r to denote E^{Routine} rounded to the nearest integer, and drawing on our previous notation, we define RCP $:= |R_r \in T|/r$. *In short, CP asks what fraction of y trips the model is able to correctly predict with y guesses, while RCP asks what fraction of r routine trips the model is able to correctly predict, with r guesses.* Returning to our example of $T = \{11, 22, 33, 44\}$ and $R_6 = \{11, 22, 32, 33, 45, 44\}$, suppose our model predicted $r = 2$. Let us now compare CP and RCP. This customer's CP will be 0.75, since of the top-4 times, 11, 22, and 33 were correct. However, the model suggests that only $r = 2$ of these times *should* have been predictable. Thus, this customer's RCP will be 1, since all of the the top-r (i.e., top-2) times were, in fact, observed.[15]

RCP-x While RCP is intuitive, it is also a very stringent measure: given r routine requests, RCP asks the model to return *exactly* r predicted day-hours. This ignores the fact that the

[15]Note that RCP cannot be defined by the benchmark, as only the full model allows for this decomposition. Hence, we will assume the routineness decomposition given by the full model, even when computing RCP for the no routine model.

predicted likelihood of seeing trips at the day-hour ranked r and $r+1$ may, in fact, be very similar, especially given the continuous nature of time.[16] Hence, we define one additional metric, which we refer to as RCP+x. Following the preceding logic, given the top $r+x$ ranked ride times for a given user, R_{r+x}, and the list of actual times the user took rides, T, RCP+x is defined as $|R_{r+x} \in T|/r$. *In short, RCP+x asks what fraction of r routine trips the model is able to correctly predict with $r+x$ guesses.* Returning one last time to our example with $T = \{11, 22, 33, 44\}$, $R_6 = \{11, 22, 32, 33, 45, 44\}$, let's consider two scenarios: first, suppose, like above, our model predicts $r = 2$. In this case, RCP+1 will still be 1, since the model already correctly identified two times. Note, however, that RCP+2 will actually *exceed* 1: given 4 guesses, the model would correctly predict 3 times, although there were only 2 expected routine requests. That is, RCP+2 = 3/2. Intuitively, RCP+x will exceed 1 if non-routine trip times happen to be captured in the top-$(r+x)$ predicted day-hours. This first example illustrates a quirk of RCP+x, but does not illustrate its most important benefit over RCP. To see the benefit of considering RCP+x, let's suppose that instead of predicting $r = 2$, our model had predicted $r = 3$. In this case, RCP $= 2/3$, but note that the rank-3 predicted time, 32, is very close to the true time, 33. RCP ignores this closeness. RCP+1, on the other hand, would consider that the next highest ranked time was, in fact, day-hour 33, and RCP+1 $= 1$, illustrating the robustness of RCP+x to the common situation where nearby hours are ranked similarly.

5.3.2. Validation Results

Table 2 displays the MAP, CP, and RCP statistics for both the full model and the baseline no routine (NR) model, averaged across customers and weeks. We see that, for all fit measures and data windows (in- and out-of-sample), the model that incorporates routine information dramatically improves ride time predictions. In a way, this is not surprising: the only part of the NR model that predicts ride times is $\mu(j)$. In that sense, the NR model is assuming the same day-hour ranking across all customers. Thus, this model is capturing the capturing the ”population routine”, corresponding to the day-hours that customers, in general, are likely to call rides. Nonetheless, by comparing the full model with this “no information” baseline,

[16]This also relates to the problem noted in Footnote 13.

	MAP (In)	MAP (Out)	CP (In)	CP (Out)	RCP (In)	RCP (Out)
Full	0.267	0.150	0.193	0.095	0.298	0.224
No Routine (NR)	0.078	0.080	0.039	0.040	0.070	0.090

Table 2: Ride timing prediction measures.
MAP is the mean average precision, CP is the conditional precision, and RCP is the routine conditional precision. For all metrics, higher values indicate a better ability to predict when people will call rides. (In) denotes in-sample performance, while (Out) denotes out-of-sample (i.e., forecast).

we corroborate that there is rich variation in the data in terms of when individual customers request rides, highlighting the predictive validity of the routine component of the model.

We also notice that, despite the improvement over the NR model, the actual values of several of the statistics in Table 2 are modest. For instance, the CP metric suggests that we are only able to accurately predict roughly 20% of the in-sample ride times, and 10% of the out-of-sample. As described previously, this is due to the fact that CP ignores that some trips are routine, while others are random, and that we only expect to be able to predict the routine trip times. The RCP metric addresses that shortcoming: when we limit ourselves to just trying to predict the routine times, we see our model does better. We are able to correctly identify 30% of routine ride times in-sample, and 22% of routine ride times out-of-sample.

These results are even more impressive when we turn our attention to the RCP-x metric, plotted in Figure 12, for $x = 0, \ldots, 6$. Recall that RCP-x essentially builds in a buffer to the RCP metric, seeing whether the model can correctly identify r ride times with its top-$(r + x)$ ranking. Indeed, we see that the model performs very well out-of-sample: for the full model (solid line), RCP-x rapidly approaches 1 as we increase x modestly. For instance, when considering as few as two additional top-ranked day-hours (i.e., RCP-2), the precision exceeds 50%. In contrast, while RCP-x also increases for the NR model, this increase is much slower, and remains far lower than for the full model, even for high values of x. Together, these results provide strong evidence that the routine part of the model is indeed capturing true routine riding behavior even in future time periods.

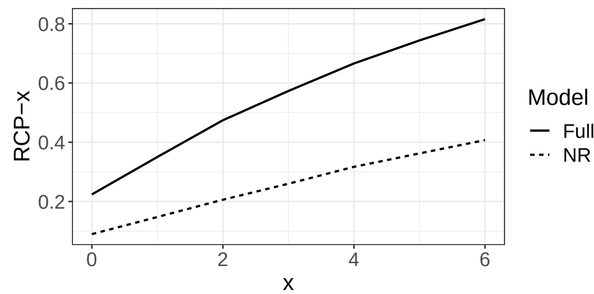

Figure 12: Values of RCP-x for different values of x.
We plot the value of RCP-x as x goes from 0 to 6, for out-of-sample data across the full and NR models. We see RCP-x rapidly converges toward 1 for the full model, suggesting good validity.

6. Routineness and Customer Management

Having established the predictive validity of the framework, we now return to the central question of the paper: Can routineness help firms better understand and manage their customers?

6.1. Routineness and Customer Value

We first consider whether routineness is an important predictor of future customer value, over and above commonly-used metrics such as the frequency and recency of usage. We consider the routineness of each user at the end of the calibration period (i.e., E_{i38}^{Routine}, as estimated by the model) and relate it to actual requests during the validation period, while controlling for other metrics that capture the overall activity of each user during the calibration sample. In particular, we consider two key dependent variables: `# Requests`, which is the number of requests a customer makes in the holdout period, and `Activity`, which is whether a customer is active at all in the holdout period, and which is a measure of 10-week retention. We then estimate several models that, in addition to the routineness metric, also include the number of requests the customer made in the last week of the calibration data, and the commonly used recency and frequency variables that capture how recently a customer last made a request and how many requests the customer has made previously. When modeling requests as the DV, we use simple OLS; for modeling activity, we use logistic regression.

Before describing the results, we note two important aspects of these regressions: first, routineness is essentially a component of the total number of requests. Hence, by controlling for

Table 3: Regressions of future activity.
We regress either number of future sessions (models 1-2), or a binary measure indicating any activity at all (models 3-4) on customer-level summary statistics. Future activity is measured either over the full holdout sample of 10 weeks (models 1 and 3), or over just the last five weeks of the holdout data (models 2 and 4).

	Dependent variable:			
	# Requests		Activity	
	OLS		*Logistic*	
	Full Holdout	Last 5 Weeks	Full Holdout	Last 5 Weeks
	(1)	(2)	(3)	(4)
Requests ($w = 38$)	2.224***	0.597***	0.383***	0.180***
	(0.223)	(0.141)	(0.108)	(0.057)
Recency	−0.189***	−0.106***	−0.140***	−0.124***
	(0.042)	(0.026)	(0.010)	(0.010)
Frequency	0.095***	0.049***	−0.00002	0.004**
	(0.007)	(0.004)	(0.002)	(0.002)
Routine ($w = 38$)	5.750***	3.284***	1.110***	0.307**
	(0.436)	(0.275)	(0.385)	(0.147)
Observations	2,000	2,000	2,000	2,000
R^2	0.567	0.425		
Note:			*p<0.1; **p<0.05; ***p<0.01	
			Intercept omitted for clarity.	

the number of requests in $w = 38$, we are in essence trying to determine whether the *shape* of within-week usage matters when predicting future customer value. As high usage can result from either random needs or routines, this specification allows us to understand whether having a higher routine component is incrementally valuable, over and above controlling for just the level of usage. Second, the inclusion of recency and frequency metrics is especially important here: a litany of models in customer base analysis have shown that these metrics are key summary statistics for predicting repeat purchasing (e.g., Schmittlein et al., 1987; Fader et al., 2005; Blattberg et al., 2008). If mere habit were the primary driving force behind customer value, we would expect these two statistics to explain much of the variation in future transactions. Thus, by incorporating these measures in the model, we can establish whether routineness matters, beyond what mere habit would already predict.

The results of these linear models are shown in Table 3. We estimate each model twice: in columns 1 and 3, we estimate the models using the DVs as measured over the entire holdout sample (10 weeks), and in columns 2 and 4, we estimate the models using the DVs as measured only over the last month of the holdout data (i.e., the last 5 weeks). The intent behind splitting the data in this way is to assess how robust routineness is in predicting short- and mid-term customer behavior. We find that a higher routineness is positively and significantly associated both with the number of requests a customer makes, and with the customer being active at all. In sum: even after controlling for how many requests a customer made at the end of the training data, and the standard recency and frequency measures from the CRM literature, we arrive at our key finding: *having more requests come from a customer's routine is positively and significantly associated with higher future customer value.*

6.2. Routineness and Other Customer Behaviors

Next, we consider whether highly routine customers behave differently in other ways, beyond spending and retention. We hypothesize that customers who are routine may not only be more likely to engage in activities that are directly valuable for the firm, but may also be more likely to forgive potential negative interactions with the firm, including price changes and service failures and disruptions. There may also be other differences between routine and non-routine users in terms of which aspects of the service are more important to them. For instance, users who routinely rely on the service may place higher importance on things like convenience of trips.

Recall that our data includes information about the rides that users requested. Some of these variables are *characteristics of the proposal*, including the cost to the user (`Price`), the time until the driver can pick the customer up (`Driver ETA`), expected time of arrival to destination (`ETA Destination`), how long the customer will have to walk to get the ride (`Pickup Walking Dist.`), total distance of the trip (from which we compute `Speed`), and the number of passengers for that request (`Passengers Req.`). We observe these characteristics for all the requests in the data. Moreover, for rides that were actually realized—that is, requests that ended up in a trip—we observe variables that capture the *quality of the ride*. These include whether the driver picked up the rider on time (`Pickup Delay`, which we measure in minutes), whether there were delays in the trip (`Dropoff Delay`), whether the rider had to walk short/long distances from their drop-off to their final destination (`Dropoff Walking Dist.`), and whether there were other passengers in the car during the trip (`# On-board (Pickup)`, `# On-board (Dropoff)`, and `Max On-board`).[17] Based on this data, in this section, we ask the following two questions: (1) whether routine customers are more likely to accept ride proposals, and particularly, less favorable proposals, and (2) whether routine customers are more likely to make a new request within 7 days of a previous ride, particularly after a service failure (i.e., a lower quality ride).

Accepting Proposals First, we consider whether routineness explains or moderates the likelihood of a customer accepting a given ride proposal. Recall that, on Via, when a user requests a ride, they are then given a proposal for that ride, which includes details like how much the ride will cost, how long it will take, and how far they will have to walk to be picked up by the driver. We hypothesize that routine customers will not only be more likely to accept proposals, but they may also be differentially affected by other aspects of the proposal, like wait time, walking distance, and price.

To examine these hypotheses, we estimate a linear probability model with user-level random effects, with a binary dependent variable that captures whether user i accepted proposal m.[18] As predictors, we include our focal (week 38) routineness metric, and the characteristics of the proposals. For this analysis, we use only independent variables about the proposal itself, since the proposals were not all accepted. We also include the interaction of routineness and proposal characteristics, which is of key interest: does routineness moderate the effect of different, potentially negative, aspects of the proposal on the likelihood of the customer accepting it? Finally, we include the total number of requests made (in week 38), to control for the customer's level of overall usage. The estimated coefficients from this specification are shown in the first column of Table 4.

First, we see that, indeed, more routine customers are more likely to accept proposals. We also find that many of the main effects from the control variables are intuitive: for example,

[17] The full data is summarized in the appendix.

[18] We use OLS here, as opposed to logistic regression, to aid in the interpretability of the interaction terms. However, the results remain consistent if we use logistic regression.

Table 4: Regressions of other CRM outcomes on routineness.

OLS regression of (1) whether a proposed ride was accepted and (2) whether, conditional on having taken a ride, a new ride was requested within 7 days. All models include a user-level random effect and control for the properties of the focal ride. The full set of control variables can be found in Table 8. The variables above the horizontal line are all variables associated with a proposal; below the line are all variables associated with only realized trips. The predictors in both models were standardized to improve readability.

	Dependent variable:	
	Accept Proposal	Request Again
	(1)	(2)
Routineness	0.084^{***}	0.208^{***}
# Requests (Week 38)	-0.021^{*}	0.321^{***}
Price	-0.031^{***}	-0.052^{**}
Driver ETA	-0.050^{***}	-0.002
ETA Destination	-0.012^{***}	-0.017
Speed	0.074^{***}	-0.034
Pickup Walking Dist.	-0.041^{***}	-0.006
# Passengers Req.	-0.003	0.025
Routineness x Price	-0.001	0.037^{**}
Routineness x Driver ETA	0.009^{***}	0.005
Routineness x ETA Destination	-0.0002	-0.009
Routineness x Speed	0.162^{***}	-0.089
Routineness x Pickup Walking Dist.	-0.008^{***}	-0.0002
Routineness x # Passengers Req.	-0.001	0.019
Pickup Delay		-0.027^{***}
Dropoff Delay		-0.010
Dropoff Walking Dist.		-0.012
# On-board (Pickup)		-0.015
# On-board (Dropoff)		0.002
Max On-board		0.012
Routineness x Pickup Delay		0.018^{**}
Routineness x Dropoff Delay		0.007
Routineness x Dropoff Walking Dist.		0.005
Routineness x # On-board (Pickup)		0.013
Routineness x # On-board (Dropoff)		0.001
Routineness x Max On-board		-0.011
Other Controls	Yes	Yes
Observations	38,166	14,704
R^2	0.052	0.068
Note:		$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

customers are less likely to accept higher price rides, less likely to accept proposals with a longer driver ETA, and less likely to accept rides with longer walking distances to their pick-up spot. However, importantly, we find that routineness *moderates* several of these. First, we find that highly routine customers are less sensitive to longer driver ETAs, suggesting they are more comfortable waiting for a ride to arrive. On the other hand, routine customers are *more* sensitive to both the pickup walking distance, and the speed of the trip. This pattern of moderation is consistent with routine customers caring more about convenience-related variables like pickup walking distance and speed. In part, this could be driven by self-selection: routine customers do not become routine if rides are inconvenient for them. Thus, routine customers self-select based on convenience factors like walking distance and speed of trip, and are thus more sensitive to these type of characteristics.

Requesting Again in the Future Another interesting question is whether routine customers are more robust to potential disruptions in service, or in general, instances of low service quality. For instance, in ride-sharing, will a routine customer be more likely to book again soon after a bad trip? To explore this, we run a similar model as above, but now restrict the analysis to occasions in which the user actually took the ride. We use as our dependent variable whether the customer requested a trip again in the next seven days. In addition to the request-level variables included previously, we also include trip-level variables, which are only realized during or after the trip has concluded. Again, our goal is to see whether routineness moderates the impact of negative aspects of the trip. The results are displayed in the second column of Table 4.

Overall, we find suggestive evidence supporting our hypotheses: first, consistent with our main CRM analysis, we find more routine customers are more likely to ride again within 7 days, in general. We also find that routineness seems to moderate the negative influence of price on re-riding. While customers are, in general, less likely to ride again after a high-priced trip, highly routine customers are less sensitive to price. We also find that routineness alleviates the negative effect of a pickup delay: similar to price, customers in general do not like waiting for rides, as captured in the significant, negative main effect of pickup delay, but routine customers are less sensitive to these delays. Taken together, this pattern of moderation suggests that highly routine customers may be more resilient to service disruptions and other negative aspects of trips, at least in the short-term.

6.3. Further Characterizations of Routines

So far, we have devised a measure of routineness at the customer-level, and established that routines are predictive of customer outcomes, including key determinants of customer value. We now turn our attention to further unpacking the construct of routineness, including analyzing which customers tend to be routine, and differentiating temporal routines from routines in terms of what someone buys (or where someone goes), and an extant measure of purchasing regularity.

6.3.1. What Makes a Routine Customer

First, we consider what types of customers tend to be routine. Understanding what makes for a routine customer can further validate our routineness metric, and suggest ways in which Via might consider acquiring routine customers, or cultivating routines. To explore this, for each user, we regress each user's week 38 routineness score on a number of variables describing that user's activity and trip types, averaged over the training period.[19] The results are shown in Table 5.

We find that more routine users have much in common: first, echoing our main CRM results, we see that routine customers have a higher probability of accepting a ride, given a request. We also find customers who take longer trips, and who request a ride with fewer other passengers (i.e., they are more likely to request a "solo trip," as opposed to bringing friends along for the ride) are more likely to have higher routineness. Partly, this may be explained by the prevalence of commuting routines, which intuitively may be more likely to be solo trips, and may be from more remote areas of the city to more central ones. A similar self-selection story may explain the positive association of routineness and driver ETA: if routine customers

[19]These results are similar if we consider the average routineness over all weeks, instead of the week 38 value. We prefer using week 38 routineness for consistency with the rest of the paper.

Table 5: Predictors of routineness.

Regression of week 38 routineness (DV) on average trip characteristics (IVs), where the average is taken over the whole training window. (These results are robust if we use the average routineness as the DV instead of its week 38 value.)

	Dependent variable:
	Routineness
Prob. Ride | Request	0.555^{***}
	(0.116)
First Week	-0.0003
	(0.001)
# Requests	0.401^{***}
	(0.008)
Price	-0.001^{***}
	(0.0001)
ETA Driver	0.047^{**}
	(0.019)
ETA Destination	-0.001
	(0.005)
# Passengers Req.	-0.131^{**}
	(0.061)
Walking Distance	-0.001^{*}
	(0.001)
Distance	0.116^{***}
	(0.025)
Airport Ride	-0.033
	(0.269)
Observations	2,000
R^2	0.605
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

are traveling at peak hours, it may take longer to find a driver. More interestingly, we find that routineness is associated with lower priced trips, and lower walking distance. While our analysis is not causal, these effects are suggestive: customers who are consistently confronted with high prices or high walking distances may stop using the platform, or never form routines.

6.3.2. "When" versus "What": Incorporating Location Information

In both our model and in the preceding discussion, we have focused exclusively on temporal routines: that is, *when* someone interacts with the firm, not *what* they do in that interaction. In a retail setting, for instance, a customer may come in at exactly the same time each week, but may buy either the same items each time, or different items. In our focal context, customers may request rides at exactly the same times each week, but may travel to either the same location each time, or different locations. Consider, for example, two work commuters, both of whom work in the same location each day. In the morning, both users may always go between the same locations, home and work. In the evening, however, one of these commuters may always return home, while the other frequently goes out for drinks or dinner. In this sense, both customers have the same "when" routine but different "what" routines. In this section, we explore to what degree "what" routines — that is, location choice — are predictive of "when" routines, and what gains there may be in accounting for "what" routines, in addition to our previously defined routineness measure.

Metrics for Location Dispersion To understand the degree to which there are "what" routines in location choice, we first need a metric of how consistent location choices are. Mathematically, it is more natural to construct measures of how dispersed (that is, how *in*consistent) trip locations are; hence, that will be our focus.[20] To understand location dispersion, we look at both pick-up and drop-off locations. In our data, locations are saved as precise latitude/longitude coordinates (or, "lat/long"), measured to five decimals. To discretize our location data to correspond to New York City street blocks, we truncate the decimal to the nearest 300th.[21] With this discretization, we then define two measures of location dispersion:

[20] Though each of our measures could be easily converted to a consistency measure by inversion.
[21] Specifically, given a raw coordinate x, we compute a truncated coordinate, $x^* = \lfloor 300x + 0.5 \rfloor / 300$.

1. **Shannon Entropy:** For this metric, we consider the empirical distribution of a user's locations (both pick-up and drop-off). For instance, if a user made ten total trips, nine to location 1 and one to location 2, the empirical distribution would be (0.9, 0.1). We then compute the Shannon entropy of that distribution, defined as:

$$\text{Entropy} = -\sum_{\ell=1}^{L} p_k \log p_k. \tag{18}$$

where p_ℓ is the empirical probability of the ℓth location, and L is the total number of locations. Intuitively, entropy captures how "predictable" a user's locations are. To illustrate, observe that $\text{Entropy}(0.9, 0.1) > \text{Entropy}(0.5, 0.5) > \text{Entropy}(0.1, 0.1, \ldots, 0.1)$.

2. **CRT Dispersion:** A feature of the entropy measure is that it does not take into account the total number of trips or locations for a given user. Hence, a user that takes 1,000 trips to the same 10 locations is just as entropic as a user who takes 10 trips to the same 10 locations. Hence, as an alternative to entropy, we consider a metric inspired by the Chinese Restaurant Table (CRT) process. The CRT process is a stochastic process, derived from the better known Chinese Restaurant Process, commonly used to model assignment to different groups (Zhou and Carin, 2013). It captures a "rich get richer" process, whereby new observations are either assigned to existing groups, with probability proportional to the sizes of those groups, or they are assigned to a new group, at a rate proportional to a dispersion parameter. Such a process could be used to model the evolution of trip location choices: for a new trip, that trip may be to an existing location, or it may be to a new location. While estimating the full CRT model is complex and cumbersome, there exists an intuitive and easily computed estimator of the CRT dispersion parameter, θ (Durrett, 2008, Chapter 1): given L unique locations in K total trips:

$$\theta \approx \frac{L}{\log K}. \tag{19}$$

In plain English, this metric is the number of unique locations divided by the log number of total trips. Contrasting this to entropy, if a user has just two unique locations in ten

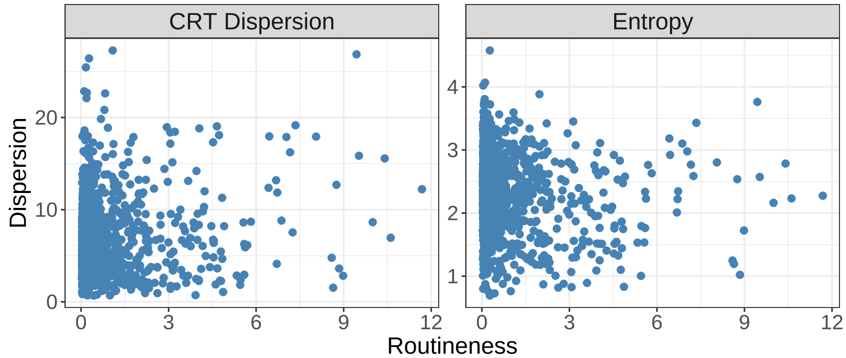

Figure 13: Joint Distributions of Routineness and the Two Location Dispersion Metrics
The joint distribution of routineness in week 38 (i.e., E_{i38}^{Routine}), and the two metrics of dispersion in location choice, entropy and CRT dispersion, showing no relationship between the two.

trips, $\theta = 2/\log(10) \approx 0.87$, regardless of whether those trips were split 5/5 or 9/1. But, unlike entropy, if a user visits 10 locations equally among 1000 trips, $\theta = 10/\log(1000) \approx 1.45$, whereas if the user visits 10 locations equally among 10 trips, $\theta = 10/\log(10) \approx 4.34$.

Location Results Having defined metrics of location dispersion, we now ask: first, to what degree is location dispersion related to temporal routineness? And second, is location choice also an important predictor of customer-level outcomes? To answer the first question, we plot the joint distribution of our routineness metric and the two metrics of location dispersion in Figure 13.[22] We see that there is no obvious relationship between temporal routineness and location dispersion. This is supported by simple regression analyses: alone, entropy and CRT dispersion explain less than 5% of the variation in routineness. When we regress routineness on both location metrics, together with other obvious individual-level controls (e.g., number of final week requests, the week the customer was acquired), we find CRT dispersion is a negative and significant predictor of routineness, while entropy is not significant. We report the full results from these regressions in the appendix. Taken together, these results suggest that location dispersion is minimally predictive of temporal routineness, albeit in an intuitive direction: customers with less dispersed location choices seem to be slightly more routine.

[22]For consistency, we focus on routineness as measured in the last week of our training data, but all the results are the same if we consider average routineness over the training data instead.

Having established that routineness and location dispersion are distinct, we now ask: does location dispersion have an impact on customer behavior over and above temporal routineness? Specifically, we return to the focal analyses in Table 3 where we explored the role of routineness in explaining the number of future requests and the likelihood of a customer being active in the future, but now also include our location dispersion metrics. We find that, when the location dispersion metrics are included alongside routineness, neither entropy nor CRT dispersion is a significant predictor of either outcome, suggesting that, from a CRM perspective, "when" matters significantly more than "what."[23]

6.3.3. Routineness versus Regularity

Finally, we consider the relationship between routineness and transaction regularity. The concept of regularity in transaction timing was introduced by Platzer and Reutterer (2016), who modeled it using a generalization of the Pareto-NBD model, the Pareto-GGG, which replaces the typical exponential-distributed inter-transaction times with gamma-distributed times, and models the parameters of that gamma distribution hierarchically. The gamma distribution is parameterized as $\text{Gamma}(k_i, k_i\lambda_i)$, and the k_i determines how regular customer i's intertransaction times are: higher values of k_i make the transactions more regularly spaced, hence the term regularity. Conceptually, there are links between regularity and routineness: in some sense, routineness can be viewed as a more specific and structured form of regularity. However, mathematically, they are distinct: we are capturing the regularity of transactions on the day-hour grid, whereas the regularity described by the Pareto-GGG purely captures consistency in intertransaction times. That means a routine customer may or may not be regular: for example, if a customer called a ride every Wednesday at 5pm, and no other time, this customer would be both regular and routine. However, there are also cases where routine customers are not regular: consider the case of Customer 44 in Figure 11, whose most likely transaction times were Saturday and Sunday in the early morning. Although this user has a consistent routine, their transactions would actually be highly irregular by the regularity metric, since their behavior features a mix of both very short and very long intertransaction times.

[23]We again report the full regression results in the appendix.

Table 6: Comparisons with regularity.
Regressions of future activity – either number of future sessions (1-2), or a binary measure indicating any activity at all (3-4) – on customer-level summary statistics, including regularity.

	Dependent variable:			
	# Requests		Activity	
	OLS		*logistic*	
	(1)	(2)	(3)	(4)
Requests ($w = 38$)	4.105***	2.285***	0.555***	0.386***
	(0.177)	(0.225)	(0.101)	(0.108)
Recency	−0.173***	−0.193***	−0.147***	−0.141***
	(0.043)	(0.042)	(0.010)	(0.010)
Frequency	0.118***	0.091***	0.001	−0.0004
	(0.007)	(0.007)	(0.002)	(0.002)
Regularity (k)	9.606***	3.771*	0.573	0.345
	(1.988)	(1.973)	(0.460)	(0.477)
Routine ($w = 38$)		5.546***		1.072***
		(0.448)		(0.387)
Observations	2,000	2,000	2,000	2,000
R^2	0.535	0.568		
Note:			*p<0.1; **p<0.05; ***p<0.01 Intercept omitted for clarity.	

To compare routineness and regularity, we estimated the Pareto-GGG model using `R`'s `BTYDplus` package, and extracted the k_i parameter for each individual. We estimated the model using hourly data, with the same training window as in the full data. We find that regularity is positively correlated with routineness ($r = 0.43$), suggesting that, in many cases, routine customers are also more regular. We then re-estimated our focal CRM regressions, explaining future transactions and activity with customer-level summaries, but now including the regularity parameter k_i as one of those predictors. The results are shown in Table 6. We find that, although regularity is, on its own, predictive of the number of future requests, when included alongside routineness, it is no longer a significant predictor, suggesting that routineness captures the important information about customer timing regularity, at least with respect to future customer outcomes. Finally, comparing the $R^2 = 0.535$ of Model 1 in Table 6 to the $R^2 = 0.567$ of Model 1 in Table 3, we observe that routineness does a somewhat better job explaining variability in the number of sessions than regularity.

7. Discussion

Our work makes two primary contributions: first, from a methodological point of view, to the best of our knowledge, this is the first paper to model customer routineness at the individual customer level. To do so, we leverage a Bayesian nonparametric Gaussian process with a unique kernel structure aimed at estimating temporal routines, nested within a Poisson process framework. This model is able to flexibly capture varying patterns of routines across customers with high level of accuracy. Additionally, it yields a customer-level decomposition of usage into a part that is routine and a part that is random, allowing to quantify the degree of a customer's routineness. Substantively, we apply the model to data from Via, a ride-sharing company, and show that we are indeed able to capture interesting, customer-level routines. We show that our model-based routineness metric is strongly predictive of customer value, insofar as it is a positive and significant predictor of both future usage and retention. Moreover, this effect is robust, even over a long time horizon, and after controlling for the *level* of usage and other typical CRM controls. Said differently, this result is noteworthy because it suggests that the temporal shape of usage matters: highly structured usage is more valuable than random usage. While we apply our model in the context of ride-sharing, the model we propose is general, and can be applied to usage or purchase data in many business settings. We encourage future research to leverage our model to explore the effect of routineness on customer behavior and customer management in other settings.

Beyond our focal focus on the relationship between routineness and customer value, we also present other results that both validate routineness as a construct, and establish its more wide-ranging importance in customer management. We show that routine customers are better customers in ways that stretch beyond just lifetime value: they appear to be generally less price sensitive, and more robust to some types of service disruptions. Conceptually, we differentiate temporal routineness from consistency of choice (i.e., the "what" of purchases), and a previously defined metric of regularity of transactions. In both cases, we show that temporal routines are more important predictors of long-run customer value. We encourage future research to further explore the similarities and differences between the related concepts of habit, regularity and routineness. Finally, more specific to our ride-sharing application, we show suggestive results about which customers and trips are more likely to be routine: we find that customers with average shorter driver arrival times and fewer passengers in the ride-share with them tended to be more routine. However, these patterns are merely suggestive: our analysis is not causal, and thus cannot established whether these are indeed drivers of routines. We leave for future research to explore more causally what are the drivers of routine behavior and how it emerges. In short, our research provides a first step in measuring and understanding the impact of routines in customers' behavior. We hope that our modeling effort and results will encourage future research in this promising area.

References

Aarts, H. and Dijksterhuis, A. (2000). Habits as knowledge structures: Automaticity in goal-directed behavior. *Journal of Personality and Social Psychology*, 78(1):53–63.

Ascarza, E. and Hardie, B. G. S. (2013). A Joint Model of Usage and Churn in Contractual Settings. *Marketing Science*, 32(February):570–590.

Ascarza, E., Iyengar, R., and Schleicher, M. (2016). The perils of proactive churn prevention using plan recommendations: Evidence from a field experiment. *Journal of Marketing Research*, 53(1):46–60.

Ascarza, E., Neslin, S. A., Netzer, O., Anderson, Z., Fader, P. S., Gupta, S., Hardie, B. G., Lemmens, A., Libai, B., Neal, D., et al. (2018). In pursuit of enhanced customer retention management: Review, key issues, and future directions. *Customer Needs and Solutions*, 5(1):65–81.

Barnard, J., McCulloch, R., and Meng, X.-L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, pages 1281–1311.

Blattberg, R Neslin, S., , and Kim, B. (2008). Database marketing: Analyzing and managing customers.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of statistical software*, 76(1):1–32.

Dew, R. and Ansari, A. (2018). Bayesian nonparametric customer base analysis with model-based visualizations. *Marketing Science*, 37(2).

Dew, R., Ansari, A., and Li, Y. (2020). Modeling Dynamic Heterogeneity Using Gaussian Processes. *Journal of Marketing Research*, 57(1).

Drolet, A. and Wood, W. (2017). Introduction to special issue: The habit-driven consumer. *Journal of the Association for Consumer Research*, 2(3):275–278.

Du, R. Y. and Kamakura, W. A. (2012). Quantitative trendspotting. *Journal of Marketing Research*, 49(4):514–536.

Du, R. Y., Netzer, O., Schweidel, D. A., and Mitra, D. (2021). Capturing marketing information to fuel growth. *Journal of Marketing*, 85(1):163–183.

Durrett, R. (2008). *Probability models for DNA sequence evolution*, volume 2. Springer.

Dynan, K. E. (2000). Habit formation in consumer preferences: Evidence from panel data. *American Economic Review*, 90(3):391–406.

Ehrenberg, A. S. C. and Goodhardt, G. J. (1968). A Comparison of American and British Repeat-Buying Habits. *Journal of Marketing Reserach*, V(February):29–33.

Fader, P. S., Hardie, B. G., and Lee, K. L. (2005). Rfm and clv: Using iso-value curves for customer base analysis. *Journal of marketing research*, 42(4):415–430.

Fuhrer, J. C. (2000). Habit formation in consumption and its implications for monetary-policy models. *American Economic Review*, 90(3):367–390.

Gopalan, P., Hofman, J. M., and Blei, D. M. (2015). Scalable recommendation with hierarchical poisson factorization. In *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence*, pages 326–335.

Guadagni, P. M. and Little, J. D. (1983). A logit model of brand choice calibrated on scanner data. *Marketing science*, 2(3):203–238.

Ho, T.-H., Park, Y.-H., and Zhou, Y.-P. (2006). Incorporating satisfaction into customer value analysis: Optimal investment in lifetime value. *Marketing Science*, 25(3):260–277.

James, W. (1890). *The Principles of Psychology*. Holt and Macmillan.

Kim, J. G., Menzefricke, U., and Feinberg, F. M. (2005). Modeling parametric evolution in a random utility framework. *Journal of Business & Economic Statistics*, 23(3):282–294.

Kingman, J. F. C. (1992). *Poisson processes*, volume 3. Clarendon Press.

Lewandowski, D., Kurowicka, D., and Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of multivariate analysis*, 100(9):1989–2001.

Loper, J., Blei, D., Cunningham, J. P., and Paninski, L. (2020). General linear-time inference for gaussian processes on one dimension. *arXiv preprint arXiv:2003.05554*.

Møller, J., Syversveen, A. R., and Waagepetersen, R. P. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, 25(3):451–482.

Neal, D. T., Wood, W., Labrecque, J. S., and Lally, P. (2012). How do habits guide behavior? Perceived and actual triggers of habits in daily life. *Journal of Experimental Social Psychology*, 48(2):492–498.

Neslin, S. A., Gupta, S., Kamakura, W., Lu, J., and Mason, C. H. (2006). Defection detection: Measuring and understanding the predictive accuracy of customer churn models. *Journal of marketing research*, 43(2):204–211.

Neslin, S. A., Taylor, G. A., Grantham, K. D., and McNeil, K. R. (2013). Overcoming the "recency trap" in customer relationship management. *Journal of the Academy of Marketing Science*, 41(3):320–337.

Ouellette, J. A. and Wood, W. (1998). Habit and Intention in Everyday Life: The Multiple Processes by Which Past Behavior Predicts Future Behavior. *Psychological Bulletin*, 124(1):54–74.

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

Rasmussen, E. and Williams, K. I. (2006). Gaussian processes for machine learning. *MIT Press*, page 248.

Schmittlein, D. C., Morrison, D. G., and Colombo, R. (1987). Counting Your Customers: Who-Are They and What Will They Do Next? *Management Science*, 33(1):1–24.

Shah, D., Kumar, V., and Kim, K. H. (2014). Managing customer profits: The power of habits. *Journal of Marketing Research*, 51(6):726–741.

Tian, L. and Feinberg, F. (2021). Broadening the horizon: Augmenting one-shot field experiments with longitudinal customer data. Working paper.

Verplanken, B., Walker, I., Davis, A., and Jurasek, M. (2008). Context change and travel mode choice: Combining the habit discontinuity and self-activation hypotheses. *Journal of Environmental Psychology*, 28(2):121–127.

White, K., Habib, R., and Hardisty, D. J. (2019). How to SHIFT consumer behaviors to be more sustainable: A literature review and guiding framework. *Journal of Marketing*, 83(3):22–49.

Wood, W., Quinn, J. M., and Kashy, D. A. (2002). Habits in everyday life: Thought, emotion, and action. *Journal of Personality and Social Psychology*, 83(6):1281–1297.

Zhang, Y., Bradlow, E. T., and Small, D. S. (2015). Predicting customer value using clumpiness: From RFM to RFMC. *Marketing Science*, 34(2):195–208.

Zhou, M. and Carin, L. (2013). Negative binomial process count and mixture modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(2):307–320.

8. Web Appendix

8.1. More Simulated Cases

In this section, we present the full set of 15 case studies. The figures are interpreted analogously as the "Decomposition" figures in the main body. Table 7 describes how each simulation was generated, and Figure 14 shows the model-based decomposition for each case. We see, in general, the correct insights are well recovered. The only exception is, in some cases where usage is purely random, the model may attribute some of that random usage to a routine (e.g., cases 1 and 2). Also of note are cases 6 and 7, where the true data generating process was a mix of two routines: one routine in the first half, and a new routine in the second half. We see that even though the model has no mechanism to learn multiple routines it still classifies these customers as fully routine. However, the routine it learns is a mixture of the two, which is a limitation of our model.

Table 7: Simulated cases.
Descriptions of the simulated customers

Case	Label	Simulation procedure
1	Random (High)	Customer makes a request at 5 day-hours each week, randomly sampled each week from the empirical distribution
2	Random (Low)	Customer makes a request at 2 day-hours each week, randomly sampled each week from the empirical distribution
3	Routine (High)	Randomly sample 5 day-hours; customer makes a request at these times, every week
4	Routine (Low)	Randomly sample 2 day-hours; customer makes a request at these times, every week
5	Commuter	Customer rides every weekday at 8 AM, and 5 PM
6	Two Routines (High)	For the first 19 weeks, the customer follows a routine, generated as in Case 1; then the customer abruptly shifts to a new routine for the remaining 19 weeks, redrawing the times at which she requests rides
7	Two Routines (Low)	For the first 19 weeks, the customer follows a routine, generated as in Case 2; then the customer abruptly shifts to a new routine for the remaining 19 weeks, redrawing the times at which she requests rides
8	Random then Routine (High)	For the first 19 weeks, the customer follows the Random (High) procedure; for the last 19 weeks, the customer follows the Routine (High) procedure
9	Random then Routine (Low)	For the first 19 weeks, the customer follows the Random (Low) procedure; for the last 19 weeks, the customer follows the Routine (Low) procedure
10	Routine then Random (High)	For the first 19 weeks, the customer follows the Routine (High) procedure; for the last 19 weeks, the customer follows the Random (High) procedure
11	Routine then Random (Low)	For the first 19 weeks, the customer follows the Routine (High) procedure; for the last 19 weeks, the customer follows the Random (High) procedure
12	Random then Dead (High)	For the first 19 weeks, the customer follows the Random (High) procedure, then stops making requests
13	Random then Dead (Low)	For the first 19 weeks, the customer follows the Random (Low) procedure, then stops making requests
14	Routine then Dead (High)	For the first 19 weeks, the customer follows the Routine (High) procedure, then stops making requests
15	Routine then Dead (Low)	For the first 19 weeks, the customer follows the Routine (Low) procedure, then stops making requests

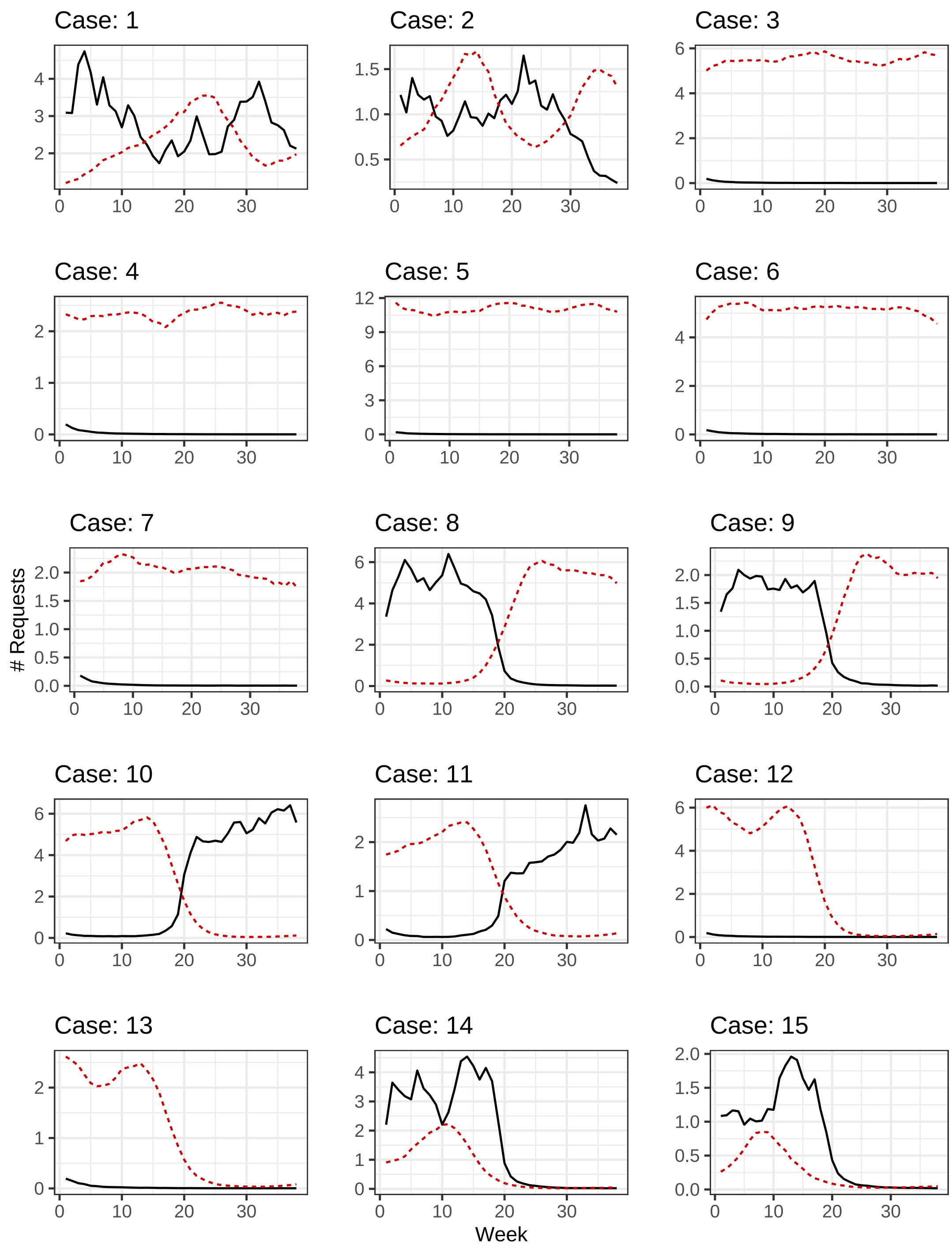

Figure 14: Full simulation results.
The model-based decomposition for all 15 simulated cases, as described in the main body o f the paper and in Table 7. The red dashed line is routine usage, while the black solid line is random usage. We see that, by and large, the model is able to correctly parse the correct data-generating pattern.

8.2. Additional Summary Statistics

Table 8: Additional summary statistics.
Summary statistics for the ride-related covariates. The variables above the horizontal line are variables about the proposal itself; those below the line are about the actual trip that was taken.

Statistic	N	Mean	St. Dev.	Min	Max
Requests Before	38,305	0.9	3.0	0.0	103.0
Ride Distance	38,305	3.3	2.5	0.1	27.9
Airport Ride	38,305	0.03	0.2	0.0	1.0
Solo Trip	38,305	0.9	0.3	0.0	1.0
Had ViaExpress Proposal	38,305	0.4	0.5	0.0	1.0
Had Shared Taxi Proposal	38,305	0.04	0.2	0.0	1.0
Sedan	38,304	0.3	0.4	0.0	1.0
Van	38,304	0.3	0.5	0.0	1.0
Ride Cost (Cents)	38,167	879.1	878.1	0.0	11,936.0
Driver ETA	38,305	7.7	4.0	0.1	70.1
ETA Destination	38,305	31.1	13.5	0.1	168.1
Speed	38,305	0.1	0.7	0.01	131.1
Pickup Walking Dist.	38,305	109.2	83.5	0.0	600.0
# Passengers Request.	38,305	1.2	0.5	1	6
Pickup Delay	16,159	1.1	2.7	−12.2	48.3
Dropoff Delay	16,159	2.6	8.9	−41.4	630.0
Dropoff Walking Dist.	19,374	88.8	70.2	0.0	577.0
# On-board (Pickup)	16,159	0.8	1.0	0.0	5.0
# On-board (Dropoff)	16,159	0.7	1.0	0.0	5.0
Max On-board	16,159	2.5	1.3	1.0	7.0

8.3. Location Regression Results

Table 9: Explaining routineness with location dispersion.
Regression of week 38 routineness (DV) on the two measures of location dispersion, along with some standard customer-level controls (IVs).

	Dependent variable:		
	Routineness		
	(1)	(2)	(3)
Entropy	-0.269^{***}		-0.037
	(0.049)		(0.059)
CRT Disp.		0.052^{***}	-0.065^{***}
		(0.008)	(0.011)
Sessions			0.005^{***}
			(0.0003)
Prob. Ride | Request			0.192^{*}
			(0.104)
First Week			0.001^{*}
			(0.001)
# Requests ($w = 38$)			0.346^{***}
			(0.008)
Observations	2,000	2,000	2,000
R^2	0.015	0.023	0.666
Note:			*p<0.1; **p<0.05; ***p<0.01
			Intercept omitted for clarity.

Table 10: Regressions of location dispersion and customer activity.
Regression of future activity (DVs) on individual-level summary statistics and the two measures of location dispersion.

	Dependent variable:	
	# Requests	Activity
	OLS	*logistic*
	(1)	(2)
Requests ($w = 38$)	2.263***	0.381***
	(0.225)	(0.108)
Recency	$-$0.193***	$-$0.140***
	(0.042)	(0.010)
Frequency	0.093***	0.0002
	(0.009)	(0.003)
Routine ($w = 38$)	5.593***	1.132***
	(0.453)	(0.396)
Entropy	$-$1.753	0.080
	(1.092)	(0.240)
CRT Disp.	0.159	$-$0.011
	(0.199)	(0.046)
Observations	2,000	2,000
R^2	0.568	
Note:	*p<0.1; **p<0.05; ***p<0.01	
	Intercept omitted for clarity.	