Understanding Customer Retrials in Call Centers: Preferences for Service Quality and Service Speed

Kejia Hu
Operations Management and Quantitative Methods, Owen School of Management, Vanderbilt University, Nashville, TN 37215, kejia.hu@owen.vanderbilt.edu,

Gad Allon
Operations, Information and Decisions Department, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104, g-allon@kellogg.northwestern.edu,

Achal Bassamboo
Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, Evanston, Il 60208, a-bassamboo@kellogg.northwestern.edu

---

A. Illustration of Retrial Identification with Three Examples

Now we illustrate our retrial identification procedure with three examples. The first example shows that our method performs well on frequent callers. The second and third examples demonstrate that our approach also correctly distinguishes retrial calls from regular callbacks for customers making limited calls.

We begin with the first example illustrated in Figure 1. In the graph, we plot a customer's calling activity over time. Each point indicates a customer calling (y-value at 1) or not calling (y-value at 0) in each hour. The number in between indicates the gap (in days) between two consecutive calls. For example, 18 days are between the first and second calls, and 3.2 days between the second and third calls. Retrial calls are highlighted with dashed lines. We can see our change point detection identifying the first retrial as a call coming 0.2 days after the previous call. This is indeed a change in the calling frequency since all the previous gaps are several days long. The second, third, and fourth retrial calls are similar, coming in 0.1 days, <0.05 days, and 0.1 days after the previous call. Then we have three consecutive calls, with the initial two showing calling gaps of <0.05 days, while the third has 0.2 days. In this case, we see that our change point detection method only identifies the first call among the three consecutive calls as a retrial but not the later ones. This is because the calling frequency is already in the high-frequency zone, and no change point is detected. Hence we need the second statistical test to pin down a retrial window, which is the length of the high-frequency zone.

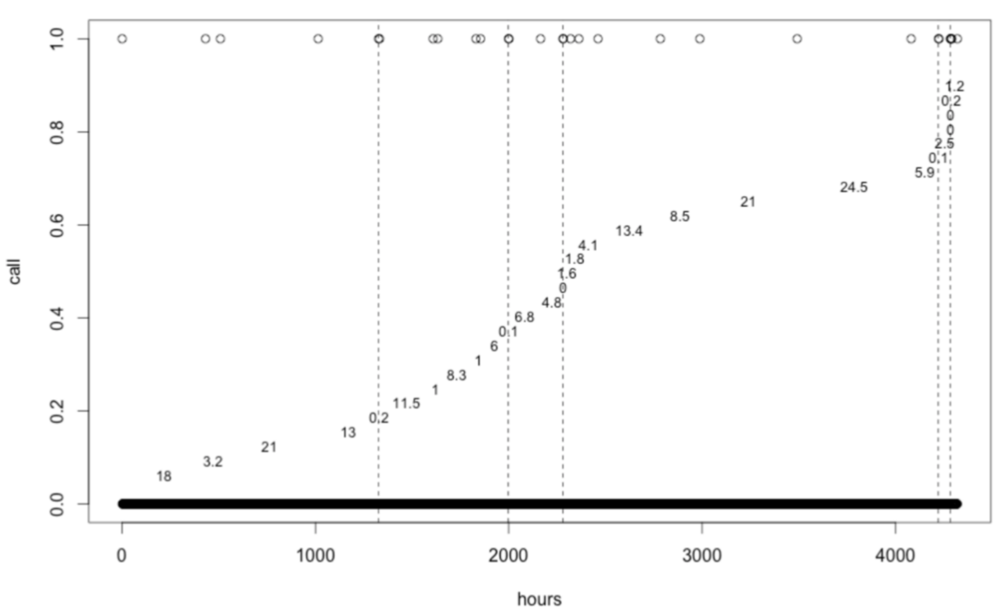

Figure 1 Example 1 for Retrial Identification

We proceed with the Kolmogorov-Smirnov test to examine the retrial window T_i^r. We want to find T_i^r which the Kolmogorov-Smirnov test concludes that $\lambda_{(0,T_i^r)} > \lambda_{(T_i^r,\infty)}$. In Table 1, we test the hypotheses one by one by changing the post-service period. For this customer, we conclude that the retrial window, T_i^r, is 2 days.

Null Hypothesis	The Alternative	P-value	Conclusion
$\lambda_{(0,1)} = \lambda_{(1,2)}$	$\lambda_{(0,1)} > \lambda_{(1,2)}$	$3.48 * 10^{-9}$	The arrival rate is **significantly** higher within one day after the previous call.
$\lambda_{(1,2)} = \lambda_{(2,3)}$	$\lambda_{(1,2)} > \lambda_{(2,3)}$	0.0017	The arrival rate is **significantly** higher within one to two days after the previous call.
$\lambda_{(2,3)} = \lambda_{(3,4)}$	$\lambda_{(2,3)} > \lambda_{(3,4)}$	0.1389	The arrival rate is **not significantly** higher within two to three days after the previous call.
$\lambda_{(3,4)} = \lambda_{(4,5)}$	$\lambda_{(3,4)} > \lambda_{(4,5)}$	0.6749	The arrival rate is **not significantly** higher within three to four days after the previous call.
$\lambda_{(4,5)} = \lambda_{(5,6)}$	$\lambda_{(4,5)} > \lambda_{(5,6)}$	0.8928	The arrival rate is **not significantly** higher within four to five days after the previous call.

Table 1 Kolmogorov-Smirnov test - Length of the illustrated customer's retrial window

We link this quantified retrial window with the retrials identified by our change point detection. A call that falls into the retrial window following an identified retrial is also considered a retrial. In Figure 2, we add the retrial window. Calls covered in the shaded area (the length of the customer's retrial window) are retrials. As stated in the formal definitions of retrials, we quantitatively define calls as retrials in two ways. They either are directly identified by the change point detection or fall into the retrial window following a retrial call.

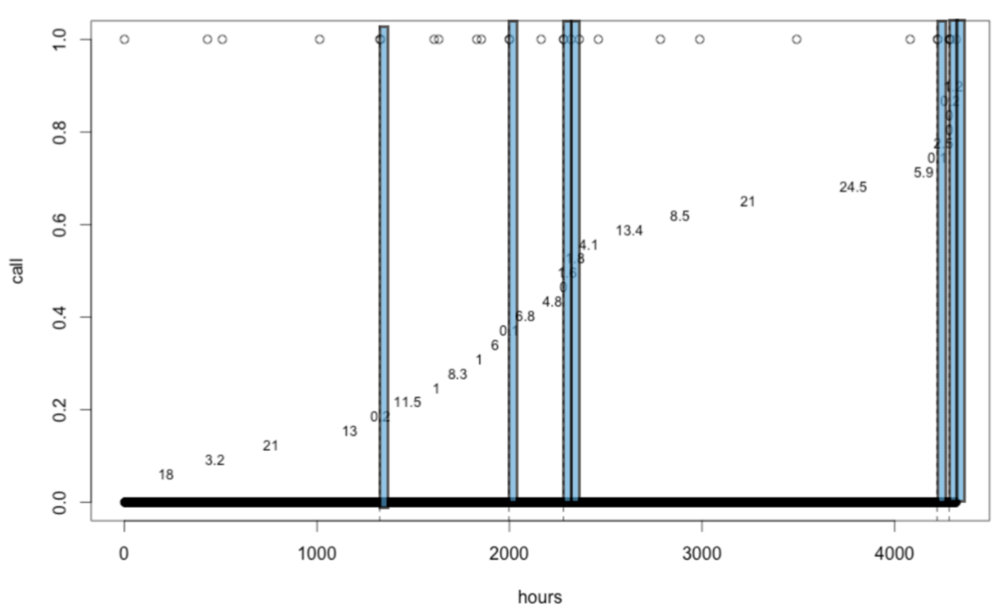

Figure 2 Example 1 for Retrial Identification with Retrial Window

Example 2 illustrated in Figure 3 is another case where calls are evenly spread out. We don't identify retrials with our algorithm. Hence we do not need a retrial window for this customer.

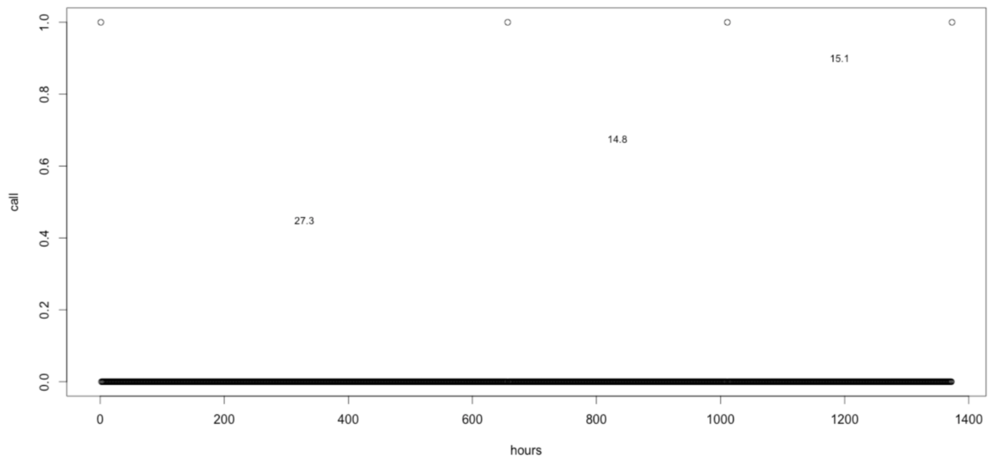

Figure 3 Example 2 for Retrial Identification — No retrial is identified

Example 3 in Figure 4 illustrates another case where the number of calls is not frequent, but we can still identify retrials. We proceed with the Kolmogorov-Smirnov test to pin down the length of the retrial window. In this example, the retrial window is 1 day.

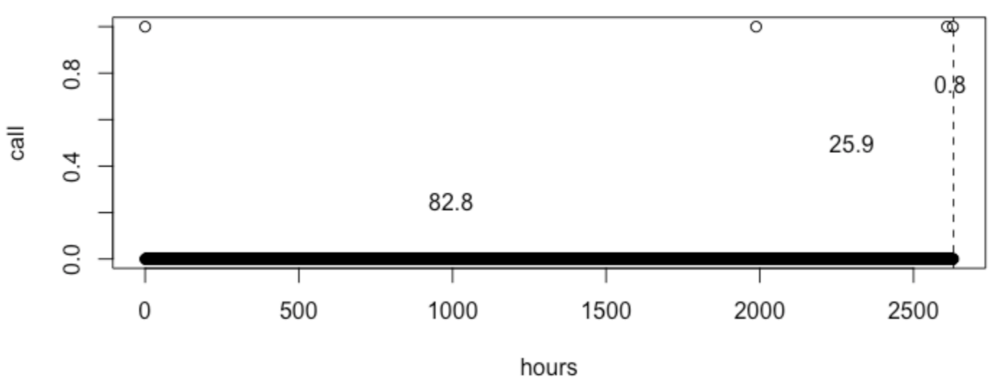

Figure 4 Example 3 for Retrial Identification — Calls are not frequent but a retrial is identified

Using these three examples, we want to achieve two purposes: illustrating our retrial identification approach and demonstrating its feasibility to identify retrials for both frequent and infrequent callers. In the next section, we further validate our retrial identification approach by using a subsample of calls with known retrial conditions.

B. Validation of Retrial Identification with Known-Request Calls

In this section, we validate our retrial identification's performance using a sub-sample of calls with known retrial conditions.

We observe customers' requests in some recorded calls. If a customer's two consecutive calls make the same request, the second call is noted as a retrial. Alternatively, if the second call is for a different request, the second call is not a retrial. The service requests in Table 2 reflect customers' actual service needs and will trigger retrials if not served appropriately. Service requests that are too ambiguous or not likely to trigger retrials, such as "general information," "current account," and "login with the initial password," are not on the list.

- Emergent mortgage	- Investment - foreign securities
- Foreign currency exchange rate, quote and deposit	- Investment - mortgages
- International Banking	- Investment - private banking
- Insurance - change policy	- Investment - provident fund
- Insurance - demands	- Investment - recruiting
- Insurance - dues inquiries	- Investment - trading room
- Insurance - endorsement	- Investment - value of the securities portfolio
- Insurance - new sales	- Loans
- Insurance - stance	- Password blocked/forgotten, release password
- Pensionary	- Postdated checks
- Securities portfolio	- Short-term deposit

Table 2 **Meaningful Service Requests**

B.1. Validating the accuracy of retrial identification

After identifying retrials qualitatively, we compare the classification with quantitatively classified retrials using our algorithm.

Qualitatively, there are 511,744 calls identified as retrials. Our algorithm identified 463,049 of those calls as retrials, an accuracy rate of 90.5%.

Qualitatively, there are 956,064 calls identified as retrials. Our algorithm identified 736,169 of those calls as retrials, an accuracy rate of 77.4%.

The overall accuracy rate of our identification approach is 82% $(= (463,049 + 736,169)/(511,744 + 956,064))$.

B.2. Validating the quantification of the retrial window

We also use calls with identifiable requests to examine the performance of our retrial window detection. In this section, we describe our validation procedure step-by-step and also show the examination results. Overall, we conclude that the retrial windows identified with our algorithm are statistically indifferent to the actual retrial window in the identified calls.

To begin with the validation, we first locate customers whose calls are recorded as meaningful service requests, as shown in Table 2. Altogether, there are 13,646 customers who have two consecutive calls that both are recorded with meaningful service requests. Among them, 52% are business customers and 48% are private customers.

Second, for each located customer, we measure his/her actual time gaps for retrials. If two consecutive calls are identified with the same meaningful service requests, we consider the second call as a retrial. By measuring the time gap between the first and second calls, we can obtain a data point for the actual retrial gap. Each customer can have several retrial gap measurements if they have multiple retrials.

In the third step, we compare customers' actual retrial gaps with our identified retrial window and summarize the results in Table 3. We first examine whether the identified retrial windows are too narrow. For each customer, we calculate the fraction of his actual retrial gaps that are less than our defined retrial window. We summarize the fraction from every customer located in the first step. The first row of Table 3 summarizes, across customers, the defined retrial windows cover the actual retrial gaps well. For more than 75% of the customers, the retrial window covers 95% of the actual retrial gaps, and for more than 50% of the customers, the retrial window's coverage is 100%. Secondly, to ease the concern that our identified retrial window is too loose, we perform another examination to check the difference between our defined retrial window and the largest actual retrial gap within the retrial window. The results are summarized in the second row of Table 3. Across all customers, the maximum time difference is 0.29 days. Since the unit of our retrial window is a day, these time differences indicate that the defined retrial windows are not loose. To summarize, when validating the defined retrial windows with the actual retrial gaps, we find that they are neither too narrow (confirmed by the first row of Table 3) nor too loose (confirmed by the second row of Table 3).

	Min	$Q1$	$Median$	$Mean$	$Q3$	Max
$Prob(l_i^R < T_i)$	57%	95%	100%	95%	100%	100%
$T_i - Max(l_i^R \| l_i^R < T_i)$	0.02d	0.05d	0.07d	0.12d	0.14d	0.29d

Table 3 Validation of the Identified Retrial Window

Note: l_i^R is a random retrial gap of customer i and T_i is his retrial window.

C. Validating the Assumption of a Constant Local Arrival Rate

In this section, we strengthen our retrial identification approach by validating its core assumption: that a customer has a local constant arrival rate when retrials are absent.

To examine the stable calling frequency prior to a retrial call, we perform a rolling-based comparison to investigate whether a customer's calling frequencies during Period t and Period $t-1$ are similar. Each period is measured as a day, i.e., the same unit used when quantifying customers' retrial windows. For each customer, we empirically obtain two time series, the calling frequency and the 1-period lagged series of it. We then perform a paired Wilcoxon test, a non-parametric test, to examine whether the difference between the two paired series is significantly different from 0. If the testing shows a small p-value, it would indicate that the calling frequencies of Period t and Period $t-1$ for this customer are significantly different. Otherwise, we would conclude that the calling frequencies during the period prior to a retrial call are stable. We summarize the testing result for each customer in Table 4. The p-values from all customers suggest that the calling frequency of Period t is statistically similar to the calling frequency of Period $t-1$ when retrials are absent.

Min	*Q*1	*Median*	*Mean*	*Q*3	*Max*
0.432	0.723	0.830	0.814	0.941	1.000

Table 4 Summary Statistics of P-values for All Customers

Null Hypo: A customer's calling frequency prior to a retrial call is stable. Reject the null if P-val is small.

We also confirm there are no changes from the external environment that may distort our change point detection method. Change point detection is used to detect an anomaly in an individual's calling frequency. If the system also simultaneously displays an anomaly in calling frequency, our detection of retrials may be biased, for example, mistaking a seasonality spike as an individual spike in his calling frequency. Through various statistical tests and visual checks, we confirm that there is no systematic change, such as seasonality, day of the week effect, or market events, that may distort our identification.

D. Robustness Checks

To check the robustness of our results, we validate our findings from three perspectives.

D.1. Distinct Preferences

First, we validate the distinct preferences between customer segments and across agent groups by demonstrating that the general model shown in our main paper fits the observations better compared to nested models where we restrict the difference in parameters across agent groups or customer segments.

The four model variations we examine are:

- Model – Different Agent Different Customer (our original model) where we consider the parameters to be different for each customer segment and agent group.
- Model – Same Agent Different Customer where we consider the parameters to be the same across agent groups but distinct between customer segments.

• Model – Different Agent Same Customer where we consider the parameters to be different across agent groups but no distinction between customer segments.

• Model – Same Agent Same Customer where we consider the parameters to be the same across agent groups and no distinction between customer segments.

The nested models' estimation results show findings consistent with our main model's regarding the distinction each can characterize between customer segments or across agent groups.

We compare the four models in Table 5 to investigate which better fits our data (using log-likelihood) and which best balances between data fit and model complexity (using AIC, the Akaike information criterion). The log-likelihood is a direct output from the estimation. AIC is computed as $2k - 2loglikelihood$, where k is the number of parameters. The best model is the one corresponding to the smallest AIC. From the calculation formula for AIC, we see that AIC rewards goodness of fit (assessed by the log-likelihood) but also penalizes models with too many parameters (such as our complex main model). In this way, AIC discourages overfitting. From Table 5, we find that the best model (based on the largest log-likelihood and smallest AIC) is our original model which distinguishes across agents and between customer segments.

Model	Diff Agent Diff Customer	Same Agent Diff Customer	Diff Agent Same Customer	Same Agent Same Customer
Log-likelihood	**-3652121**	-4645198	-7020930	-7148419
Number of Parameters	26	14	13	7
AIC	**7304295**	9290424	14041886	14296852

Table 5 Comparison across Four Candidate Models. The first column shows the main model in the paper. The following three columns show three nested models.

D.2. Frequent Callers

In this subsection, we examine whether our findings are sensitive to customers' knowledge of the call center's service design by using sub-samples of calls from frequent callers.

First, we sort customers based on their calling frequency in our observation period, then construct the following sub-samples:

• Top 1%: this sample is composed of calls from customers whose calling frequency is ranked as the top 1% in the corresponding segment. The sample size is 244,079 calls for the private segment and 12,920 calls for the business segment.

• Top 5%: this sample is composed of calls from customers whose calling frequency is ranked as the top 5% in the corresponding segment. The sample size is 600,912 calls for the private segment and 25,593 calls for the business segment.

- Top 10%: this sample is composed of calls from customers whose calling frequency is ranked as the top 10% in the corresponding segment. The sample size is 830,653 calls for the private segment and 32,002 calls for the business segment.
- Top 25%: this sample is composed of calls from customers whose calling frequency is ranked as the top 25% in the corresponding segment. The sample size is 1,263,334 calls for the private segment and 45,376 calls for the business segment.

Second, we estimate our structural model using the four sub-samples and compare their results with our main model's. All key findings we highlight in Section 6 remain the same.

D.3. Known Retrials and Non-Retrials

Third, we examine our findings by using sub-samples of calls with qualitatively defined retrials and non-retrials.

First, we extract sub-samples from the dataset. The unit of analysis of our structural model is one call episode. To construct a sub-sample covering only correctly identified retrials, we pick episodes where 1) all the retrials and non-retrials are labeled based on service requests; 2) our algorithm correctly labels all retrials and non-retrials. To construct a sub-sample with wrongly identified retrials, we pick episodes where 1) all the retrials and non-retrials are labeled based on service requests, and 2) our algorithm incorrectly labels at least one non-retrial as a retrial or one retrial as a non-retrial. For the correctly identified sub-sample, we have 43,986 private customer episodes and 6,416 business customer episodes. For the wrongly identified sub-sample, we have 30,916 private customer episodes and 1,589 business customer episodes.

Second, we estimate our structural model using the two sub-samples of correctly and incorrectly identified calls and compare the results with our main model's. We find that the key findings we highlight in the paper remain the same.