Working Paper

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 22, 2020

Abstract

Routines are central to consumer behavior in many industries, including ride-sharing, where consumers may use the same app to take the same trips on a regular basis. While prior work has established the importance of repeat behavior for marketing, little work has been done to understand the implications of routines, which we define as repeated behavior with a distinct, recurring, temporal structure. Partly, this lack of research stems from the statistical problem of estimating routines. In this paper, we propose a new approach to measuring routine usage, which we apply in the context of ride-sharing. Specifically, we model usage of the platform as an individual-level inhomogeneous Poisson point process, where the rate of usage is determined partly by a Bayesian nonparametric Gaussian process. In estimating this rate function, we leverage a unique cyclical kernel structure, that allows for precise estimation of recurrent behavior. We then use this model to estimate individual-level routines in usage of a ride-sharing service. We show that more routine users tend to be more valuable customers, with high individual-level "routineness" being significantly associated with higher future usage and lower churn rates.

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

*Corresponding author. Email: ryandew@wharton.upenn.edu

1. Introduction

Routines are an integral feature of daily life: 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 a routine. Such routines often involve consumption, like the brand of coffee purchased each morning, or the mode of transportation taken to and from work, and are thus of key importance to marketers looking to entrench their products in those routines. Yet, while routines are intuitively important drivers of consumer behavior, prior research has not studied their implications for customer management.

In this work, we build a statistical model to capture routine behavior, which we then link to 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] 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 is in a routine.

While habit formation and its implications have been studied extensively, little research has been done on understanding the impact of routines more specifically. Yet, there are many reasons routines may matter, and why firms may wish to understand their existence, antecedents, and consequences. For instance, from a demand forecasting perspective, knowing a consumer is in a routine can aid firms in making more precise forecasts of demand. In this paper, we focus on another specific implication of routines: customer value. In particular, we hypothesize that customers who use a product as part of their routine are higher value customers, insofar as they may (1) consume the product more often, and (2) may be less likely to churn than non-routine customers. We hypothesize that the effect of a routine exists over and above the effect of mere

[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 we will not capture by focusing on weekly routines.

habitual, but non-routine usage, as is already captured in many existing CRM models.

The statistical model we develop allows us to isolate the share of consumption of a focal product that can be attributed to a routine. 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, such that certain days will exhibit similarities in consumption, and that consumption within a day exhibits 24-hour cycle, to precisely estimate individual-specific variation in recurring behavior. Based on the routine component of the rate, we constructe an individual-specific "routineness" metric that measures to what degree an individual's behavior is structured around a routine.

We apply our model and routineness metric to data from 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 show that, as hypothesized, users who are more routine in their behavior, as captured by our routineness metric, are also more valuable, in terms of both higher future usage rates and lower propensity to churn, even after controlling for overall usage. The effect is robust to numerous specifications. Having established the value of routineness in customer behavior, we then explore how early and with what degree of certainty a firm can identify the emergence of a routine, and what factors lead customers to develop and exit routines.

The rest of the paper is organized as follows: in Section 2, we discuss the prior literature on habits, routines, and the intersection of habits and CRM. In Section 3, we describe our model for capturing and measuring customer routines. In Section 4, we describe the ride-sharing data we use to test our model. In Section 5, we discuss some simulations that validate our model. Finally, in Sections 6 - 7, we describe the results of applying our model and routineness metric to data. We conclude in Section 8 with discussion and directions for future research.

2. Literature

While routines have been relatively understudied, 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 habits to simply indicate repeatedly buying the same product or repeatedly buying from the same company, without suggesting 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 central to to other important marketing constructs, including brand and store loyalty and brand inertia, 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 significantly 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 without conscious thought (Ouellette and Wood, 1998; Wood et al., 2002), and often in goal-directed behavior (Aarts and Dijksterhuis, 2000), or triggered by contextual cues (Neal et al., 2012). Of particular relevance for our empirical application of ride-sharing, habits have recently come under scrutiny 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).

In this research, we move beyond studying repeated behavior or habits, to studying specifically routines. We show that beyond just repeatedly purchasing, the *structure* of when customers interact with the firm also matters for predicting long-run customer value. In some sense, routines are a specific type of habits, where a habitual behavior is built into a temporal structure. Thus, many of the predictions about habitual behavior carry over to routines: we suspect, for instance, that routines reflect automatic choices, and will thus be more difficult to break, leading to sticky long-run behavior. However, we argue that routines are a special kind of habit, exactly because they structure consumers' daily lives, and thus a customer who is *routinely* consuming a focal product or service will be even more valuable than one who is merely habitually (i.e., repeatedly) consuming the product.

In incorporating additional information about timing into a model of customer value, our work expands a growing literature on extending the traditional RFM framework 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. Routineness can be viewed as a specific form of regularity, where the regular transactions happen with a specific temporal structure.

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 (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 been used as the aggregate-level to model customer base dynamics (Dew and Ansari, 2018), and at the individual-level to model dynamics in brand choice (Dew et al., 2020).

3. Model

In this section, we specify a model of weekly usage that captures routines, which yields a natural metric for how routine a customer's behavior is. We then apply this model to ride-sharing data in Sections 4 - 7. By "usage," we mean the consumer interacting with the firm in some way. In our application, the dependent variable will be requesting rides (i.e., using the service), but our model can be applied in any context where the focal outcome is the incidence of a behavior (e.g., making a purchase with the firm, visiting the firm's website). Specifically, we consider a dependent variable T_{ij} which indicates the time that user i interacted for the jth time with the firm. We aggregate those time stamps into a count variable, y_{it}, which equals the number of times user i interacts with the firm during time period t. The time intervals we consider are weeks, which we further subdivide into days and hours, such that $t = (w, d, h)$, with w indexing weeks (i.e., $w = 5$ is week 5 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. Finally, we often consider the unit of "day-hours," the combination of days and hours which we denote simply as $j = (d, h)$.

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:

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

The term α_{iw} captures the baseline level of *random* (non-routine) usage for person i in week w. μ_j accounts for differences in expected usage across day-hours, common to everyone. γ_{iw} captures the baseline level of *routine* usage for person i in week w. Finally, η_{ij} captures an individual's routine: the part of usage that is specific to that user, specified over day-hours, consistent over weeks. 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 a sum of two count processes, $y_{it} = y_{it}^{\text{Random}} + y_{it}^{\text{Routine}}$, 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 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 happened due to the user's routine. These metrics will be our focus in Sections 4-7.

3.2. Specifying the Components of the Rates

Our usage model specifies individual-level, time-varying usage as a function of two rates, each with two terms: two baseline terms, α_{iw} and γ_{iw}, and two day-hour rates, μ_j, and η_{ij}. To model α_{iw}, we consider a straightforward autoregressive model, such that:

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

capturing the fact that usage in week w is likely related to usage the previous week, but imposing 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 $\eta_i(j)$, and then model these functions using Gaussian processes.

Gaussian processes (GPs) provide a way of specifying a prior distribution over a space of functions. 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). 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 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 $f(x) \sim \mathcal{GP}(m(x), k(x, x'))$. For a finite, fixed set of inputs, $x = (x_1, \dots, x_N)$, $f(x) \sim \mathcal{GP}(m(x), k(x, x'))$ is equivalent to:

$$f(x_1, \dots, x_N) \sim \mathcal{N}(m(x_1, \dots, x_N), K), \; s.t. \; K_{ij} = k(x_i, x_j). \tag{7}$$

Mathematically, the matrix K is the gram matrix of the kernel $k(x, x')$, evaluated over all inputs, and is called the kernel matrix. Intuitively, a GP provides 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 dependences beteween the inputs 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 its gram matrix be positive definite.

Returning to our specification, we model:

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

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

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

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

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

This kernel captures functions that are assumed to be relatively smooth, with a smoothness parameter ρ (also called the length-scale), and an amplitude parameter σ.

Novel to our model, $k_{\mathrm{DH}}(j, j'; \phi)$ is a day-hour (DH) kernel specified 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 related than other days (e.g., weekends versus weekdays, or work days versus days off). To capture these properties, we fuse a periodic kernel [CITE] with an unstructured estimate of the connectedness of different days of the week, such that:

$$\begin{aligned} k_{\mathrm{DH}}(j, j'; \phi = \{\sigma, \rho, \Omega\}) &= \sigma^2 \Omega_{d,d'} \times k_{\mathrm{Per}}(h, h'; \rho) && (12) \\ &= \sigma^2 \Omega_{d,d'} \times \exp\left\{\frac{1}{2\rho^2} \sin^2\left(\frac{\pi |h - h'|}{24}\right)\right\}, && (13) \end{aligned}$$

where the matrix Ω is a correlation matrix over days of the week. The right-hand side of this product, $k_{\mathrm{Per}}(h, h'; \rho)$, is the periodic variant of the squared exponential kernel, defined with a

24-hour cycle. The kernel matrix implied by our DH kernel is given by,

$$K_{\mathrm{DH}} = \sigma^2 \Omega \otimes K_{\mathrm{Per}}. \tag{14}$$

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 an LKJ prior for correlation matrices, such that $\Omega \sim \mathrm{LKJ}(2)$, which puts a weak prior toward the identity matrix (Barnard et al., 2000).

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 $k_{\mathrm{Per}}(h, h')$, 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 one additional assumption: the baseline term for the routine part of usage, $\gamma_i(w)$, is assumed to evolve 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, and a baseline that is assumed to evolve slowly, capturing the consistent nature of routines. In constrast, 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. To enforce this smoothness in our model, we assume that the length-

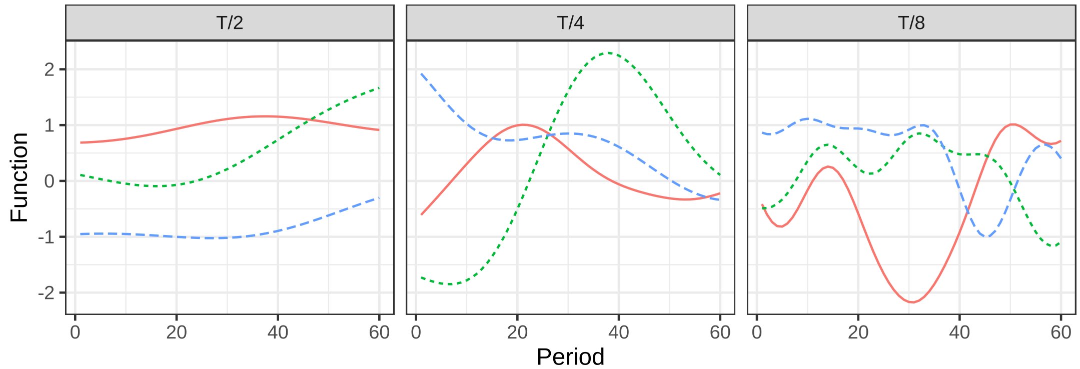

Figure 1: The effect of the length-scale parameter on draws from a GP with the squared exponential kernel. Each panel represents a different length-scale (ρ). Each line is an independent draw from a GP with an SE kernel with that length-scale.

scale parameter, ρ, of $\gamma_i(w)$'s squared exponential kernel is fixed to a relatively high value. Specifically, if T is the total number of time periods in the data, we assume $\rho = T/4$. In Figure 1, we plot draws from a GP with a squared exponential kernel, across several values of ρ, to illustrate this choice. Provided ρ is large enough, we have not found the model to be very sensitive to the specific value. 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.

Finally, since both γ and η are defined at the individual-level, we further assume that $\eta_i(0) = 0$, effectively making the interpretation of this part of the rate relative to Sunday at 12 AM.

3.4. Inference

We estimate the model parameters in a fully Bayesian fashion using NUTS, a gradient-based MCMC sampler, as implemented in Stan. 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, in nearly all cases, customers 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 of our model for all observations from a single customer i can be decomposed into two terms:

$$\log p(y_i \mid \lambda_i) = \sum_{y_{it} \neq 0} y_{it} \log(\lambda_{it}) - \sum_t \lambda_{it} + C, \tag{15}$$

where C does not depend on 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.

4. Application: Ride-sharing

We apply our model to data from a popular NYC-based ride-sharing company. The data contain detailed records on a subsample of customers for whom complete histories were available. Specifically, in our application, we focus on a subsample of 2,000 customers, with data spanning 100 weeks. Of these 100 weeks, we discard the first 30 weeks, which is the period in which our sample of customers was acquired, and in which not many customers are active. Of the remaining 70 weeks, we reserve the final 10 weeks as a holdout period, and train the model on the first 60 weeks.

The dependent variable we focus on is the number of request sessions in a given hour. A request is when a customer opens the ride-sharing app and requests a ride with the company. For this particular ride-sharing service, a request results in a proposal, which the rider can then accept or reject. Occassionally, a rider requests and then rejects the proposal many times, looking for a better proposal. To handle situations like this, the company defines request sessions, by grouping together back-to-back requests as a session. The vast majority of riders have either zero or one request session per hour. To simplify language in the rest of the paper, we will refer to request sessions simply as requests. Summary statistics for our request data are presented in Table 1.

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. As our goal is to capture routines and

Table 1: Summary statistics for our ride-sharing data

# Customers	2,000
# Weeks	60
# Requests	116,627
Mean Reqs / Customer	58.3
Median Reqs / Customer	35
Mean Reqs / Customer / Week	3.1
Median Reqs / Customer / Week	2
Mean Weeks / Customer	19.1

underlying patterns of customers needs, rather than whether the company was actually able to fulfill those needs, requests makes a natural dependent variable.

5. Simulations

As there are many moving components to our model, in this section, we briefly describe some simulated examples of usage patterns, together with their individual-level model estimates. We simulated the usage of these customers following specific, managerially meaningful patterns, to show how our model detects different aspects of customer behavior. To estimate the model parameters, we combine the data from these simulated individuals with a random sample of 100 real customers, to ensure that the population-level parameters are consistent with reality. In the remainder of this section, we highlight a few of these case studies, and we report the results for the remainder in Appendix A.

5.1. Case Study: Routine Customer

In Figure 2, we plot several of the estimated model parameters from a simulated individual with a high-level of routine usage. Specifically, this individual was simulated by first drawing five day-hours at random, then assuming the individual makes one request at these five day-hours every week. There are five panels in Figure 2: (1) at the left/top, "Random Scale" plots the posterior mean of α_{iw}; (2) at the left/middle, "Routine Scale" plots the posterior mean of $\gamma_i(w)$; (3) at the left/bottom, "Decomposition" plots the posterior median of E_{iw}^{Random} and E_{iw}^{Routine}; (4) in the center, the posterior mean of $\eta_i(j)$; and (5) at right, the expected number of

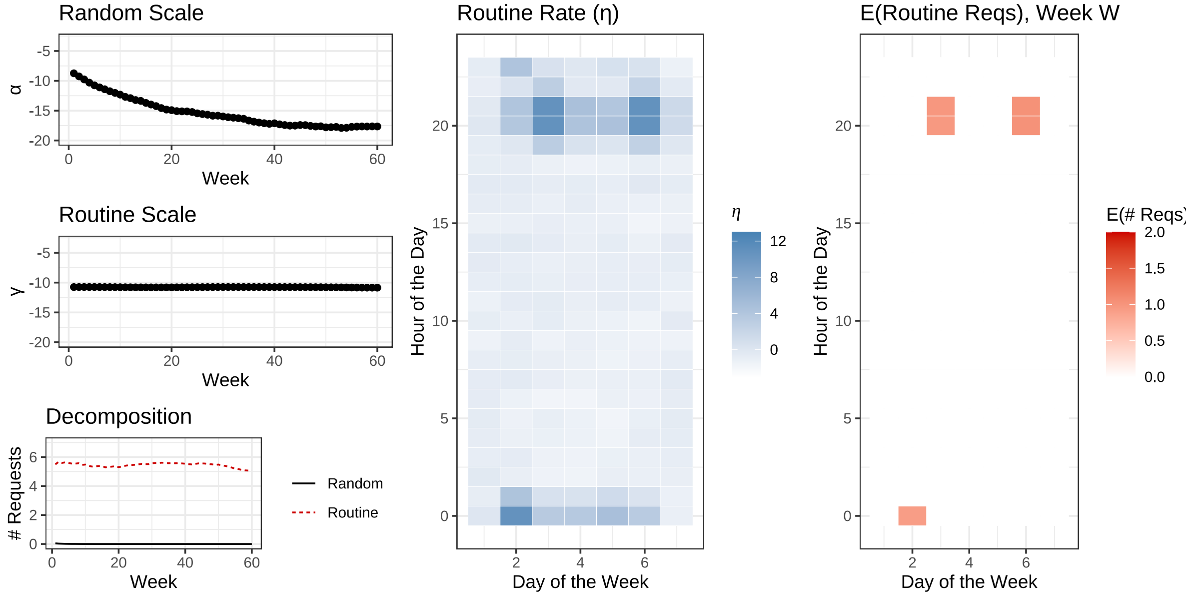

Figure 2: Model estimates for a simulated individual with a high-level of purely routine usage.

requests for each day/hour (i.e., the exponentiated sum of (2) and (4)) during the last week of the data.

From Figure 2, we see the model is correctly able to parse out that this is a routine user: in the Decomposition panel, we see the routine component is capturing roughly five requests per week, each week, while the random requests is essentially zero. When we look at the variation in the baselines, we see α_{iw} always decreasing and strongly negative, while $\gamma_i(w)$ remains flat. The reason α_{iw} decreases is because α_{i1} is pooled toward the population-level starting point; in effect, all of these values of α_{iw} imply essentially no random requests. The routine rate, $\eta_i(j)$ peaks at five different spots: day 2 at hour 0 (Monday at 12 AM), day 3 at hours 20 and 21 (Tuesday at 8 and 9 PM), and day 6 at hours 20 and 21 (Friday at 8 and 9 PM). This, combined with the estimate for $\gamma_i(w)$ yields the expected requests shown in the right panel, corresponding to exactly five requests at those five times each week.

5.2. Case Study: Random Customer

In contrast, in Figure 3, we see the results for a simulated user with no routine at all. For this person, we simulated the data by drawing five random day-hours from the empirical distribution of request times, each week. In the model estimates, we see essentially the opposite

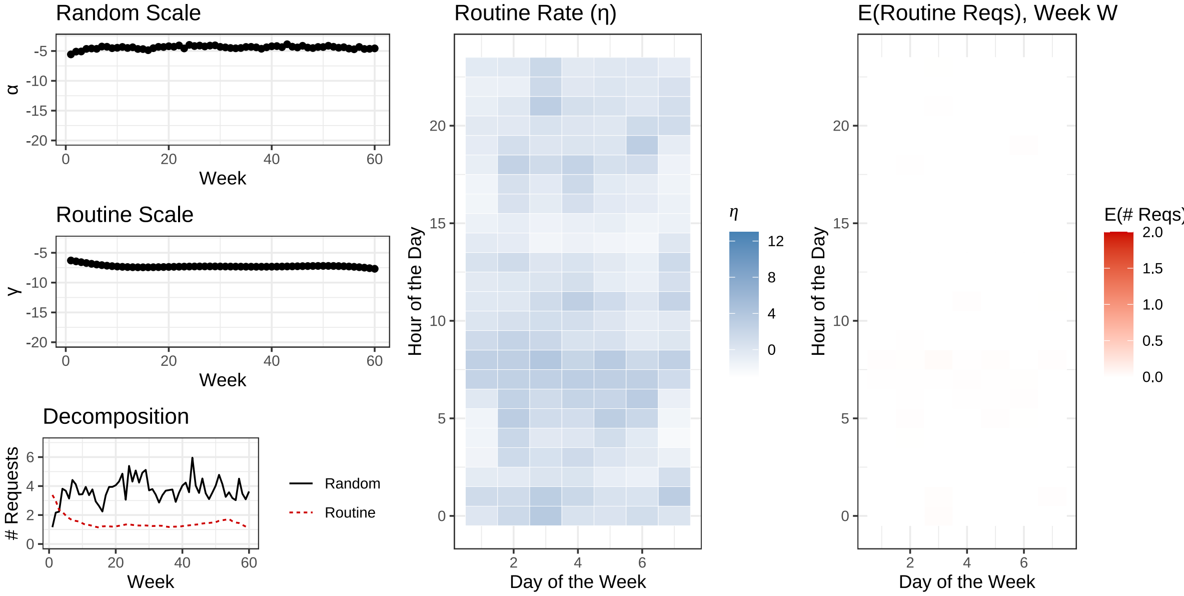

Figure 3: Model estimates for a simulated individual with a high-level of purely random usage.

patterns as the previous case study: the person has a relatively high random scale, α, which is consistent across weeks. The routine scale, γ is again consistent across weeks, but when combined with a very small and diffuse η, implies this user is essentially making no routine requests, as shown both in the decomposition, and in the rightmost panel.

5.3. Case Study: Random then Routine

In Figure 4, we show the ability of the model to detect the emergence of a routine. For this person, the routine emerges in a very clean way: after week 30, the person is immediately in a routine, whereas before week 30, the person is using the service randomly. We see that η captures the five peaks in this person's usage. However, these peaks play no role in the actual expected behavior for this person until week 30, due to the baseline terms α and γ: before week 30, α is high and γ is low, implying essentially zero routine usage. After week 30, they reverse, leading to essentially the same behavior as in the first case study.

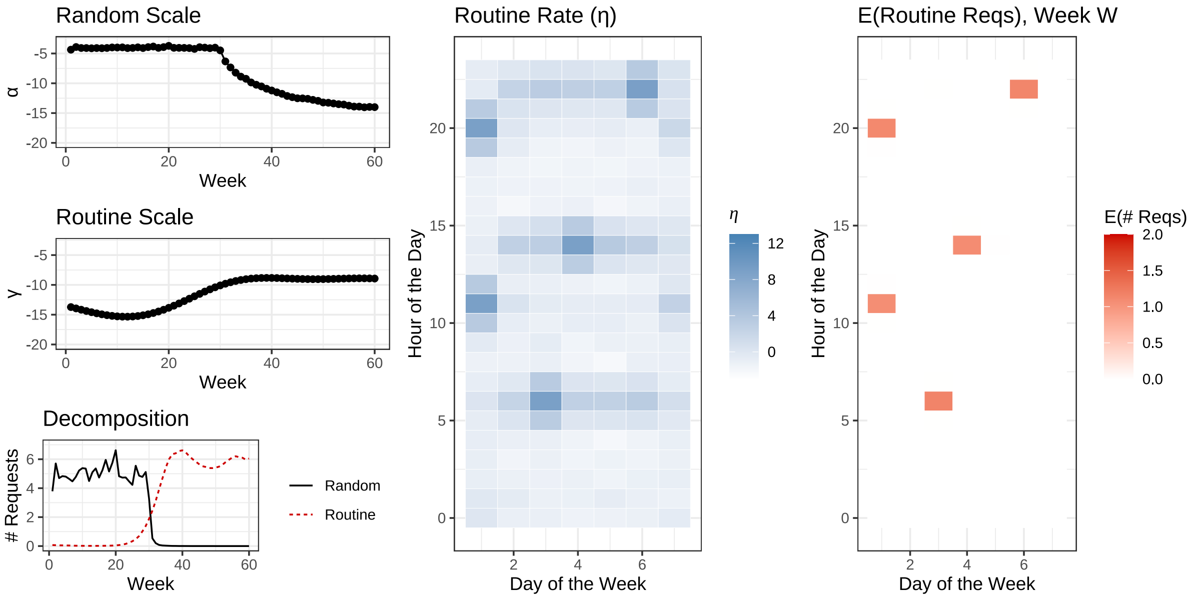

Figure 4: Model estimates for a simulated individual who first uses the service randomly, then switches at week 30 to a routine.

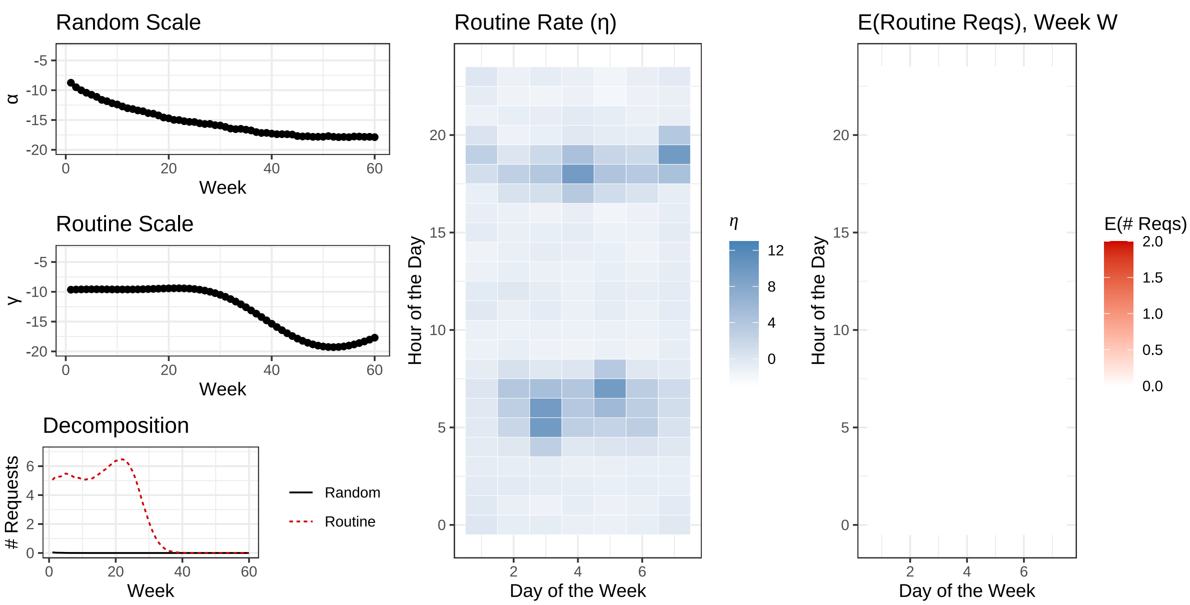

Figure 5: Model estimates for a simulated individual who first uses the service in a routine, then suddenly churns in week 30.

5.4. Case Study: Routine then Churn

Finally, although our model is not explicitly designed to detect churn, churn can be captured in our framework when both baseline terms become very negative, essentially implying zero expected requests. In Figure 5, we plot the model estimates for a simulated customer who uses the service routinely for 30 weeks, then suddenly churns. We see that the model detects this, and the routine scale slowly decays to a very negative value after week 30, leading to zero expected requests at the end of the data.

6. Model Estimates

We now turn to describing the model estimates from the real data, estimated on the full sample of 2,000 customers. In this section, we first describe some of the population-level parameter estimates, characterizing request patterns broadly, before describing some individual case studies and predictive exercises. We then link these model estimates to CRM outcomes in the next section.

6.1. Population Parameters

There are two main population-level parameters of interest: the common population-level rate parameter, $\mu(j)$, which governs when people 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 in routines, a priori. We plot the posterior mean of $\mu(j)$ in Figure 6, and visualize the posterior mean of Ω as a correlation plot in Figure 7.

From these two plots, we see some intuitive patterns emerge: first, from Figure 6, we see the most common times to randomly request a ride are during the commuting hours and the evenings, broadly defined, whereas people do not tend to make requests in the middle of the night (hours 2-5, or 2 AM to 5 AM). This pattern is moderated somewhat on the weekends, where there's a noticeable drop at 4 AM, corresponding to the closing time of many bars in New York City. Similarly, the correlation matrix in Figure 7 captures an intuitive pattern: weekdays tend to be more similar to one another than weekends. Saturday and Sunday are somewhat

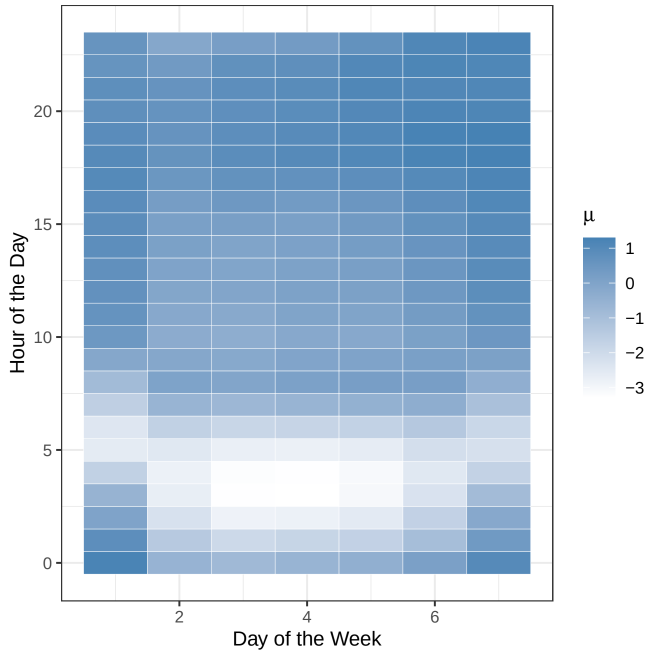

Figure 6: Posterior mean of $\mu(j)$, the common rate of usage across people at the day-hour level.

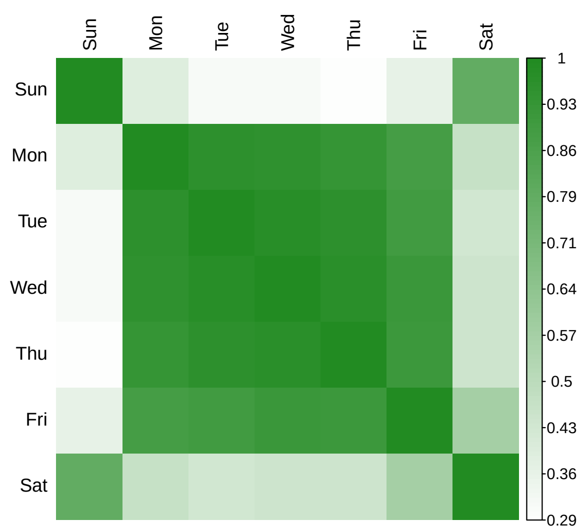

Figure 7: Visualization of the posterior mean of Ω, the correlation matrix across days for routines. Darker colors indicate higher correlation.

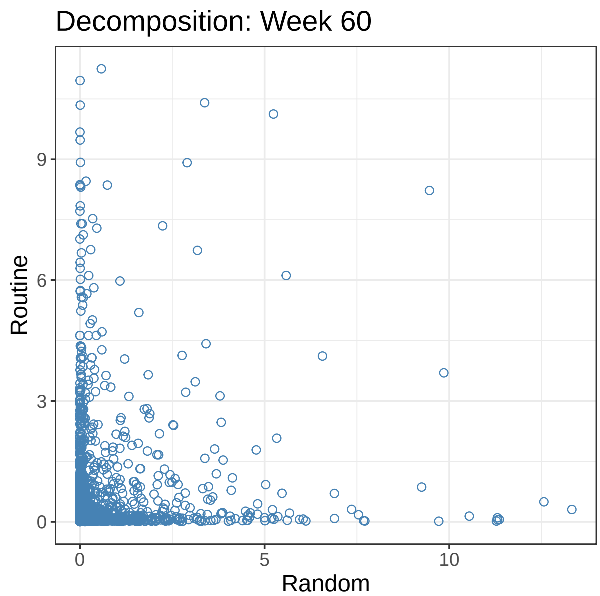

Figure 8: The joint distribution of the posterior medians of E_{iW}^{Routine} and E_{iW}^{Random}, where $W = 60$, the last week of the data.

correlated, as are Friday and Saturday.

6.2. Decomposition

The primary output of our model is the decomposition of usage into routine and random requests. In Figure 8, we show this joint distribution of the two parts of our decomposition, E_{iW}^{Routine} and E_{iW}^{Random}, for all customers, in the last week of our data. While this decomposition evolves week to week as customer behavior changes, the patterns seen in Figure 8 tend to hold generally: we find a typically L-shaped decomposition, where users tend to either be routine, or random, both not necessarily heavy on both components. The vast majority of people fall in the lower left, with few requests per week, balanced between random and routine.

To assess whether this decomposition actually captures the true data patterns, we compute $E(y_{iw}|\lambda_{iw}) = E_{iw}^{\text{Routine}} + E_{iw}^{\text{Random}}$, and plot this against the actual number of requests made in the training data in Figure 9. We see a strong correlation between our model's expectation and reality, reflecting good in-sample fit.

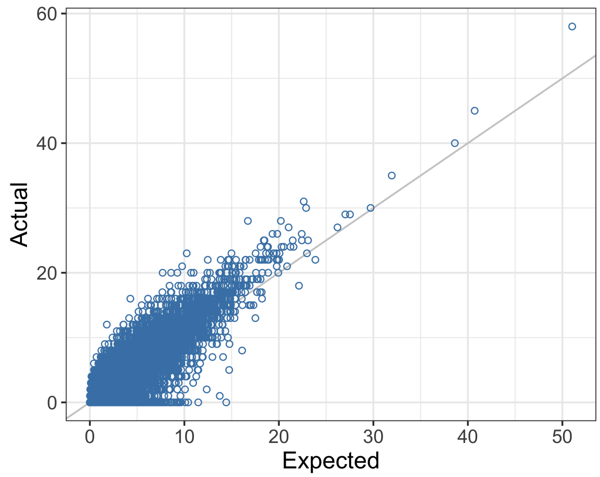

Figure 9: In-sample fit, where expected is $E(y_{iw}|\lambda_{iw}) = E_{iw}^{\text{Routine}} + E_{iw}^{\text{Random}}$, and actual is the actual number of requests a person made.

6.3. Case Studies

Having established the validity of our model both through simulations and in in-sample fit, we now illustrate the insights possible through the model by examining two case studies of real users. Relative to the simulated examples, the results on real users are less clean cut in their interpretation, but still offer valuable customer-level insights for managers.

In Figure 10, we show the same posterior estimates and decompositions as we did in Figures 2 - 5, but for a representative, actual customer. We term this individual an "occasional commuter," as this person tends to use the service during commuting hours, especially after work in the early part of the week, and on Friday mornings. Note that this person's random usage fluctuates quite a bit week over week.

In contrast, in Figure 11, we show the same plots but for a person with a fairly consistent high level of random usage, and only minimal routine usage. The only consistent aspect of this person's usage is a fairly routine call to the service at 4 AM on Saturdays. For this reason, we label this person the "random partier."

Similar to the random partier, routines may exist at odd hours. Our last case study is displayed in Figure 12, which shows the model estimates for a person with an emerging evening routine: this person's routine usage has grown stronger over time, as indicated by γ. The η plot indicates that this person tends to call rides in the evening hours on weekdays, as well as in the

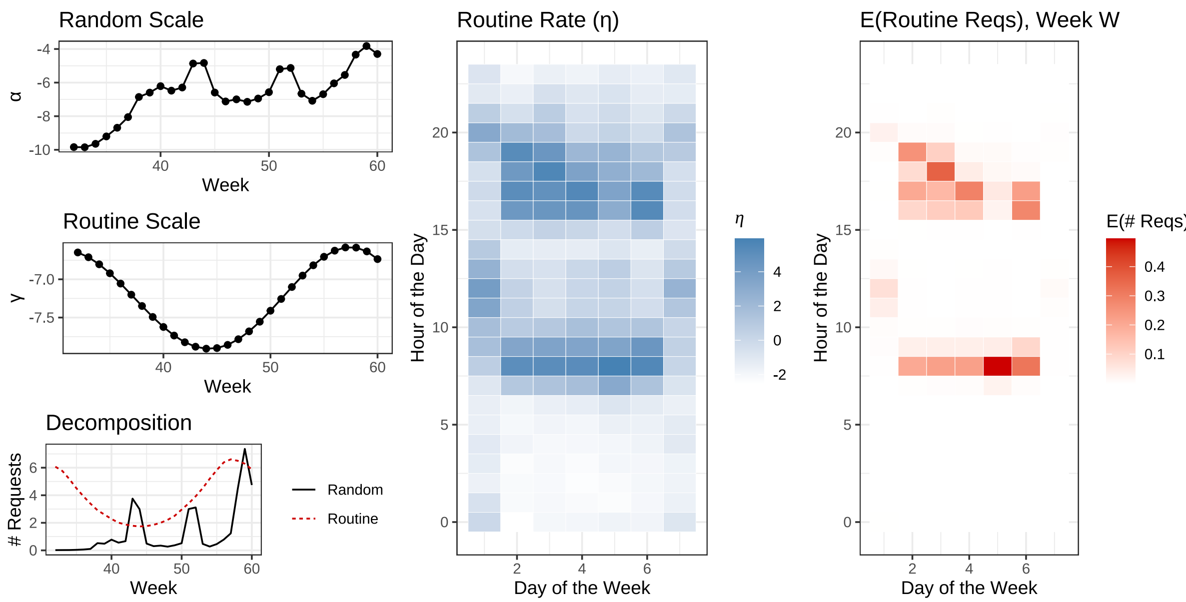

Figure 10: Model estimates for an individual who uses the service in an evolving routine, typically in commuting hours. We label this person the "occasional commuter."

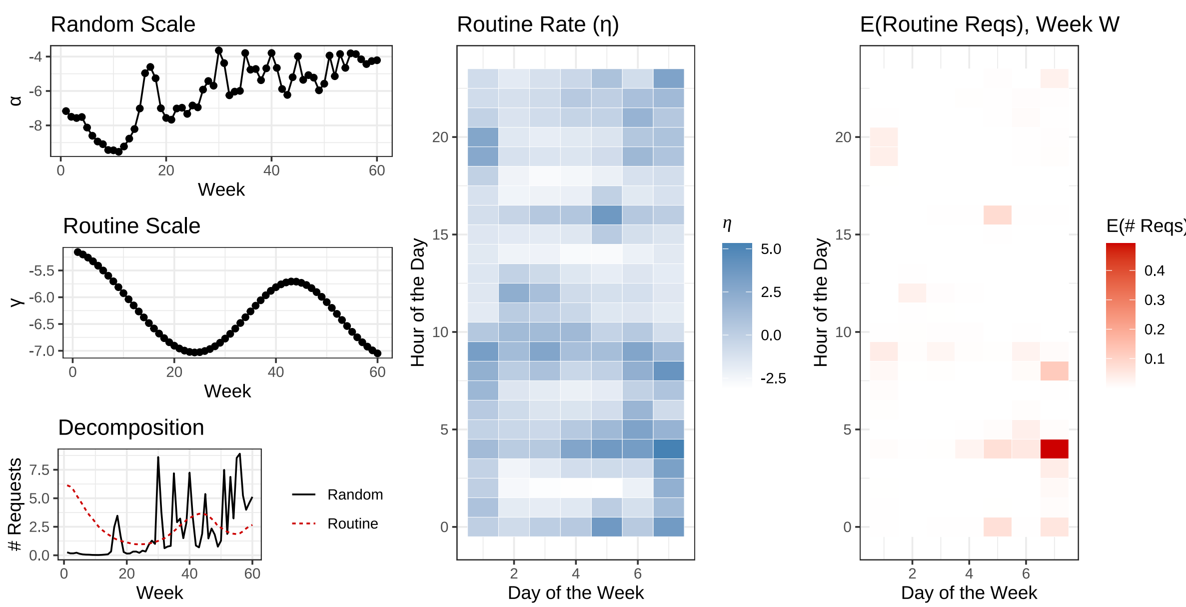

Figure 11: Model estimates for an individual who mostly uses the service at random, except for on Saturday mornings at 4 AM. We label this person the "random partier."

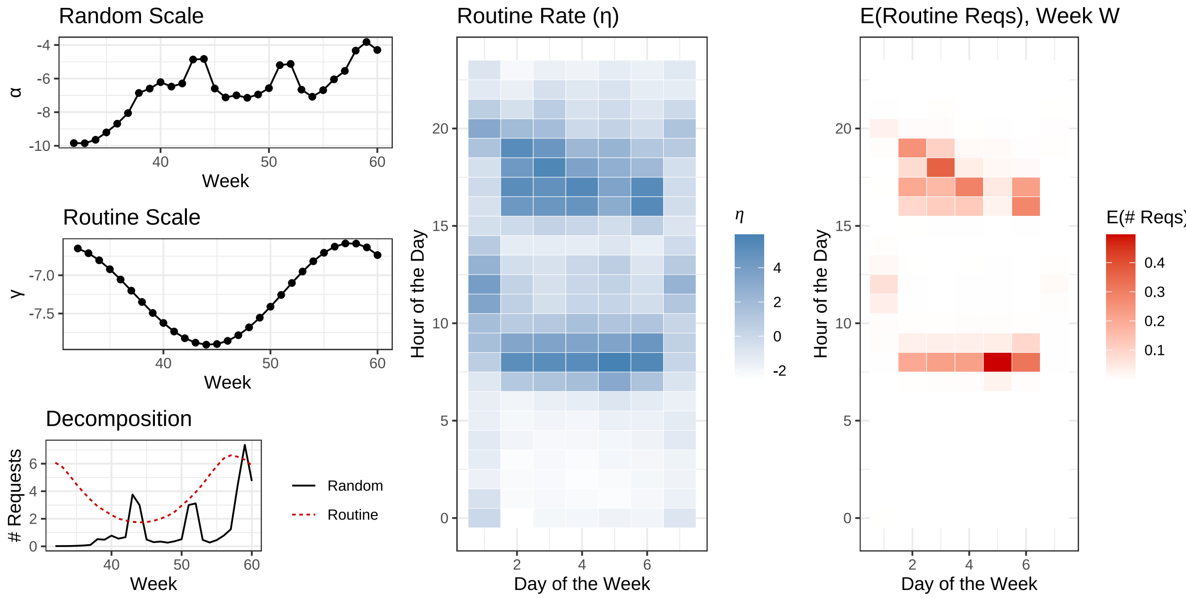

Figure 12: Model estimates for an individual who uses the service in an emerging evening routine. We label this person the "night owl."

early mornings on Saturdays. While the α plot indicates that this user's random usage has also grown somewhat, the decomposition makes clear that this growth is slight: most of this user's increased usage has been within his or her routine.

7. Routineness and CRM

Now, we return to the central question of the paper: having identified a metric for routineness from usage data, is routineness an important predictor of customer value?

In particular, we consider two simple, holdout period dependent variables: the number of requests a customer makes, and whether that customer is active at all. To understand the relationship between routineness, as measured at the end of our training data (E_{iW}^{Routine}), and these measures, we estimate several linear models, controlling for both (1) the number of requests the person made in week W, and (2) classic recency and frequency controls, capturing how recently a person last made a request, and how many requests the person has made previously. For modeling holdout requests as the DV, we use simple OLS; for modeling activity, we use logistic regression.

Before describing the results, it is important to note that routineness is essentially a component of the total requests. Hence, by controlling for the number of requests in measuring the effect of routineness, we are in essence trying to determine whether the *shape* of usage matters. As a high level of usage can be arrived at either through random needs or routine usage, 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.

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

8. Discussion

Our work makes two primary contributions: first, we specify an inhomogeneous Poisson point process to capture customer activity, with a dynamic rate modeled with a Bayesian nonparametric Gaussian process with a unique kernel structure. While ours is not the first to combine Poisson and Gaussian processes, the kernel we develop for this task is new, and tailored to the setting of estimating temporal routines. This model yields a customer-level decomposition of usage into a part that is routine, and a part that is random. Then, in an application to ride-sharing data, we show that this model produces intuitive insights about customer behavior, and more importantly, that the routineness metric distilled from the model results is strongly predictive of customer value. We find that routineness is a positive and significant predictor of

Table 2: Regression results: linking routineness to CRM outcomes.

	Holdout ($t > 60$)		Last Month ($t > 65$)	
DV	Requests	Active	Requests	Active
Model	OLS	Logit	OLS	Logit
	(1)	(2)	(3)	(4)
Routine ($t = 60$)	5.190***	1.515***	2.606***	0.390***
	(0.371)	(0.404)	(0.229)	(0.147)
Requests ($t = 60$)	8.150***	1.796***	3.583***	0.793***
	(0.359)	(0.298)	(0.222)	(0.129)
Recency	−1.234***	−1.129***	−0.551***	−1.134***
	(0.301)	(0.090)	(0.186)	(0.092)
Frequency	2.858***	0.064	1.683***	0.095
	(0.327)	(0.104)	(0.202)	(0.080)
Constant	11.832***	2.047***	5.859***	0.606***
	(0.282)	(0.189)	(0.174)	(0.074)
R^2	0.552		0.425	
Log Likelihood		−783.321		−999.580
Note:		*p<0.1; **p<0.05; ***p<0.01		

both future usage level, and future retention. Moreover, this effect is robust, even over a long time horizon, after controlling for the *level* of usage, and 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 our application is to ride-sharing, the model we propose is general, and can be applied to usage or purchase data in many business settings.

Our research on routines and CRM is on-going. In particular, we are currently studying how firms can cultivate routines, and what may lead customers to abandon their routines. By understanding what factors lead to the formation and dissolution of routines, our results on routineness and customer value can give firms an additional lever to create customer value by fostering the adoption of routines. We are also in the process of studying the role location plays in our ride-sharing data. So far, we have only looked at temporal aspects of routines, yet what customers do may be equally interesting in informing our understanding of their routines. Customers may always use the service at the same time to go to the same places, or they may use the service at the same time to go to many different places. Understanding if, or how these customers differ is an important aspect of continuing study.

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.

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.

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).

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.

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.

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

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.

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.

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.

Appendices

A. More Simulated Cases

In this section, we present the full set of 15 case studies. The figures are interpreted analogously as those in Section 5. Table 3 describes how each simulation was generated.

Table 3: Descriptions of the simulated customers

Case	Label	Simulation procedure
1	Routine (High)	Randomly sample 5 day-hours; customer makes a request at these times, every week
2	Routine (Low)	Randomly sample 2 day-hours; customer makes a request at these times, every week
3	Commuter	Customer rides every weekday at 8 AM, and 5 PM
4	Random (High)	Customer makes a request at 5 day-hours each week, randomly sampled each week from the empirical distribution
5	Random (Low)	Customer makes a request at 2 day-hours each week, randomly sampled each week from the empirical distribution
6	Random then Routine (High)	For the first 30 weeks, the customer follows the Random (High) procedure; for the last 30 weeks, the customer follows the Routine (High) procedure
7	Random then Routine (Low)	For the first 30 weeks, the customer follows the Random (Low) procedure; for the last 30 weeks, the customer follows the Routine (Low) procedure
8	Routine then Random (High)	For the first 30 weeks, the customer follows the Routine (High) procedure; for the last 30 weeks, the customer follows the Random (High) procedure
9	Routine then Random (Low)	For the first 30 weeks, the customer follows the Routine (High) procedure; for the last 30 weeks, the customer follows the Random (High) procedure
10	Random then Dead (High)	For the first 30 weeks, the customer follows the Random (High) procedure, then stops making requests
11	Random then Dead (Low)	For the first 30 weeks, the customer follows the Random (Low) procedure, then stops making requests
12	Routine then Dead (High)	For the first 30 weeks, the customer follows the Routine (High) procedure, then stops making requests
13	Routine then Dead (Low)	For the first 30 weeks, the customer follows the Routine (Low) procedure, then stops making requests
14	Two Routines (High)	For the first 30 weeks, the customer follows a routine, generated as in Case 1; then the customer abruptly shifts to a new routine for the remaining 30 weeks, redrawing the times at which she requests rides
15	Two Routines (Low)	For the first 30 weeks, the customer follows a routine, generated as in Case 2; then the customer abruptly shifts to a new routine for the remaining 30 weeks, redrawing the times at which she requests rides

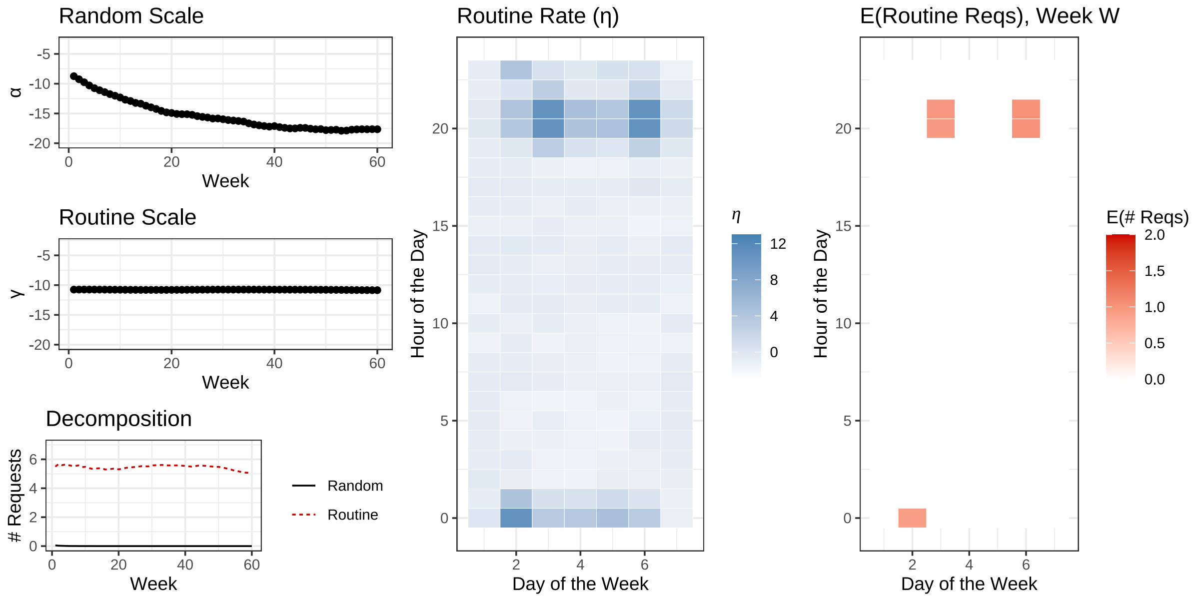

Figure 13: Simulated case 1.

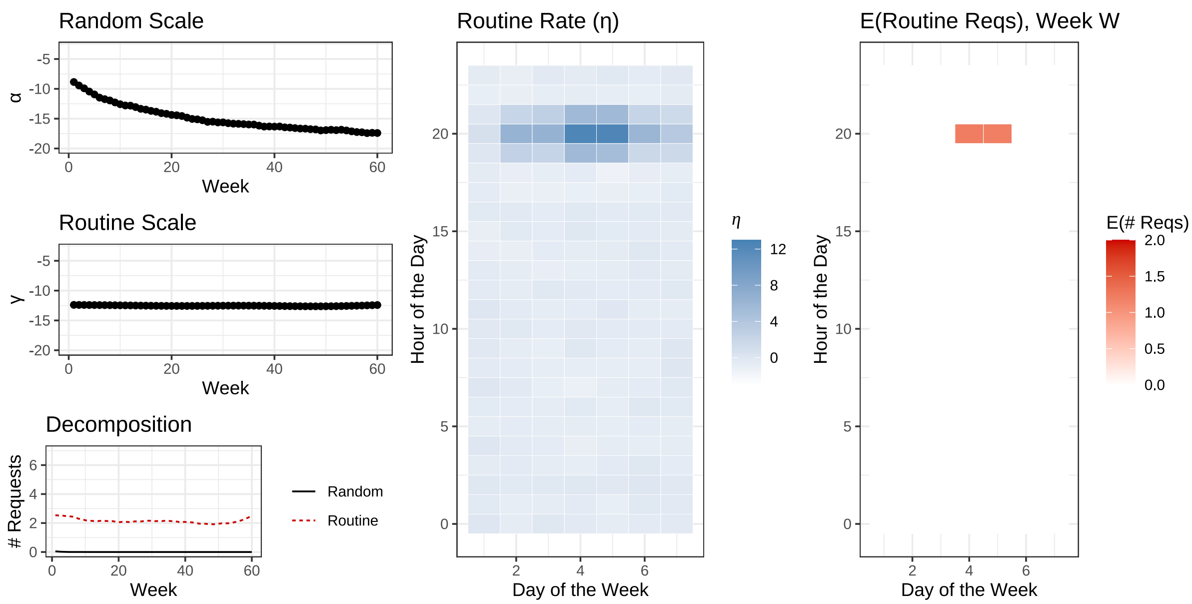

Figure 14: Simulated case 2.

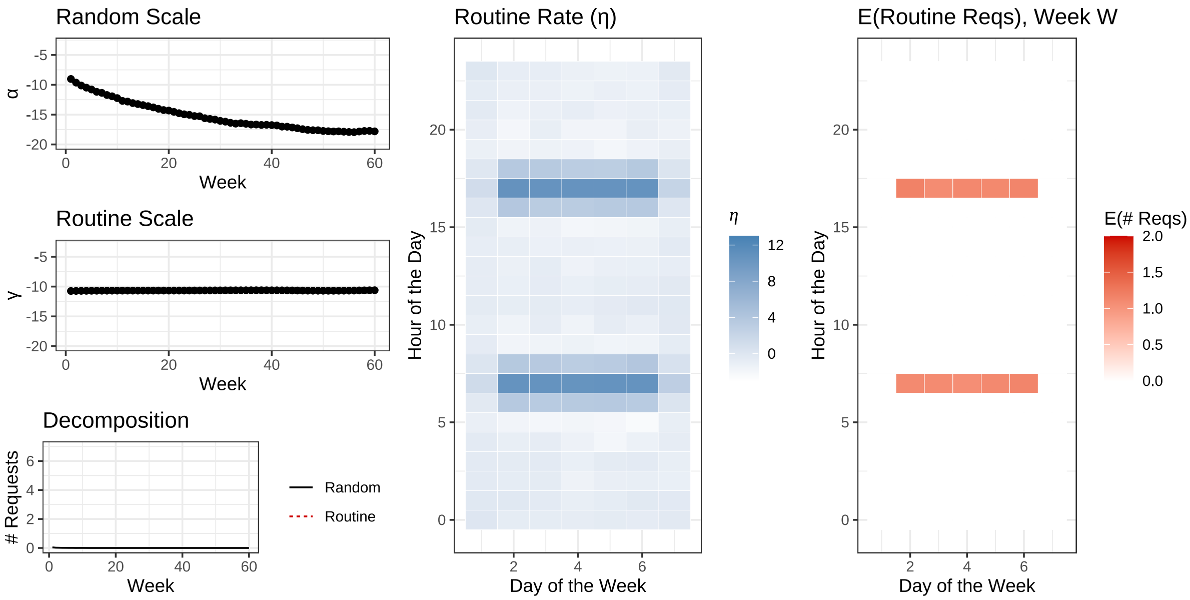

Figure 15: Simulated case 3.

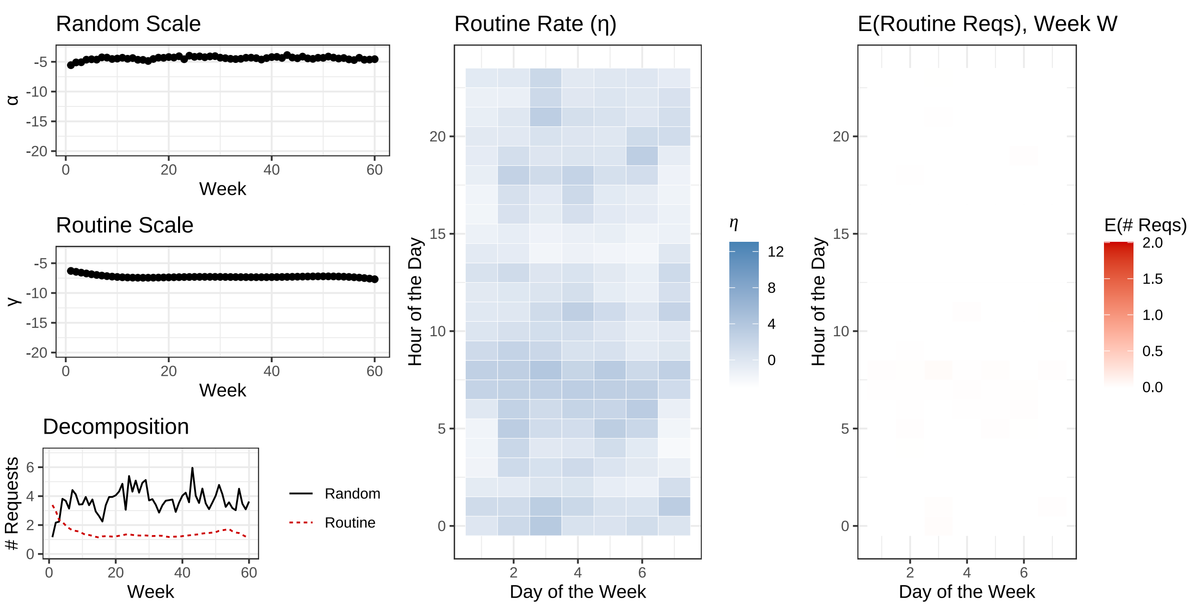

Figure 16: Simulated case 4.

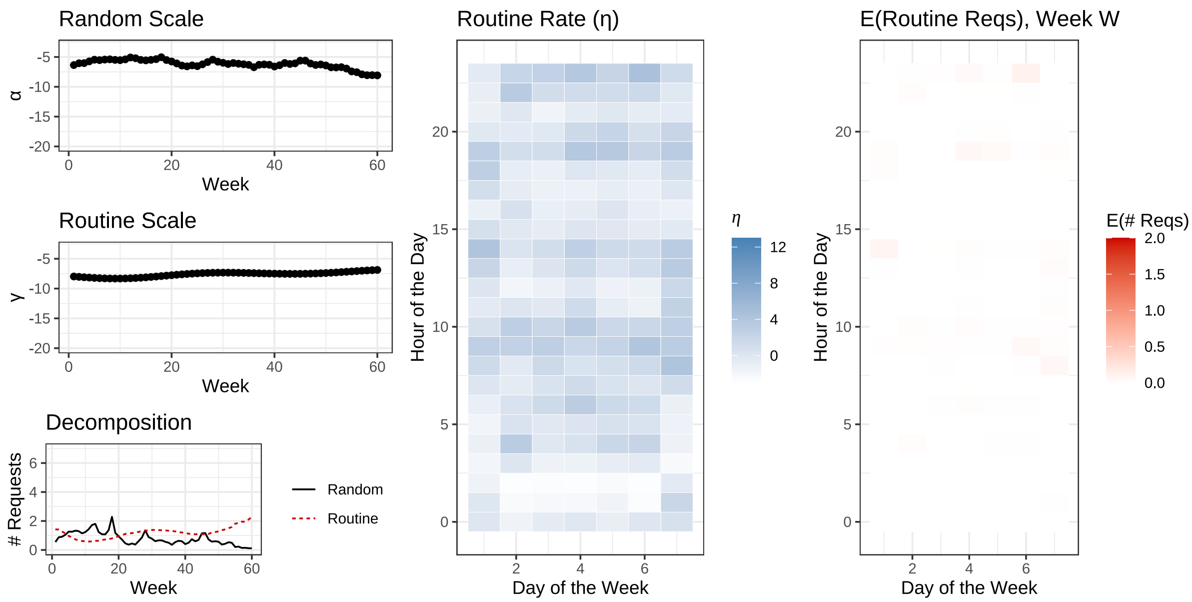

Figure 17: Simulated case 5.

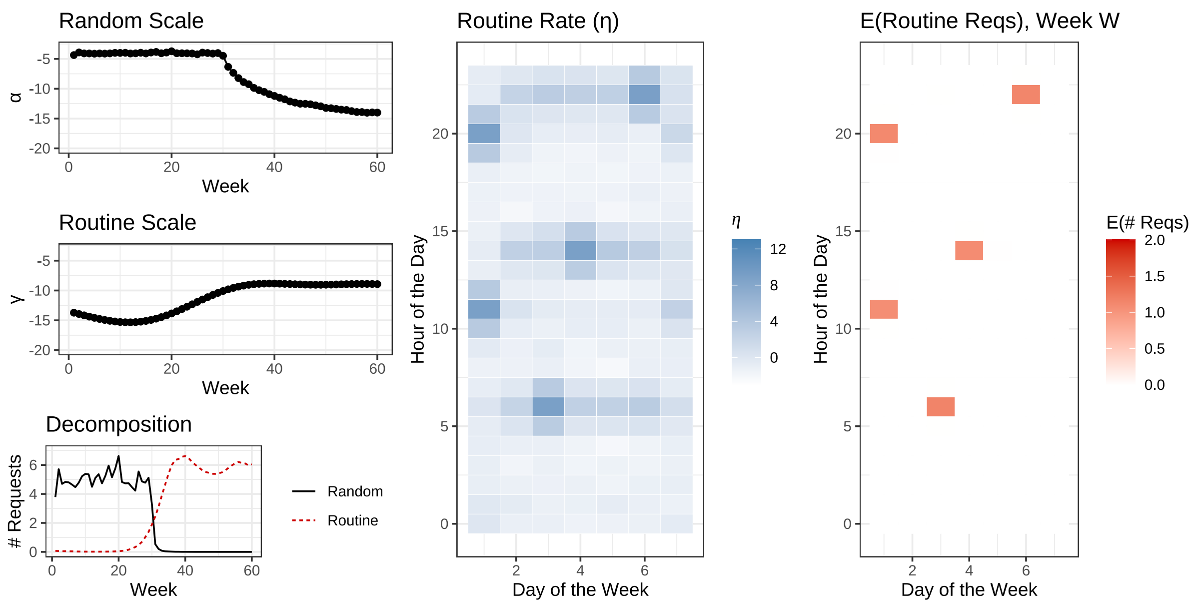

Figure 18: Simulated case 6.

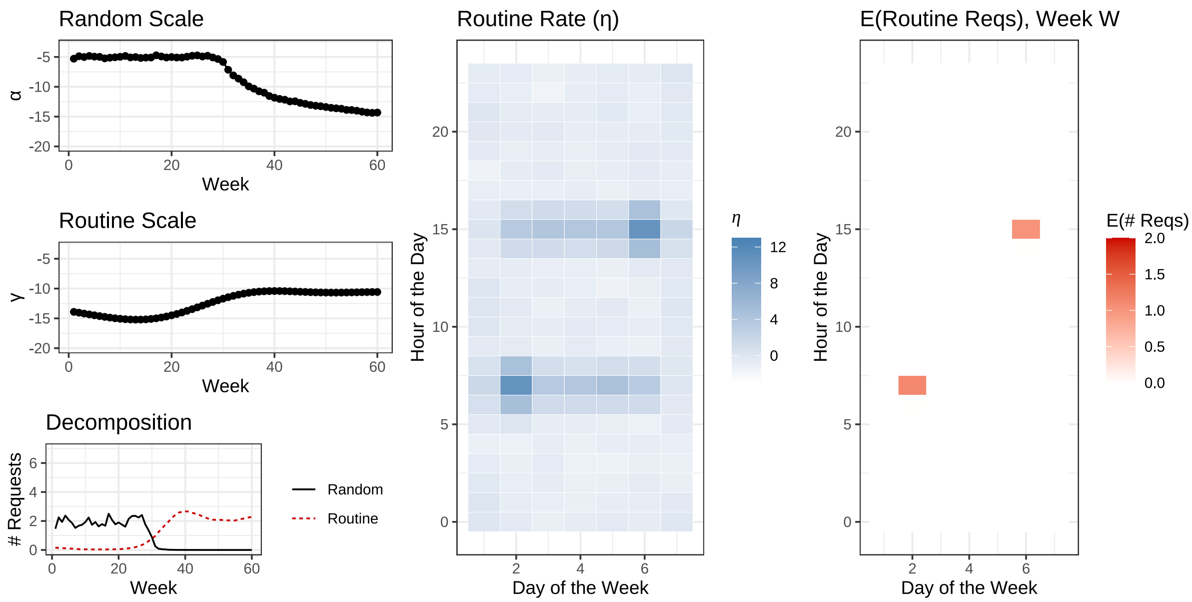

Figure 19: Simulated case 7.

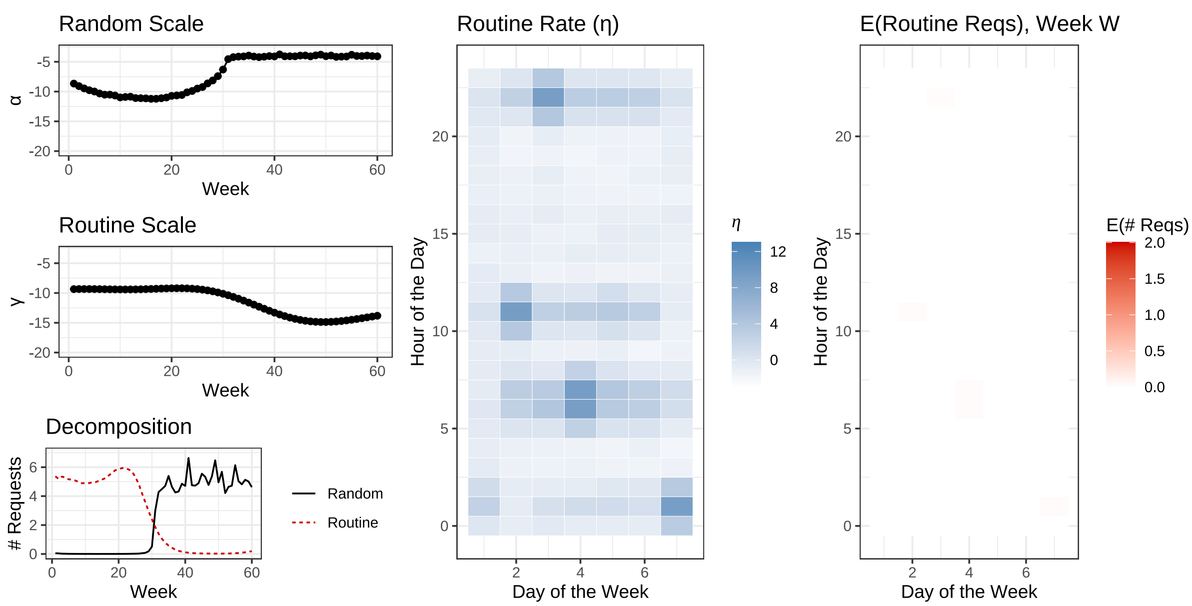

Figure 20: Simulated case 8.

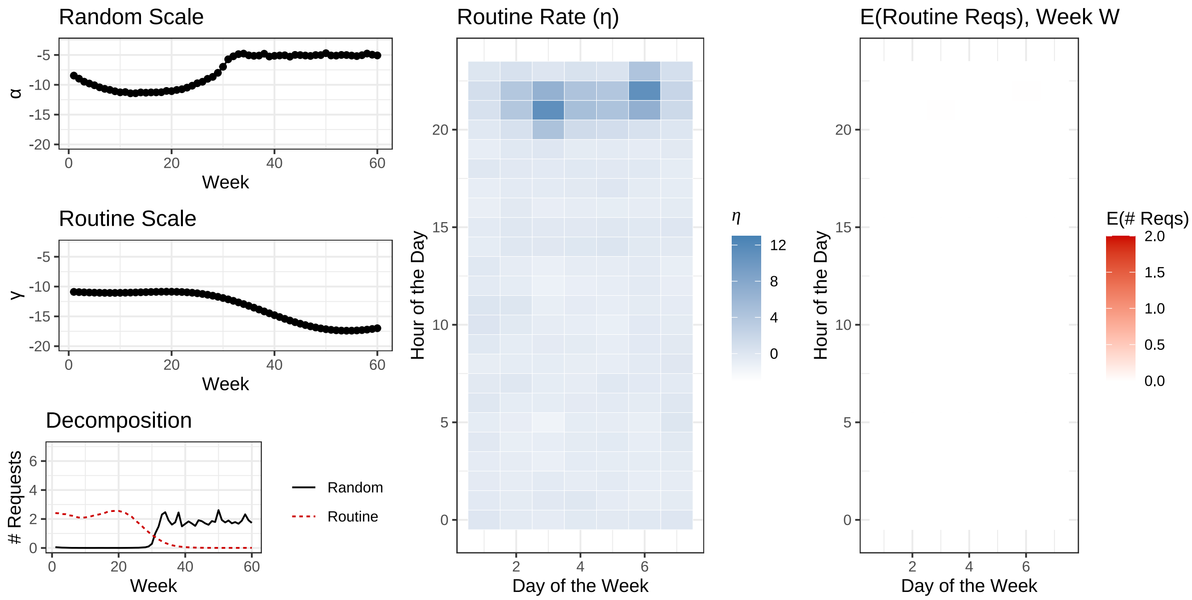

Figure 21: Simulated case 9.

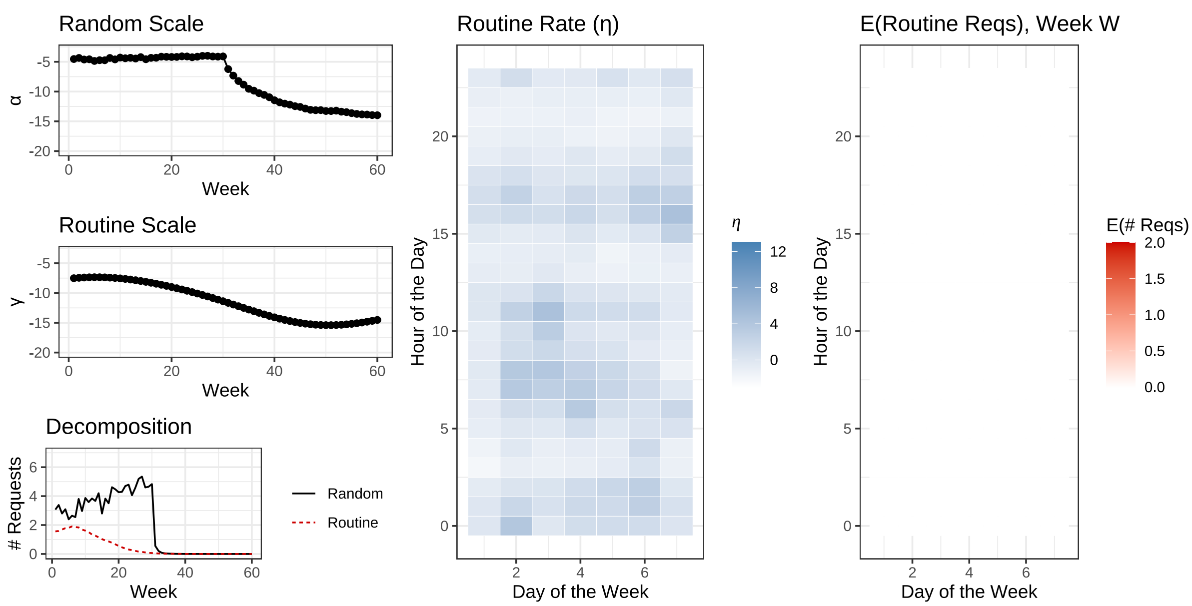

Figure 22: Simulated case 10.

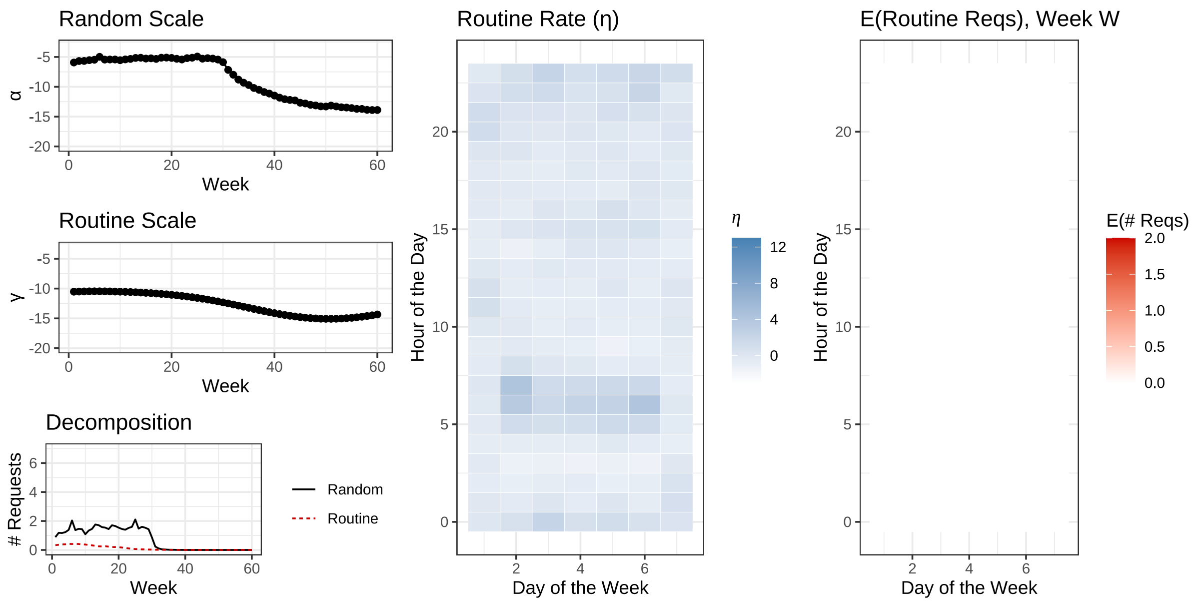

Figure 23: Simulated case 11.

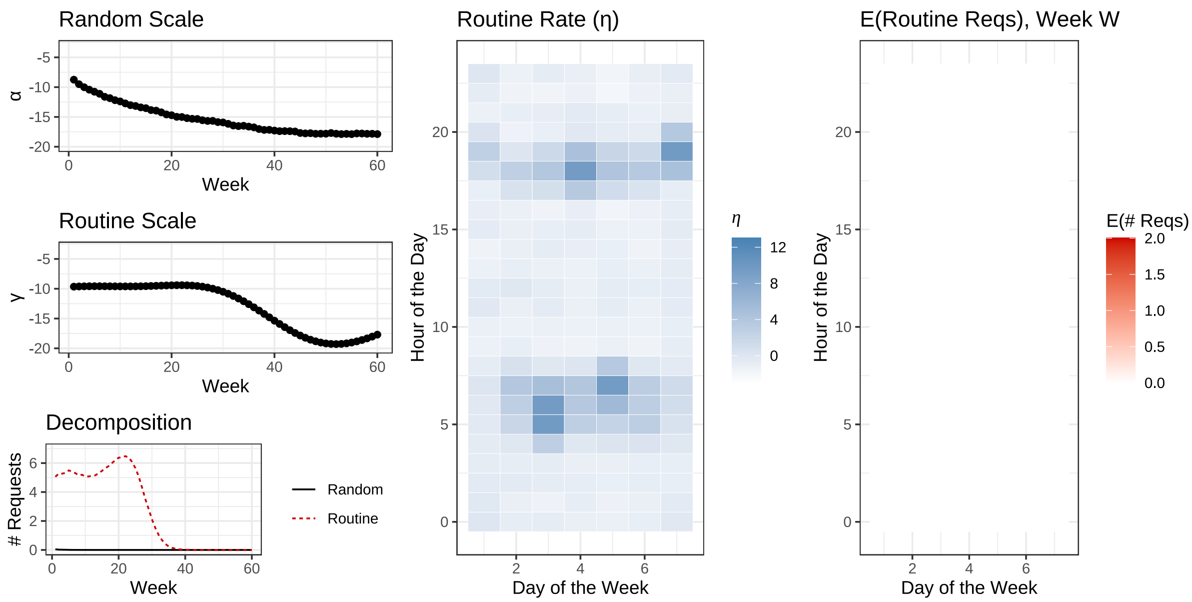

Figure 24: Simulated case 12.

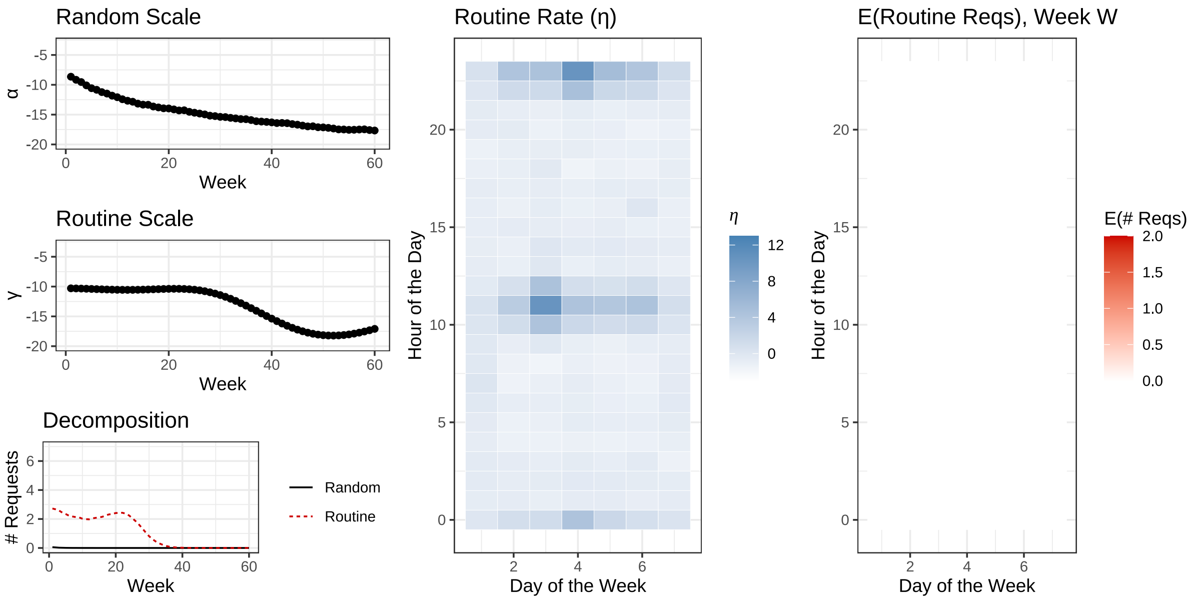

Figure 25: Simulated case 13.

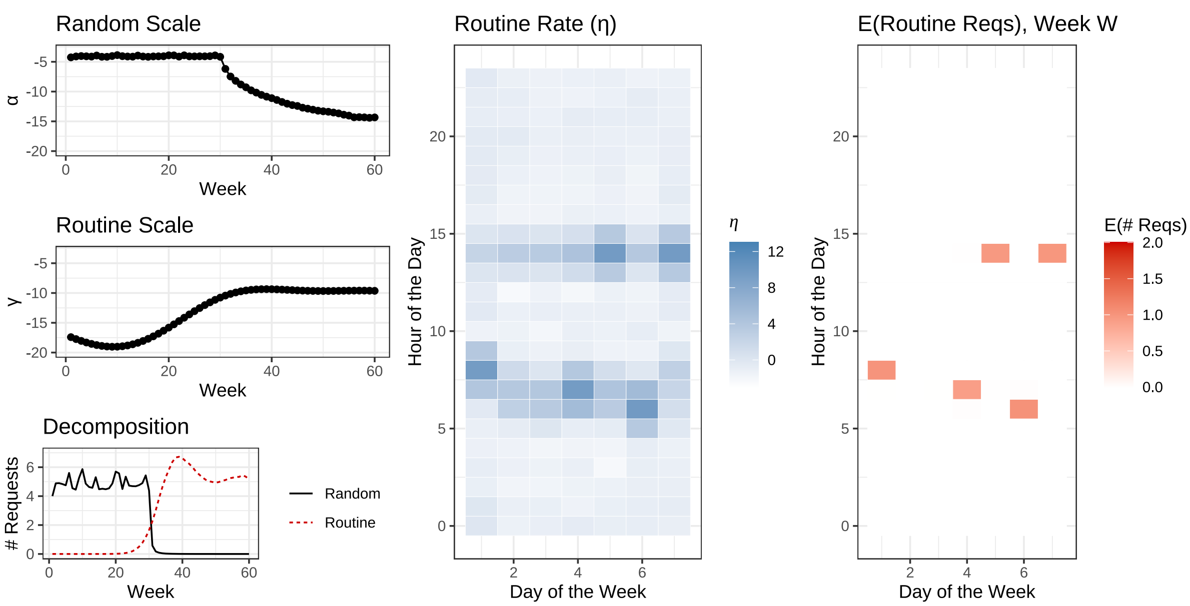

Figure 26: Simulated case 14.

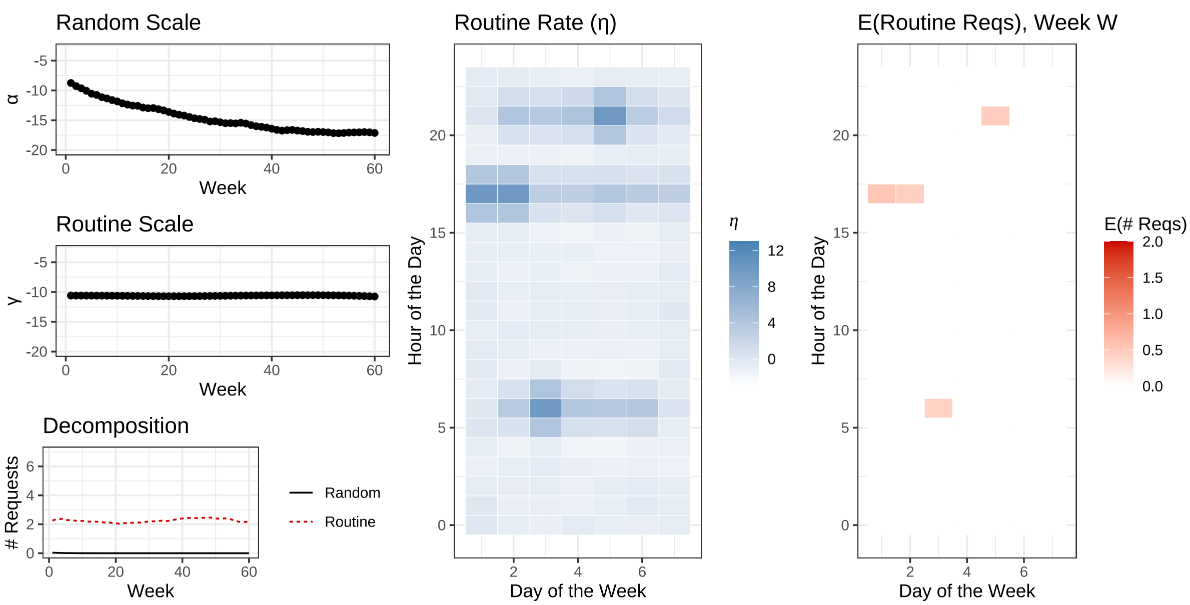

Figure 27: Simulated case 15.