Estimating Millions of Dynamic Timing Patterns in Real-Time

DIANE LAMBERT, JOSÉ C. PINHEIRO, and DON X. SUN*

Abstract

In some business applications, the transaction behavior of each customer is tracked separately with a *customer signature*. A customer's signature for buying behavior, for example, may contain information on the likely place of purchase, value of goods purchased, type of goods purchased, and timing of purchases. The signature may be updated whenever the customer makes a transaction, and, because of storage limitations, the updating may be able to use only the new transaction and the summarized information in the customer's current signature. Standard sequential updating schemes, such as exponentially weighted moving averaging, can be used to update a characteristic that is observed at random, but timing variables like day-of-week are not observed at random, and standard sequential estimates of their distributions can be badly biased. This paper derives a fast, space-efficient sequential estimator for timing distributions that is based on a Poisson model that has periodic rates that may evolve over time. The sequential estimator is a variant of an exponentially weighted moving average. It approximates the posterior mean under a dynamic Poisson timing model and has good asymptotic properties. Simulations show that it also has good finite sample properties. A telecommunications application to a random sample of 2,000 customers shows that the model assumptions are adequate and that the sequential estimator can be useful in practice.

Keywords: Customer profile; Day-of-week data; Event history; Exponentially weighted moving average; Hour-of-day data; Transaction data.

1 Background

Massive databases of customer transactions are common throughout business and industry. Telecommunications providers keep a detailed record of each call placed over their network. Credit card providers keep a

*The authors are in the Statistics Research Department of Bell Labs, Lucent Technologies. They thank Rick Becker for many helpful discussions about an early version of this problem and Fei Chen for help with computations.

detailed record of each charge to a card. Companies record visits to their website. Network managers track login commands on their networks. Online catalogs and stock brokers keep information about each sale. In many contexts, the goal is to track the behavior of each customer in real-time from transaction data.

Building a summary of each customer's behavior from transaction data and updating it with each new transaction that the customer makes is not trivial, especially when there are millions of transactions per day, millions of customers who might make transactions, thousands of new customers each day, thousands of departing customers each day, many variables to track per customer, and many if not most customers change their behavior over time. Probably the simplest kind of summary is simply a rolling history of transactions for each customer. Rolling histories are easy to update, but they have several disadvantages. First, rolling histories over a fixed length of time may hold too little data for infrequent customers and too much data for active customers, while rolling histories over a fixed number of calls may cover too long a period for inactive customers and too short a period for active customers. Second, rolling histories with variable numbers of calls can be difficult to manage in large databases, especially with a changing set of customers. Third, retrieving a large number of calls at the time of a transaction may be too slow to affect the outcome of the transaction. Thus, there is a need for a short, fixed-length summary of a customer's data that captures the important features of the customer's behavior, can be initialized for new customers quickly and reliably, and can be updated sequentially with each transaction that the customer makes, without requiring access to the details of previous transactions.

Relative frequency distributions or histograms are obvious candidates for summarizing customer behavior. They are easily understood by the programmers who maintain databases of customer summaries. They are fixed-length, so the database of summaries is easier to manage. They are nonparametric so they are effective with a highly variable customer base. Histograms are also appropriate for categorical data, discrete data, and discretized continuous data, so type of transaction, timing pattern, place of transaction, and size of duration of the transaction can all be summarized by histograms.

Updating histograms sequentially is not difficult when observations are randomly sampled. If $\widehat{\boldsymbol{\pi}}_n$ is the vector of current histogram probabilities and $\boldsymbol{X}_{n+1}$ is a characteristic of the current transaction, represented as a vector of 0's except for a 1 in the cell that contains the observed value, then the sequentially updated vector of histogram probabilities is

$$\widehat{\boldsymbol{\pi}}_{n+1} = (1 - w_{n+1})\widehat{\boldsymbol{\pi}}_n + w_{n+1}\boldsymbol{X}_{n+1}, \tag{1}$$

where $w_{n+1} = 1/(n+1)$ and $\widehat{\boldsymbol{\pi}}_0 = \mathbf{0}$. Updating thus requires only the most recent transaction, the number of transactions made so far and the current summary. We call this an *unweighted average,* because $\widehat{\boldsymbol{\pi}}_n$ weights each observed transaction $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ equally.

If the customer's behavior changes over time, then a histogram of relative frequencies is inappropriate because recent transactions have no more influence on the histogram than old transactions do. Evolving behavior is tracked better by an *exponentially weighted moving average (EWMA)*. The updated EWMA vector $\widehat{\boldsymbol{\pi}}_{n+1}$ is given by equation (1) with $w_{n+1} = w$ for a fixed weight w, $0 < w < 1$, that controls the extent to which $\widehat{\boldsymbol{\pi}}_{n+1}$ is affected by a new transaction and the speed with which a previous transaction is "aged out." The initial probability estimate $\widehat{\boldsymbol{\pi}}_0$ must be specified, perhaps from historical data on other customers. (See Abraham and Ledolter (1983) for more information about EWMA.) Under some conditions, the EWMA approximates the posterior mean under a Bayesian dynamic model (West and Harrison, 1989).

Unweighted averages and EWMA estimates are appropriate when variables are randomly sampled. Timing variables like day-of-week are not randomly sampled, however. If the transaction rate on Monday is high and the most recent transaction occurred early Monday morning, then the next transaction is likely to occur on Monday and unlikely to occur on Tuesday or any day of the week other than Monday. Because unweighted averaging and exponentially weighted moving averaging increase the estimated probability for a histogram cell every time that cell is observed, the estimated probability for Monday first rises with every transaction made on Monday and then falls with every transaction made before the following Monday. The evolution of the unweighted and EWMA sequential estimators when transactions are made according to a Poisson process with equal day-of-week rates is illustrated in the first two panels of Figure 1 for simulated data. The periodic behavior of these estimators is an artifact that reflects only the order in which transactions are made.

Better sequential estimates of timing distributions that are nearly as simple to compute as exponentially weighted moving averages are possible, however. The key idea is to estimate the *transaction rate* $\lambda_{j,n}$ for period j at the time of call n and then estimate the *period probability* $\pi_{j,n}$ for period j by $\lambda_{j,n}/\sum_k \lambda_{k,n}$. This paper shows that the approximate posterior mean, $\widehat{\lambda}_{j,n}$, for $\lambda_{j,n}$ under a simple dynamic Poisson model satisfies

$$\widehat{\lambda}_{j,n}^{-1} = \begin{cases} (1-w)\widehat{\lambda}_{j,n-1}^{-1} + wZ_{j,n}, & \text{if transaction } n \text{ falls in period } j \\ \widehat{\lambda}_{j,n-1}^{-1} + \frac{w}{1-w} Z_{j,n}, & \text{if transaction } n \text{ does not fall in period } j, \end{cases} \tag{2}$$

where $Z_{j,n}$ is the time that has elapsed in a period j since the previous transaction $n-1$. We call this the *event-driven estimator* or *EDE* because it is updated whenever there is a transaction (event). Figure 1 shows the EDE for the probability of Monday on the simulated data used to illustrate the performance of the unweighted average and the EWMA. Clearly, the EDE avoids the extreme periodic bias of the unweighted average and EWMA. Moreover, the EDE is very close to the exact maximum likelihood estimator (MLE) for the simulated Poisson process. The MLE, however, is not appropriate for sequentially estimating timing distributions for a volatile database because it cannot be computed until a full week after the first transaction has passed and does not adapt to evolving behavior, as Section 3 shows.

The EDE shares the good characteristics of the EWMA: simplicity, ability to adapt to evolving param-

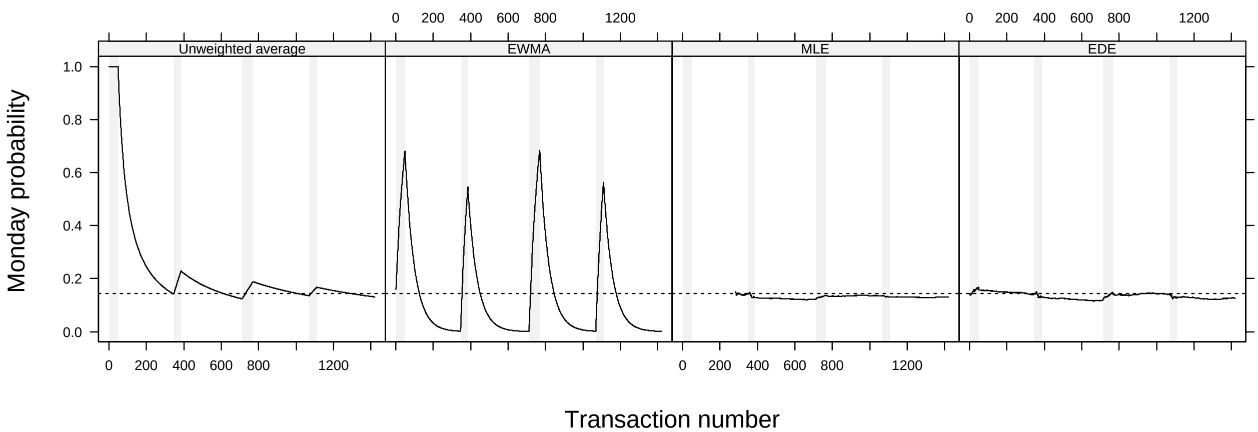

Figure 1: Sequential probability estimates for Monday during a four week period, when transactions for each day of the week follow a Poisson(50) distribution. The nominal probability of each day, 1/7, is indicated by a dashed line. Shaded bars highlight transactions made on Mondays. The first panel shows the unweighted average; the second panel shows the EWMA with $w = .02$ and an initial probability of 1/7 for each weekday. Both the unweighted and EWMA estimated probability for Monday peaks at the end of Monday then slowly falls during the rest of the week until the first transaction made on the following Monday. The pattern dampens for unweighted averaging over time because eventually each observation is given only a small uniform weight. For comparison, the third panel shows the maximum likelihood estimates for the simulated Poisson model. These estimates cannot be computed without a full week of data and do not adapt as behavior evolves. The fourth panel shows the event-driven estimates proposed in this paper, using weight $w = .02$ and initial rates $\lambda_{j,0} = 40,\ j = 1, \ldots, 7$.

eters, and minimal storage requirements. Moreover, the EDE is approximately optimal under a dynamic Poisson timing model, in the sense that it approximates the posterior mean under a particular dynamic Poisson timing model. Computing the EDE is similar to computing an EWMA, except that the rate for each period, not just the rate for the observed period, is updated whenever there is a transaction. The updated estimate for the observed period is a weighted average of the previous reciprocal rate and the time since the last transaction in period j. The updated reciprocal rate estimate for any other period increases by the amount of time that has been spent in the period since the last transaction. The reciprocal rate $\widehat{\lambda}_{j,n}^{-1}$ estimates the average time between transactions in period j.

This paper is organized as follows. Section 2 sets up a Poisson process with transaction rates that are constant within a period j or evolve deterministically over time. Section 3 derives maximum likelihood estimators (MLEs) under the Poisson model with constant periodic rates and gives their asymptotic behavior. Section 4 extends the model to allow dynamic transaction rates that evolve randomly, even within a period, and Section 5 shows that the EDE approximates the posterior mean for this model. An application to sequentially tracking the day-of-week calling patterns of a random sample of about 2,000 telecommunications customers over a three month period is presented in Section 6. The large sample behavior of the EDE is considered in Section 7, and its finite sample behavior is explored through simulations in Section 8. Readers who are not interested in the derivation of the EDE or its distributional properties may move directly from the model in Section 4 to the application in Section 6 and then on to the simulation in Section 8. Conclusions are given in Section 9.

2 The Poisson Timing Model

Suppose that calendar time is broken into a sequence of cycles, each of duration τ, and that each cycle is broken into J periods with fixed, but possibly unequal, lengths. For example, a cycle may last a week and have the $J = 3$ periods $\{Monday, \dots, Friday\}$, $\{Saturday\}$ and $\{Sunday\}$. Transaction i occurs at a calendar time T_i during a period $R_i \in \{1, \dots, J\}$ of cycle C_i. The time required to complete period j is δ_j, and the time required to complete one cycle is $\tau = \sum_{j=1}^{J} \delta_j$. Thus, a transaction at time T_i falls in cycle $C_i = \lfloor T_i/\tau \rfloor + 1$, where $\lfloor x \rfloor$ denotes the greatest integer in x. It falls in period $R_i = j$ if $\Delta_{j-1} < T_i - (C_i - 1)\tau \leq \Delta_j$, where $\Delta_j = \sum_{m=1}^{j} \delta_m$ is the time from the start of the cycle to the end of period j and $\Delta_0 = 0$. Let $S_i = T_i - (C_i - 1)\tau - \Delta_{R_i - 1}$ be the time spent in period R_i up to T_i. We are interested in the probabilities of periods $1, \ldots, J$ during cycle k when the times of transactions in each period of each cycle follow a Poisson process.

Let $N_{k,j}(u)$ be the number of transactions up to time u within period j of cycle k; *i.e.*, the number of transactions between calendar times $(k-1)\tau + \Delta_{j-1}$ and $(k-1)\tau + \Delta_{j-1} + u$. Suppose that $\{N_{k,j}(u)\}$ is a truncated Poisson process with rate $\lambda_{k,j}$. That is, if the interval $[u, u+v]$ falls entirely in period j of cycle k, then

$$N_{k,j}(u+v) - N_{k,j}(u) \sim Poisson\left(\lambda_{k,j} v\right).$$

If $N_{k,j}(u)$ is independent of $N_{k',j'}(v)$ for $k \neq k'$ or $j \neq j'$, then the number of transactions $N_{k,j} = N_{k,j}(\delta_j)$ during period j of cycle k has a $Poisson\left(\lambda_{k,j}\delta_j\right)$ distribution. Also, the number of transactions $N_k = \sum_j N_{k,j}$ within cycle k has a $Poisson\left(\lambda_k\right)$ distribution, where $\lambda_k = \sum_{j=1}^{J} \lambda_{k,j}\delta_j$. We call this the *Poisson timing model* because the number of transactions N_t up to a calendar time t that fall in period j_t of cycle k_t has a $Poisson$ distribution with parameter $\sum_{k<k_t} \lambda_k + \sum_{j<j_t} \lambda_{k,j}\delta_j + \lambda_{k_t,j_t} s$ where $s = t - (k_t - 1)\tau - \Delta_{j_t - 1}$ is the time spent in period j_t. The conditional distribution of $N_{k,j}$ given N_k is $Binomial\left(N_k, \lambda_{k,j}\delta_j/\lambda_k\right)$ because

$$\begin{aligned}
Pr\left(N_{k,j} = m | N_k = M\right) &= \frac{Pr\left(N_{k,j} = m, N_k - N_{k,j} = M - m\right)}{Pr\left(N_k = M\right)} \\
&= \frac{Pr\left(N_{k,j} = m\right) Pr\left(N_k - N_{k,j} = M - m\right)}{Pr\left(N_k = M\right)} \\
&= \begin{pmatrix} M \\ m \end{pmatrix} \left(\frac{\lambda_{k,j}\delta_j}{\lambda_k}\right)^m \left(1 - \frac{\lambda_{k,j}\delta_j}{\lambda_k}\right)^{M-m}, \quad m = 0, \ldots, M.
\end{aligned}$$

The *period probabilities* $\pi_{k,1}, \ldots, \pi_{k,J}$ for cycle k then satisfy

$$\pi_{k,j} = Pr\left(N_{k,j} = 1 | N_k = 1\right) = \frac{\lambda_{k,j}\delta_j}{\lambda_k}.$$

The model described so far assumes that the transaction rate in each period of each cycle is an arbitrary constant $\lambda_{k,j}$. It is often more reasonable to assume that the transaction rates are constant across cycles, so that $\lambda_{k,j} = \lambda_{m,j}$ for all k, m, or to assume that the transaction rates evolve smoothly. For example, $\lambda_{k+1,j} = (1-\alpha)\lambda_{k,j} + \alpha$ for a small α in $(0,1)$. This kind of assumption justifies a sequential updating scheme in which transactions from previous cycles affect the estimates for the current cycle.

3 Maximum Likelihood Estimation under the Poisson Timing Model

Let n be the total number of observed transactions and let $N_{k,j}$ be the number of transactions that fall in period j of cycle k. Under the Poisson timing model with arbitrary piecewise constant transaction rates, only $N_{k,j}$ has information about $\lambda_{k,j}$, and there is no information in the sample about $\lambda_{k,j}$ if $k > C_n$ or if $k = C_n$ and $j > R_n$. When the last transaction in period j of cycle k is known, so that $C_n > k$ or $C_n = k$ and $R_n > j$, the likelihood for $\lambda_{k,j}$ is proportional to the probability of the observed value of $N_{k,j}$. When $(C_n, R_n) = (k, j)$, so there is at least one transaction in the period but the last transaction in the period may not have occurred yet, the likelihood of $\lambda_{k,j}$ is proportional to the probability of the observed value of $N_{k,j}(S_n)$, where S_n is again the time spent in (C_n, R_n). Thus, the log-likelihood of $\lambda_{k,j}$ given that $C_n \geq k$ or that $C_n = k$ and $R_n \geq j$ is

$$\ell\left(\lambda_{k,j}|N_{k,j}\right) = \begin{cases} N_{k,j}\log\left(\lambda_{k,j}\right) - \lambda_{k,j}\delta_j + const, & C_n > k, \text{ or } C_n = k \text{ and } R_n > j \\ N_{k,j}(S_n)\log\left(\lambda_{k,j}\right) - \lambda_{k,j}S_n + const, & (C_n, R_n) = (k, j). \end{cases}$$

It follows that the MLE of $\lambda_{k,j}$ for $k < C_n$ or $k = C_n$ and $j \leq R_n$ is

$$\widehat{\lambda}_{k,j} = \begin{cases} \frac{N_{k,j}}{\delta_j}, & C_n > k, \text{ or } C_n = k \text{ and } R_n > j \\ \frac{N_{k,j}(S_n)}{S_n}, & (C_n, R_n) = (k, j). \end{cases} \tag{3}$$

The invariance of maximum likelihood estimation implies that the MLE of π_{kj} is $\widehat{\pi}_{k,j} = \widehat{\lambda}_{k,j}\delta_j / \sum_{m=1}^{J} \widehat{\lambda}_{k,m}\delta_m$. Note that the MLEs of the period probabilities $\pi_{k,j}$ can only be calculated after the first observation in the last period of a cycle or after the cycle is completed, if no transactions occur in the last period. A Bayesian formulation of the Poisson timing model, with prior distributions on the $\lambda_{k,j}$, would allow $\pi_{k,j}$ to be estimated any time, even before any transaction occurs. Bayesian estimation under the Poisson timing model is not considered in this paper, however.

It is possible to have a hierarchy of periodicities. For example, periods could be broken into phases (or days into hours). Then the MLEs for phases would be based on counts for phases rather than counts over periods, and MLEs for periods and cycles would be obtained by summing up MLEs for phases.

The MLE of $\lambda_{k,j}$ given by equation (3) makes no use of the transactions outside period j of cycle k. This is inefficient when transaction rates do not vary across cycles or vary smoothly from one cycle to the next, especially when transaction rates are low. If the rate for each period is constant across cycles, then the

MLE of the period rate λ_j is

$$\widehat{\lambda}_j = \begin{cases} \sum_{k \le C_n} N_{k,j} / \left(C_n \delta_j\right), & R_n > j \\ \sum_{k < C_n} N_{k,j} / \left[\left(C_n - 1\right) \delta_j\right], & R_n < j \\ \left[\sum_{k < C_n} N_{k,j} + N_{C_n, R_n}(S_n)\right] / \left[\left(C_n - 1\right) \delta_j + S_n\right], & R_n = j. \end{cases}$$

The MLEs of the period probabilities are $\widehat{\pi}_j = \widehat{\lambda}_j \delta_j / \sum_{m=1}^{J} \widehat{\lambda}_m \delta_m$, which cannot be calculated until either a transaction occurs in the last period of the first cycle or the first cycle is completed. Thus, the MLEs are not appropriate for tracking timing behavior sequentially.

Instead of assuming that transaction rates are constant across cycles, we could allow them to evolve deterministically. For example, assume that $\lambda_k = \beta k$ and $\pi_{k,j} = \pi_j$ for all k, so that the transaction rate changes linearly at each cycle but the relative frequency distribution for the periods is stable across cycles. Then the MLEs are

$$\widehat{\beta} = 2\left[\sum_{j > R_n} \frac{\sum_{k < C_n} N_{k,j}}{\delta_j C_n \left(C_n - 1\right)} + \sum_{j < R_n} \frac{\sum_{k \le C_n} N_{k,j}}{\delta_j C_n \left(C_n + 1\right)} + \frac{\sum_{k < C_n} N_{k,j} + N_{C_n, R_n}(S_n)}{C_n \left[\left(C_n - 1\right)\delta_j + 2S_n\right]}\right]$$

$$\widehat{\pi}_j = \begin{cases} 2\sum_{k \le C_n} N_{k,j} / \left(\delta_j C_n \left(C_n + 1\right) \widehat{\beta}\right), & j < R_n \\ 2\sum_{k < C_n} N_{k,j} / \left(\delta_j C_n \left(C_n - 1\right) \widehat{\beta}\right), & j > R_n \\ \left(2\sum_{k < C_n} N_{k,j} + N_{C_n, j}(S_n)\right) / \left(C_n \left[\left(C_n - 1\right)\delta_j + 2S_n\right] \widehat{\beta}\right), & j = R_n. \end{cases}$$

Section 4 describes a more general model in which transaction rates evolve randomly.

4 A Dynamic Poisson Timing Model

As in Section 2, let $N_{k,j}$ be the number of transactions during period j of cycle k. We now assume that the transaction rate is dynamic, so a different rate applies to each transaction. Let $T_{k,j,i},\ i = 1, \ldots, N_{k,j}$, be the time of the i^{th} transaction that falls in period j of cycle k. Let $T_{k,j,0} = (k-1)\tau + \Delta_{j-1}$ be the start of period j in cycle k and $T_{k,j,N_{k,j}+1} = (k-1)\tau + \Delta_j$ be the end of period j in cycle k. The structure of the data is illustrated in Figure 2.

The waiting times $Y_{k,j,i} = T_{k,j,i} - T_{k,j,i-1}, i = 1, \ldots, N_{k,j}$, between transactions in the same period of the same cycle are uncensored. At time $T_{k,j,N_{k,j}+1}$, the waiting time $Y_{k,j,N_{k,j}+1}$ from the last transaction in period j until the end of period j can be considered to be a censored waiting time because the period could no longer be observed. Let $Z_{k,j,N_{k,j}+1}$ be the uncensored waiting time for the end of period j in cycle k, that

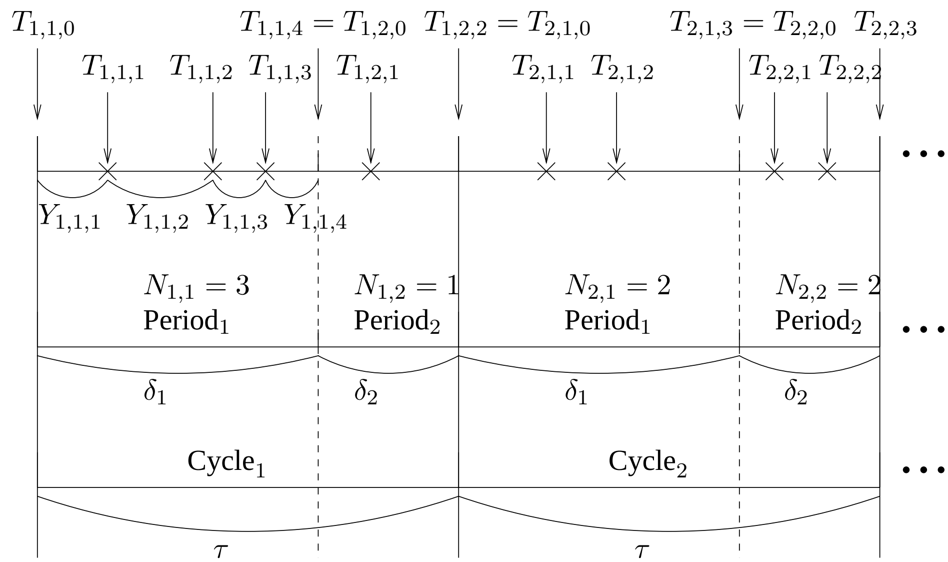

Figure 2: The structure of the data for a Poisson timing process with two periods. Note that $T_{i,j,k}$ is either the start of a period, the time of a transaction, or the time from the last transaction in the period to the end of the period.

is, the transaction time that would have been observed if the period had continued to be observed. Including $Y_{k,j,N_{k,j}+1}$ in the model allows the information that no transaction was seen in a period to be used to update the rate for that period.

Under a Poisson timing model, $Y_{k,j,1}, \ldots, Y_{k,j,N_{k,j}}$ are independent exponential random variables and $Y_{k,j,N_{k,j}+1}$ is a right-censored exponential. Under a dynamic Poisson timing model, the mean of the exponential distribution evolves randomly. Here we assume that the mean in effect at time $T_{k,j,i}$ is a random perturbation of the rate in effect at the time of the last transaction in period j prior to $T_{k,j,i}$. Thus, the transaction rate changes when and only when there is a transaction. To be more precise, the *dynamic Poisson* *timing model* is defined as follows:

$$
\begin{aligned}
Y_{k,j,i} &\sim \text{exponential}(\lambda_{k,j,i}), \quad i = 1, \ldots, N_{k,j} \\
Y_{k,j,N_{k,j}+1} &= \min(Z_{k,j,N_{k,j}+1}, \, (k-1)\tau + \Delta_j - T_{k,j,N_{k,j}}), \\
Z_{k,j,N_{k,j}+1} &\sim \text{exponential}(\lambda_{k,j,N_{k,j}+1}) \\
\lambda_{k,j,i} &= \begin{cases} \epsilon_{k,j,i}\lambda_{k,j,i-1}, & i = 2, \ldots, N_{k,j}+1, \ N_{k,j} \geq 1, \quad \text{or} \quad i = 1, \ k = 1 \\ \lambda_{k,j,i-1}, & i = 1, \ k > 1 \end{cases} \\
\epsilon_{k,j,i} &\sim \Gamma(\alpha, \alpha), \quad \alpha > 0,
\end{aligned}
\tag{4}
$$

where $\lambda_{k,j,0} = \lambda_{k-1,j,N_{k-1,j}+1}$ and $\lambda_{1,j,0}$ is a known constant. Note that the transaction rate for a period stays constant until a transaction occurs in that period, and then the rate for the next period (which controls the waiting time until the next transaction) is generated. Thus, the transaction rate that applies to the first transaction in a period j after time 0 is generated at time 0. Model (4) is an example of a Bayesian Dynamic Model (see West and Harrison (1989)).

Because the mean of $\epsilon_{k,j,i}$ is one and the variance is $1/\alpha$, the transaction rate is more stable for larger α. The current transaction rate is also more variable when the previous transaction rate was high than when it was low. To see this, let $C_{k,j,(i)}$ denote the cycle of the last transaction in period j prior to time $T_{k,j,i}$, that is,

$$
C_{k,j,(i)} = \begin{cases} \max\{l < k : N_{l,j} \geq 1\}, & N_{k,j}(T_{k,j,i} - T_{k,j,0}) \leq 1 \\ k, & N_{k,j}(T_{k,j,i} - T_{k,j,0}) \geq 2, \end{cases}
$$

where $N_{k,j}(T_{k,j,i} - T_{k,j,0})$ represents the number of transactions in period j of cycle k up to time $T_{k,j,i}$. By convention, $C_{k,j,(i)} = 0$ if no transactions occurred in period j prior to $T_{k,j,i}$. Let $\lambda_{k,j,(i)}$ denote the rate in effect at the time of the last transaction in period j of cycle $C_{k,j,(i)}$. It follows from model (4) that $(\lambda_{k,j,i} \,|\, \lambda_{k,j,(i)}) \sim \Gamma(\alpha, \alpha/\lambda_{k,j,(i)})$, which has variance $\lambda^2_{k,j,(i)}/\alpha$.

The posterior distribution of the current transaction rate can be found as follows. Let $W_{k,j,i}$ denote the time elapsed in period j between the last transaction in period j and time $T_{k,j,i-1}$, so

$$
W_{k,j,i} = \begin{cases} \sum_{l=C_{k,j,(i)}}^{k-1} Y_{l,j,N_{l,j}+1}, & C_{k,j,(i)} < k \\ 0, & C_{k,j,(i)} = k, \end{cases}
$$

with the convention that $Y_{0,j,1} = 0$. For $1 \leq i \leq N_{k,j}$, the posterior distribution of $\lambda_{k,j,i}$ given the previous transaction rate $\lambda_{k,j,(i)}$ and all the observed data $\boldsymbol{Y}_{k,j,i} = \{Y_{1,j,1}, \ldots, Y_{k,j,i}\}$ up to and including time $T_{k,j,i}$

is $\Gamma(\alpha+1, \alpha/\lambda_{k,j,(i)} + W_{k,j,i} + Y_{k,j,i})$. When $i = N_{k,j}+1$, the time $Y_{k,j,i}$ after the last transaction in a period until the end of the period is a censored exponential and the posterior distribution of $\lambda_{k,j,i}$ given $\lambda_{k,j,(i)}$ and the $\boldsymbol{Y}_{k,j,i}$ is $\Gamma(\alpha, \alpha/\lambda_{k,j,(i)} + W_{k,j,i} + Y_{k,j,i})$. Thus,

$$\mathrm{E}\left(\lambda_{k,j,i} \,|\, \lambda_{k,j,(i)}, \boldsymbol{Y}_{k,j,i}\right) = \begin{cases} \left[(1-w)\lambda_{k,j,(i)}^{-1} + w\left(W_{k,j,i} + Y_{k,j,i}\right)\right]^{-1}, & 1 \le i \le N_{k,j} \\ \left[\lambda_{k,j,(i)}^{-1} + \frac{w}{1-w}\left(W_{k,j,i} + Y_{k,j,i}\right)\right]^{-1}, & i = N_{k,j}+1, \end{cases}$$

where $w = (\alpha+1)^{-1}$. This posterior mean is not directly useful for our sequential estimation problem because it depends on the unknown period rate $\lambda_{k,j,(i)}$ and all the past observations. It does, however, resemble the EDE given by equation (2) of Section 1.

5 Derivation of the EDE

5.1 An Approximation to the Posterior Mean of A Transaction Rate

The posterior mean of $\lambda_{k,j,i}$ given the initial rate $\lambda_{j,0}$ and $\boldsymbol{Y}_{k,j,i}$ can be written as

$$\begin{aligned} \mathrm{E}\left(\lambda_{k,j,i} \,|\, \lambda_{j,0}, \boldsymbol{Y}_{k,j,i}\right) &= \mathrm{E}\left(\mathrm{E}\left(\lambda_{k,j,i} \,|\, \lambda_{k,j,(i)}, \lambda_{j,0}, \boldsymbol{Y}_{k,j,i}\right) \,|\, \lambda_{j,0}, \boldsymbol{Y}_{k,j,i}\right) \\ &= \mathrm{E}\left(\mathrm{E}\left(\lambda_{k,j,i} \,|\, \lambda_{k,j,(i)}, \boldsymbol{Y}_{k,j,i}\right) \,|\, \lambda_{j,0}, \boldsymbol{Y}_{k,j,i}\right) \\ &= \begin{cases} \mathrm{E}\left(\left[(1-w)\lambda_{k,j,(i)}^{-1} + w\left(W_{k,j,i} + Y_{k,j,i}\right)\right]^{-1} \,|\, \lambda_{j,0}, \boldsymbol{Y}_{k,j,i}\right), & 1 \le i \le N_{k,j} \\ \mathrm{E}\left(\left[\lambda_{k,j,(i)}^{-1} + \frac{w}{1-w}\left(W_{k,j,i} + Y_{k,j,i}\right)\right]^{-1} \,|\, \lambda_{j,0}, \boldsymbol{Y}_{k,j,i}\right), & i = N_{k,j}+1, \end{cases} \end{aligned}$$

where $w = (1+\alpha)^{-1}$.

A first-order Taylor expansion of $f\left(\lambda_{k,j,(i)}\right) = [(1-w)\lambda_{k,j,(i)}^{-1} + w(W_{k,j,i} + Y_{k,j,i})]^{-1}$ around $\mathrm{E}\left(\lambda_{k,j,(i)} | \lambda_{j,0}, \boldsymbol{Y}_{k,j,i-1}\right)$ gives

$$\begin{aligned} \mathrm{E}\left(\lambda_{k,j,i} \,|\, \lambda_{j,0}, \boldsymbol{Y}_{k,j,i}\right) &= \mathrm{E}\left(\left[(1-w)\,\lambda_{k,j,(i)}^{-1} + w\left(W_{k,j,i} + Y_{k,j,i}\right)\right]^{-1} \,|\, \lambda_{j,0}, \boldsymbol{Y}_{k,j,i}\right) \\ &\simeq \mathrm{E}\left(\left[(1-w)\,\mathrm{E}\left(\lambda_{k,j,(i)} \,|\, \lambda_{j,0}, \boldsymbol{Y}_{k,j,i-1}\right)^{-1} + w\left(W_{k,j,i} + Y_{k,j,i}\right)\right]^{-1} \,|\, \lambda_{j,0}, \boldsymbol{Y}_{k,j,i}\right) + \\ &\quad \mathrm{E}\left(f'\left(\mathrm{E}\left(\lambda_{k,j,(i)} | \lambda_{j,0}, \boldsymbol{Y}_{k,j,i-1}\right)\right)\left[\lambda_{k,j,(i)} - \mathrm{E}\left(\lambda_{k,j,(i)} | \lambda_{j,0}, \boldsymbol{Y}_{k,j,i-1}\right)\right] \,|\, \lambda_{j,0}, \boldsymbol{Y}_{k,j,i}\right) \\ &= \left[(1-w)\,\mathrm{E}\left(\lambda_{k,j,(i)} \,|\, \lambda_{j,0}, \boldsymbol{Y}_{k,j,i-1}\right)^{-1} + w\left(W_{k,j,i} + Y_{k,j,i}\right)\right]^{-1} \end{aligned}$$

for $1 \le i \le N_{k,j}$, where f' denotes the first derivative of f. A similar Taylor expansion of $f(\lambda_{k,j,(i)}) = [\lambda_{k,j,(i)}^{-1} + w(1-w)^{-1}(W_{k,j,i} + Y_{k,j,i})]^{-1}$ around $\mathrm{E}\left(\lambda_{k,j,(i)} | \lambda_{j,0}, \boldsymbol{Y}_{k,j,i-1}\right)$ gives

$$\mathrm{E}\left(\lambda_{k,j,i} \,|\, \lambda_{j,0}, \boldsymbol{Y}_{k,j,i}\right) \simeq \left[\mathrm{E}\left(\lambda_{k,j,(i)} \,|\, \lambda_{j,0}, \boldsymbol{Y}_{k,j,i-1}\right)^{-1} + \frac{w}{1-w}\left(W_{k,j,i} + Y_{k,j,i}\right)\right]^{-1}$$

for $i = N_{k,j} + 1$.

Together, the Taylor approximations suggest an iterative procedure for updating $\lambda_{k,j,i}$. Namely, take

$$\widehat{\lambda}^{-1}_{k,j,i} = \begin{cases} (1-w)\widehat{\lambda}^{-1}_{k,j,(i)} + w\left(W_{k,j,i} + Y_{k,j,i}\right), & k \geq 1,\ 1 \leq i \leq N_{k,j} \\ \widehat{\lambda}^{-1}_{k,j,(i)} + \frac{w}{1-w}\left(W_{k,j,i} + Y_{k,j,i}\right), & k \geq 1,\ i = N_{k,j} + 1, \end{cases}$$

where $\widehat{\lambda}_{1,j,1} = \mathrm{E}\left(\lambda_{1,j,1} | \lambda_{j,0}, Y_{1,j,1}\right)$. For $k > 1$ or $i > 1$, $\widehat{\lambda}_{k,j,i}$ approximates the posterior mean $\mathrm{E}\left(\lambda_{k,j,i} | \lambda_{j,0}, \boldsymbol{Y}_{kji}\right)$. Because the transaction rates $\widehat{\lambda}_{k,j,i}$ are updated at the end of each period j or at each new transaction in period j, it is easy to verify that the last equations can be re-written as

$$\widehat{\lambda}^{-1}_{k,j,i} = \begin{cases} (1-w)\widehat{\lambda}^{-1}_{k,j,i-1} + wY_{k,j,i}, & k \geq 1,\ 1 \leq i \leq N_{k,j} \\ \widehat{\lambda}^{-1}_{k,j,i-1} + \frac{w}{1-w}Y_{k,j,i}, & k \geq 1,\ i = N_{k,j} + 1, \end{cases} \quad (5)$$

which uses only the previous estimate and the current transaction to update the estimators. The iterative estimators for the period probabilities are then $\widehat{\pi}_{k,j,i} = \widehat{\lambda}_{k,j,i}\delta_j / \left(\sum_{m=1}^{J} \widehat{\lambda}_{k,m,i}\delta_m\right)$.

The Taylor approximations suggest that the weight w in the iterative estimate should be proportional to the variance of the multiplicative noise factor in the dynamic model (4): $w = \alpha(\alpha+1)^{-1}\mathrm{var}\left(\epsilon_{k,j,i}\right)$. If $\mathrm{var}\left(\epsilon_{k,j,i}\right)$ is large, then $\lambda_{k,j,i}$ evolves quickly and more weight must be given to the current observation $Y_{k,j,i}$, which leads to a noisier sequence of estimates $\left\{\widehat{\lambda}_{k,j,i}\right\}$. Conversely, if $\mathrm{var}\left(\epsilon_{k,j,i}\right)$ is small, then $\lambda_{k,j,i}$ evolves slowly and more weight should be given to previous waiting times, giving more stable $\left\{\widehat{\lambda}_{k,j,i}\right\}$. The simulations in Section 8, however, suggest that larger w, on the order of $1/\sqrt{\alpha}$, may be better.

The iterative estimate (5) is updated whenever there is a transaction in period j or a period j ends. We call this *full iterative estimation* because each transaction rate is updated as soon as it changes or as soon as a period to which it applies ends. Thus, full iterative estimation requires updating the signature for each customer at the end of every period. Retrieving all signatures at the end of every period can be impractical, however. Section 5.2 shows that the *event-driven* estimator *EDE*, which is updated only when there is a transaction, is equivalent to the full iterative estimator (5) whenever there is a transaction. The EDE is out-of-date whenever no transaction occurs in a period, but this limited staleness has not been a concern in the applications that we have seen, however.

5.2 Updating at Transaction Times Only

Now suppose the transaction rates can be updated only at the time of a transaction, and whenever there is a transaction the estimated transaction rates for all periods since the last transaction are updated. For example, suppose that $J = 3$ and the current transaction falls in period 2 of cycle 10. If the previous transaction fell in period 2 of cycle 8, then $\lambda_{9,3,1}$ and $\lambda_{10,1,1}$ as well as $\lambda_{10,2,1}$ would be updated at time $T_{10,3,1}$. Note that missed periods are "found" only at the first transaction time in a period. Because updating occurs only at transaction times, we now subscript by transaction number rather than by cycle and period. The transaction times are denoted by $T_1, \ldots, T_n$, their periods by $R_1, \ldots, R_n$, the time from the start of period R_n to T_n is S_n, and the transaction rates in effect at transaction n are $\lambda_{1,n}, \ldots, \lambda_{J,n}$.

Let $M_{j,n} \geq 0$ be the number of missed periods of type j between T_{n-1} and T_n, *i.e.*, the number of cycles during which the transaction rate for period j was not updated. In the example above, $M_{3,n} = M_{1,n} = 2$ and $M_{2,n} = 1$. Let $Z_{j,n}$ be the waiting time for period j at transaction n, so that

$$Z_{j,n} = \begin{cases} T_n - T_{n-1}, & R_n = R_{n-1} = j \text{ and } C_n = C_{n-1} \\ M_{j,n}\delta_j + S_n + \delta_j - S_{n-1} & R_n = R_{n-1} = j \text{ and } C_n > C_{n-1} \\ M_{j,n}\delta_j + \delta_j - S_{n-1}, & R_{n-1} = j,\ R_n \neq j \\ M_{j,n}\delta_j + S_n, & R_{n-1} \neq j,\ R_n = j \\ M_{j,n}\delta_j, & R_{n-1} \neq j,\ R_n \neq j. \end{cases}$$

Finally, define the updated *event-driven estimator* or *EDE* for period j at transaction n by

$$\widehat{\lambda}_{j,n}^{-1} = \begin{cases} (1-w)\widehat{\lambda}_{j,n-1}^{-1} + wZ_{j,n}, & R_n = j \\ \widehat{\lambda}_{j,n-1}^{-1} + \frac{w}{1-w}Z_{j,n}, & R_n \neq j \end{cases} \tag{6}$$

for $j = 1, \ldots, J$.

With equation (6), all estimated transaction rates are brought up-to-date at each transaction time, but some estimated rates may be out-of-date between transactions because inactive periods have not yet been accounted for. The reciprocal of the updated rate for period j is a weighted average of the previous estimated reciprocal rate for period j and the time $Z_{j,n}$ spent in period j since the last transaction when the current transaction falls in period j. If the current transaction does not fall in period j, then the updated reciprocal rate for period j is not a weighted average. Instead, the previous estimate is increased by a term that is proportional to the time in period j with no activity.

Because all transaction rates are brought up-to-date at the time of a transaction, regardless of which period the transaction falls in, the EDE of the period probability for period j is

$$\widehat{\pi}_{j,n} = \frac{\widehat{\lambda}_{j,n}\delta_j}{\sum_{i=1}^{J} \widehat{\lambda}_{i,n}\delta_i}. \tag{7}$$

A case-by-case analysis shows that at time T_n the J event-driven estimators from equation (6) are identical to the J iterative estimates from equation (5). For example, suppose that $R_n = j$, $R_{n-1} = i \neq j$ and $M_{j,n} > 0$. Then the transaction at time T_n is the first in period j of cycle C_n. With the full iterative estimator, the estimated reciprocal transaction rate for period j would have been updated by a censored exponential of length δ_j at the end of each of the previous $M_{j,n}$ intervals of type j. Therefore, just before time T_n, the full iterative estimator would be the reciprocal of

$$\widehat{\lambda}_{j,n-1}^{-1} + \frac{w}{1-w} M_{j,n}\delta_j.$$

At time T_n, this estimator would be updated to

$$(1-w)\left(\widehat{\lambda}_{j,n-1}^{-1} + \frac{w}{1-w} M_{j,n}\delta_j\right) + wS_n = (1-w)\widehat{\lambda}_{j,n-1}^{-1} + wZ_{j,n},$$

which is identical to the reciprocal of the EDE after transaction n.

The EDE (6) is computed from the current estimated rate $\widehat{\lambda}_{j,n}$ and the time T_{n-1} of the last transaction. This is almost as memory-efficient as the EWMA estimate (1), which requires storing only the first $J-1$ of the $\widehat{\pi}_j$'s.

6 An Application: Estimating Day-of-Week Calling Patterns

Simply stated, fraud detection is the discrimination of legitimate transactions from fraudulent ones. Because customers are extremely diverse, the discrimination must be tailored to each customer separately. An important step in fraud detection, then, is tracking the legitimate behavior of each customer in real-time, so that there is an accurate basis of comparison for discriminating fraud. (See Cahill, Lambert, Pinheiro and Sun (2000) for an overview of one approach to real-time fraud detection.) In this section, we focus on tracking the day-of-week calling patterns for a random sample of about 2,000 callers who made between 50 and 800 completed calls during peak hours over a three month period. Peak hours are here defined to be 9:00 a.m. to 11:00 a.m. and 1:00 p.m. to 4:00 p.m. Monday through Friday. Not all customers made calls for the entire three months; the time between a customer's first and last calls ranged from eight to thirteen weeks.

The dynamic Poisson model holds for a customer if the times $Y_{k,j,i}$ between calls i and $i-1$ during peak hours on day-of-week j in week k behave as independent $\mathrm{exponential}(\lambda_{k,j,i})$ random variables. (The day-of-week calling patterns for non-peak calls may be different. To avoid the complications of nesting hours-within-days, we ignore the non-peak calls here.) The calling rates $\lambda_{k,j,i}$ for the customer are unknown and might change throughout the three months, but here we test the stronger model that the rates for each customer are approximately constant over the three month period, so $Y_{k,j,i} \sim \mathrm{exponential}(\lambda_j)$.

A chi-squared goodness-of-fit test (Conover, 1980) for the exponential distribution can be applied to the waiting times for each day of week for each customer separately, giving a total of 10,000 *p*-values. (We used the S-PLUS function chisq.gof with 6 equi-depth cells to compute each *p*-value (MathSoft, 1996).) Under the null hypothesis that the Poisson timing model is adequate, these *p*-values are uniformly distributed in the interval $(0, 1)$. Each of the five panels in Figure 3 plots the 2,000 goodness-of-fit *p*-values for one day of the week against the $U(0, 1)$ quantiles. If the Poisson timing model with constant transaction rates is adequate for this set of customers, then the points in each plot should lie on a straight line.

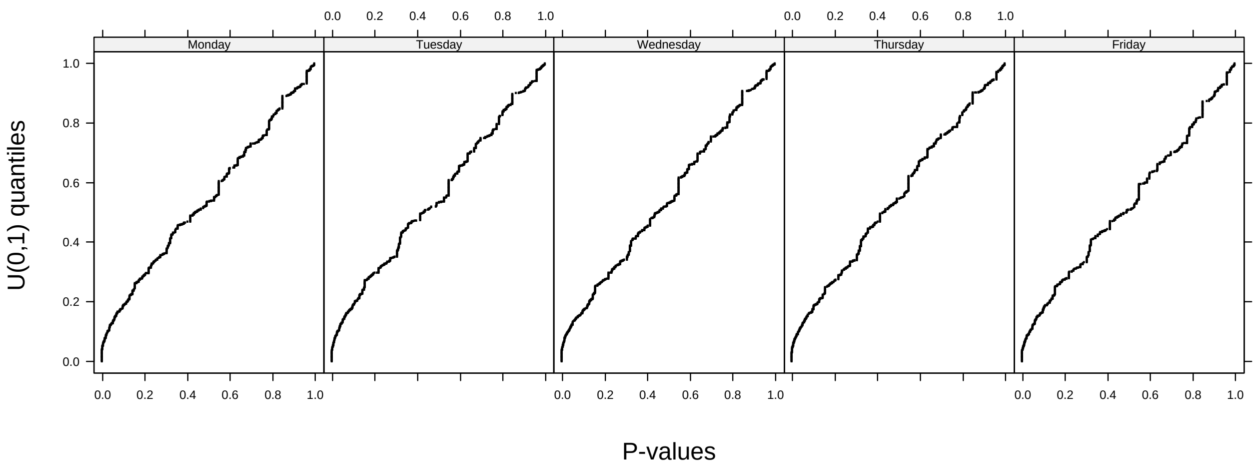

Figure 3: Quantile-quantile plots of the day-of-week goodness-of-fit *p*-values for the Poisson timing model for a random sample of 2,000 customers. Under the null assumption, the *p*-values should lie on a straight line.

The points in all panels of Figure 3 nearly lie on a straight line, except near zero. This suggests that the Poisson timing model with constant peak calling rates for each day of week is reasonable for most customers. The curvature near zero that is observed in all panels suggests that a small set of *p*-values are smaller than expected under the null hypothesis. For these customers, either the Poisson distribution is wrong or the assumption of constant peak calling rates for that day-of-week is wrong. For example, vacations, important deadlines requiring many calls, or habits such as placing all calls at the start or end of the business day violate the stationarity assumption. (They could be accommodated under the more general dynamic Poisson timing model (4), however.) Taking the intercept from the regression of the theoretical $U(0,1)$ quantiles on the empirical quantiles as an estimate of the fraction of customers for whom the constant rate Poisson model is inadequate shows that the simple model does not fit between 5.5% and 7.5% of customers across weekdays. We consider this degree of lack-of-fit acceptable in our application.

Even if the Poisson model with constant rates does not fit, the EDE estimates of the day-of-week probabilities may still be adequate. To check if that is the case here, we take the empirical distribution $\boldsymbol{\pi}$ of a customer's calls per weekday over the three month period as the underlying day-of-week distribution for the customer and compute the average absolute relative error of the EDE $\widehat{\boldsymbol{\pi}}_n$ with respect to $\boldsymbol{\pi}$ as a function of call number by taking

$$\epsilon_n = 100 \sum_{j=1}^{5} \frac{|\pi_j - \widehat{\pi}_{j,n}|}{5\pi_j}.$$

Percentiles of the average error at call n over all customers that make at least n calls then describe the performance of the EDE. The median, .25 quantile and .75 quantiles of the average absolute relative curves for the EDE and the EWMA estimates with $w = .02$ and uniform initial probabilities ($\widehat{\pi}_{j,0} = .2,\ j = 1, \ldots, 5$), are shown as a function of call number in Figure 4.

The curves in the left panel of Figure 4 are based on the original sample. Considering the entire sample of customers as a whole, the EDE clearly outperforms the EWMA estimate. After about 100 calls, the median EDE average absolute relative error is smaller than the .25 quantile of the EWMA average absolute relative error; after 300 calls, the .75 EDE quantile is smaller than the .25 EWMA quantile. The median average absolute relative error stabilizes around 5% for the EDE and around 15% for the EWMA.

The curves in the right panel of Figure 4 are based on the 264 customers for which the simple Poisson timing model is inadequate, in the sense that their *p*-values are in the bottom 5% of the *p*-values obtained on at least one day. This gives us a sense of the performance of the day-of-week probability estimates when the constant rate Poisson model is inadequate. As expected, the average absolute relative errors for this subset of customers tend to be larger than those for all 2,000 customers considered as one group (left panel), but the EDE still clearly dominates the EWMA. The median average absolute relative error curve oscillates around 8% for the EDE and around 22% for the EWMA. In this application, then, the EDE is relatively robust to departures from the Poisson timing model.

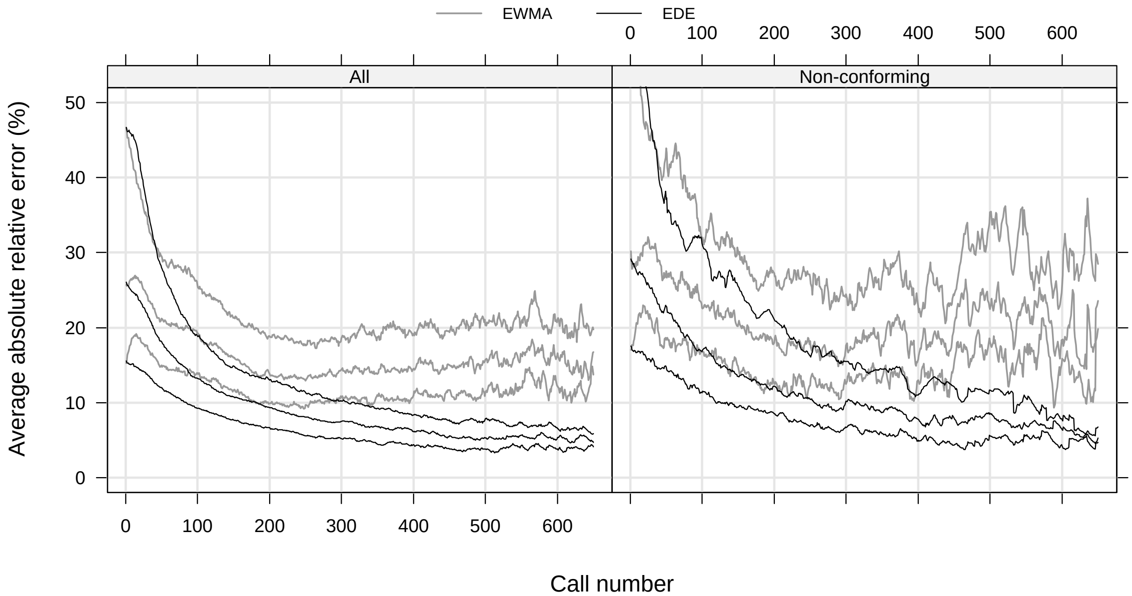

Figure 4: Average absolute relative error curves for the EDE and EWMA estimates of the day-of-week probabilities. The three curves for each type of estimate refer to the .25, .50, and .75 pointwise quantiles of the average absolute relative errors across customers. The left panel includes all customers in the sample, while the right panel includes only those who fail the goodness-of-fit test for the Poisson timing model.

7 Approximate Behavior of the EDE

7.1 Approximate Conditional Moments

Equation (6) is convenient for updating the rate estimates, but not for establishing their properties. For that, it is more convenient to re-write the EDE as a sum of terms that are defined at the transactions that fall in period j, plus an additional term that reflects the inactive time spent in period j, if the current period is unequal to j. To simplify the notation, we suppress the subscript j in this section.

Let $N_{+,i}$ denote the number of transactions that fall in period j up to and including time T_i and $N_{-,i}$ denote the number of transactions in period j up to but not including time T_i, so

$$N_{+,i} = \sum_{m=1}^{i} I(R_m = j) \quad \text{and} \quad N_{-,i} = N_{+,i} - I(R_i = j),$$

where $I(A) = 1$ if condition A is true and 0 otherwise. Let L_i be the index of the i^{th} transaction in period j, so $L_i = \min\{m : N_{+,m} = i\}$, with the convention that $L_0 = 0$. Note that L_i is only defined for $i = 0, \ldots, N_{+,n}$ at transaction n. Finally, let $W_{+,i}$ be the total time spent in period j between time T_i and the previous transaction in a period j (or since time 0, if there are no previous transactions in period j). Then

$$W_{+,i} = \sum_{m=L_{N_{-,i}}+1}^{i} Z_{j,m}.$$

In particular, $W_{+,L_1}, \ldots, W_{+,L_{N_{+,n}}}$ are the waiting times between transactions that fall in period j up to time T_n.

Equation (6) shows that when the current transaction does not fall in period j, the reciprocal of the EDE is increased by a constant multiple of the inactive time since the last transaction. Averaging occurs only when there is a transaction in period j. It is thus possible to accumulate the inactive time and update the transaction rate for period j only at the time of a transaction in period j. Thus, the EDE can be rewritten as

$$\widehat{\lambda}_{j,n}^{-1} = \begin{cases} (1-w)\widehat{\lambda}_{j,L_{N_{-,n}}}^{-1} + wW_{+,n}, & R_n = j \\ \widehat{\lambda}_{j,L_{N_{-,n}}}^{-1} + \frac{w}{1-w} W_{+,n}, & R_n \neq j. \end{cases}$$

Since $\widehat{\lambda}_{j,L_i}^{-1} = (1-w)\widehat{\lambda}_{j,L_{i-1}}^{-1} + wW_{+,L_i}$ (with $\widehat{\lambda}_{j,L_0} = \lambda_{j,0}$), it follows that

$$\widehat{\lambda}_{j,n}^{-1} = \begin{cases} \sum_{i=1}^{N_{+,n}} (1-w)^{i-1} wW_{+,L_i} + (1-w)^{N_{+,n}} \lambda_{j,0}^{-1}, & R_n = j \\ \sum_{i=1}^{N_{+,n}} (1-w)^{i-1} wW_{+,L_i} + (1-w)^{N_{+,n}} \lambda_{j,0}^{-1} + \frac{w}{1-w} W_{+,n}, & R_n \neq j. \end{cases} \tag{8}$$

To summarize, the rate estimates obtained by (i) updating at every transaction in period j and at the end of every period j, or (ii) updating at the time of each transaction, regardless of which period it falls in, but not at the end of each period j, or (iii) updating only when there is a transaction in period j are identical at the time of a transaction in period j. Updating scheme (ii) is most useful for computing, but (iii) is most useful for establishing properties of the rate estimator.

We next derive approximations to the moments and the distribution of $\widehat{\lambda}_{j,n}$. Our main interest here is to understand the forecasting properties of $\widehat{\lambda}_{j,n}$ given a sequence of transaction rates $\boldsymbol{\lambda}_{j,n} = \{\lambda_{j,1}, \ldots, \lambda_{j,N_{+,n}}\}$ and $N_{+,n}$. To simplify the notation, we suppress the dependence on the conditioning variables, but all moments and densities should be understood as conditional on $\boldsymbol{\lambda}_{j,n}$ and $N_{+,n}$. For example, $\mathrm{E}\left(\widehat{\lambda}_{j,n}^{-1}\right)$ denotes $\mathrm{E}\left(\widehat{\lambda}_{j,n}^{-1} | \boldsymbol{\lambda}_{j,n}, N_{+,n}\right)$.

Given $\boldsymbol{\lambda}_{j,n}$, the waiting times $W_{+,L_i}, i = 1, \ldots, N_{+,n}$ between transactions in period j are independent $\mathrm{exponential}\,(\lambda_{j,L_i})$ random variables under the dynamic Poisson timing model. The proof is given in the Appendix. The only other waiting time that needs to be considered is $W_{+,n}$ when $R_n \neq j$, where

$$W_{+,n} = \delta_j - S_{L_{N_{+,n}}} + \delta_j \sum_{m=L_{N_{+,n}}+1}^{n} M_{j,m}.$$

Conditional on $\boldsymbol{\lambda}_{j,n}$ and $N_{+,n}$, the last sum has a $\mathrm{geometric}\left(1 - exp(-\lambda_{j,L_{N_{+,n}+1}}\delta_j)\right)$ distribution and is independent of $S_{L_{N_{+,n}}}$. If the transaction rate for period j changes so slowly that it is approximately constant within cycles, then $\delta_j - S_{L_{N_{+,n}}}$ is approximately distributed as a truncated $\mathrm{exponential}\left(\lambda_{j,L_{N_{+,n}+1}}\right)$ in the interval $[0, \delta_j]$. As shown in the Appendix, combining these results implies that $W_{+,n}$ given $\boldsymbol{\lambda}_{j,n}$ and $N_{+,n}$ is approximately $\mathrm{exponential}\left(\lambda_{j,L_{N_{+,n}+1}}\right)$ when $R_n \neq j$.

When $R_n = j$, it follows from equation (8) that, given $\boldsymbol{\lambda}_{j,n}$ and $N_{+,n}$,

$$\begin{aligned} \mathrm{E}\left(\widehat{\lambda}_{j,n}^{-1}\right) &= \sum_{i=1}^{N_{+,n}} (1-w)^{i-1}\, w\lambda_{j,L_i}^{-1} + (1-w)^{N_{+,n}}\, \lambda_{j,0}^{-1} \\ \mathrm{var}\left(\widehat{\lambda}_{j,n}^{-1}\right) &= \sum_{i=1}^{N_{+,n}} (1-w)^{2(i-1)}\, w^2\lambda_{j,L_i}^{-2}. \end{aligned} \tag{9}$$

Similarly, when $R_n \neq j$,

$$\begin{aligned} \mathrm{E}\left(\widehat{\lambda}_{j,n}^{-1}\right) &\simeq \sum_{i=1}^{N_{+,n}} (1-w)^{i-1}\, w\lambda_{j,L_i}^{-1} + (1-w)^{N_{+,n}}\, \lambda_{j,0}^{-1} + \frac{w}{1-w}\lambda_{j,L_{j,N_{+,n}+1}}^{-1} \\ \mathrm{var}\left(\widehat{\lambda}_{j,n}^{-1}\right) &\simeq \sum_{i=1}^{N_{+,n}} (1-w)^{2(i-1)}\, w^2\lambda_{j,L_i}^{-2} + \left(\frac{w}{1-w}\right)^2 \lambda_{j,N_{+,n}+1}^{-2}. \end{aligned} \tag{10}$$

If $\lambda_{j,n} = \lambda_j$ for all n, then for $R_n = j$

$$\begin{aligned} \mathrm{E}\left(\widehat{\lambda}_{j,n}^{-1}\right) &= \left[1 - (1-w)^{N_{+,n}}\right] \lambda_j^{-1} + (1-w)^{N_{+,n}}\, \lambda_{j,0}^{-1} \quad \xrightarrow{N_{+,n}\to\infty} \quad \lambda_j^{-1} \\ \mathrm{var}\left(\widehat{\lambda}_{j,n}^{-1}\right) &= \frac{w}{2-w}\left[1 - (1-w)^{2N_{+,n}}\right] \lambda_j^{-2} \quad \xrightarrow{N_{+,n}\to\infty} \quad \frac{w}{2-w}\lambda_j^{-2}. \end{aligned} \tag{11}$$

When $R_n \neq j$, the approximations in (10) become exact and simplify to

$$\begin{aligned} \mathrm{E}\left(\widehat{\lambda}_{j,n}^{-1}\right) &= \frac{1}{1-w}\left[1 - (1-w)^{N_{+,n}+1}\right] \lambda_j^{-1} + (1-w)^{N_{+,n}}\, \lambda_{j,0}^{-1} \quad \xrightarrow{N_{+,n}\to\infty} \quad \frac{1}{1-w}\lambda_j^{-1} \\ \mathrm{var}\left(\widehat{\lambda}_{j,n}^{-1}\right) &= \frac{w}{(2-w)(1-w)^2}\left[1 - (1-w)^{2(N_{+,n}+1)}\right] \lambda_j^{-2} \quad \xrightarrow{N_{+,n}\to\infty} \quad \left[\frac{w}{(2-w)(1-w)^2}\right] \lambda_j^{-2}. \end{aligned} \tag{12}$$

7.2 Approximate Conditional Distribution Under a Poisson Model With Constant Rates

If the transaction rate for period j is constant across cycles, then, given $\boldsymbol{\lambda}_{j,n}$ and $N_{+,n}$, the EDE of a reciprocal rate is a linear combination of independent, identically distributed exponential random variables, plus an exponentially decreasing term that depends on $\lambda_{j,0}$. If all coefficients in the linear combination were the same, the linear combination would be conditionally distributed as a gamma random variable. This suggests approximating the conditional distribution of $\widehat{\lambda}_{j,n}^{-1}$ by an exponentially decreasing constant plus a gamma distribution with parameters that match the first two moments of $\widehat{\lambda}_{j,n}^{-1}$. That is,

$$\widehat{\lambda}_{j,n}^{-1} \dot{\sim} \Gamma\left(\frac{\left[\mathrm{E}\left(\widehat{\lambda}_{j,n}^{-1}\right) - (1-w)^{N_{+,n}}\lambda_{j,0}^{-1} \right]^2}{\mathrm{var}\left(\widehat{\lambda}_{j,n}^{-1}\right)}, \frac{\mathrm{E}\left(\widehat{\lambda}_{j,n}^{-1}\right) - (1-w)^{N_{+,n}}\lambda_{j,0}^{-1}}{\mathrm{var}\left(\widehat{\lambda}_{j,n}^{-1}\right)} \right) + (1-w)^{N_{+,n}}\lambda_{j,0}^{-1}.$$

From equations (11) and (12),

$$\widehat{\lambda}_{j,n}^{-1} \dot{\sim} (1-w)^{N_{+,n}}\lambda_{j,0}^{-1} + \begin{cases} \Gamma\left(\frac{2-w}{w} \frac{\left[1-(1-w)^{N_{+,n}}\right]^2}{1-(1-w)^{2N_{+,n}}}, \frac{2-w}{w} \frac{1-(1-w)^{N_{+,n}}}{1-(1-w)^{2N_{+,n}}}\lambda_j \right), & R_n = j \\ \Gamma\left(\frac{2-w}{w} \frac{\left[1-(1-w)^{N_{+,n}+1}\right]^2}{1-(1-w)^{2\left(N_{+,n}+1\right)}}, \frac{(2-w)(1-w)}{w} \frac{1-(1-w)^{N_{+,n}+1}}{1-(1-w)^{2\left(N_{+,n}+1\right)}}\lambda_j \right), & R_n \neq j. \end{cases} \tag{13}$$

Letting $N_{+,n} \longrightarrow \infty$, the right side of (13) asymptotically simplifies to

$$\widehat{\lambda}_{j,n}^{-1} \dot{\sim} \begin{cases} \Gamma\left(\frac{2-w}{w}, \frac{2-w}{w}\lambda_j\right), & R_n = j \\ \Gamma\left(\frac{2-w}{w}, \frac{(2-w)(1-w)}{w}\lambda_j\right), & R_n \neq j. \end{cases} \tag{14}$$

Simulations suggest that the approximation (13) is remarkably good for a wide range of w, even for small $N_{+,n}$. It is sufficient to simulate the behavior of $\widehat{\lambda}_{j,n}$ at $\lambda_j = 1$ under the Poisson timing model because, with constant rates, the distribution of $\lambda_j \sum_{i=1}^{N_{+,n}} (1-w)^{i-1} wW_{+,L_i}$ in equation (8) is free of λ_j and the contribution of the initial rate $\lambda_{j,0}$ decreases exponentially. Using the statistical language S (Becker, Chambers and Wilks, 1988), we generated 2500 simulated values of $\widehat{\lambda}_{j,n}^{-1} - (1-w)^{N_{+,n}}\lambda_{j,0}^{-1}$ with $w = .02$, $N_{+,n} = 10$, for both $R_n = j$ and $R_n \neq j$ and compared the empirical distribution of the 2500 values of $\lambda_j[\widehat{\lambda}_{j,n}^{-1} - (1-w)^{N_{+,n}}\lambda_{j,0}^{-1}]$ with the gamma approximation. Figure 5 shows that the empirical cumulative distribution function (CDF) is close to the approximate gamma CDF. Similar results, not shown here, were obtained for other values of w and $N_{+,n}$.

The approximate distribution of $\lambda_j[\widehat{\lambda}_{j,n}^{-1} - (1-w)^{N_{+,n}}\lambda_{j,0}^{-1}]$ can be used as a pivot to construct approximate conditional prediction intervals for the transaction rates. Let $\gamma_{j,n}(\alpha, w)$ denote the $100\alpha^{th}$ percentile of the gamma approximation (13) for $\widehat{\lambda}_{j,n}^{-1}$ with $\lambda_j = 1$. Then an approximate conditional prediction interval of level $1-\alpha$ for λ_p based on $\widehat{\lambda}_{j,n}^{-1}$ is

$$\left[\frac{\gamma_{j,n}(\alpha/2, w)}{\widehat{\lambda}_{j,n}^{-1} - (1-w)^{N_{+,n}}\lambda_{j,0}^{-1}}, \frac{\gamma_{j,n}(1-\alpha/2, w)}{\widehat{\lambda}_{j,n}^{-1} - (1-w)^{N_{+,n}}\lambda_{j,0}^{-1}} \right].$$

For large $N_{+,n}$, similar intervals can be obtained using (14).

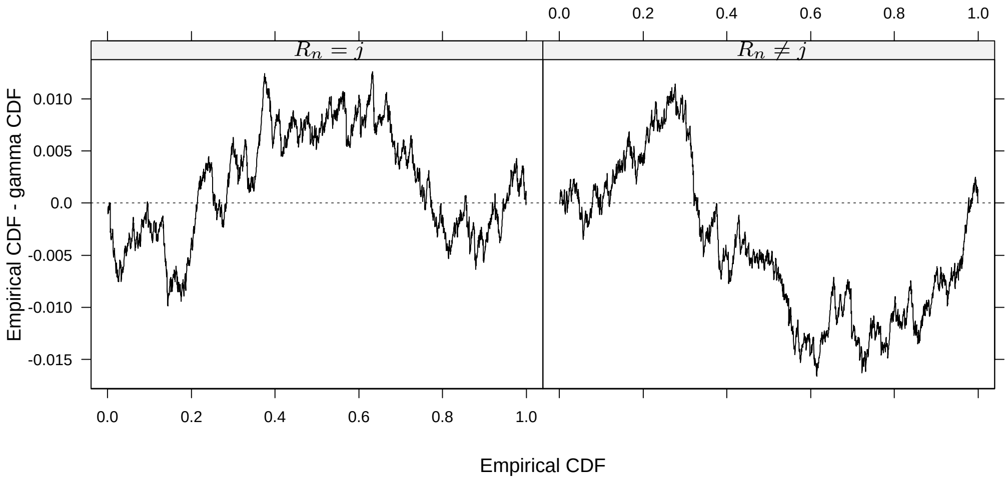

Figure 5: Difference between the empirical CDF and the gamma CDF for 2500 simulated values of $\widehat{\lambda}_{j,n}^{-1} - (1-w)^{N_{+,n}}\lambda_{j,0}^{-1}$ with $w = .02$ and $N_{+,n} = 10$, for $R_n = j$ and $R_n \neq j$.

Approximate conditional moments for the estimated period probabilities $\widehat{\pi}_{j,n}$ can be obtained from equations (9) and (10) (or equations (11) and (12)) by applying the *delta method* (Efron and Tibshirani, 1993) to equation (7). When transaction rates are constant across cycles, better approximate prediction intervals for period probabilities can be obtained from a parametric bootstrap using the approximate distributions (13) or (14) and the independence of the transaction rate estimators (Efron and Tibshirani, 1993).

In practice, transaction rates tend to vary slowly with time, reflecting changes in the environment and the individual, so the assumption of constant transaction rates across cycles may be reasonable for only a limited number of cycles. The approximations continue to hold locally, but $N_{+,n}$ needs to be redefined as the number of transactions during period j for the past κ cycles and $\lambda_{j,0}$ needs to be redefined as the estimated transaction rate for period j at the end of cycle $\max(0, C_n - \kappa)$ where κ is the number of cycles over which $\lambda_{j,n}$ is nearly constant.

8 Finite Sample Performance of the EDE

We use simulation to compare the performance of the EDE with updating weights of $w = .02, .05,$ and $.1$ with the performance of the MLE for the constant rate Poisson model of Section 3. In each simulation, there are 12 cycles (or weeks) of 7 periods (or days), and each period has the same transaction rate. Thus, the period probabilities are all 1/7. The Poisson timing model with constant period rates is simulated in Section 8.1, and a dynamic Poisson timing model is simulated in Section 8.2. Finally, a Poisson timing model with a deterministic shift in the period rates at the end of cycle 12 is simulated in Section 8.3. At the end of cycle 12, one period rate is increased by 25%, another decreased by 25%, and the rates for the other periods are unchanged. All period rates then remain constant for another 12 cycles.

The performance of the estimators changes with time, as more transactions are made and, under the dynamic model or shift model, as the underlying transaction rates change. If $\theta(t)$ is a parameter of interest (a rate or a period probability) at time t and $\widehat{\theta}(t)$ its estimate, then we define the *absolute relative error* at time t as

$$\epsilon\left[\widehat{\theta}(t)\right] = 100 \text{ E}\left(\left|\widehat{\theta}(t) - \theta(t)\right|\right)/\theta(t).$$

Under each scenario 10,000 paths of the transaction process were simulated under the different Poisson timing models, each path covering 12 cycles. For each simulated path, period rates and probabilities were estimated at each transaction and the corresponding absolute relative errors were evaluated on a fixed, equally spaced grid of 100 time points, using linear interpolation. This allows the absolute relative error curves from different transaction paths to be combined. The mean of the 10,000 absolute relative error curves are reported. The simulation standard deviations of the mean absolute relative error curves are also reported to give an assessment of the precision in the simulation results.

8.1 Poisson Model with Constant Rates

Four scenarios, with identical transaction rates λ fixed at constant values of $1, 5, 10,$ or 20 for each of seven periods, giving period probabilities of 1/7 for each period, were simulated. Estimated transaction rates for each period were initialized for each of the seven periods independently by randomly selecting a value in the interval $[.75\lambda, 1.25\lambda]$ according to a uniform distribution. This initialization procedure implies that, on average, the transaction rates are initialized at the correct value, but the initial rate for any period can be off by as much as 25%. In practice, initial transaction rates are derived from past customer data, so initial rates averaged over customers tend to be unbiased. Therefore, the simulated initialization procedure is realistic.

The absolute relative error curves for the MLE and the EDE, which are shown in Figure 6, suggest the following conclusions. (Table 1 gives the sample standard deviations of the 10,000 simulated absolute relative error curves for the MLE and the EDE for the different transaction rates.)

- The simulated absolute relative error of the MLE for the Poisson timing model decreases as the number of transactions increases.

- The smallest updating weight, $w = .02$, gives the best EDE under the constant rate model. During the first 12 cycles, the EDE with $w = .02$ is better than the MLE for $\lambda = 1, 5$; better or equivalent to the MLE for $\lambda = 10$; and worse than the MLE for $\lambda = 20$, after 4 cycles.

- The EDE with any of the three weights is better than the MLE during the initial cycles, even though the initial value assigned to λ is only correct on average. In contrast, as mentioned in Section 3, the MLEs for the period probabilities are undefined until the first transaction in the last period of the first cycle is observed or the first cycle ends.

- The absolute relative error rates for the EDE appear to stabilize at a value that does not depend on λ.

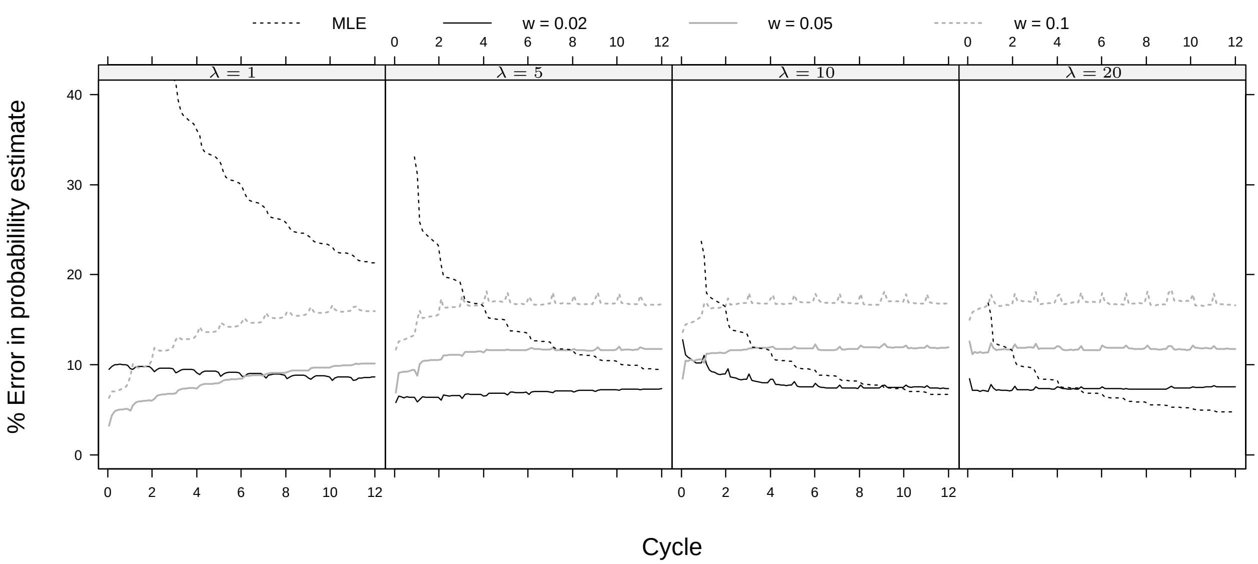

Figure 6: Absolute relative error curves for the first period probability for the MLE and the EDE with weights $w = .02, .05$, and $.1$ over 12 cycles of a Poisson timing model with constant rates. Each panel corresponds to a different transaction rate, which is held fixed for all periods.

These conclusions remain valid when the transaction rates for each period are constant across cycles, but are different among periods. Figure 7 shows the simulated absolute relative error curves for the estimated probability of period 1 under the same scenarios used to produce the results in Figure 6, but with the period

		EDE		
λ	MLE	$w = .02$	$w = .05$	$w = .1$
1	.25 (.16,.57)	.04 (.01,.05)	.07 (.03,.08)	.11 (.06,.13)
5	.11 (.07,.24)	.05 (.02,.06)	.09 (.06,.1)	.13 (.08,.14)
10	.08 (.05,.17)	.05 (.03,.06)	.09 (.08,.1)	.13 (.09,.14)
20	.05 (.04,.12)	.05 (.03,.06)	.09 (.07,.09)	.13 (.11,.14)

Table 1: Simulation standard deviations for the MLE and the under the Poisson timing model with constant rates. Reported values are averaged over the time grid; the ranges of the simulation standard deviations over the time grid are given in parentheses.

transaction rates $\lambda_1, \ldots, \lambda_7$ set to $\lambda_1 = .8\lambda$, $\lambda_2 = \lambda$, $\lambda_3 = 1.2\lambda$, $\lambda_4 = .5\lambda$, $\lambda_5 = 3\lambda$, $\lambda_6 = 2\lambda$, $\lambda_7 = .5\lambda$. As before, the *base* transaction rates are $\lambda = 1, 5, 10,$ and 20. The absolute relative error curves shown in Figure 7 are nearly identical to those in Figure 6, and the simulation standard deviations are nearly identical to those in Table 1.

8.2 Dynamic Poisson Timing Model

Four scenarios with initial transaction rates of $\lambda_0 = 1, 5, 10,$ and 20 transactions per period are considered. The EDE is initialized as in Section 8.1. The multiplicative noise term responsible for the evolution of the transaction rates in the dynamic Poisson model is simulated from a $\Gamma(\alpha, \alpha)$ distribution with large α ($\alpha \geq 400$), which means that the transaction rates evolve slowly. For large α, $\Gamma(\alpha, \alpha) \simeq \mathcal{N}(1, 1/\alpha)$, so the endpoints of a 99% prediction interval for the relative change are approximately $\pm 2.58/\sqrt{\alpha}$. The endpoints of the approximate 99% prediction intervals for the values of α used in the simulation, corresponding to high ($\alpha = 400$), moderate ($\alpha = 1600$), and low ($\alpha = 4000$) noise levels are, respectively, $\pm .129$, $\pm .064$, and $\pm .041$.

Figure 8 presents the absolute relative error rate curves for the MLE and the EDE for $j = 1$ for each combination of λ_0 and α. The main conclusions are as follows.

- When $\alpha = 400$ (high noise), the EDE performs better than the MLE. The absolute relative errors for the EDE stabilize around 20%, independent of λ_0, but the absolute relative error for the MLE appears to increase with time. The EDE with $w = .05$ is best, but the other values of w are nearly as good.

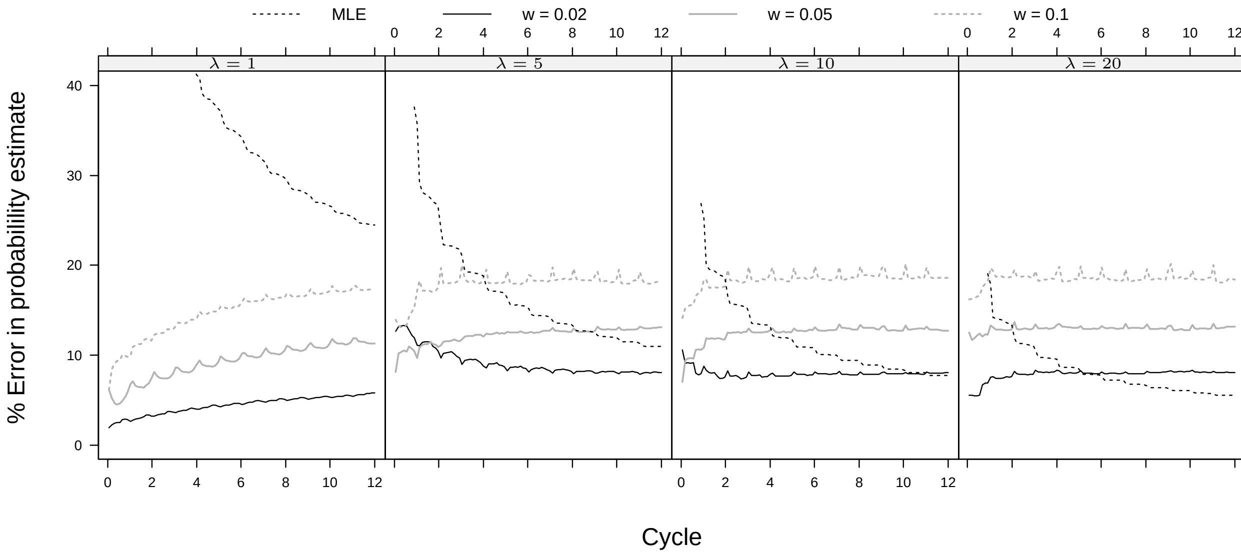

Figure 7: Absolute relative error curves for the first period probability for the MLE and the EDE with weights $w = .02, .05$, and $.1$ over 12 cycles of a Poisson timing model with constant rates, which differ among periods. Each panel corresponds to a different base transaction rate.

- When $\alpha = 1600$ (moderate noise), the EDE with $w = .02$ and the EDE with $w = .05$ have similar absolute relative error curves, but the other estimators have larger absolute relative errors. The absolute relative error for the MLE increases over time for $\lambda_0 = 20$.
- When $\alpha = 4000$ (low noise), the EDE with $w = .02$ has the best absolute relative error curve. The MLE has a similar absolute relative error curve (flat around 10%) for $\lambda_0 = 10,\ 20$.

The simulation standard deviations for the absolute relative error curves corresponding to $\alpha = 400$ are presented in Table 2. The simulation standard deviations for the absolute relative error curves corresponding to $\alpha = 1600$ and $\alpha = 4000$ fall between the values reported in Tables 1 and 2.

8.3 Poisson Model with Deterministic Shifts in Rates

Under this scenario, the period rates for the Poisson timing model are identical and constant for the first 12 cycles. At the end of the 12^{th} cycle, the first period rate is increased by 25% and the rate for the fourth period is decreased by 25%, so the cycle rate remains the same. The period rates then stay constant for the next 12 cycles. The rates used for the first 12 cycles are $\lambda = 1, 5, 10$, and 20 transactions per period. Each

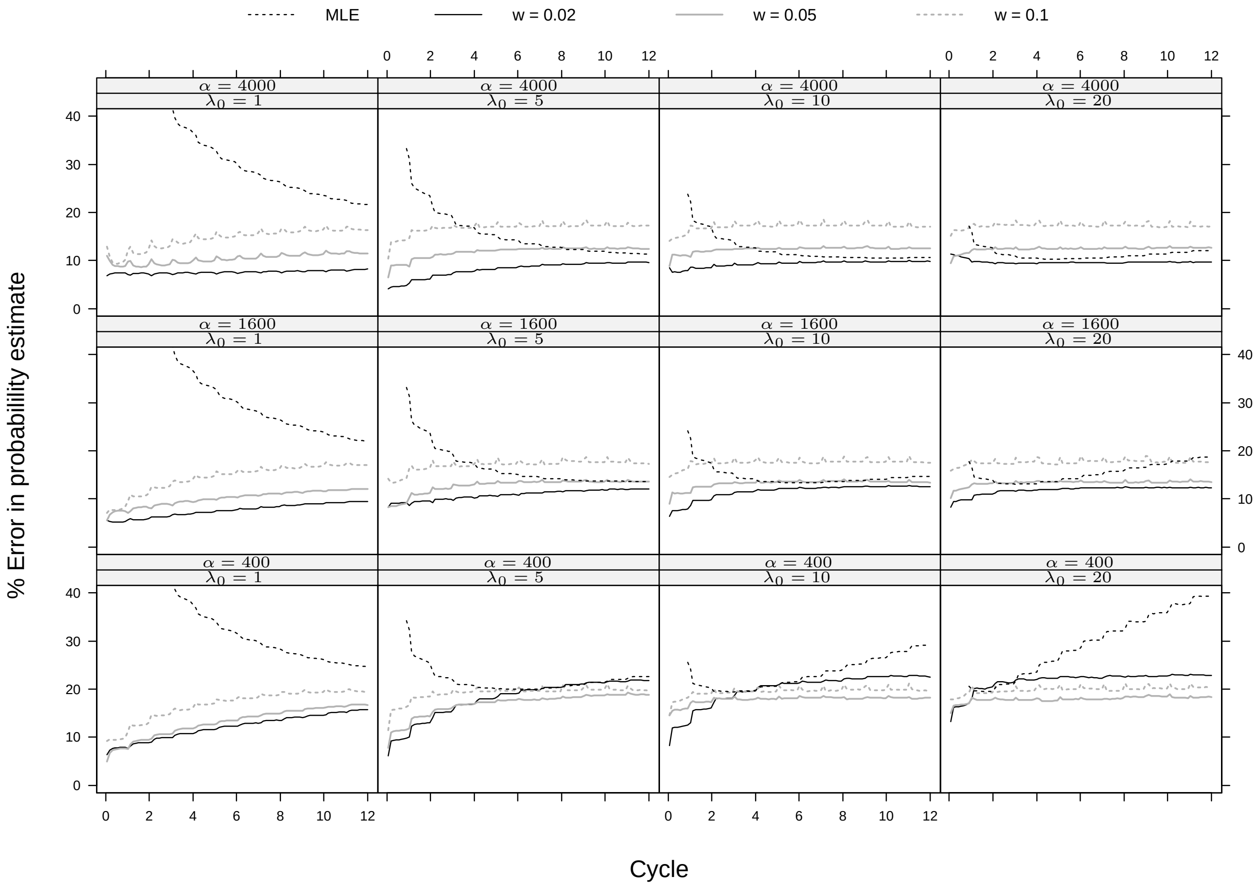

Figure 8: Absolute relative error curves for the first period probability for the MLE and the EDE with weights $w = .02, .05$, and $.1$ over 12 cycles of a dynamic Poisson timing model. Each panel corresponds to a different initial transaction rate λ_0, which is the same for all periods, and a different noise parameter α.

EDE is initialized using the procedure described in Section 8.1.

Figure 9, which displays the simulated absolute relative error curves for estimators of the first period probability, suggests the following.

- The EDE is able to absorb the shift in the rate considerably faster than the MLE.
- Estimators with larger w adapt faster, but the absolute relative error rate stabilizes at a large value.
- The EDE with $w = .02$ offers the best trade-off between ability to adapt to change quickly and low average absolute relative error over the period.

The simulation standard deviations for the absolute relative error curves are very similar to the ones reported

λ	MLE	EDE $w = .02$	$w = .05$	$w = .1$
1	.27 (.20,.58)	.11 (.04,.15)	.1 (.03,.13)	.13 (.05,.16)
5	.18 (.17,.26)	.15 (.05,.19)	.14 (.07,.16)	.16 (.08, .18)
10	.21 (.16,.28)	.18 (.08,.21)	.15 (.08,.16)	.16 (.13,.18)
20	.30 (.17,.49)	.20 (.11,.21)	.15 (.09,.16)	.16 (.13,.18)

Table 2: Simulation standard deviations for the MLE and the EDE under the dynamic Poisson timing model with $\alpha = 400$, corresponding to the highest relative change in the transaction rates. Reported values are averaged over the time grid. The ranges of the simulation standard deviations over the time grid are included in parenthesis.

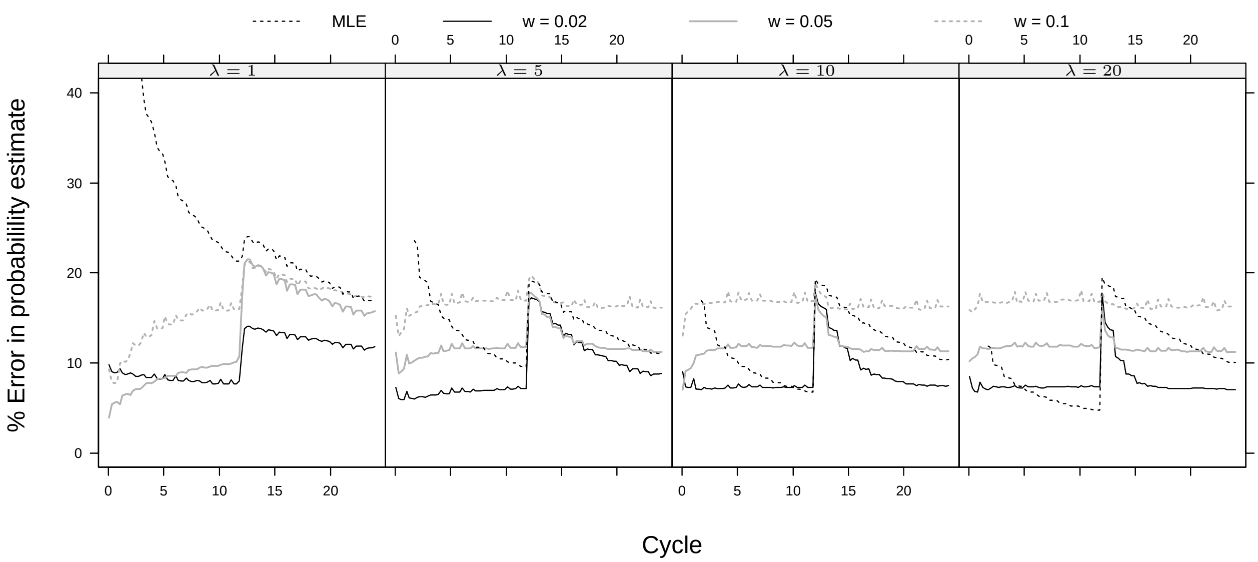

Figure 9: Absolute relative error curves for the first period probability for the MLE and the EDE with weights $w = .02, .05$, and $.1$ over 24 cycles of a Poisson timing model with deterministic shifts in the period rates at the end of cycle 12. Each panel corresponds to a different transaction rate for the first 12 cycles, which is the same for all periods.

in Table 1.

In summary, the MLE under the Poisson timing model with constant rates is not appropriate under dynamic Poisson models with high or moderate levels of noise and under Poisson models with shifts in the transaction rates. The EDE with $w = .02$ performs well under a variety of scenarios.

9 Discussion

Real-time applications involving very large databases of transactions may use customer signatures to track each individual's behavior. These signatures use fixed, generally small, storage space, so that they can be easily retrieved from the database for analysis or updating. Timing variables, such as day of week and hour of day, are an important part of customer signatures.

We describe a space-efficient, fast procedure for sequentially estimating timing distributions. Our procedure uses only the times of the current and the last transaction to update the transaction rates and the period probabilities and, therefore, can be used with *stream* transaction data that are continuously collected over time. The proposed estimation method is competitive with maximum likelihood under a Poisson timing model with constant rates and considerably better than the maximum likelihood estimates for the constant rate Poisson model when transaction times follow a dynamic Poisson timing model. In the application to tracking the calling behavior of 2,000 customers described in Section 6, the Poisson timing model was shown to be reasonable for most individuals. The event-driven estimate (EDE) of the period probabilities performed well even for those customers for which the Poisson timing model may not be adequate. We have used the EDE to estimate literally millions of timing distributions in several applications, and in each it has tracked the patterns of most customers well.

Further research is needed on methods for finding optimal updating weights for the EDE under a dynamic Poisson model. The simulation study suggests that a weight of about $1/\sqrt{\alpha}$ is appropriate, but more work is needed to develop methods that reliably produce the best updating weights.

The EDE presented in this paper uses fixed updating weights, but the methods we describe can be extended to incorporate updating weights that change with time. For example, if a transaction rate cannot be initialized reliably, then the updating weight can be larger for the first several transactions to allow the estimated rates and probabilities to move rapidly away from the initial value. Later, if the transaction rates are stable, the weight can be decreased to reduce the variability in the EDE. Or we might let the updating weight depend on the current estimate of the transaction rate. Even without extensions such as adaptive weighting, however, the EDE studied in this paper is attractive for estimating and forecasting timing patterns for many individuals in real time.

References

Abraham, B. and Ledolter, J. (1983). *Statistical Methods for Forecasting*, John Wiley & Sons, New York, NY.

Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988). *The New S Language: A Programming Environment for Data Analysis and Graphics*, Wadsworth, Belmont, CA.

Cahill, M. H., Lambert, D., Pinheiro, J. C. and Sun, D. X. (2000). Detecting fraud in the real world, *Technical report*, Bell Labs, Lucent Technologies.

Conover, W. J. (1980). *Practical Nonparametrix Statistics*, John Wiley & Sons, New York.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*, Chapman & Hall, New York.

MathSoft (1996). *Version 3.4 for Unix Supplement*, Data Analysis Products, MathSoft Inc., Seattle, WA.

West, M. and Harrison, J. (1989). *Bayesian forecasting and dynamic models*, Springer-Verlag.

APPENDIX

Result 1 *Given $\boldsymbol{\lambda}_{j,n}$, the waiting times $W_{+,L_i}, i = 1, \ldots, N_{+,n}$ between transactions in period j are independent* $\text{exponential}\,(\lambda_{j,L_i})$ *random variables under the dynamic Poisson timing model.*

By definition, $W_{+,L_i} = \sum_{m=L_{i-1}+1}^{L_i} Z_{j,m}$ and, from the assumptions of the dynamic Poisson timing model and conditional on $\boldsymbol{\lambda}_{j,n}$, the $Z_{j,m}$ corresponding to disjoint time intervals are independent. Because sums of disjoint independent variables remain independent, the W_{+,L_i} are independent.

The proof that $W_{+,L_i} \sim \text{exponential}(\lambda_{j,L_i})$ consists of showing that the corresponding survival function $P(W_{+,L_i} > t|\boldsymbol{\lambda}_{j,n})$ coincides with that of an $\text{exponential}(\lambda_{j,L_i})$, that is, $\exp(-\lambda_{j,L_i} t)$. First consider $0 \leq t < \delta_j - S_{L_{i-1}}$. In this case, because $Z_{j,L_{i-1}+1}$ is distributed as the minimum between an $\text{exponential}(\lambda_{j,L_i})$ random variable and the time left in period j, $\delta_j - S_{L_{i-1}}$,

$$P(W_{+,L_i} > t|\boldsymbol{\lambda}_{j,n}) = P\left(Z_{j,L_{i-1}+1} > t|\boldsymbol{\lambda}_{j,n}\right) = \exp\left(-\lambda_{j,L_i} t\right).$$

Now consider the case $t \geq \delta_j - S_{L_{i-1}}$ and let $p_t = \lfloor (t - \delta_j + S_{L_{i-1}})/\delta_j \rfloor$ denote the number of intervals of size δ_j in $t - \delta_j + S_{L_{i-1}}$. Then, the event $[W_{+,L_i} > t]$ is equivalent to the intersection of the events $\left[Z_{j,L_{i-1}+1} = \delta_j - S_{L_{i-1}}\right]$, $\left[\sum_{m=L_{i-1}+1}^{L_i} M_{j,m} = p_t\right]$, and $\left[S_{L_i} > t - (p_t+1)\delta_j + S_{L_{i-1}}\right]$. The first event is equivalent to *no transactions between $T_{L_{i-1}}$ and the end of period j* and, from the definition of $Z_{j,L_{i-1}+1}$,

$$P\left(Z_{j,L_{i-1}+1} = \delta_j - S_{L_{i-1}}|S_{L_{i-1}}\right) = \exp\left[-\lambda_{j,L_i}\left(\delta_j - S_{L_{i-1}}\right)\right].$$

The second event is *no transactions in period j for p_t successive intervals*, which has probability $\exp(-\lambda_{j,L_i} p_t \delta_j)$. Finally, if $T_{L_{i-1}}$ is the time of the last transaction in period j and no transactions occurred in the next p_t periods j, S_{L_i} is the minimum of the time elapsed in period j until the next transaction in that period and δ_j. Therefore, conditional on the first two events, the last event has probability $\exp[-\lambda_{j,L_i}(t - (p_t+1)\delta_j + S_{L_{i-1}})]$. Combining these results gives, for $t \geq \delta_j - S_{L_{i-1}}$

$$\begin{aligned} P\left(W_{+,L_i} > t|\boldsymbol{\lambda}_{j,n}, S_{L_{i-1}}\right) &= \exp\left\{-\lambda_{j,L_i}\left(\delta_j - S_{L_{i-1}}\right) - \lambda_{j,L_i} p_t \delta_j - \lambda_{j,L_i}\left[t - (p_t+1)\,\delta_j + S_{L_{i-1}}\right]\right\} \\ &= \exp\left(-\lambda_{j,L_i} t\right) = P\left(W_{+,L_i} > t|\boldsymbol{\lambda}_{j,n}\right), \end{aligned}$$

where the last equality follows from the fact that the conditional probability does not depend on $S_{L_{i-1}}$. Therefore, $P(W_{+,L_i} > t|\boldsymbol{\lambda}_{j,n}) = \exp(-\lambda_{j,L_i} t)$, for all $t \geq 0$, and the result follows.

Result 2 *If the transaction rate for period j is constant within a cycle, then, given $\boldsymbol{\lambda}_{j,n}$ and $N_{+,n}$, $W_{+,n}$ is independent of $W_{+,L_i},\ i = 1,\ldots,N_{+,n}$ and follows an* $\text{exponential}(\lambda_{j,L_{N_{+,n}+1}})$ *distribution when $R_n \neq j$.*

When $R_n \neq j$,

$$W_{+,n} = \delta_j - S_{L_{N_{+,n}}} + \delta_j \sum_{m=L_{N_{+,n}}+1}^{n} M_{j,m}. \tag{15}$$

Independence from $W_{+,L_i},\ i = 1,\ldots,N_{+,n}$ follows from the assumptions of the dynamic Poisson timing model because $W_{+,n}$ and W_{+,L_i} are functions of number of transactions in disjoint time intervals.

If the transaction rate for period j does not change within a cycle, $X_1 = \delta_j - S_{L_{N_{+,n}}}$ is distributed as a truncated exponential with parameter $\lambda_{j,L_{N_{+,n}+1}}$ in the interval $[0,\delta_j]$. That is, its density function is

$$f_{X_1}(t) = \begin{cases} \frac{\lambda_{j,L_{N_{+,n}+1}} \exp\left(-\lambda_{j,L_{N_{+,n}+1}} t\right)}{1-\exp\left(-\lambda_{j,L_{N_{+,n}+1}}\delta_j\right)}, & 0 \leq t \leq \delta_j \\ 0, & \text{otherwise} \end{cases}.$$

The summation on the right hand side of (15), $X_2 = \sum_{m=L_{N_{+,n}}+1}^{n} M_{j,m}$, follows a $\text{geometric}\left(1 - exp(-\lambda_{j,L_{N_{+,n}+1}}\delta_j\right)$ distribution and is independent of X_1. It is then easy to verify that, for $t \in [(k-1)\delta_j, k\delta_j) \quad k = 1,2,\ldots,$

$$\begin{aligned} P\left(W_{+,n} > t | X_1 = x_1\right) &= \begin{cases} P\left(X_2 > k\right), & (k-1)\delta_j \leq t < (k-1)\delta_j + x_1 \\ P\left(X_2 \geq k\right), & (k-1)\delta_j + x_1 \leq t < \delta_j \end{cases} \\ &= \begin{cases} \exp\left[-(k-1)\lambda_{j,L_{N_{+,n}+1}}\delta_j\right], & (k-1)\delta_j \leq t < (k-1)\delta_j + x_1 \\ \exp\left(-k\lambda_{j,L_{N_{+,n}+1}}\delta_j\right), & (k-1)\delta_j + x_1 \leq t < \delta_j. \end{cases} \end{aligned}$$

It follows that, for $t \in [(k-1)\delta_j, k\delta_j) \quad k = 1,2,\ldots,$

$$\begin{aligned} P\left(W_{+,n} > t\right) = \mathrm{E}\left(P\left(W_{+,n} > t | X_1\right)\right) =& \frac{\exp\left(-k\lambda_{j,L_{N_{+,n}+1}}\delta_j\right)}{1-\exp\left(-\lambda_{j,L_{N_{+,n}+1}}\delta_j\right)} \int_0^{t-(k-1)\delta_j} \exp\left(-\lambda_{j,L_{N_{+,n}+1}} x_1\right) dx_1 + \\ & \frac{\exp\left[-(k-1)\lambda_{j,L_{N_{+,n}+1}}\delta_j\right]}{1-\exp\left(-\lambda_{j,L_{N_{+,n}+1}}\delta_j\right)} \int_{t-(k-1)\delta_j}^{\delta_j} \exp\left(-\lambda_{j,L_{N_{+,n}+1}} x_1\right) dx_1 \\ =& \exp\left(-\lambda_{j,L_{N_{+,n}+1}} t\right). \end{aligned}$$

That is, $W_{+,n} \sim \text{exponential}(\lambda_{j,L_{N_{+,n}+1}})$.