Enhancing the Predictive Power with Time-Adjusted Residual Effect Simulations: Recency-Frequency and Time Regularity Measurements

Doyeon Kwak
AI Center
CJ
Seoul, South Korea
doyeon.kwak@cj.net

Se-Un Park
AI Center
CJ
Seoul, South Korea
su.park5@cj.net

Sunbum Youn
AI Center
CJ
Seoul, South Korea
sunbum.youn@cj.net

Emma Bergqvist
AI Center
CJ
Seoul, South Korea
e.bergqvist@cj.net

***Abstract*— This paper introduces a novel methodology to enhance the prediction of new events in a future period by developing more suitable indicators that address the limitations of traditional Recency-Frequency-Monetary (RFM) frameworks. By differentially simulating the residual effects of identical (or similar) events based on their occurrence times, we propose an efficient approach to predicting the likelihood of new events by supplementing the Recency-Frequency (RF) function. This methodology mitigates the information loss in the Recency measure and accommodates variations in event occurrence times. We further develop the Time Regularity measurement, leveraging residual effect simulations of the events to overcome the limitations of conventional periodic indicators, such as clumpiness, which solely rely on event occurrence frequency and time intervals between the occurrences. The Time Regularity function provides a more comprehensive measure of event occurrence density within an observed period, as the residual effects increment over time. Our proposed methodologies are optimized for calculating the probability of event occurrences in a given future time frame, via a more comprehensive set of information regarding individual event occurrence times and resulting residual effects. They have potential for broad applications across various domains, including customer lifetime value analysis in marketing, medication residual rate and intake timing analysis in pharmaceutical contexts, predictive maintenance analysis such as prognostics and health monitoring, and many other domains under time series analysis. This study contributes to the advancement of predictive analytics by offering more accurate and robust information for anticipating future events.**

Keywords*— *clumpiness, RFM, feature engineering, predictive analysis

I. Introduction

The ability to predict short-term future outcomes by inference from the occurrence of repeated events for subjects throughout the experimental period is of significant importance. Researchers collect data on occurrence time, frequency, and intensity to accomplish this and employ the RFM analysis (Recency, Frequency, and Monetary) [1,2]. Moreover, entropy-based methods, such as clumpiness, have been used to quantify event distribution within a specific time frame [3]. As well as, a model that leverages timing irregularities to improve predictions of customer activities—whether they exhibit regular, random, or clumpy timing patterns—has been proposed using probabilistic approaches such as Markov Chain Monte Carlo methods, demonstrating the relationship between clumpiness and the regularity parameter [4]. These models show broad applications especially in customer behavior analytics and predictions in marketing [1,3] and other domains such as prognostics and health monitoring [5,6].

However, both conventional RFM analysis and the clumpiness index have limitations in preserving the detailed time-series information of transaction data. Recency element in the widely used RFM methodology is constrained by its narrow focus on the most recent event occurrence time, leading to significant information loss. Many researchers redefine their Recency values to fit their research purpose, for instance, averaging the time distances of multiple events to the final observed time, or accounting only events that exceed a designated threshold [7,8].

Additionally, the clumpiness index, which relies solely on event frequency and the inter-event times, or time intervals between event occurrences, is best suited for events that are independent of one another [9]. Consequently, the clumpiness index faces challenges in both statistical and practical aspects [10]. Also, when an event occurrence influences the subject or residual effects persist after the event, the likelihood of future events will change accordingly. If event occurrence times within an observed period can be obtained during the experiment, it is essential to implement a methodology that simulates posterior effects due to events based on their occurrence times.

To address the limitations of existing methods, we propose a novel Recency-Frequency (RF) metric that preserves the time-related information of all events, not just the most recent. This approach blends the RFM's Recency with Frequency, significantly improving upon current measures that neglect the occurrence time information from earlier events. Moreover, we aim to develop a Time Regularity (TR) measurement that accommodates the time-shifts information on event occurrences, allowing the analysis to consider when event aggregations or "clumps" appear within the observation period.

To achieve this, we will incorporate the concept of residual effects for each recorded event, incrementally adjusting the impact of an event based on the event occurrence time relative to the observed duration. The proposed residual effects increase over time, with the decay rate diminishing for more recent events, applying the Recency attribute. Utilizing this technique, the proposed Recency-Frequency function can effectively measure Recency of all events by aggregating their residual effects that change according to the various event occurrence times (Fig. 1). Additionally, the Time Regularity function indicates the density of event occurrences across designated time intervals by calculating the area under the curve for a series of residual effects (Fig. 1). These methodologies demonstrate the potential to significantly enhance the foundational information for predicting the likelihood of event occurrences in the future.

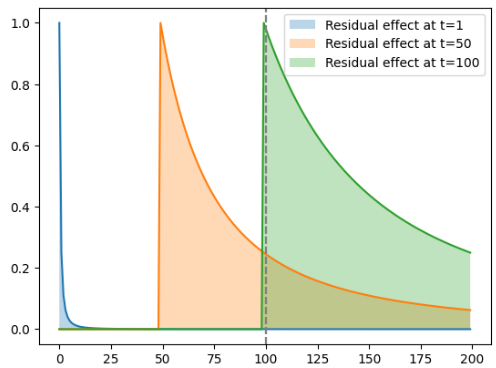

Fig. 1. Simulated residual effects of three event occurrences at t=1, 50, and 100 with observed time duration of 100 time-units. As the event occurrence times shift, the exponential decay of the residual effects is reduced, resulting in distinct residual impacts for events at various timestamps. Note that the time duration used to simulate residual effects is intentionally set twice the length of the observed time duration, at 200 time-units to maximize the residual effect for the latest possible event occurrence at T.

II. Addressing the limitations of Recency in RFM

The RFM analysis, widely used for predicting the likelihood of event occurrences in future periods, is based on extracting the elements of Recency, Frequency, and Monetary value (amount or event intensity attribute) from observed subjects during a designated time frame [11]. Recency measures the time interval between the most recent event occurrence and the latest observed time, while Frequency accounts for the number of event occurrences for a subject within the observed duration. Monetary value represents the cumulative impact attributes of the events. Subjects with more recent, more frequent, and greater attribute values within the observed duration are assumed to have a higher probability of event reoccurrence in the future [2].

Among these elements, Recency concentrates exclusively on the information from the most recent event occurrence, neglecting data related to all other events. While numerous experiments have shown that the timing of the most recent event significantly influences the likelihood of future event occurrences as noted by Wei [8], the conventional Recency variable has clear limitations due to the loss of information associated with the occurrence times of the other events.

III. Problems with the existing periodicity indicator, clumpiness

The clumpiness index has been developed to measure the distribution of events occurring within a specific period to infer periodicity [9]. There are various models for calculating clumpiness, including entropy and log-utility models, but all of them rely solely on the occurrence frequency and the time intervals between occurrences. Clumpiness application is limited to cases where the effects of each event are independent of one another. When event occurrences influence the observed subjects and residual effects persist after the events, potentially impacting future event occurrences, the effectiveness of the clumpiness index diminishes. If the event occurrence times within an observed period can be obtained during the experiment, it becomes possible to simulate their residual effects, taking into account the varying occurrence times.

To resolve these issues, we aim to create a periodicity measurement that incorporates residual effect simulations, which vary based on the event occurrence times, in predictive analyses with residual effects. In doing so, we can effectively account for the interconnectedness of events.

IV. Formulating Recency-Frequency and Time Regularity

Although the conventional RFM methodology may assign identical Recency and Frequency values to some subjects, their future characteristics could still vary due to overlooked factors. Particularly, when extracting Recency data, it is within our reach for us to acquire the observed occurrence times for all events, rather than just the most recent one. The differences in occurrence times lead to variations in the overall residual effects, which take into account the proportionate influence of individual events calculated using our Recency-Frequency measurement, and the relative distribution density of the repeated event occurrences, or Time Regularity.

For each subject, the residual effect of an event occurring during the observed period (T) is calculated by summing the exponent distributions, where the stationary event occurrence time (t_j) is divided by incrementally increasing time (i). As the occurrence time is divided by the incrementally increasing denominator through iteration ($i = 1 \rightarrow 2T$) and then raised to the power of two, this process creates the decay of the residual effect. This exponent is an adjustable parameter depending on the degree of residual effect decay. As shown in Equation 1, this allows the initial residual effect at the occurrence time to start at 1, and decay at a rate based on the point in time (i). This calculation compensates for Recency of each event, with later occurrence times yielding a higher numerator value, resulting in an increased area under the exponential decay curve (Fig. 1). Despite the fact that the earlier event would have a longer participation period, its residual effects would be less than that of the possible most recent event, over the duration of 2T.

The summation of residual effects of all observed events within the observation period accounts for the Frequency (n) of the events. This creates the Recency-Frequency (RF) measure, which captures not only the Recency of a single recent event, but also for all events occurring at different times (Fig. 2). The RF value can be divided by the frequency of events (n) to normalize the results, providing a new Recency indicator (RF/F). For example, for event j, we compute the RF integration of the time residual effects of $(t_j/i)^2$ over $2T$, with the observed time duration T or current time T, only after event j occurred at t_j. Then, we consider all events $j=1,..,n$.

Furthermore, by multiplying the potential (P_j), which represents the appropriate magnitude for each residual effect calculation for an event j, the combined effect of RF and P can be determined. As residual effect is calculated for each event, the influence of each event is incorporated differently. In the context of customer value analysis, this approach enables the integration of RFM into a single value.

[redacted] (1)

RF = Recency-Frequency index, n = number of events, j is iteration from 1 to n, T = observed time duration, i is iteration from 1 to 2T, t_j is occurrence time of event j, P_j is potential at t_j (in terms of RFM, Monetary value. If repeated events, set P_j=1 to negate effects)

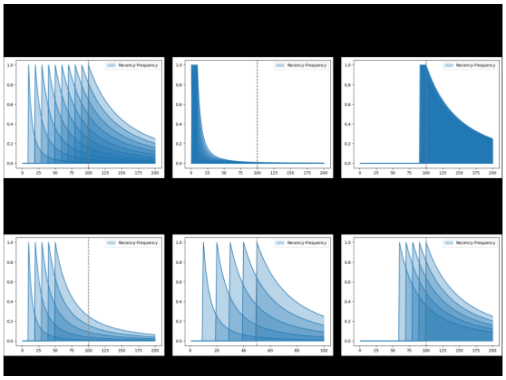

Fig. 2. Recency-Frequency of various cases of event occurrences within observed time duration (T) and the visualization of the residual effects that make up the RF value. The iteration is extended to 2T in order to maximize the residual effects for the latest possible event occurrence at T. P described in Equation 1 is set to 1 for all cases.

Clumpiness, one of the most commonly used irregularity measurements, calculates the sum of inter-event times among a series of events by employing an entropy function as a means to quantify the uncertainty in event distribution [9]. Though it is a straightforward technique to generate a summary variable for a series of events over time, its application hinges on a critical assumption: that the events do not impact the subjects or their future intentions concerning the timing of occurrences.

Furthermore, although clumpiness has the advantage of measuring relative clustering of event occurrences to infer the periodicity of the measured events, the fact that all participating subjects are assumed to have the same participation duration makes only relative comparisons possible: when there are subjects with different entry and exit points during the observation period (e.g., online platform registration times, various drug administration times, etc.), it is impossible to make an absolute comparison as only the ratio of event aggregation is shown between at least two events. In other words, if it is not applied to subjects with similar participation periods, the results may skew.

The newly devised Time Regularity function (TR) aggregates the magnitude of residual effects differently depending on their event occurrence times and inter-event time durations over the overall observed duration, taking into account the relative entry and exit points of participating subjects (Eq. (2)).

The residual effect calculation works the same way as the proposed Recency-Frequency measurement. Time-unit (i) iterations are carried out within the observed period, and the residual effect of each event is quantified using the latest event occurrence time at t_i divided by i and raised by the exponent. The event occurrence time (t_i) divided by i at i is fixed to have a value of 1, as the numerator and the denominator are equal to one another in that instance. The calculation decays over the iteration, with the latest event occurrence time remaining constant until a new event occurs. This allows the residual effects to persist for the inter-event time duration or when the considered duration ends at 2T. The summation of these residual effect fluctuations during the inter-event times quantifies event density over time, inversely similar to prior clumpiness derivations. The main distinction is that our method differentiates the residual decays based on timing of the occurrences, emphasizing the reward for later event residual effects while diminishing the impacts of earlier event residual effects when measuring the event distribution. This method highly depends on the time-unit used to calculate the Time Regularity, which we can normalize by dividing by the maximum possible residual density outcome (R_{max}), the condition where every time-unit iteration has a new event occurrence.

Subsequently, the residual effect decay calculation, which is based on individual event occurrence times, resets upon a new event occurrence, as shown in Fig. 3. The TR function utilizes this reset mechanism to penalize clumpy phenomena. The overall distribution of residual effects along the observed period, the TR measurement, is determined by adding the residual effects of each event to express only during the corresponding inter-event time. This method promotes having higher frequency, evenly distributed events to maximize the expression of residual effects to increase the value of Time Regularity. However, since the residual decay is minimized, and the effect is maximized as the observed period approaches its end, late event aggregations lead to early termination of high residual effects through the resetting process.

The core principle of the TR function is to simulate residual effects by calculating their cumulative amount during a specific observed period, factoring in the element that mitigates the decay rate using the time elapsed up to the current iterated time (i) as the denominator. In other words, it serves as a time irregularity indicator that accommodates various participation durations of the subjects by altering the residual effects based on the event occurrence time from the subject's perspective. This sets it apart from existing clumpiness indicators that only permit relative comparisons between participants with the same participation period, without taking the occurrence times and the residual effects into account.

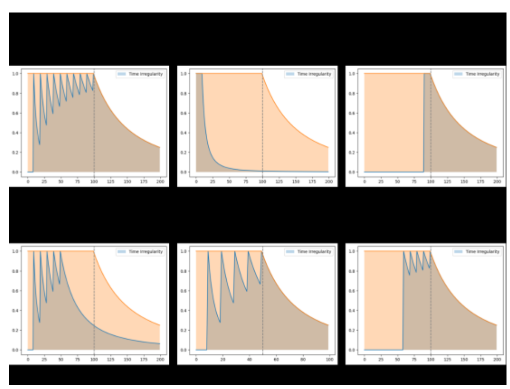

Fig. 3. Time Regularity of various cases of event occurrences within observed time duration (T) and the visualization of the curve resulting from residual effects that make up the Time Regularity value (in blue). The orange areas represent R_{max} for each case. The iteration is extended to 2T in order to maximize the residual effects for the latest possible event occurrence at T.

[redacted]

[redacted] (2)

TR = Time Regularity index, R = Cumulative residual effects, R_{max} = Cumulative residual effects when there are event occurrences for all $t_i \leqslant T$, T = observed time duration, i is time iteration from 1 to 2T, t_i is occurrence time of an event

V. Designing Experimental Cases

To validate our proposed methods for calculating Recency-Frequency and Time Regularity measurements, we have designed nine experimental cases featuring extreme event occurrence distributions, as demonstrated in Table 1 and visualized in Fig. 4. Cases A through E display a shift in event aggregations from the beginning to the end of the observed duration, with Case C presenting an even distribution of events. Case F mirrors Case E but includes a late entry point, resulting in an even distribution. Case G and H offer a slight modification to Case C's even distribution, with the second event occurring at a time index of 19 rather than 20 for Case G, and the second event occurring at a time index of 21 instead of 20 for Case H, in order to test the continuity of the measurements. Case I exhibits an even distribution with an inter-event time of 9 instead of 10, allowing us to assess the convergence of the Time Regularity indicator.

Furthermore, we compare the characteristics of Time Regularity with the four clumpiness property checks, as proposed by Zhang, Bradlow, and Small [9], to determine the alignment of our approach with existing methodologies.

TABLE I. Experimental Cases

	Observed duration (T)	*Occurrence time Indices*	*Number of events (n)*
A	200	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20]	20
B	200	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, **200**]	20
C	200	[10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200]	20
D	200	[**1**, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200]	20
E	200	[181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200]	20
F	20	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20]	20
G	200	[10, **19**, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200]	20
H	200	[10, **21**, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200]	20
I	200	[19, 28, 37, 46, 55, 64, 73, 82, 91, 100, 109, 118, 127, 136, 145, 154, 163, 172, 181, 190]	20

Nine experimental cases for validating Recency-Frequency and Time Regularity measurements. Abnormal occurrence times are highlighted in bold.

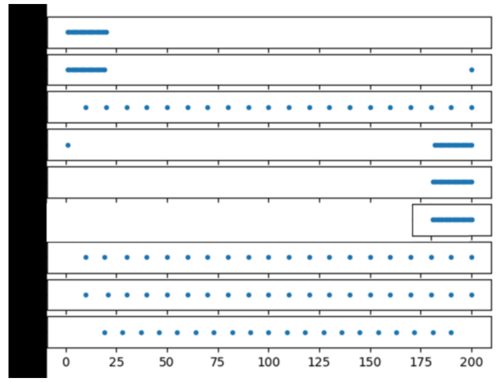

Fig. 4. Visualization of data points in each experimental case from Table 1.

VI. Recency-Frequency Validity Tests

The Recency-Frequency measurement is maximized when event aggregations occur at the end of the observation period and minimized when they take place at the beginning of the observation period, assuming the event frequencies remain the same. This effect should be apparent even by the minimal shift in occurrence times.

Moreover, an earlier entry point reduces the residual effects by increasing the iterated time duration, represented by the denominator i. For subjects with identical event distributions, a late entry point significantly enhances the relative residual effect due to the shorter observation time duration.

Within the nine experimental cases, if the following relationships are satisfied, the Recency-Frequency measurement conforms to the designated properties:

- A < B < C < D < E: The time-shift in event aggregations throughout the observed duration positively affects residual effects, causing an increase in Recency-Frequency.
- E < F: Late entry points for subjects should enable calculations based on the relative observation duration.
- G < C < H: Continuity of the RF value should be present incrementally increasing the RF value as event occurrence time is delayed.

TABLE II. Recency-Frequency Test for Experimental Cases

	Recency-Frequency (RF)	*Recency-Frequency/F (RF/F)*	*Recency*	*Average IET*	*Average Recency*	*Time Regularity (TR)*
A	1.2734	0.0637	180	1	189.5	0.1285
B	1.7574	0.0879	0	10	180.5	0.4550
C	8.3149	0.4157	0	10	95	0.8910
D	11.3855	0.5693	0	10	18.5	0.4014
E	11.9707	0.5985	0	10	9.5	0.3993
F	149.7113	7.4856	0	1	9.5	1
G	8.3095	0.4155	0	10	95.05	0.8912
H	8.3202	0.4160	0	10	94.95	0.8906
I	8.4746	0.4237	10	9.5	95.5	0.8489

Values for Recency-Frequency (RF), normalized Recency-Frequency (RF/F), conventional Recency (R), average inter-event time, average time interval from all events to the end of the observation time (Average Recency), and Time Regularity are measured.

Event frequency directly contributes to the number of residual effects that comprise the Recency-Frequency measurement. As a result, it is possible to normalize the RF value by dividing it by the event frequency to obtain a new normalized Recency value, RF/F.

While Recency-Frequency (RF) and normalized Recency (RF/F) increase as events occur later in the observed duration, conventional Recency and average inter-event time lose occurrence time information in Cases A through E. The average time interval from events to the end of the observation time (average Recency), proposed by Hsieh [8], displays a downward trend as events are shifted in time. However, the single outlier in Case B compared to Case A and the single outlier in Case D compared to Case E contribute vastly different relative proportions to Case B and Case D, even though the numerical difference is the same at 9. In the comparison between Cases A and B, the outlier causes 4.75% decrease, while in the comparison between Cases E and D, the outlier creates 94.74% increase in the value. Time Regularity assesses the distribution of the events and is not suitable for accurately evaluating the Recency and Frequency aspects of the cases.

In the more detailed condition to observe the continuity property, we can see an incremental increase in the order of Cases G < C < H for both RF and RF/F. Additionally, there is an inverse decreasing relationship for the average time interval from events to the end of the observation time (Average Recency) and Time Regularity.

The late entry of a participant should be considered within the participant's relative time frame and assess the Recency and Frequency within the shorter observed duration. Comparing Case E and Case F reveals the same events with different observed durations. Having a higher value for Case F with more events per time duration holds true for Recency-Frequency and RF/F, as well as Time Regularity. Conventional Recency, average IET, and the average time interval from events to the end of the observation time (Average Recency) cannot account for the relative observed duration.

The normalized Recency value, along with the RF value and RFP value, which includes the magnitude P, can be used in conjunction with conventional RFM analysis to enhance the available information for making predictions.

VII. Time Regularity validity tests

The four clumpiness properties proposed by Zhang, Bradlow and Small [9] in their research are as follows:

- Minimum. The measure should be the minimum, if the events are equally spaced.
- Maximum. The measure should be the maximum if all of the events are gathered together.
- Continuity. Shifting event times by a very small amount should only change the measure by a small amount.
- Convergence. As events move closer (further apart), the measure should increase (decrease).

The Time Regularity index, which increases when more residual effects are expressed throughout the observed period due to regularized event distribution, has an opposite direction to the conventional clumpiness indices including entropy-like clumpiness. Therefore, the minimum and maximum properties of clumpiness should be inversely related to Time Regularity.

The Time Regularity only approximately satisfies the minimum property of clumpiness, as the residual effects increase over the observed duration, requiring slightly larger inter-event time spacing for later events to maximize Time Regularity. However, Time Regularity conforms to the maximum property for the events clustered at the beginning of the observed duration, because the residual effects are minimal for the earliest times. These are further curtailed by the reset caused by subsequent residual effects except for the latest one, as seen in Case A and B in our experiments.

The continuity property aligns with Time Regularity, as its residual effect shifts the same amount for both the numerator and denominator sides. The convergence property of clumpiness is consistent with the inverse of Time Regularity, as the reset mechanism diminishes the expression of residual effects when events are clustered together.

These properties are proven through the relationship between Time Regularity values of the nine experimental cases described in Table 1.

- Among Cases A through E, C with an even event distribution should exhibit the highest value for Time Regularity (lowest clumpiness).
- A < E, B < D: In terms of future likelihood, Time Regularity should favor E over A concerning the Recency effect (A > E for clumpiness).
- C < F: Despite having even event distributions with the same event frequency, a shorter observed duration window should increase the relative contribution of each event.
- G < C < H: The minor shifts of an event one-unit time should display minor shifts in the regularity value satisfying the continuity property.
- I < C: While showing even event distributions, I brings events closer together, thus decreasing Time Regularity (increase clumpiness).

TABLE III. Time Regularity test for experimental cases

	ENTL	*ENT*	*2M*	*LU*	*3LC*	*Time Regularity (TR)*	*Recency-Frequency (RF)*
A	0.8327	-0.5033	0.0005	100.6680	0.015	0.1285	1.2734
B	0.8100	-0.5672	0.8195	95.4695	0.915	0.4550	1.7574
C	0.0532	-2.8459	0.0475	56.9189	0.150	0.8910	8.3149
D	0.8100	-0.5672	0.8195	95.4695	0.915	0.4014	11.3855
E	0.8327	-0.5033	0.0005	100.6680	0.015	0.3993	11.9707
F	0.0805	-2.8459	0.0475	56.9189	0.150	1	149.7113
G	0.0533	-2.8454	0.0476	56.9290	0.155	0.8912	8.3095
H	0.0533	-2.8454	0.0476	56.9290	0.155	0.8906	8.3202
I	0.1179	-2.6514	0.0385	58.9208	0.135	0.8489	8.4746

Time Regularity test for nine experimental cases. Values for Time Regularity (TR) and Recency-Frequency (RF), as well as conventional clumpiness indicators are measured.

Table 3 shows that the Time Regularity index is in line with most of the clumpiness properties. The maximum Time Regularity for a specific event frequency occurs when the residual effects are spaced out to maximize the area under the curve. Among Cases A through E, Case C of an even distribution of events, closely approximates the ideal spread of residual effects. This is also true for entropy-like (ENTL), entropy (ENT), log-utility (LU) clumpiness measurements introduced by Zhang, Bradlow, and Small [9]. However, the second moment (2M) and three-largest-component (3LC) measurements are not designed to accommodate the time duration before the first event and after the last event. As a result, Cases A and E exhibit lower clumpiness (higher Time Regularity) due to their shorter, evenly spaced intervals.

The two cases with the most aggregated events over their respective observed durations are Case A and E, with Case A featuring early occurrences and Case E late occurrences of events. Recency is one of the most crucial elements for predicting future events (Wei et al., 2010), so the occurrence time is useful. Time Regularity incorporates this concept as residual effects increase over time, resulting in Case E of a higher Time Regularity value than A. However, conventional clumpiness indices are not designed for individual event information, but for summation of inter-event times, without the time-directional information. Therefore, they cannot distinguish between Cases A and E or B and D.

In comparing Cases B and D, Time Regularity also marginally fails to evaluate Case D over B, because of the greater penalty for late event aggregations, as larger potential residual effects resulting from later occurrence times are curtailed by the reset mechanism. We address it using a parameter to the numerator i, limiting the maximum peaks of residual effects of earlier events. Setting the parameter variable equal to the observed duration (T) results in a numerator i+T, which prevents any earlier event's residual effect from having a greater area than the reset later event's residual effect. However, this approach may overly empower the residual effects of late events. Instead, we can combine the Recency-Frequency measurement with Time Regularity to better distinguish between event aggregation time shifts.

Case F is a condensed version of Case C, where the influence of each event contribution is highest in the shortest observed duration. This maximizes Time Regularity as the residual effect expression R is at R_{max}. However, clumpiness cannot differentiate the time scale changes. As long as the event distribution remains the same, the scale is not considered, resulting in the same clumpiness value to the Case C for ENT, 2M, LU and 3LC. This limitation allows for only relative comparison with similar observed durations.

Cases G, C and H differ only in the second event occurrence, which shifts by a single time-unit difference, respectively. The slight shifts in both directions from C, resulting in Cases G and H, should cause a relatively small increase in clumpiness. This is true for all clumpiness indices as well as Time Regularity and Recency-Frequency measurements. However, since the reset mechanism compensates larger areas for earlier event aggregations, the time-position of the clumps inversely affects the Recency tendencies. This issue can be resolved by using the Recency-Frequency measurement, which promotes later event aggregations. Clumpiness cannot differentiate Cases G and H, as they do not account for changes due to time-shifts.

Case I clusters all Case C events to the center by reducing the inter-event times by one time-unit without scaling down the observed duration. This leads to a decrease in Time Regularity, as fewer residual effects are expressed in smaller inter-event time windows. Log-based clumpiness measurements, such as ENTL, ENT and LU, successfully capture the increased clumpiness as intended. However, 2M and 3LC methods fail to account for the clustering behavior, resulting in lower clumpiness values.

Overall, Recency-Frequency index effectively represents the time-positional information of all events during the observed duration. Time Regularity performs well compared to conventional clumpiness indicators, providing information on the density distribution of the events. However, it cannot measure the time-positional information regarding event aggregations due to the reset mechanism, slightly reducing the influence for later event aggregations. Time Regularity should be either complemented with the Recency-Frequency to provide additional data on time-positional valuation, or adjusted by adding the abovementioned parameter.

We observe that our RF metric can be computed in $O(nT)$, order of number of total event n multiplied by the observed duration T and time regularity is computed in $O(T)$. This assures that the proposed methods are efficiently computed.

VIII. Conclusion

We proposed Recency-Frequency and Time Regularity measurements to enhance the prediction of new events by addressing limitations of conventional RFM and clumpiness. By simulating numerous event times and residual effects, we demonstrated that these mitigate the information loss in the Recency measure, allow event time variations, and leverage residual effect simulations and events density, thus promising for calculating the probability of new events. They have potential for many applications in marketing, pharmaceutical areas, predictive maintenance analysis, etc., with more accurate and robust information for future events.

References

[1] Malthouse, Edward C., and Robert C. Blattberg. "Can we predict customer lifetime value?." Journal of interactive marketing 19.1 (2005): 2-16.

[2] Birant, Derya. "Data mining using RFM analysis." Knowledge-oriented applications in data mining. IntechOpen, 2011.

[3] Zhang, Yao, Eric T. Bradlow, and Dylan S. Small. "Predicting customer value using clumpiness: From RFM to RFMC." Marketing science 34.2 (2015): 195-208.

[4] Michael Platzer and Thomas Reutterer, "Ticking Away the Moments: Timing Regularity Helps to Better Predict Customer Activity." Marketing Science, (2016).

[5] G. Zhou, Z. Wang and X. Wang, "Study on the time variability of load characteristics based on Markov Chain Monte Carlo simulation," International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT), Changsha, China, (2015), 830-834.

[6] S. Wang and Y. Li, "A Novel Nonlinear Analysis Tool: Multi-scale Symbolic Sample Entropy and Its Application in Condition Monitoring of Rotary chinery," Asia-Pacific International Symposium on Advanced Reliability and Maintenance Modeling (APARM), Vancouver, BC, Canada, (2020): 1-5.

[7] Hsieh, Nan-Chen. "An integrated data mining and behavioral scoring model for analyzing bank customers." Expert systems with applications 27.4 (2004): 623-633.

[8] Wei, Jo-Ting, Shih-Yen Lin, and Hsin-Hung Wu. "A review of the application of RFM model." African Journal of Business Management 4.19 (2010): 4199.

[9] Zhang, Yao, Eric T. Bradlow, and Dylan S. Small. "New measures of clumpiness for incidence data." Journal of Applied Statistics 40.11 (2013): 2533-2548.

[10] Junichiro Niimi and Takahiro Hoshino, "Prediction of the Customer Activity with Using the Improved Measure of Clumpiness." Kodo Keiryogaku (The Japanese Journal of Behaviormetrics), Volume 47, Issue 1, Pages 27-40, (2020).

[11] Yeh, I-Cheng, King-Jang Yang, and Tao-Ming Ting. "Knowledge discovery on RFM model using Bernoulli sequence." Expert Systems with Applications 36.3 (2009): 5866-5871.

[12] Dua, D. and Graff, C. "Online Retail II Data Set". UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science, (2019).