Reimaging Consumer Creditworthiness by Harnessing E-Commerce Digital Footprints

Yijun LI[a], Albert Di WANG[b], Qi WU[*,a]

[a]*Department of Data Science, City University of Hong Kong*
[b]*McCombs School of Business, The University of Texas at Austin*

This version: March 9, 2025

Abstract

BigTech firms have reshaped consumer credit markets by seamlessly embedding Fintech lending services into their ecosystems. Unlike traditional financial institutions, these companies possess vast digital footprints generated from consumer activities, ranging from product browsing and transaction execution to post-purchase credit management. However, few studies have investigated the value of digital footprints in credit risk evaluation, particularly from a dynamic perspective, thereby overlooking critical insights into how credit risk evolves over time. To address this gap, we propose a novel framework that models consumer footprints in continuous time. By accommodating irregular and sporadic footprints, our model enables real-time monitoring of default risk. Additionally, the proposed end-to-end loss function jointly estimates the Probability of Default and Exposure at Default, offering a more comprehensive view of credit risk than traditional scoring approaches. We evaluate our model using longitudinal data from a leading BigTech platform. Empirical results demonstrate that our method significantly outperforms established benchmarks. Our findings underscore the critical role of modeling digital footprints in credit risk assessment, providing valuable guidance for more inclusive, adaptive, and data-driven financial services.

Keywords: Financial Technology; Consumer Credit Risk; Digital Footprints; Irregular Time Series.

[*] This paper absorbs and significantly revises the previous version titled *Neural Learning of Online Credit Risk*, and supersedes it with substantial methodological and empirical advancements. For inquiries, please correspond with Qi WU at `qi.wu@cityu.edu.hk`. All questions and comments are welcome.

1. Introduction

The rise of BigTech firms in financial services has fundamentally reshaped consumer credit markets. Companies such as Amazon and Alibaba, which initially operated as digital marketplaces, have seamlessly integrated financial products like buy-now-pay-later (BNPL) and digital lending into their platforms. By leveraging their extensive ecosystems of user interactions, these platforms have successfully embedded credit solutions within their business models, lowering barriers to purchase and increasing transaction volumes. By 2020, BigTech lenders had issued over $700 billion in credit globally (Cornelli et al., 2023), highlighting their growing influence in shaping modern financial markets.

Unlike traditional financial institutions that assess creditworthiness based on historical income records, employment status, and credit bureau scores, BigTech lenders possess an alternative and extensive source of financial insight: the vast digital footprints generated by millions of consumers interacting with their platforms. The economic value of these digital traces has been demonstrated in applications such as recommendation systems (Wang et al., 2024), user profiling (Valanarasu, 2021), and urban management (Traunmueller et al., 2018). As depicted in Figure 1, every online activity, such as browsing a product, making a purchase, or managing an existing credit, leaves behind a trace that can provide valuable signals of creditworthiness. These digital footprints offer a real-time, behavior-driven perspective on financial decision-making, potentially enabling more inclusive and adaptive credit risk assessment models.

While prior researches have demonstrated that non-traditional financial indicators, such as social media activity (Ge et al., 2017) or social network relationships (Lin et al., 2013), can offer valuable credit risk signals, particularly for individuals with limited formal credit histories, these indicators primarily capture static attributes of consumer behavior. They provide point-in-time insights but lack the ability to track how financial decision-making unfolds over time (Berg et al., 2020). In contrast, digital footprints from BigTech platforms provide a dynamic and continuous record of consumer financial behavior, covering the entire decision-making chain: from initial intent (browsing and product selection) to transaction execution (purchase and payment choice) and

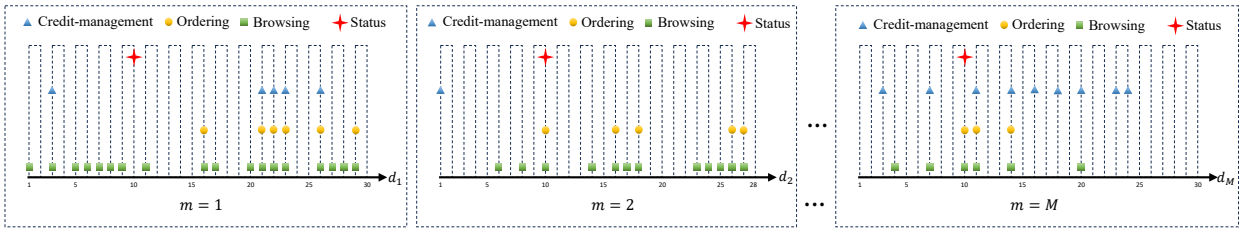
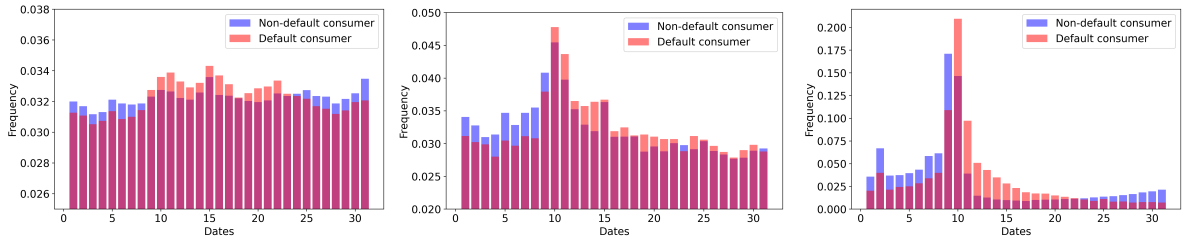


Fig. 1. Example digital footprints for a consumer. This consumer has $M = 12$ monthly bills issued on the last day of each month (i.e., statement dates) and the payment due by the 10th of next month (i.e., due dates). Red crosses indicate due dates, containing the default status and amount for each bill. The browsing, ordering, and credit-management footprints are denoted by green squares, yellow circles, and blue triangles, respectively. The predictive task involves using sporadic digital footprints of the last month to predict default outcomes for the current bill.

post-transaction credit management (credit repayment and refund). This ability to track consumer financial interactions at every stage offers a richer view of credit risk.

Despite the recognized value of digital footprints, existing studies have largely overlooked a critical aspect: the occurrence timing of digital footprints (Wang et al., 2018, Liang et al., 2021). Our empirical analysis reveals that consumer credit behavior can be inferred not only by the recorded attributes of digital footprints, but also by their precise timing. Specifically, we find distinct temporal patterns in how consumers engage with BigTech services. For instance, as illustrated in Figure 2, those at higher risk of default tend to show heightened financial activity, such as increased borrowing and spending, immediately after due dates, while consumers with lower credit risk are more likely to make proactive repayments before deadlines (see detailed analysis in Appendix Appendix B). These patterns suggest that creditworthiness is not just a function of static financial standing but is dynamically reflected in the way consumers interact with the platform over time (Agarwal et al., 2021).

To unlock the full potential of digital footprints for credit risk assessment, we introduce NeuCredit, a sequential modeling approach that integrates consumer footprints in a continuous-time framework. Unlike traditional Recurrent Neural Networks, which process information in fixed time steps, NeuCredit is built on the continuous-time domain, allowing it to model consumer activity with a finer temporal granularity. By continuously tracking the evolution of consumer behaviors, NeuCredit provides a more responsive and real-time risk assessment, offering a significant



(a) Number of browse on products. **(b)** The borrow amount. **(c)** The repaid amount.

Fig. 2. The temporal fluctuation in the frequency of non-default (blue) and default (red) consumer behaviors over a one-month period. (a) Default consumers exhibit higher browsing activity, particularly around the middle of the month. (b) A sharp increase in borrowing activity occurs immediately after the due date (10th day), with default consumers borrowing more frequently than non-default consumers. (c) Repayment activity peaks around the due date, with non-default consumers tending to make payments earlier, whereas default consumers exhibit delayed repayment behavior.

advantage over conventional sequential models.

In addition, the evaluation of credit risk in modern financial systems extends beyond merely estimating the Probability of Default (PD). A more comprehensive assessment incorporates Exposure at Default (EAD)—the amount a consumer is expected to owe at the time of default. Traditional credit scoring frameworks often treat EAD as a secondary concern, assuming relatively static credit exposures. However, in revolving credit products like BNPL and fintech loans, where consumers can repeatedly borrow and repay within a short period, EAD becomes even more critical for risk assessment. Unlike installment loans with fixed repayment schedules, the outstanding balance in revolving credit products fluctuates dynamically based on consumer borrowing and repayment behavior, making it essential to accurately estimate EAD to assess potential credit losses. NeuCredit jointly models both PD and EAD, providing a more holistic measure of risk that enables financial institutions to make better-informed lending decisions.

We validate NeuCredit using transaction data from a leading BigTech platform, comprising 49,223 active consumers over a 12-month period. The data encompasses a wide array of digital footprints captured as consumers interact with the platform, including activities related to browsing, ordering, and credit management. The results show that our approach significantly outperforms various baseline models across three distinct baseline streams. To further elucidate

the source of this performance enhancement, we visually represent the continuous-time updating process of latent default risk, which not only clarifies the improvements made by NeuCredit but also illustrates how BigTech lenders can leverage this model for real-time credit risk assessment, allowing them to design more adaptive and inclusive credit policies.

This study makes four contributions to the advancement of financial technology and consumer credit risk assessment as follows:

- We develop a structured framework for leveraging digital footprints from BigTech platforms to enhance consumer credit risk evaluation. By demonstrating how to use digital footprints, such as browsing, purchasing, and credit management activities, to serve as reliable indicators of financial responsibility, we provide a data-driven approach to credit assessment in digital finance.

- We highlight the critical role of digital footprint timing in credit risk prediction, demonstrating its potential value in capturing insights that extend beyond traditional models. To effectively capture this overlooked perspective, we introduce a continuous-time attentive neural network model that accurately represents the irregular and sporadic nature of digital footprints, enabling a real-time assessment of consumer credit risk.

- We propose an end-to-end training loss function that jointly evaluates PD and EAD within a unified framework. By integrating these two key risk components, our approach provides a more comprehensive assessment of credit risk, enabling BigTech firms to improve consumer segmentation, optimize credit allocation strategies, and deliver more personalized financial services based on dynamic risk profiles.

- We conduct extensive empirical experiments using real-world data from a major BigTech platform, demonstrating that our model significantly outperforms existing baselines - achieving a 3.9% improvement in AUC-ROC and a 15.7% increase in AUC-PR metrics.

2. Literature Review

2.1. Consumer Credit Risk Assessment

The evolution of consumer credit has undergone significant transformation over the years, largely driven by technological advancements and the emergence of innovative business models. Historically, consumer credit is dominated by commercial banks, utilizing empirical models such as logistic regression and linear discriminant analysis to assess creditworthiness (Edward, 1968, Wiginton, 1980). Although they offer a standardized framework for evaluating risk, they frequently exclude consumers with limited financial records.

The advent of peer-to-peer lending platforms, such as Lending Club and Prosper, introduces a new paradigm in consumer credit by facilitating direct lending between individuals, bypassing traditional financial institutions. Unlike banks, these platforms often cater to borrowers who lack credit histories, necessitating the exploration and integration of alternative data sources to enhance credit risk prediction. Such alternative data includes social network relationships (Lin et al., 2013), social media activities (Ge et al., 2017), soft information (Iyer et al., 2016), online records (Berg et al., 2020), telecommunication behaviors (Zhou et al., 2021b), semantic text analysis (Wang et al., 2020, Netzer et al., 2019), transaction data (Lee et al., 2024), and even microexpressions of applicants (Chang et al., 2024).

The incorporation of these alternative data sources into credit scoring models has necessitated the adoption of more sophisticated machine-learning techniques. Methods such as support vector machines (Luo et al., 2020), ensemble models (Bai et al., 2022), and deep neural networks (Cheng et al., 2022) have become increasingly prevalent due to their ability to process large datasets and uncover complex, nonlinear patterns that traditional statistical models might overlook (Li et al., 2024a).

More recently, BigTech firms have further revolutionized consumer credit through the advent of FinTech lending. They introduce unconditional revolving credit services that enable consumers to make purchases on their platforms without immediate payment. Consumers can access this credit as long as their spending remains within the assigned limit and typically benefits from a one-month interest-free period, which incentivizes higher spending (Li et al., 2024b). However,

the unconditional nature of this credit inherently increases financial risk, yet limited research has investigated how consumer digital footprints can be utilized for credit risk assessment, particularly from a dynamic perspective that captures the evolving patterns of financial behavior over time.

Beyond estimating the PD as a measure of credit risk, financial institutions also explore various alternative measures, including EAD, estimating the outstanding exposure at the time of default. For revolving credit products like credit cards and FinTech loans, where borrowing and repayment are continuous, EAD estimation becomes more challenging due to fluctuating balances and credit limits. Accurately estimating EAD in this context is essential for effective risk management, as it influences expected loss calculations and informs capital allocation and pricing decisions. To tackle these challenges, various methods have been proposed, including mixture models (Leow and Crook, 2016, Tong et al., 2016) and generalized cohort methods (Gürtler et al., 2018).

Despite these efforts, most approaches tend to focus on a single aspect of credit risk (i.e., PD or EAD) without modeling these factors simultaneously. This isolated approach can lead to suboptimal risk assessments, as it fails to capture the interactions between different risk components. Recognizing this gap, Papouskova and Hajek (2019) introduced a two-stage credit risk model that first employs class-imbalanced ensemble learning to predict PD, followed by a regression for EAD estimation. While this method begins to integrate multiple risk components, it still treats PD and EAD sequentially rather than concurrently.

Building on these developments, our study advances the literature by integrating digital footprints into a dynamic credit risk modeling framework that jointly predicts PD and EAD. By highlighting the significance of timing in capturing default tendencies and refining exposure estimates, our approach improves the precision and reliability of credit risk assessment. This enables BigTech institutions to implement more targeted and responsive credit strategies, ultimately paving the way for adaptive, data-driven decision-making in digital finance.

2.2. Multivariate Time Series Modeling

Evaluating credit risk is a complex task involving multivariate time series classification and forecasting. Traditional models like Autoregressive Moving Average and Exponential Smoothing have long been staples in time series analysis (Hamilton, 2020). While effective for simpler

datasets, these methods often fall short when dealing with the complexity of modern data, especially in capturing non-linear relationships and high-order dependencies that are prevalent in real-world scenarios.

The introduction of Recurrent Neural Networks (RNNs), including architectures like Gated Recurrent Unit (GRU) (Cho et al., 2014) and Long Short-Term Memory networks (LSTM) (Hochreiter and Schmidhuber, 1997), marked a significant leap forward in handling temporal dependencies in data. These models have demonstrated excellence across various tasks such as forecasting (Li et al., 2025), classification (Wang et al., 2018), imputation (Cao et al., 2018), and anomaly detection (Lin et al., 2024). Their versatility has led to widespread applications in fields ranging from finance (Chen et al., 2024) to marketing (Yin et al., 2024) and healthcare (Ray et al., 2023).

Despite their strengths, standard RNN face limitations when processing long sequences. They can struggle to retain information from earlier time steps, which is critical for accurate predictions in time series data. To address this, the attention mechanism is introduced, allowing models to focus on relevant parts of the input sequence and thereby improving performance (Bahdanau, 2014).

However, traditional RNNs and attention mechanisms are not inherently equipped to handle the irregular and sporadic nature of dynamic consumer footprints. These models are typically designed for sequential data with uniform time steps, such as natural language, where the timing between words is consistent. In contrast, consumer behavior data can be highly irregular, with events occurring at unpredictable intervals. This irregularity means that the precise timing of events carries essential contextual information that standard RNNs may overlook.

To address these challenges, literature has diverged into two main streams. The first stream enhances temporal sensitivity by incorporating explicit time information as additional features (Choi et al., 2016). Further advancements like Time2Vec embed time into high-dimensional vectors, providing a more expressive representation of temporal information (Kazemi et al., 2019). Other models learn time embeddings that capture trends and seasonality, enriching the model's understanding of temporal patterns (Zhou et al., 2021a).

The second stream involves adapting RNNs to operate in a continuous-time framework, aiming to align the model more closely with the natural flow of time. Unlike discrete-time RNNs that

assume constant hidden states between observations, continuous-time models introduce mechanisms that allow the model's memory to evolve over time. For instance, GRU-D incorporates decay mechanisms that enable the model's hidden state to diminish exponentially as time progresses, better reflecting the temporal decay of information (Che et al., 2018). Similarly, the T-LSTM model divides the LSTM's memory into short-term and long-term components, integrating decay into the short-term memory to capture the diminishing influence of former data (Baytas et al., 2017).

Our paper aligns with this second approach by employing a continuous-time attentive GRU implemented through Neural ODEs (Chen et al., 2018). We recognize that the decay mechanisms used in models like GRU-D are specific instances of ODEs, making them subsets of our broader modeling framework. By integrating continuous-time dynamics with attention mechanisms, our model can better handle irregular time intervals and focus on the most relevant events in a consumer's activity sequence.

3. Methodology

3.1. Overview

In this section, we outline the data structure and notations used in our analysis. We employ lowercase letters to signify scalar values (e.g., x), bold lowercase letters to represent vectors (e.g., $\mathbf{x}$), and bold uppercase letters to denote matrices (e.g., $\mathbf{X}$). For clarity, a comprehensive list of all variable notations, including the dimensions of vectors and matrices, is provided in Appendix Appendix A.

Consumer footprints on BigTech platforms are captured at two levels of data granularity: month-level and daily-level. We collect data from N consumers over M months, focusing on their browsing, ordering, and credit-management activities. For each consumer i ($i \in \{1, \cdots, N\}$) consumer in the m-th ($m \in \{1, \cdots, M\}$) month, these activities can occur on any day within the month, reflecting the irregular and sporadic nature of online consumer interactions. To effectively handle the variability in timing of these activities, we record the occurrence timestamps of browsing, ordering, and credit-management footprints for each consumer i at each month m as vectors

$\mathbf{t}_m^{i,b}$, $\mathbf{t}_m^{i,o}$, and $\mathbf{t}_m^{i,c}$:

$$\mathbf{t}_m^{i,b} = [t_{m,1}^{i,b}, \cdots, t_{m,K_m^{i,b}}^{i,b}] = 1 \leq t_{m,1}^{i,b}, \cdots, t_{m,K_m^{i,b}}^{i,b} \leq T_m,$$
$$\mathbf{t}_m^{i,o} = [t_{m,1}^{i,o}, \cdots, t_{m,K_m^{i,o}}^{i,o}] = 1 \leq t_{m,1}^{i,o}, \cdots, t_{m,K_m^{i,o}}^{i,o} \leq T_m,$$
$$\mathbf{t}_m^{i,c} = [t_{m,1}^{i,c}, \cdots, t_{m,K_m^{i,c}}^{i,c}] = 1 \leq t_{m,1}^{i,c}, \cdots, t_{m,K_m^{i,c}}^{i,c} \leq T_m,$$

where $K_m^{i,b}$, $K_m^{i,o}$, and $K_m^{i,c}$ represent the total number of records that browsing, ordering, and credit-management for consumer i at month m occur. T_m stands for the number of days in month m. In the sequel, unless specifically noted, we will drop the consumer upper script i and month lower script m to lighten the notations. Therefore, $\mathbf{t}_m^{i,b}$, $\mathbf{t}_m^{i,o}$, and $\mathbf{t}_m^{i,c}$ shall become $\mathbf{t}^b$, $\mathbf{t}^o$, and $\mathbf{t}^c$.

Then, at specific times for browsing, ordering, and credit-management, identified respectively as $t_k^b \in \mathbf{t}^b$, $t_k^o \in \mathbf{t}^o$, and $t_k^c \in \mathbf{t}^c$, we use $\mathbf{x}_{t_k^b}^b \in \mathbb{R}^{D_b}$, $\mathbf{x}_{t_k^o}^o \in \mathbb{R}^{D_o}$, and $\mathbf{x}_{t_k^c}^c \in \mathbb{R}^{D_c}$ to denote the features of occurred browsing, ordering, and credit-management footprints. D_b, D_o, and D_c indicate the feature dimensions for browsing, ordering, and credit-management, respectively. Therefore, the sequences of these activities for each month m are compiled as follows:

$$\mathbf{X}^b = [\mathbf{x}_{t_1^b}^b, \cdots, \mathbf{x}_{t_{K^b}^b}^b], \quad \mathbf{X}^o = [\mathbf{x}_{t_1^o}^o, \cdots, \mathbf{x}_{t_{K^o}^o}^o], \quad \mathbf{X}^c = [\mathbf{x}_{t_1^c}^c, \cdots, \mathbf{x}_{t_{K^c}^c}^c].$$

In this paper, our objective is to predict the default amount y_m at each month m and default probability $p(y_m > 0)$ for each consumer at the end of each month m by leveraging their historical digital footprints of browsing $\{\mathbf{X}_1^b, \cdots, \mathbf{X}_m^b\}$, ordering $\{\mathbf{X}_1^o, \cdots, \mathbf{X}_m^o\}$, and credit-management $\{\mathbf{X}_1^c, \cdots, \mathbf{X}_m^c\}$.

3.2. Sporadic Sequential Behaviors Encoding

The central component of our NeuCredit is a recurrent unit designed to capture the dynamic patterns of consumer behavior over time. Recurrent Neural Networks (RNNs) are widely recognized for their effectiveness in processing sequential data and have been extensively used to model consumer behaviors (Yin et al., 2024). In an RNN, each input vector is processed in order, with its information incorporated into the network's hidden state. This hidden state acts as the network's internal memory, essential for retaining and transferring information across time steps (Goodfellow et al., 2016).

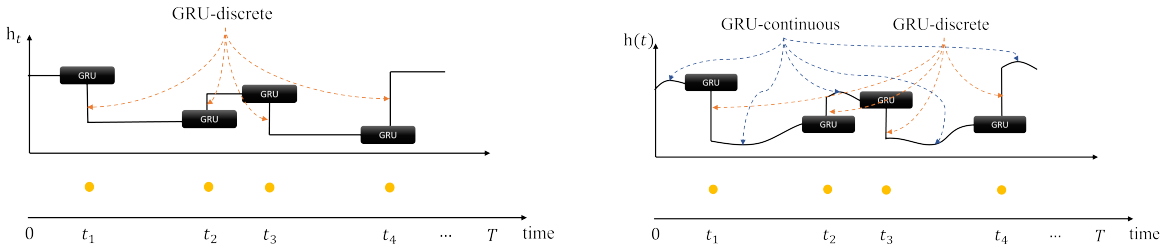
Vanilla RNN often encounter challenges such as the vanishing or exploding gradient problem, which inhibits their capacity to model long-term dependencies. To overcome these limitations, advanced variants such as LSTM (Hochreiter and Schmidhuber, 1997) and GRU (Cho et al., 2014) have been developed. These architectures introduce gating mechanisms that regulate the flow of information. For instance, the GRU updates its hidden state using the following equations:

$$\begin{aligned}
\mathbf{r}_t &= \sigma(\mathbf{W}_r \mathbf{x}_t + \mathbf{U}_r \mathbf{h}_{t-1} + \mathbf{b}_r), \\
\mathbf{z}_t &= \sigma(\mathbf{W}_z \mathbf{x}_t + \mathbf{U}_z \mathbf{h}_{t-1} + \mathbf{b}_z), \\
\tilde{\mathbf{h}}_t &= \tanh(\mathbf{W}_h \mathbf{x}_t + \mathbf{U}_h (\mathbf{r}_t \odot \mathbf{h}_{t-1}) + \mathbf{b}_h), \\
\mathbf{h}_t &= (1 - \mathbf{z}_t) \odot \tilde{\mathbf{h}}_t + \mathbf{z}_t \odot \mathbf{h}_{t-1}
\end{aligned} \tag{1}$$

where $\mathbf{x}_t$ is the input vector of variables at time t, and $\mathbf{h}_t$ is the memory of GRU that aggregates information accumulated up to timestamp t. $\mathbf{r}_t$ serves as the reset gate, determining the extent to which the previous memory should be disregarded when generating the current memory. $\mathbf{z}_t$ acts as the update gate, dictating the trade-off between retaining the previous memory and incorporating the new input at each timestamp. $\tilde{\mathbf{h}}_t$ represents the candidate update gate, capturing the new information that needs to be added to the current memory. $\{\mathbf{W}_r, \mathbf{U}_r, \mathbf{b}_r\}$, $\{\mathbf{W}_z, \mathbf{U}_z, \mathbf{b}_z\}$, and $\{\mathbf{W}_h, \mathbf{U}_h, \mathbf{b}_h\}$ are the trainable network parameters. Additionally, $\odot$ is the Hadamard product operator that implements the element-wise multiplication. $\sigma(\cdot)$ and $\tanh(\cdot)$ are activation functions that introduce non-linearity.

Despite these advancements, GRU still has limitations when modeling consumer footprints, particularly due to their inadequate handling of the exact timing of events and the intervals between them. For instance, consider a GRU used to model a sequence of consumer orders within a month, with events occurring on days $\mathbf{t}_1^o = [16, 21, 22, 23, 26, 29]$. While the GRU captures the order of these events, it does not account for their specific timestamps. If the first event occurred on day 17 instead of day 16, with all other features remaining the same, the GRU would produce the same prediction. This inability to distinguish between events occurring at different times can lead to significant inaccuracies, especially when the timing of consumer actions provides crucial insights into financial behavior and potential risks.

To address this challenge, it is imperative to recognize that RNN encounter limitations due



(a) Discrete-time memory evolution.

(b) Continuous-time memory evolution.

Fig. 3. The difference between discrete-time and continuous-time memory evolution.

to their exclusive focus on events at discrete-time events. For clarity, consider the ordering sequence in Figure 3a, where the GRU updates its hidden state only at the times of observable events t_1, t_2, t_3, t_4, and keeps the hidden state constant during periods without observations such as $t_1 - t_2, t_2 - t_3$, and $t_3 - t_4$. As a result, predicting values at a specific time point, such as t_4, relies solely on the memory updated at t_3, ignoring the actual duration between t_3 and t_4.

The intrinsic limitation of GRU necessitates a solution that accounts for the continuous-time dynamics present in the data instead of discrete-time dynamics (De Brouwer et al., 2019). Starting from the GRU update Eqn. (1), if we subtract the memory $\mathbf{h}_{t-1}$ from both sides, we can derive the following expression:

$$\mathbf{h}_t - \mathbf{h}_{t-1} = (1 - \mathbf{z}_t) \odot (\tilde{\mathbf{h}}_t - \mathbf{h}_{t-1}).$$

Then, by taking the limit as the time difference between $t - 1$ and t tends to zero, we arrive at the continuous-time dynamics of memory as an ODE

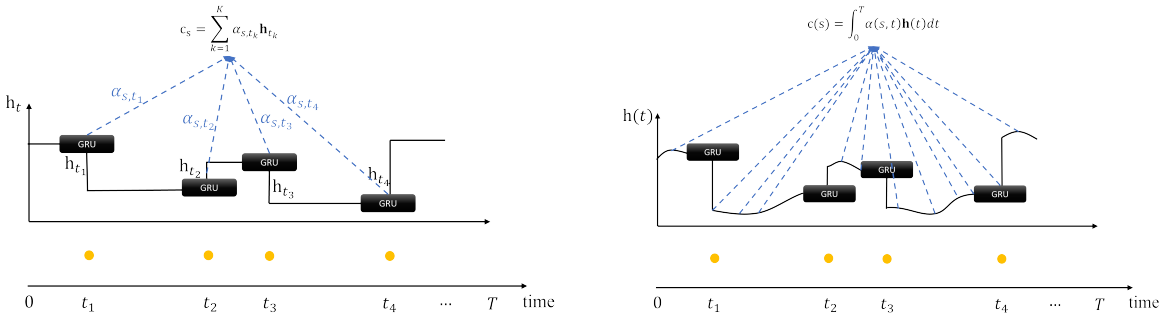
$$\frac{d\mathbf{h}(t)}{dt} = (1 - \mathbf{z}(t)) \odot \left(\tilde{\mathbf{h}}(t) - \mathbf{h}(t)\right). \tag{2}$$

This ODE, referred to as *GRU-Continuous*, allows the hidden state to evolve continuously over time, even during intervals without observations. When new inputs become available, we use the original GRU update equations (Eqn. (1)), now referred to as *GRU-Discrete*, to update the hidden state based on the new data. By integrating the continuous-time and discrete-time formulations, we create a hybrid model called *GRU-Hybrid*. This model effectively bridges the gaps between discrete events and provides a more robust representation of the hidden state's evolution over time, accommodating irregular intervals and sporadic behaviors. This process is visually illustrated in

Figure 3b. In practical implementation, the continuous dynamics described by the ODE can be integrated into neural network architectures (Chen et al., 2018).

3.3. Continuous-time Attention Mechanism

The attention mechanism fundamentally enhances the capacity of RNNs to process long sequences effectively, which is first introduced by Bahdanau et al. (2015) and generalized by Vaswani et al. (2017). By allowing models to dynamically focus on particular segments of an input sequence for generating representations, attention mechanisms alleviate the constraints imposed by the fixed-sized hidden states of traditional RNNs. The adoption of this technique has led to substantial advancements in a range of applications, such as natural language processing and image processing. Furthermore, the use of attention weights significantly enhances the interpretability of neural network decisions, providing clear insights into which aspects of the input data are prioritized during the prediction process (Guo et al., 2019). This level of transparency is especially valuable in domains such as credit risk evaluation.



(a) Discrete-time attention. **(b)** Continuous-time attention.

Fig. 4. The difference between discrete-time and continuous-time attention mechanism.

The attention mechanism fundamentally relies on constructing a context vector, $\mathbf{c}_s$, for a specific time point s. In discrete-time scenarios, it is achieved by computing a weighted sum of the hidden states $\mathbf{h}_{t_1}, \cdots, \mathbf{h}_{t_K}$, as illustrated by Figure 4a and following equation:

$$\mathbf{c}_s = \sum_{k=1}^{K} \alpha_{s,t_k} \mathbf{h}_{t_k}, \quad \text{where } \alpha_{s,t_k} = \frac{e_{s,t_k}}{\sum_{l=1}^{K} e_{s,t_l}}. \tag{3}$$

Here, α_{s,t_k} represents the attention score, which quantifies the relevance of the query $\mathbf{q}_s$ at time s and with each hidden state $\mathbf{h}_{t_k}$ at each observed time point t_k. The attention score is derived from

a normalized similarity measure, typically calculated through an exponential function of the dot product, $e_{s,t_k} = \exp(\mathbf{q}_s^\top \mathbf{h}_{t_k})$. However, this discrete-time method poses challenges for generalization to continuous-time, where the evaluation of α_{s,t_k} would require potentially infinite computations (Jhin et al., 2021).

To address this limitation and apply attention within our proposed continuous-time GRU framework, it is crucial to extend the discrete-time attention concept into a continuous-time setting. This extension is depicted by Figure 4b and the following formulation:

$$\mathbf{c}(s) = \int_0^T \alpha(s,t)\mathbf{h}(t)dt, \quad \text{where } \alpha(s,t) = \frac{e(s,t)}{\int_0^T e(s,t)dt}. \tag{4}$$

where $e(s,t) = \exp(\mathbf{q}(s)^\top \mathbf{h}(t))$ represents the continuous-time equivalent of the similarity score. The computation of the context vector $\mathbf{c}(s)$ involves evaluating two integrals: one to normalize the similarity scores over the observation period, $\int_0^T e(s,t)dt$, and another to compute the weighted average of hidden states across the same period. To simplify the computation, consider:

$$\mathbf{c}(s) = \int_0^T \alpha(s,t)\mathbf{h}(t)dt = \int_0^T \frac{e(s,t)}{\int_0^T e(s,t)dt}\mathbf{h}(t)dt = \frac{1}{\int_0^T e(s,t)dt}\int_0^T e(s,t)\mathbf{h}(t)dt$$

To efficiently compute the context vector for the continuous-time attention mechanism, we try to compute the denominator and the numerator involved separately. For the denominator, denote $\mathcal{E}(T) = \int_0^T e(s,t)dt$, and thus

$$\frac{d\mathcal{E}(t)}{dt} = \frac{d\int_0^t e(s,t')dt'}{dt} = e(s,t). \tag{5}$$

Similarly, for the numerator, we assume $C(T) = \int_0^T e(s,t)\mathbf{h}(t)dt$. By applying the Leibniz integral rule, we can derive:

$$\frac{dC(t)}{dt} = \frac{d\int_0^t e(s,t')\mathbf{h}(t')dt'}{dt} = e(s,t)\mathbf{h}(t) + \int_0^t \frac{\partial e(s,t')\mathbf{h}(t')}{\partial t}dt' = e(s,t)\mathbf{h}(t). \tag{6}$$

By integrating the Eqn. (2) for the hidden state dynamics and attention components, Eqns. (5) and (6), we can construct an augmented ODE system that captures the continuous dynamics of the attention mechanism in conjunction with the GRU's memory updates. This augmented ODE

system is represented as:

$$\frac{d}{dt}\begin{bmatrix} \mathbf{h}(t) \\ C(t) \\ \mathcal{E}(t) \end{bmatrix} = \begin{bmatrix} (1-\mathbf{z}(t)) \odot \left(\tilde{\mathbf{h}}(t) - \mathbf{h}(t)\right) \\ e(s,t)\mathbf{h}(t) \\ e(s,t) \end{bmatrix}. \tag{7}$$

Solving this ODE from an initial observation period up to time T involves integrating each component of the system over the time interval from 0 to T, obtaining $\mathbf{h}(T)$, $C(T)$, and $\mathcal{E}(T)$. The initial values, $\mathbf{h}(0)$, $C(0)$, and $\mathcal{E}(0)$, are often set to zero in practical applications to simplify the integration process.

In the context of predicting credit risk, our goal is to assess the probability and amount of default at the end time T of each month, taking into account the comprehensive digital footprints of consumers throughout that month. To enhance the predictive power of the final hidden state $\mathbf{h}(T)$, we combine it with the context vector using an additive formulation $\mathbf{z}(T) = \mathbf{h}(T) + \frac{C(T)}{\mathcal{E}(T)}$. This enriched state $\mathbf{z}(T)$ leverages both the dynamic memory encoded by the GRU-Hybrid, which captures sequential dependencies and temporal patterns, and the context vector, which distills the essence of the consumer's behavior across the observed period into a focused snapshot.

3.4. Hierarchical Network

Given that the consumer footprints are available at the daily level while the billing is calculated at the monthly level, we employ a hierarchical architecture, as illustrated in Figure 5. This architecture is specifically designed to evaluate credit risk at the end of each billing month, utilizing the encoded memories derived from browsing, ordering, and credit-management footprints.

In the bottom-level layers, we utilize an attentive GRU-Hybrid to process daily consumer footprints, specifically browsing $\mathbf{x}_m^b$, ordering $\mathbf{x}_m^o$, and credit-management $\mathbf{x}_m^c$, into distinct representations $\mathbf{z}_m^b(T_m) \in \mathbb{R}^{D_z}$, $\mathbf{z}_m^o(T_m) \in \mathbb{R}^{D_z}$, and $\mathbf{z}_m^c(T_m) \in \mathbb{R}^{D_z}$ respectively. Additionally, demographic features such as age, location, and gender are integrated using a feedforward neural network to form a static demographic representation $\mathbf{z}^d \in \mathbb{R}^{D_z}$. To enhance prediction, we introduce a gated fusion layer to combine $\mathbf{z}_m^b(T_m)$, $\mathbf{z}_m^o(T_m)$, and $\mathbf{z}_m^c(T_m)$ and fuse them into a vector $\mathbf{z}_m$. The formula for this layer is given by:

$$\mathbf{z}_m = \sum \text{softmax}(f([\mathbf{z}^d; \mathbf{z}_m^b(T_m); \mathbf{z}_m^o(T_m); \mathbf{z}_m^c(T_m)])) \odot [\mathbf{z}^d; \mathbf{z}_m^b(T_m); \mathbf{z}_m^o(T_m); \mathbf{z}_m^c(T_m)], \tag{8}$$

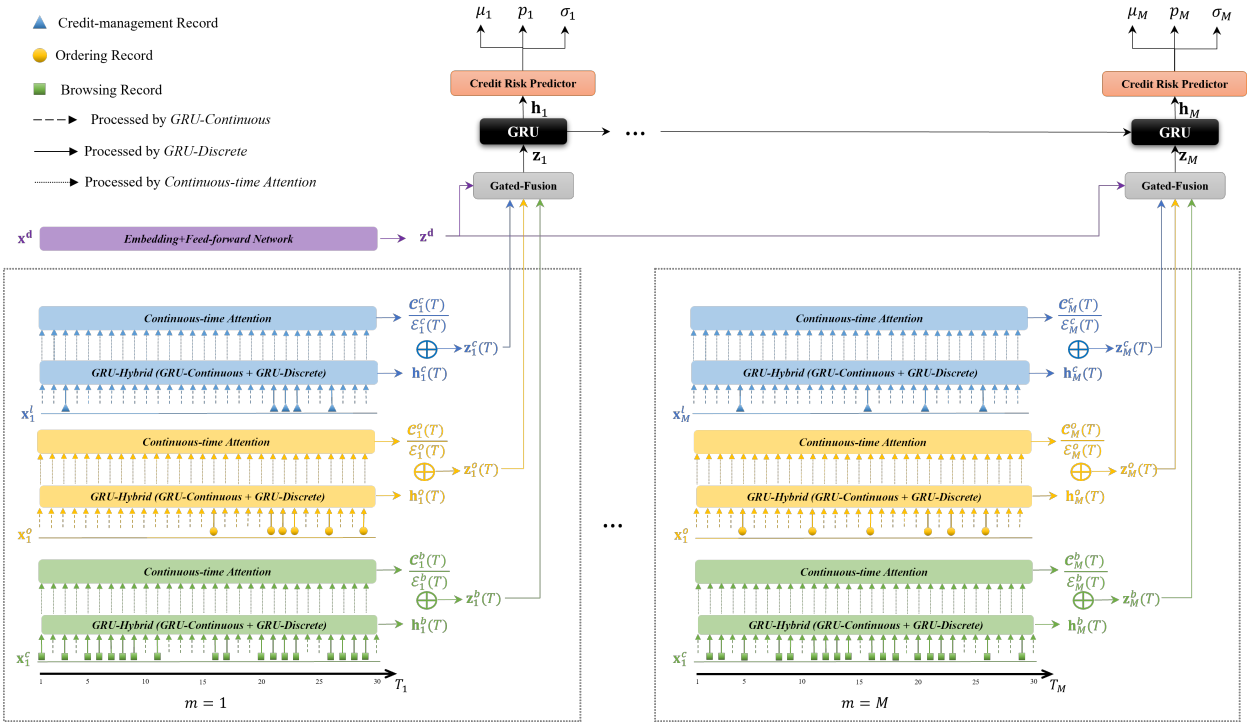


Fig. 5. The model framework of NeuCredit.

where $f : \mathbb{R}^{D_{dz}+D_{bz}+D_{oz}+D_{cz}} \to \mathbb{R}^4$ are trainable function, capturing the relative importance of static, browse, order, and credit-management information in credit risk assessment.

In the upper-level layer, where the time interval between two consecutive months is uniformly spaced, we thus employ a vanilla GRU to capture the serial dependencies at the month level. Specifically, the input to the monthly GRU is the fused memories $\mathbf{z}_m$ at month m, and the monthly GRU updates the monthly memories $\mathbf{h}_m \in \mathbb{R}^{D_h}$ according to:

$$\mathbf{h}_m = \text{GRU}(\mathbf{z}_m, \mathbf{h}_{m-1}). \tag{9}$$

This ensures that the memory at month m is dependent on consumer behaviors at the daily level in month m and the memory from the previous month $\mathbf{h}_{m-1}$.

3.5. Expected Loss

In the existing literature on credit risk, the primary focus has traditionally been on predicting the PD (Oskarsdottir and Bravo, 2021). While this provides a useful measure of risk, it does not

account for the variance in the amounts owed by different consumers, which can significantly impact the potential losses for BigTech platforms. For example, the financial impact of a default by a consumer owing $10,000 is substantially more severe than that of a consumer with a $100 debt.

To address this limitation, our model extends beyond only predicting the PD to also estimating the potential default amounts, referred to as EAD. By incorporating both PD and EAD, we can provide a more comprehensive measure of credit risk.

Our loss function ($\mathcal{L}_y$), termed Expected Loss, is formulated to minimize the negative log-likelihood of default behaviors for all consumers across all months. i.e.,

$$\begin{aligned}\mathcal{L}_y &= -\log\left(\prod_{i=1}^{N}\prod_{m=1}^{M}[1-p_d(y_m^i=0)]^{\mathbf{1}_{\{y_m^i=0\}}}\cdot[p_d(y_m^i>0)p_a(y_m^i;\mu_m^i,\sigma_m^i)]^{\mathbf{1}_{\{y_m^i>0\}}}\right),\\ &= -\sum_{i=1}^{N}\sum_{m=1}^{M}\mathbf{1}_{\{y_m^i=0\}}\log[1-p_d(y_m^i=0)]+\mathbf{1}_{\{y_m^i>0\}}\cdot\log[p_d(y_m^i>0)p_a(y_m^i;\mu_m^i,\sigma_m^i)]. \qquad (10)\end{aligned}$$

Here, y_m^i represents the amount of default for consumer i at month m. The indicator functions $\mathbf{1}_{\{y_m^i=0\}}$ and $\mathbf{1}_{\{y_m^i>0\}}$ signify whether the consumer has default behavior. If the amount is 0 (non-default), then $\mathbf{1}_{\{y_m^i=0\}}=1$ and $\mathbf{1}_{\{y_m^i>0\}}=0$; if the amount is greater than 0 (default), then $\mathbf{1}_{\{y_m^i=0\}}=0$ and $\mathbf{1}_{\{y_m^i>0\}}=1$. Moreover, $p_d(y_m^i>0)=1-p_d(y_m^i=0)$ represents the probability of default for the consumer, while $p_a(y_m^i;\mu_m^i,\sigma_m^i)$ represents the distribution of the potential default amount given the parameters μ_m^i,σ_m^i. It's worth noting that $p_a(y_m^i;\mu_m^i,\sigma_m^i)$ only appears in the second term when $\mathbf{1}_{\{y_m^i>0\}}=1$, as default amount of the non-default consumers always equal to 0.

The Eqn. (10) can be further simplified as follows

$$\mathcal{L}_y = -\sum_{i=1}^{N}\sum_{m=1}^{M}\mathbf{1}_{\{y_m^i=0\}}\log[1-p_d(y_m^i=0)]+\mathbf{1}_{\{y_m^i>0\}}\cdot\log[p_d(y_m^i>0)]+\mathbf{1}_{\{y_m^i>0\}}\cdot\log[p_a(y_m^i;\mu_m^i,\sigma_m^i)].$$

We can observe that the first two terms represent the cross-entropy loss ($\mathcal{L}_{CE}$) for predicting whether the consumer will be default. The third term is to fit the distribution of default amounts only when consumers exhibit default behaviors. We assume the default amount for each consumer follows the Log-normal distribution with varying μ_m^i,σ_m^i. The reason is that the log-normal distribution is capable of modeling skewed distributions well, particularly those exhibiting non-negativity and heavy-tailed characteristics (see Figure D.13 in Appendix Appendix D). Therefore,

the loss function can be finally expressed as

$$\mathcal{L}_y = \sum_{i=1}^{N}\sum_{m=1}^{M} \mathcal{L}_{CE} + \mathbf{1}_{\{y_m^i>0\}} \cdot \left[\log(y_m^i \sigma_m^i \sqrt{2\pi}) + \frac{(\log y_m^i - \mu_m^i)^2}{2(\sigma_m^i)^2}\right]. \tag{11}$$

We can then estimate the heterogeneous probability of default $p_d(y_m^i > 0)$, the distribution parameters of default amount μ_m^i and σ_m^i conditional on the $\mathbf{h}_m^i$ learned from Hierarchical Network at each month, i.e.,

$$\left(p_d(y_m^i > 0), \mu_m^i, \sigma_m^i\right) = g_y(\mathbf{h}_m^i), \tag{12}$$

where $g_y : \mathbb{R}^{D_z} \to \mathbb{R}^3$ is a function constructed by a two-layer feed-forward network.

Finally, we can estimate the expected loss (EL_m^i) for consumer i at month m as the product of PD and EAD, i.e.,

$$EL_m^i = p_d(y_m^i > 0) \cdot \mathbb{E}[y_m^i] = p_d(y_m^i > 0) \cdot \exp\left(\mu_m^i + \frac{(\sigma_m^i)^2}{2}\right). \tag{13}$$

Once the expected loss of consumers is determined, the platform can classify consumers into different categories. For example, for consumers with high default amounts and probability of default, the platform should monitor their behavior in a timely manner and adjust their credit limits to reduce losses. For consumers with low default amounts and probability of default, platforms should consider increasing their credit limits to encourage their spending.

4. Experiment

4.1. Research Context

We obtain a proprietary dataset from a leading Chinese BigTech platform listed on U.S. stock exchanges. This platform boasts over 500 million active consumers and generates annual revenue exceeding $100 billion. Offering a wide array of product categories, including electronics, daily necessities, food items, and pharmaceuticals, it enables consumers to engage in various online shopping activities such as searching, browsing, and ordering. In addition to retail services, the platform provides financial offerings like consumer credit, allowing customers to finance their purchases first and pay later. The associated BigTech lender offers significant flexibility in payment

arrangements, including options for installment payments on orders and bills, as well as minimum payments on bills. While this flexibility empowers consumers to manage their finances effectively, it also introduces complexity to credit risk assessment due to the diversity of payment behaviors.

Our dataset includes comprehensive data on browsing, ordering, and credit-management activities of 49,223 active consumers over the course of 12 monthly billing cycles. The browsing data provides insights into the device type used and specific product information such as the shop, brand, and product category. The ordering data captures details about the products purchased, quantities, total order amounts, and the use of coupons, cash, or credit cards, along with shipping addresses. Additionally, the credit-management data encompasses a range of actions that affect the consumer's outstanding credit balance, including credit used for purchases, choices of installment plans, applicable installment interest fees, and adjustments to credit balances due to refunds or payments. The richness and granularity of this dataset provide an excellent opportunity to model consumer behavior and assess credit risk in a real-world setting. For a detailed and statistical description of the dataset, please refer to Appendix Appendix D.

4.2. Experiment Setting

To rigorously evaluate the generalizability of our model, we partition the dataset into training and testing sets. Specifically, we allocate 10,000 individuals for training and reserve the remaining 39,223 for testing, ensuring that the test set is substantially larger to provide a robust assessment under real-world conditions. Within the training cohort, we employ a 5-fold cross-validation strategy to optimize hyperparameters and prevent overfitting. In each of the five experiments, 8,000 individuals are used for training and 2,000 for validation.

Given the significant imbalance between positive and negative samples in our dataset, where the defaults are relatively rare (only around 10% of consumers), we select the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) and the Area Under the Precision-Recall Curve (AUC-PR) as evaluation metrics. The AUC-ROC metric provides insight into the trade-off between true positive and false positive rates across various threshold settings, aiding in understanding the model's overall discriminative ability. Additionally, in imbalanced datasets, AUC-PR is often more informative as it focuses on the model's performance in predicting the minority class

by assessing precision and recall.

To evaluate the performance in predicting EAD, we rank defaulting consumers according to their actual default amounts and their predicted default amounts separately. We then compute the Spearman's Rank Correlation (SRC) coefficient between these two rankings as the evaluation metric (Chamberlain et al., 2017). We chose this metric over Mean Squared Error (MSE) or Mean Absolute Error (MAE) because these metrics are sensitive to extreme values, which can disproportionately influence results and skew the assessment of overall performance (Hyndman and Koehler, 2006). In contrast, SRC measures the strength and direction of the association between the two rankings, providing valuable practical insights. By accurately ranking consumers based on their likelihood of incurring significant delinquencies, the platform can implement more targeted and efficient risk management strategies.

In our experiments, the input feature sizes for the browsing, ordering, and credit-management footprints are 11, 19, and 10, respectively, reflecting the dimensionality of data collected in each category. To fine-tune our model, we employ a random search within predefined parameter ranges, as detailed in Appendix Appendix C. This method allows for efficient exploration of the hyperparameter space without the computational expense of exhaustive grid search. We implement an adaptive learning rate strategy, starting with an initial rate of 0.005 and halving it when the validation loss does not decrease for five consecutive epochs. This approach accelerates convergence and helps avoid local minima. We use the Adam optimization method (Kingma and Ba, 2014) as our optimizer, leveraging its advantages in handling sparse gradients and adjusting learning rates during training.

4.3. Results on Modeling Timing of Digital Footprints

In the previous section, we highlight the critical importance of accurately modeling the sporadic digital footprints that consumers leave on platforms. Incorporating the timing of these footprints is essential for gaining comprehensive insights into consumer behavior and enhancing the precision of credit risk predictions. To validate, we conduct a series of experiments within the Hierarchical Network framework. In these experiments, we test the performance of various baseline models by replacing our proposed attentive GRU-Hybrid unit with alternative units, while keeping

the overall framework and training loss consistent to ensure a fair comparison.

We categorize the baseline models into three groups based on how they handle timing information. The first stream includes traditional GRU cells (Cho et al., 2014), which primarily focus on capturing the order of sequences but ignoring the actual timing of the footprints. This model serves as a fundamental baseline, illustrating the impact of sequence order without explicit timing information on the accuracy of predictions.

The second stream aims to enhance temporal sensitivity by incorporating the timing of footprints directly into the input features, termed GRU-time. Building on this, the GRU-Time2Vec model embeds the scalar time values into high-dimensional vectors using the Time2Vec method (Kazemi et al., 2019). This approach provides a more expressive representation of temporal information. These enriched time vectors are then concatenated with the original input data and processed through the GRU. The objective of these models is to empirically demonstrate the benefits of including detailed timing information in improving prediction performance.

The third stream involves adapting the traditional RNN architecture to suit a continuous-time framework. This adaptation is exemplified by models such as GRU-Decay (GRU-D) (Che et al., 2018), Time-aware LSTM (T-LSTM) (Baytas et al., 2017), and Multirepresentational Attention GRU (MA-GRU) (Yin et al., 2024), which integrate an exponential decay mechanism into the hidden state. This mechanism dynamically adjusts based on the time intervals between events, aiming to more accurately reflect the true dynamics of consumer interactions by accounting for the decay in information relevance over time. We provide a detailed description of these streams of baselines in Appendix Appendix E.

We rigorously train each baseline model and then evaluate their performance using the same test set. By keeping the overall framework and training loss the same across all models, we ensure that any differences in performance could be attributed to the modeling of timing information rather than other factors. The results, detailed in Table 1, provide insightful comparisons among the various models.

Specifically, the standard GRU model demonstrates the lowest performance metrics with an AUC-ROC of 0.727 and an AUC-PR of 0.228. Incorporating time as an additional feature in the GRU-Time model led to noticeable improvements, increasing the AUC-ROC to 0.739 and

Table 1 Evaluation Metrics (mean±standard deviation) on Test Dataset in Five-fold Cross-validation. Bold highlights the top-performing results.

Recurrent Unit	AUC-ROC	AUC-PR	SRC
GRU	0.727 ± 0.003***	0.228 ± 0.007***	0.642 ± 0.015***
GRU-Time	0.739 ± 0.006***	0.244 ± 0.011***	0.650 ± 0.015**
GRU-Time2Vec	0.765 ± 0.017***	0.286 ± 0.030***	**0.667 ± 0.005**
GRU-D	0.745 ± 0.004***	0.255 ± 0.007***	0.642 ± 0.013***
T-LSTM	0.737 ± 0.012***	0.239 ± 0.015***	0.548 ± 0.031***
MA-GRU	0.773 ± 0.003***	0.319 ± 0.007***	0.609 ± 0.011***
NeuCredit (GRU-Hybrid)	**0.803 ± 0.008**	**0.369 ± 0.010**	0.661 ± 0.008

Note: * denote models significantly lower than the best-performing models (t-test $^{***}p < 0.01$,** $p < 0.05$,* $p < 0.1$)

the AUC-PR to 0.244. This demonstrates that even a simple inclusion of timing information can enhance predictive performance. Further enhancements are observed with the GRU-Time2Vec model, where time is embedded into a high-dimensional vector, achieving an AUC-ROC of 0.765 and an AUC-PR of 0.286, underscoring the importance of effectively representing temporal information.

Similarly, models that adapt the classical GRU or LSTM to a continuous-time framework—such as GRU-D, T-LSTM, and MA-GRU—demonstrated superior performance compared to the standard GRU. For instance, GRU-D achieved an AUC-ROC of 0.745 and an AUC-PR of 0.255, while MA-GRU attained an AUC-ROC of 0.773 and an AUC-PR of 0.319. These results suggest that integrating mechanisms to handle the decay of memory over time significantly improves prediction accuracy, aligning more closely with the dynamic nature of consumer behaviors.

On the other hand, while the prediction of PD benefits significantly from incorporating the timing of digital footprints, the estimation of EAD appears less dependent on these temporal dynamics. Specifically, the SRC metric for the standard GRU model stands at 0.642, and shows a modest increase to 0.667 when timing is integrated as an additional feature using GRU-Time. Interestingly, in models adopting a continuous-time framework, the SRC either remains similar or decreases compared to the standard GRU. This suggests that the increased complexity of continuous-time models might not necessarily translate to better EAD estimation. One possible explanation is that EAD estimation relies more on cumulative behavior over time rather than the

precise timing of individual events.

At last, NeuCredit, our proposed model, outperforms all baseline units with the highest scores in AUC-ROC at 0.808 and AUC-PR at 0.376 while also achieving a competitive SRC score of 0.661. These results clearly illustrate the advantage of NeuCredit's approach in leveraging continuous-time dynamics and a sophisticated attention mechanism, providing the most effective solution for predicting credit risk using consumer digital footprints.

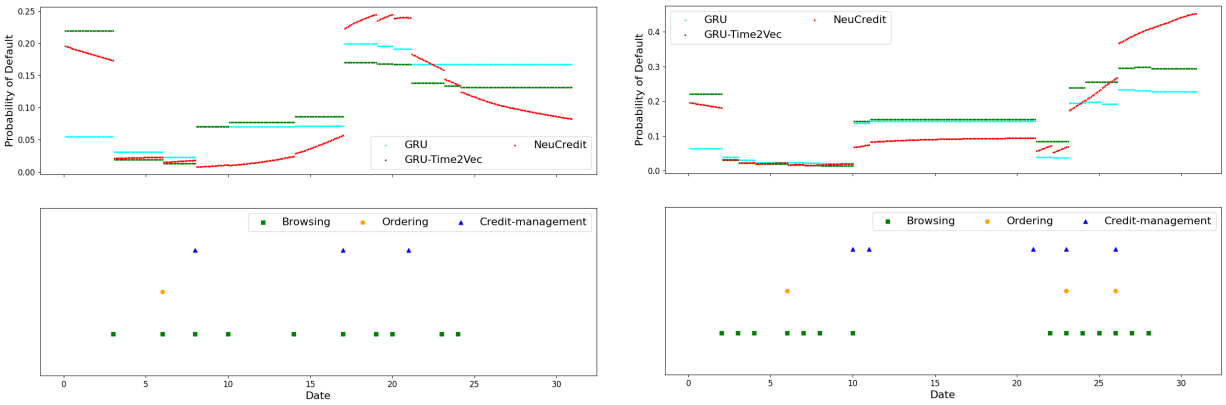
4.4. Source of Improvement

In this section, we delve into the enhancements provided by our model, NeuCredit, which leverages continuous-time dynamics to improve credit risk prediction. We compare the performance of NeuCredit with two baseline models: the standard GRU (representing models that consider only the sequence order of events) and GRU-Time2Vec (representing models that include timing information as additional features). Our objective is to understand how effectively modeling the timing of digital footprints contributes to enhanced predictive capabilities.

To illustrate the differences among these models, we analyze the dynamics of predicting the probability of default within a month for two consumers, as depicted in Figure 6. Subplot (a) presents data for a non-defaulting consumer, while Subplot (b) shows data for a defaulting consumer. In each subplot, the lower graph displays the occurrence times of browsing, ordering, and credit-management activities throughout the month. The upper graph plots the real-time predictions of default probability generated by the three models.

The standard GRU model, serving as our foundational baseline, processes sequences without explicitly accounting for the timing between events. As a result, its predictions of default probability remain constant between observed activities, regardless of time intervals. The model updates its predictions only when new digital footprints are recorded, leading to a stepwise prediction pattern that may overlook important temporal dynamics.

Building upon the GRU, the GRU-Time2Vec model incorporates timing information by embedding timestamps into high-dimensional vectors using the Time2Vec method (Kazemi et al., 2019). These time embeddings are added as features to the model, allowing for more responsive updates at the times when events occur and improving performance compared to the standard



(a) The predicted default probability for a non-default consumer.

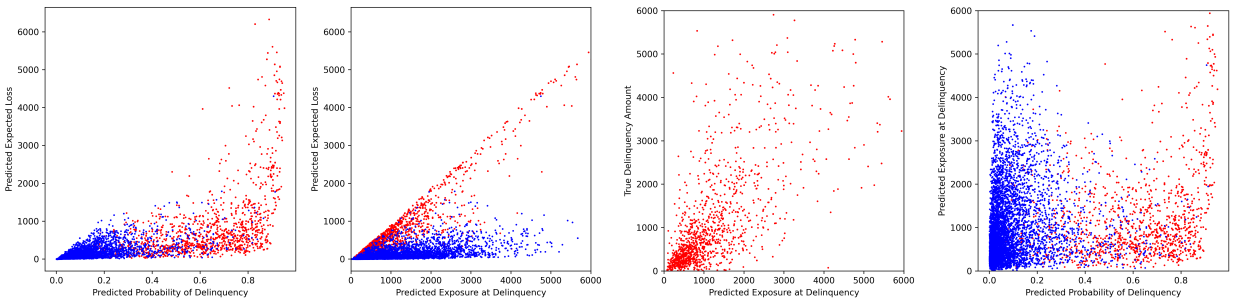
(b) The predicted default probability for a default consumer.

Fig. 6. The predicted default probability for a default consumer and a non-default consumer via GRU (blue), GRU-Time2Vec (green) and proposed NeuCredit (red).

GRU. However, GRU-Time2Vec still operates within a discrete-time framework, resulting in predictions that update in steps.

In contrast, NeuCredit employs a continuous-time framework that dynamically updates its predictions not only in response to new activities but also based on the time elapsed since the last event. This means that NeuCredit can adjust its assessment of a consumer's default probability continuously over time, providing a more accurate and real-time reflection of risk as time progresses. The model effectively captures the decay or escalation of risk during periods of inactivity, when discrete-time models may fail to recognize.

For instance, by the end of the month, the standard GRU predicts a default probability of approximately 0.17 for the non-defaulting consumer and 0.22 for the defaulting consumer—a relatively small difference that may not effectively distinguish between the two. The GRU-Time2Vec model offers some improvement, predicting probabilities of 0.12 and 0.30 for the non-defaulting and defaulting consumers, respectively. NeuCredit, on the other hand, predicts a default probability of 0.08 for the non-defaulting consumer and 0.43 for the defaulting consumer. These more distinct probabilities demonstrate NeuCredit's enhanced ability to differentiate between consumer behaviors and accurately assess credit risk.



(a) PD vs. EL. **(b)** EAD vs. EL. **(c)** Predicted vs. True EAD. **(d)** PD vs. EAD.

Fig. 7. The relationship among PD, EAD, and EL. Blue/red dots represent non-default/default consumers.

4.5. Results for Expected Loss (EL)

In this section, we examine how our model estimates the PD and EAD to calculate the EL for consumers. Specifically, Figures 7a and 7b illustrate the relationships between PD and EL, and between EAD and EL, respectively. In Figure 7a, we observe that consumers with higher predicted PD values correspond to higher expected losses. Similarly, Figure 7b shows that increased predicted EAD values are associated with higher expected losses. These patterns underscore the importance of jointly modeling PD and EAD, as both components contribute significantly to the estimation of EL.

Furthermore, Figure 7c explores the relationship between the estimated EAD and the actual exposure or default amount for default consumers. The positive proportional relationship indicates that our model not only predicts the likelihood of default but also provides a reliable estimate of the potential financial loss if a default occurs. This capability is crucial for lenders to assess the severity of potential defaults and allocate resources appropriately.

Figure 7d further presents the relationship between PD and EAD, providing critical insights for targeted credit management strategies. For example, consumers positioned in the upper right quadrant—those with high default probability and high exposure at default—should have their credit limits reduced and be subject to more stringent monitoring to mitigate potential losses. Conversely, consumers in the lower left quadrant—characterized by low default probability and low exposure at default—could be encouraged to increase consumption through higher credit limits.

4.6. Results for Continuous-time Attention

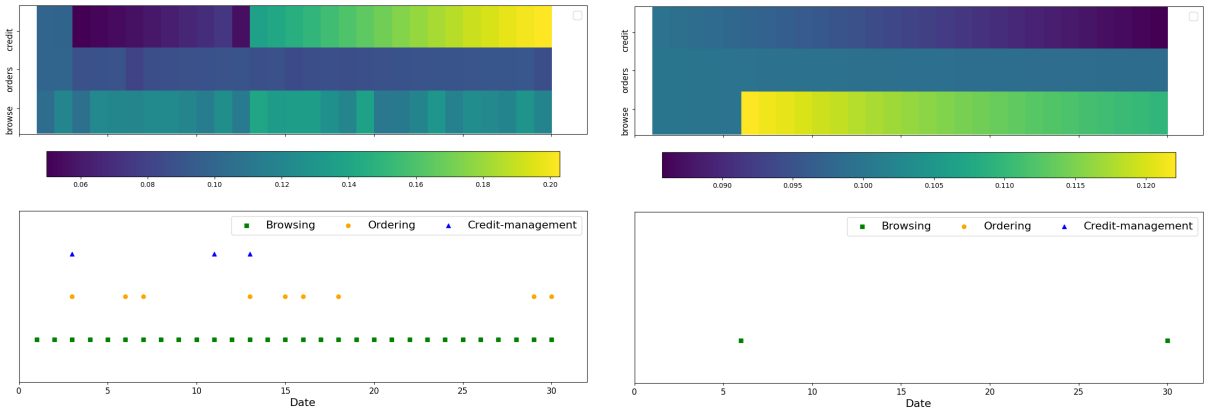
In this section, we explore the interpretative capabilities of our model by examining how it applies the continuous-time attention mechanism across different consumer profiles. Figure 8 illustrates the continuous-time attention weights for two consumers: one with dense digital footprints and another with sparse digital activity. This visualization highlights where the model concentrates its focus during the prediction process. The upper panels depict the evolution of attention weights throughout the billing cycle, while the lower panels display the corresponding digital footprints for each consumer.

For the consumer with rich digital footprints (Figure 8a), the model displays a pronounced focus on credit-management activities, which is intuitive and logical, as financial data typically contains the most essential information for assessing credit risk. Conversely, the consumer with sparse digital footprints (Figure 8b) presents a different scenario, with limited data on ordering and credit-management activities. In response, the model shifts its focus predominantly towards the sparse browsing data available. This adaptation highlights the model's capability to leverage whatever data is available to optimize risk assessment, ensuring robust predictive performance even in data-scarce environments.

Moreover, the attention weights further highlight the effectiveness of the GRU-Hybrid framework. Specifically, the attention weights undergo significant shifts at each occurrence timing of digital footprint, reflecting how the GRU-Hybrid updates its hidden state with new data. Additionally, the attention weights also change slightly even in the absence of new observations. For example, in Figure 8(a), the attention allocate to credit-management data continues to increase despite the absence of new footprints. This indicates that the GRU-Hybrid can extract valuable temporal information from the passage of time itself and enhance the inference of credit risk. As a result, the attention mechanism captures more informative aspects of the data, improving the model's overall predictive performance.

4.7. Heterogeneous Analysis

In this section, we examine the heterogeneous effect of NeuCredit compared to the baseline GRU model, focusing on identifying which consumer groups benefit the most from incorporating



(a) Consumer with dense digital footprints.

(b) Consumer with sparse digital footprints.

Fig. 8. The continuous-attention of two distinct consumers.

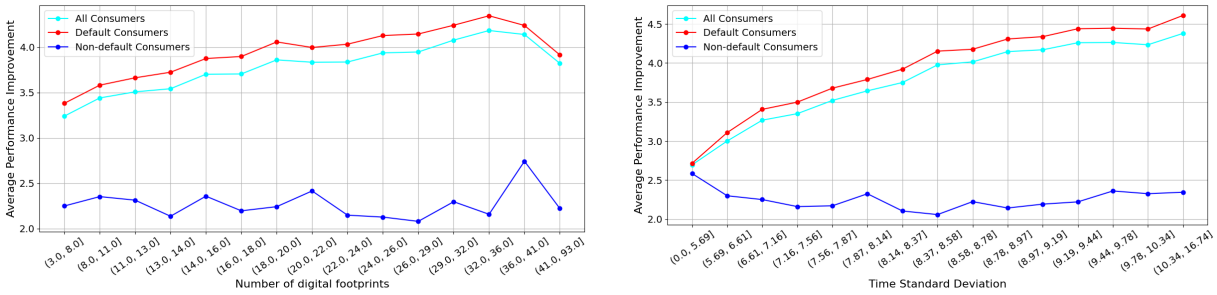
timing information into credit risk assessment. Since the primary distinction between NeuCredit and GRU lies in their treatment of event timing, this analysis allows us to evaluate the extent to which capturing temporal dynamics enhances predictive performance.

To quantify the performance improvement of NeuCredit over GRU, we evaluate the predicted default probability for each consumer in the test set. Performance improvement is defined based on changes in the odds ratio between NeuCredit and GRU. Specifically, if a consumer defaults, a higher predicted probability of default is considered better, as it reflects stronger predictive accuracy. Conversely, if a consumer does not default, a lower predicted probability is preferable, as it reduces false positives. The relative improvement is thus computed as follows:

$$\text{Relative Performace Improvement} = \begin{cases} \exp\left(\text{logit}_{NeuCredit} - \text{logit}_{GRU}\right), & \text{if default,} \\ \exp\left(\text{logit}_{GRU} - \text{logit}_{NeuCredit}\right), & \text{if not default,} \end{cases}$$

where $\text{logit}(p) = \log(\frac{p}{1-p})$ represents the log odds transformation of the predicted default probability p.

To structure this analysis, we segment consumers based on two key temporal factors: the frequency of digital footprints, measured as the total number of recorded digital footprints in a one-month period, and the variability in footprint occurrence, quantified as the standard deviation of the timestamps of online activities within the same period. Consumers are categorized into 15 groups along these two dimensions, allowing us to assess how NeuCredit's performance improvements



(a) The occurrence frequency of digital footprints.

(b) The occurrence variability of digital footprints.

Fig. 9. The heterogeneous analysis of occurrence frequency and irregularity of digital footprints. All test sets (39,223) are equally clustered into 15 groups and the performance improvements are averaged.

over GRU vary across different consumer behaviors.

As shown in Figure 9a, NeuCredit consistently outperforms GRU across all frequency groups, as performance improvement values consistently greater than 1, confirming the benefit of incorporating event timing. While the performance gains for non-default consumers remain relatively stable, default consumers initially show a slight increase in improvement as footprint frequency rises. However, when the number of digital footprints becomes excessively high, the performance gap narrows, likely because the abundance of footprints provides sufficient predictive signals, reducing the relative impact of precise timing information.

Additionally, Figure 9b illustrates that NeuCredit outperforms GRU across all levels of footprint variability. However, the extent of improvement differs between default and non-default consumers. Default consumers with higher timing variance benefit significantly more from NeuCredit, suggesting that their online activities are more time-sensitive and may be often triggered by external factors, such as liquidity constraints or repayment deadlines. In contrast, non-default consumers exhibit stable improvements, implying that their behaviors follow more stable and structured patterns.

4.8. Ablation Studies

In this section, we perform two ablation studies to evaluate the effectiveness of key components in NeuCredit and to understand their contributions to the model's overall performance. In the first experiment, we assess the role of the continuous-time attention mechanism in enhancing predictive

accuracy. We modify NeuCredit by removing the continuous-time attention component, resulting in a model that relies solely on the GRU-Hybrid to predict credit risk.

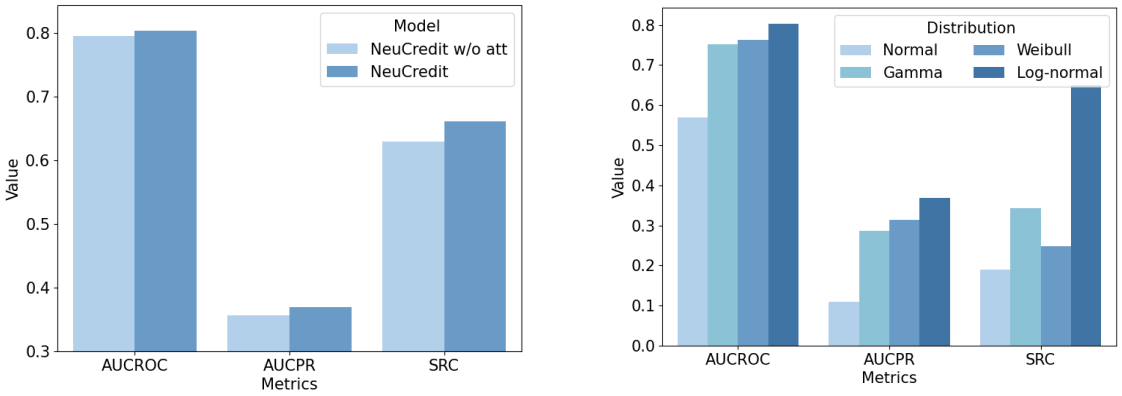
The results of this experiment are presented in Figure 10a. We observe that all evaluation metrics, including the AUC-ROC, AUC-PR, and SRC, experience a slight decline compared to the full NeuCredit model. These findings suggest that while the GRU-Hybrid effectively captures the temporal patterns in consumer behavior, the continuous-time attention mechanism provides additional benefits. By allowing the model to focus on the most relevant information at different times, the attention mechanism enhances the representation of the consumer's credit risk profile.

In the second experiment, we investigate the impact of different distributional assumptions for the default amount in the Expected Loss calculation. In our original model, we assume that the default amounts follow a log-normal distribution, which accounts for the positively skewed nature of financial loss data and ensures that predicted amounts are non-negative. In this experiment, we replace the log-normal distribution with three alternative distributions: Normal, Gamma, and Weibull distributions. We then evaluate the model's performance using each of these distributions in the EL loss function.

The results are depicted in Figure 10b. We find that the model assuming a Normal distribution for default amounts performs the worst among the alternatives, with significant decreases in both AUC-ROC and AUC-PR metrics. This poor performance is likely due to the Normal distribution's inability to handle the skewness and non-negativity constraints of the default amount data. The models using Gamma and Weibull distributions show better performance than the Normal distribution but still do not match the effectiveness of the log-normal assumption. The model with the log-normal distribution assumption continues to deliver the best performance, with the highest AUC-ROC and AUC-PR scores. This outcome suggests that the log-normal distribution aligns most closely with the empirical distribution of default amounts in our dataset

5. Conclusion

The rapid expansion of BigTech firms into financial services has fundamentally transformed consumer credit markets. In this study, we demonstrate the potential of using digital footprints to improve consumer credit risk assessment. While traditional credit scoring models primarily rely



(a) Ablation results in continuous-time attention.

(b) Ablation results on the alterative distribution of EL loss.

Fig. 10. Results of Ablation Studies.

on static features, we show that digital footprints, comprising browsing behaviors, purchasing decisions, and credit management activities, offer a dynamic and real-time perspective on consumer creditworthiness.

We highlight the importance of not only the presence of digital footprints but also their timing in evaluating the credit risk. To effectively capture this temporal information, we introduce NeuCredit, a continuous-time attentive neural network designed to model the irregular and sporadic nature of digital footprints. It overcomes the limitations of traditional RNNs, which process information at fixed intervals and fail to capture variations in consumer behavior over time. Additionally, we propose an end-to-end training loss function that jointly estimates both the Probability of Default and Exposure at Default within a unified framework. This dual-component risk assessment enables financial institutions to improve credit allocation strategies, mitigate potential losses, and personalize credit offerings.

Our empirical analysis, conducted on a real-world dataset, validates the effectiveness of NeuCredit. The results show that our model significantly outperforms baseline methods, achieving notable improvements in AUC-ROC and AUC-PR metrics. Furthermore, our heterogeneous analysis reveals that NeuCredit's performance gains are particularly pronounced among consumers with frequent and irregular digital footprints, suggesting that default-prone consumers exhibit more time-sensitive behavioral patterns.

These findings have important implications for the future of FinTech lending. By incorporating digital footprints and their timing into credit risk assessment, BigTech lenders can develop more accurate and adaptive risk models, fostering greater financial inclusion while managing default risk more effectively. Future research can extend our approach by incorporating additional behavioral signals, exploring causal relationships in consumer credit behavior, and refining continuous-time modeling techniques to further enhance predictive accuracy.

References

Agarwal, S., Bubna, A., and Lipscomb, M. (2021). Timing to the statement: Understanding fluctuations in consumer credit use. *Management Science*, 67(8):5124–5144.

Bahdanau, D. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Bahdanau, D., Cho, K. H., and Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. In *3rd International Conference on Learning Representations, ICLR 2015*.

Bai, M., Zheng, Y., and Shen, Y. (2022). Gradient boosting survival tree with applications in credit scoring. *Journal of the Operational Research Society*, 73(1):39–55.

Baytas, I. M., Xiao, C., Zhang, X., Wang, F., Jain, A. K., and Zhou, J. (2017). Patient subtyping via time-aware lstm networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 65–74.

Berg, T., Burg, V., Gombović, A., and Puri, M. (2020). On the rise of fintechs: Credit scoring using digital footprints. *The Review of Financial Studies*, 33(7):2845–2897.

Cao, W., Wang, D., Li, J., Zhou, H., Li, L., and Li, Y. (2018). Brits: Bidirectional recurrent imputation for time series. *Advances in Neural Information Processing Systems*, 31.

Chamberlain, B. P., Cardoso, A., Liu, C. B., Pagliari, R., and Deisenroth, M. P. (2017). Customer lifetime value prediction using embeddings. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1753–1762.

Chang, X., Dai, L., Feng, L., Han, J., Shi, J., and Zhang, B. (2024). A good sketch is better than a long speech: evaluate delinquency risk through real-time video analysis. *Review of Finance*, page rfae044.

Che, Z., Purushotham, S., Cho, K., Sontag, D., and Liu, Y. (2018). Recurrent neural networks for multivariate time series with missing values. *Scientific Reports*, 8(1):6085.

Chen, L., Pelger, M., and Zhu, J. (2024). Deep learning in asset pricing. *Management Science*, 70(2):714–750.

Chen, R. T., Rubanova, Y., Bettencourt, J., and Duvenaud, D. K. (2018). Neural ordinary differential equations. *Advances in Neural Information Processing Systems*, 31.

Cheng, D., Niu, Z., Li, J., and Jiang, C. (2022). Regulating systemic crises: Stemming the contagion risk in networked-loans through deep graph learning. *IEEE Transactions on Knowledge and Data Engineering*, 35(6):6278–6289.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. (2014). Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

Choi, E., Bahadori, M. T., Schuetz, A., Stewart, W. F., and Sun, J. (2016). Doctor ai: Predicting clinical events via recurrent neural networks. In *Machine Learning for Healthcare Conference*, pages 301–318. PMLR.

Cornelli, G., Frost, J., Gambacorta, L., Rau, P. R., Wardrop, R., and Ziegler, T. (2023). Fintech and big tech credit: Drivers of the growth of digital lending. *Journal of Banking & Finance*, 148:106742.

De Brouwer, E., Simm, J., Arany, A., and Moreau, Y. (2019). Gru-ode-bayes: Continuous modeling of sporadically-observed time series. *Advances in Neural Information Processing Systems*, 32.

Edward, A. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*, 23(4):589–609.

Ge, R., Feng, J., Gu, B., and Zhang, P. (2017). Predicting and deterring default with social media information in peer-to-peer lending. *Journal of Management Information Systems*, 34(2):401–424.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.

Guo, T., Lin, T., and Antulov-Fantulin, N. (2019). Exploring interpretable lstm neural networks over multi-variable data. In *International Conference on Machine Learning*, pages 2494–2504. PMLR.

Gürtler, M., Hibbeln, M. T., and Usselmann, P. (2018). Exposure at default modeling–a theoretical and empirical assessment of estimation approaches and parameter choice. *Journal of Banking & Finance*, 91:176–188.

Hamilton, J. D. (2020). *Time series analysis*. Princeton university press.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.

Hyndman, R. J. and Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688.

Iyer, R., Khwaja, A. I., Luttmer, E. F., and Shue, K. (2016). Screening peers softly: Inferring the quality of small borrowers. *Management Science*, 62(6):1554–1577.

Jhin, S. Y., Jo, M., Kong, T., Jeon, J., and Park, N. (2021). Ace-node: Attentive co-evolving neural ordinary differential equations. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 736–745.

Kazemi, S. M., Goel, R., Eghbali, S., Ramanan, J., Sahota, J., Thakur, S., Wu, S., Smyth, C., Poupart, P., and Brubaker, M. (2019). Time2vec: Learning a vector representation of time. *arXiv preprint arXiv:1907.05321*.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Lee, J. Y., Yang, J., and Anderson, E. T. (2024). Using grocery data for credit decisions. *Management Science*.

Leow, M. and Crook, J. (2016). A new mixture model for the estimation of credit card exposure at default. *European*

Journal of Operational Research, 249(2):487–497.

Li, T., Kou, G., Peng, Y., and Yu, P. S. (2024a). Feature selection and grouping effect analysis for credit evaluation via regularized diagonal distance metric learning. *INFORMS Journal on Computing*.

Li, Y., Leung, C. H., Sun, X., Wang, C., Huang, Y., Yan, X., Wu, Q., Wang, D., and Huang, Z. (2024b). The causal impact of credit lines on spending distributions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 180–187.

Li, Y., Leung, C. H., and Wu, Q. (2025). Probabilistic learning of multivariate time series with temporal irregularity. *IEEE Transactions on Knowledge and Data Engineering*.

Liang, T., Zeng, G., Zhong, Q., Chi, J., Feng, J., Ao, X., and Tang, J. (2021). Credit risk and limits forecasting in e-commerce consumer lending service via multi-view-aware mixture-of-experts nets. In *Proceedings of the 14th ACM international conference on web search and data mining*, pages 229–237.

Lin, H., Liu, G., Wu, J., and Zhao, J. L. (2024). Deterring the gray market: Product diversion detection via learning disentangled representations of multivariate time series. *INFORMS Journal on Computing*, 36(2):571–586.

Lin, M., Prabhala, N. R., and Viswanathan, S. (2013). Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending. *Management science*, 59(1):17–35.

Luo, J., Yan, X., and Tian, Y. (2020). Unsupervised quadratic surface support vector machine with application to credit risk assessment. *European Journal of Operational Research*, 280(3):1008–1017.

Netzer, O., Lemaire, A., and Herzenstein, M. (2019). When words sweat: Identifying signals for loan default in the text of loan applications. *Journal of Marketing Research*, 56(6):960–980.

Oskarsdottir, M. and Bravo, C. (2021). Multilayer network analysis for improved credit risk prediction. *Omega*, 105:102520.

Papouskova, M. and Hajek, P. (2019). Two-stage consumer credit risk modelling using heterogeneous ensemble learning. *Decision Support Systems*, 118:33–45.

Ray, A., Jank, W., Dutta, K., and Mullarkey, M. (2023). An lstm+ model for managing epidemics: Using population mobility and vulnerability for forecasting covid-19 hospital admissions. *INFORMS Journal on Computing*, 35(2):440–457.

Tong, E. N., Mues, C., Brown, I., and Thomas, L. C. (2016). Exposure at default models with and without the credit conversion factor. *European Journal of Operational Research*, 252(3):910–920.

Traunmueller, M. W., Johnson, N., Malik, A., and Kontokosta, C. E. (2018). Digital footprints: Using wifi probe and locational data to analyze human mobility trajectories in cities. *Computers, Environment and Urban Systems*, 72:4–12.

Valanarasu, M. R. (2021). Comparative analysis for personality prediction by digital footprints in social media. *Journal of Information Technology*, 3(02):77–91.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017).

Attention is all you need. *Advances in Neural Information Processing Systems*, 30.

Wang, C., Han, D., Liu, Q., and Luo, S. (2018). A deep learning approach for credit scoring of peer-to-peer lending using attention mechanism lstm. *IEEE Access*, 7:2161–2168.

Wang, C., Shi, Y., Guo, X., and Chen, G. (2024). Probing digital footprints and reaching for inherent preferences: A cause-disentanglement approach to personalized recommendations. *Information Systems Research*.

Wang, Z., Jiang, C., Zhao, H., and Ding, Y. (2020). Mining semantic soft factors for credit risk evaluation in peer-to-peer lending. *Journal of Management Information Systems*, 37(1):282–308.

Wiginton, J. C. (1980). A note on the comparison of logit and discriminant models of consumer credit behavior. *Journal of Financial and Quantitative Analysis*, 15(3):757–770.

Yin, J., Feng, Y., and Liu, Y. (2024). Modeling behavioral dynamics in digital content consumption: An attention-based neural point process approach with applications in video games. *Marketing Science*.

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., and Zhang, W. (2021a). Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on Artificial Intelligence*, volume 35, pages 11106–11115.

Zhou, J., Wang, C., Ren, F., and Chen, G. (2021b). Inferring multi-stage risk for online consumer credit services: an integrated scheme using data augmentation and model enhancement. *Decision Support Systems*, 149:113611.

Appendix A. Notations for Main Variables

Table A.2 Summary of Main Notations.

Sections	Notations	Description	Range/Shape
Data	N	the total number of consumers	$\mathbb{N}^+$
	M	the total number of months	$\mathbb{N}^+$
	T_m	the total number of days in month m	$\mathbb{N}^+$
	$\mathbf{t}_m^{i,b}$	the browsing timestamps vector for consumer i at month m	$(K_m^{i,b}, 1)$
	$K_m^{i,b}$	the total number of browsing timestamps for consumer i at month m	$\mathbb{N}^+$
	$\mathbf{t}_m^{i,o}$	the ordering timestamps vector for consumer i at month m	$(K_m^{i,o}, 1)$
	$K_m^{i,o}$	the total number of ordering timestamps for consumer i at month m	$\mathbb{N}^+$
	$\mathbf{t}_m^{i,c}$	the credit-management timestamps vector for consumer i at month m	$(K_m^{i,c}, 1)$
	$K_m^{i,c}$	the total number of credit-management timestamps for consumer i at month m	$\mathbb{N}^+$
	$\mathbf{x}^b_{m,t^b_{m,k}}$	the browsing features for consumer i at timestamp $t^b_{m,k}$ in month m	$(D_b, 1)$
	$\mathbf{x}^o_{m,t^o_{m,k}}$	the ordering features for consumer i at timestamp $t^o_{m,k}$ in month m	$(D_o, 1)$
	$\mathbf{x}^c_{m,t^l_{m,k}}$	the credit-management features for consumer i at timestamp $t^c_{m,k}$ in month m	$(D_c, 1)$
	$\mathbf{X}_m^b$	the browsing feature matrices for consumer i in month m	$(D_b, K_m^{i,b})$
	$\mathbf{X}_m^o$	the ordering feature matrices for consumer i in month m	$(D_o, K_m^{i,o})$
	$\mathbf{X}_m^c$	the credit-management feature matrices for consumer i in month m	$(D_c, K_m^{i,c})$
Sequential Encoding	$\mathbf{x}_t$	the input vector of GRU at timestamp t	$(D_x, 1)$
	$\mathbf{h}(t)$	the continuous-time memory vector of GRU at t	$(D_h, 1)$
	$\mathbf{W}_{(r,z,h)}$	the trainable input-to-hidden matrices of GRU	(D_h, D_x)
	$\mathbf{U}_{(r,z,h)}$	the trainable hidden-to-hidden matrices of GRU	(D_h, D_h)
	$\mathbf{b}_{(r,z,h)}$	the trainable bias vector of GRU	$(D_h, 1)$
Continuous -time Attention	$\mathbf{h}(T)$	the memory vector at end time T	$(D_h, 1)$
	$\mathcal{C}(T)$	the numerator of context vector at time T	$(D_h, 1)$
	$\mathcal{E}(T)$	the denominator of context vector at time T	$\mathbb{R} \in (0, 1)$
	$\mathbf{z}(T)$	the enhanced representation memory vector at time T	$(D_z, 1)$
Hierarchical Network	$\mathbf{z}_m^b(T_m)$	the encoded memory vector of browsing features in month m	$(D_z, 1)$
	$\mathbf{z}_m^o(T_m)$	the encoded memory vector of ordering features in month m	$(D_z, 1)$
	$\mathbf{z}_m^c(T_m)$	the encoded memory vector of credit-management features in month m	$(D_z, 1)$
	$\mathbf{z}^d$	the encoded vector of static features in month m	$(D_z, 1)$
	$\mathbf{U}_{(b,o,c)}$	the trainable fusion weight matrices	$(D_z, D_{bh/oh/ch})$
	$\mathbf{b}_{(b,o,c)}$	the trainable fusion bias vectors	$(D_z, 1)$
	$\mathbf{z}_m$	the fused memory vector in month m	$(D_z, 1)$
	$\mathbf{h}_m$	the memory vector of monthly-GRU in month m	$(D_h, 1)$
Expected Loss	$p_d(y_m^i > 0)$	the default probability for consumer i at month m	$[0, 1]$
	μ_m^i, σ_m^i	the parameters of predicted default distribution for consumer i at month m	$\mathbb{R}^+$

Appendix B. Empirical Evidence for the Essential of Capturing Timing of Digital Footprints in Consumer Credit Risk Prediction

Consumer behaviors, especially online behaviors, often deviate from traditional rational models, exhibiting patterns influenced by timing and psychological triggers rather than strict utility optimization. As Agarwal et al. (2021) demonstrate, there is a 15% surge in credit card spending during the week immediately following a statement issuance, a pattern that suggests behavioral responses tied to financial cycles rather than purely liquidity needs. This insight aligns with a growing body of literature that highlights how consumers may not always follow strict economic rationality, often influenced by psychological factors and situational salience rather than objective financial logic.

In our main paper, we analyze behavioral changes in browsing, borrowing, and repayment patterns across a monthly billing cycle using basic statistical measures, as shown in Figure 2. The findings reveal distinct temporal patterns in financial behaviors between consumers who eventually default and those who do not. To further substantiate the statistical significance of these patterns, we conduct a regression analysis inspired by Agarwal et al. (2021). Consumers are categorized into two groups—default and non-default—and we first estimate the following model for each group:

$$y_{it} = \alpha_i + \sum_{j=2}^{n} \beta_{t-j} I_{i,t-j} + \gamma' X_t + \epsilon_{it}$$

where the coefficients $\beta_2, \ldots, \beta_{31}$ measure the incremental effect on spending each day after the issuance of the credit statement, with the first day (omitted from the model) serving as the baseline. Here, X_t is a vector of controls, which include day-of-the-week, day-of-the-month, and standard errors clustered by the consumers. y_{it} represents various consumer behaviors, such as the number of product browses, borrowing amount, and repayment on each day. In the top panels of Figure 10, we plot the coefficient fluctuations for both default (red) and non-default (blue) consumers, with shaded areas indicating the confidence intervals for the estimated coefficients.

Then, we combine the two groups and employ an interaction term to capture differential effects

between the default and non-default groups for each day. Specifically, we estimate:

$$y_{it} = \alpha_i + \sum_{j=2}^{n} \beta_{t-j} I_{i,t-j} + \sum_{j=2}^{n} \theta_{t-j} I_{i,t-j} \cdot s_i + \phi s_i + \gamma' X_t + \epsilon_{it}$$

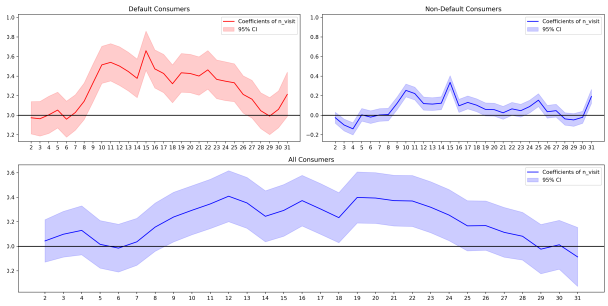
where s_i indicates whether the consumer is a defaulter (denote as 1) or non-defaulter (demnote as 0), and the interaction coefficients $\theta_2, \ldots, \theta_{31}$ quantify the statistical differences between the two groups on each day. In Figure 10, the lower panels depict the fluctuations in these interaction coefficients.

Our findings reveal that, compared to non-default consumers, default-prone consumers display similar browsing patterns at the beginning and end of each billing cycle but tend to browse more products in the mid-cycle and after the due date (around the 10th day of each cycle), suggesting a propensity for impulsive spending. For borrowing behavior, defaulting consumers initially show no difference but tend to borrow more after the due date. In terms of repayment, non-default consumers are more likely to repay 1-3 days before the due date, whereas defaulting consumers often delay repayment, suggesting potential liquidity constraints or a lower repayment priority among defaulters.

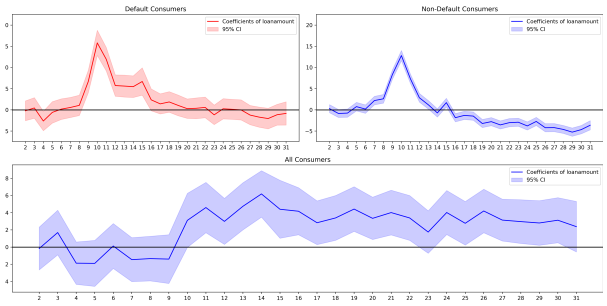
This empirical evidence underscores the critical importance of incorporating timing in analyzing consumer digital footprints for credit risk prediction. It illustrates how specific temporal markers within the billing cycle can significantly impact consumer financial behaviors. Incorporating temporal dimensions of digital interactions can thus enhance credit risk models, potentially leading to more accurate predictions and targeted interventions.

Appendix C. Hyper-parameters Selection

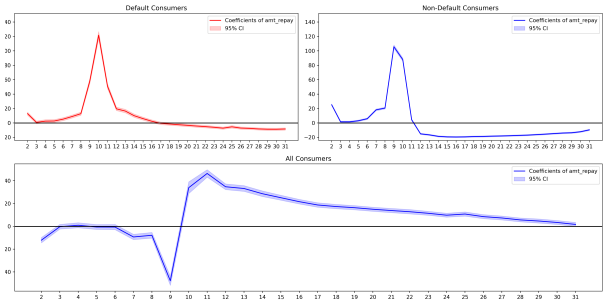
Neural networks, owing to their complex architectures, involve the selection of various hyper-parameters critical to their performance. The optimization of these hyper-parameters is an indispensable aspect of training neural networks, as it directly influences the model's ability to capture intricate patterns within the data. The conventional methodology for hyper-parameter tuning involves cross-validation. In this approach, neural networks with different hyper-parameter configurations are trained on the training set, and their performance is evaluated based on metrics such as log-likelihood on the validation set.



(a) Coefficients of Number of browse on products.



(b) Coefficients of the borrow amount.



(c) Coefficients of the repay amount.

Fig. B.11. The fluctuation of the regression coefficients for three key variables across one billing month.

Given the inherent computational demands of training recurrent neural network (RNN) models, characterized by their sequential processing nature, we adopt a random search strategy for hyper-parameter selection rather than a grid search. This strategy allows for a more efficient exploration of the hyper-parameter space. In our investigation, we focus on the three most pivotal hyper-parameters: learning rate, batch size, and the hidden size of the Gated Recurrent Unit (GRU). The hyper-parameter search space is defined to cover a range of values deemed crucial for the optimal functioning of the model. Specifically, the learning rate varied across [0.05, 0.01, 0.005, 0.001], the batch size explored options such as [256, 512, 1024, 2048], and the hidden size of the GRU spanned [30, 50, 80, 100].

Among these hyper-parameters, the learning rate emerged as particularly influential. An excessively high learning rate could lead to erratic fluctuations in training loss, hindering effective convergence. Conversely, a small learning rate might impede the learning process. To address this, we implement an adaptive learning rate strategy. If the validation loss fails to decrease over five consecutive epochs, the learning rate is halved.

Ultimately, after a systematic exploration and evaluation, our chosen hyperparameters are set as follows: initial learning rate equals 0.005, batch size equals 512, and hidden size of GRU equals 50. These selections are informed by a nuanced consideration of model performance, computational efficiency, and convergence stability.

Appendix D. Descriptive Statistics

Among the 49,223 sampled consumers, 11,862 experienced at least one billing month with a delay of 1 to 29 days (0+ default), 2,023 had a delay of 30 to 59 days (30+ default), 849 were delayed for 60 to 89 days (60+ default), and 1,506 were late for more than 90 days (90+ default). Remarkably, 32,983 consumers maintained a pristine record with no delinquencies during the entire observation period.

Table D.3 The descriptive statistics of important variables in the dataset (all the footprint variables are aggregated at the daily level).

Dimensions	Variables	Mean	Std	q=5%	q=25%	q=50%	q=75%	q=95%
Demographics	Gender (Male=1)	0.591	0.492	0	0	1	1	1
	Age	29.05	7.19	20	23	28	33	43
Ordering	Orders amounts before discount	461.22	1041.98	20.4	76	179.6	399	1779
	Orders amounts after discount	333.42	821.14	15.8	54	121.1	272	1270
	Orders discounts	161.01	1500.16	0	8.6	48.29	156.13	639
	Orders payments	300.21	1646.52	2.69	45	105.32	231.5	1117.94
	Avg num of products per order	3.67	8.91	1	1	2	4	11
	Avg num of orders per day	2.45	2.94	1	1	1	3	7
	Intervals of orders	8.72	12.20	1	2	4	10	31
Browsing	Num of platform visits	20.43	35.14	1	3	9	23	77
	Num of product browsing	10.86	16.87	1	2	5	13	39
	Num of shop visits	6.79	8.95	1	2	4	8	23
	Num of brand visits	6.15	8.09	1	1	3	8	21
	Num of mobile visits	19.94	34.67	1	3	9	22	76
	Num of pc visits	0.43	5.12	0	0	0	0	0
	Intervals of browsing	2.22	3.15	1	1	1	2	7
Credit-management	Amount of credit borrows	398.89	932.85	10.4	55.88	124.9	300.13	1789
	Installment period of credit	2.31	3.69	1	1	1	1	12
	Installment cost	9.94	66.66	0	0	0	0	31.44
	Amount of credit repays	805.84	1160.66	33.03	170	423.57	954.82	2962.78
	Amount of credit refunds	458.87	1118.11	5.9	49.58	126.4	319.4	2296.99
	Intervals of credit borrowing	8.16	11.23	1	2	4	10	31

We present a detailed analysis of the key variables within our dataset, summarized in Table D.3. The descriptive statistics offer insights into four main dimensions: demographics, ordering, browsing, and credit management. Starting with demographics, our dataset indicates a male majority, with males comprising 59% of the consumer base. The age demographic skews young,

with an average age of just 29.1 years, and more than half of the consumers are under 30 years old, suggesting a youthful consumer base primarily engaging with e-commerce platforms.

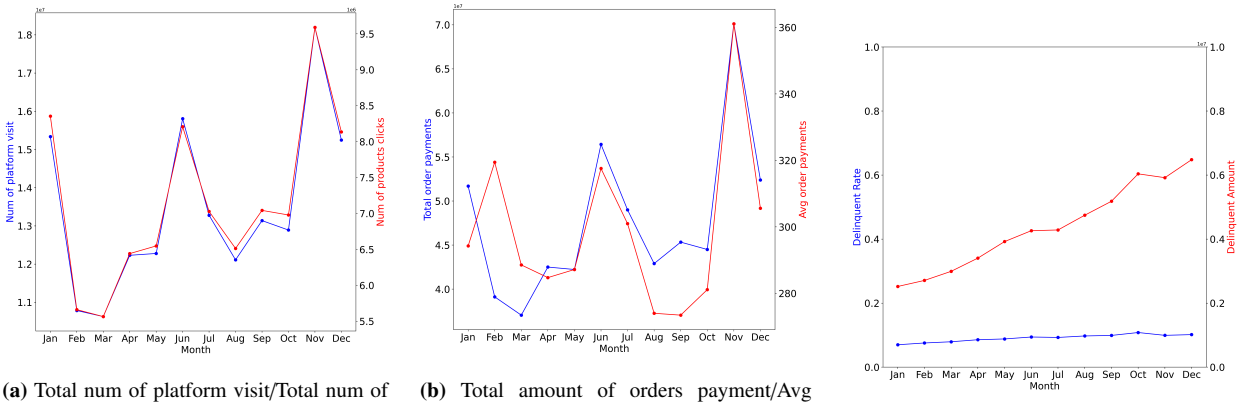
In the ordering dimension, we focus on several important features: the total orders amounts before and after discount, the discount of orders, the actual order payments, the average number of products per order, and the average number of orders per day. The data shows a high degree of variability in these aspects, as evidenced by standard deviations that are significantly higher than the means, reflecting the diverse range of prices and product types involved in different orders.

The browsing data highlights key features such as the number of platform visits, the number of products, shops, and brands browsed, and the usage of mobile versus PC devices for these activities. The data show that consumers are highly active in engaging in product exploration, indicating that consumers compare products across different shops and brands before purchasing. The data also emphasizes the dominance of mobile devices in e-commerce, with most browsing and purchasing actions conducted on mobile, while PC usage remains comparatively minimal.

In the credit management dimension, we examine features like the total amount of credit used, the installment period, the installment cost, and the amount of repayment and refund. Our findings suggest a prevalent use of credit among consumers, with a significant preference for the one-month repayment option. This preference is likely influenced by the interest-free period offered by the platform, highlighting consumer tendencies towards short-term financing options.

Last, we present statistics on the temporal intervals between successive ordering, browsing, and credit-management footprints for all consumers. On average, consumers exhibit a pattern of placing a order every 8.72 days, browsing a product every 2.22 days, managing credit every 8.16 days. Notably, the standard deviations for the temporal intervals in all three types of footprints surpass their respective means, indicating sporadic consumer footprints and diverse intervals between successive consumer footprints.

We further illustrate the temporal dynamics of some important variables in Figure D.12, including the total number of platform visits and product browsing, the total amount of order payments, the average number of order payments, the default rate, and the total default amount, over the 12-month observation period. The visualization reveals prominent peaks in both browsing and order placement during the months of June and November, which are attributed to promotional



(a) Total num of platform visit/Total num of products clicked

(b) Total amount of orders payment/Avg amount of orders payment

(c) The default rate/Total default amount.

Fig. D.12. The dynamics of browsing, order payments, and default rate/amount across 12 months.

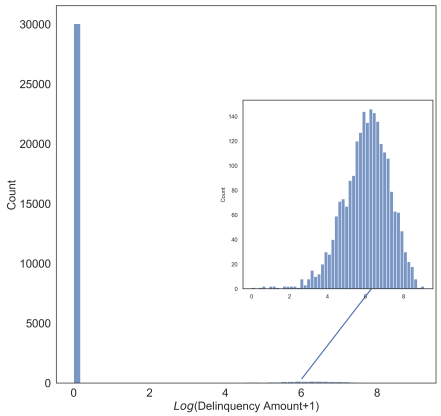


Fig. D.13. The default amount distribution in one month.

activities. Regarding credit risk, the default rate remains stable around 10%, while the total default amount exhibits a consistent upward trend. This trend may be attributed to factors such as an increase in individual default amounts and the potential accumulation of interest on overdue balances, emphasizing the need for proactive credit risk management by the platform.

At last, we plot the default amount distribution in our data, as in Figure D.13. Specifically, the default records pertaining to consumer behaviors exhibit an extreme imbalance, as illustrated in Figure D.13, where the distribution of log-transformed default amounts for one month is depicted. Notably, consumers without any default behavior account for 90% of the total consumer base, while consumers with default records constitute only 10%. Upon logarithmic transformation, the default distribution approximates a normal distribution, indicating that the original default amounts

follow an approximately log-normal distribution.

Appendix E. Introduction to the baseline models

Appendix E.1. GRU-D

GRU-D (Che et al., 2018) introduces decay factors to adapt the model to time series data where observations may be missing or occur at irregular intervals. These decay factors help in adjusting the model's hidden states based on the elapsed time since the last observation, thereby allowing the model to handle gaps in data effectively.

The decay rate for the hidden state, denoted as γ_h, is crucial for moderating the influence of previous observations on the current state. This rate is a function of the time gap δ_t between the current and the last observed time steps, usually defined as:

$$\gamma_h(t) = \exp(-\max(0, \mathbf{W}_\gamma \delta_t + \mathbf{b}_\gamma))$$

where $\mathbf{W}_\gamma$ and $\mathbf{b}$ are model parameters that we train jointly with all the other parameters of the GRU. This formulation ensures that the influence of the past states decays exponentially with the increase in time gap. Then, the decay factor integrates with hidden state of standard GRU equations as:

$$\mathbf{h}_{t-1} \leftarrow \gamma_h(t) \odot \mathbf{h}_{t-1}.$$

At last, the decayed hidden state $\mathbf{h}_{t-1}$ is feed directly into the standard GRU for updating with input $\mathbf{x}_t$.

Appendix E.2. T-LSTM Networks

The Time-Aware LSTM (T-LSTM) introduces a novel approach to handling time decay in LSTM models, which is essential for managing the varying intervals between observations in longitudinal patient records. The T-LSTM model modifies the standard LSTM architecture by introducing a time decay component into the cell state updates. Detailed mathematical expressions

of the proposed T-LSTM architecture are given below:

$$
\begin{aligned}
\mathbf{C}^{S}_{t-1} &= \tanh\left(\mathbf{W}_d \mathbf{C}_{t-1} + \mathbf{b}_d\right) && \text{(Short-term memory)} \\
\hat{\mathbf{C}}^{S}_{t-1} &= \mathbf{C}^{S}_{t-1} * g\left(\Delta_t\right) && \text{(Discounted short-term memory)} \\
\mathbf{C}^{T}_{t-1} &= \mathbf{C}_{t-1} - \mathbf{C}^{S}_{t-1} && \text{(Long-term memory)} \\
\mathbf{C}^{*}_{t-1} &= \mathbf{C}^{T}_{t-1} + \hat{\mathbf{C}}^{S}_{t-1} && \text{(Adjusted previous memory)} \\
\mathbf{f}_t &= \sigma\left(\mathbf{W}_f \mathbf{x}_t + \mathbf{U}_f h_{t-1} + \mathbf{b}_f\right) && \text{(Forget gate)} \\
\mathbf{i}_t &= \sigma\left(\mathbf{W}_i \mathbf{x}_t + \mathbf{U}_i h_{t-1} + \mathbf{b}_i\right) && \text{(Input gate)} \\
\mathbf{o}_t &= \sigma\left(\mathbf{W}_o \mathbf{x}_t + \mathbf{U}_o h_{t-1} + \mathbf{b}_o\right) && \text{(Output gate)} \\
\tilde{\mathbf{C}}_t &= \tanh\left(\mathbf{W}_c x_t + U_c h_{t-1} + b_c\right) && \text{(Canditate memory)} \\
\mathbf{C}_t &= \mathbf{f}_t \odot \mathbf{C}^{*}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{C}}_t && \text{(Current memory)} \\
\mathbf{h}_t &= \mathbf{o}_t \odot \tanh\left(\mathbf{C}_t\right) && \text{(Current hidden state)}
\end{aligned}
$$

where $\mathbf{x}_t$ represents the current input, $\mathbf{h}_{t-1}$ and $\mathbf{h}_t$ are previous and current hidden states, and $\mathbf{C}_{t-1}$ and $\mathbf{C}_t$ are previous and current cell memories. $\{\mathbf{W}_f, \mathbf{U}_f, \mathbf{b}_f\}$, $\{\mathbf{W}_i, \mathbf{U}_i, \mathbf{b}_i\}$, $\{\mathbf{W}_o, \mathbf{U}_o, \mathbf{b}_o\}$, and $\{\mathbf{W}_c, \mathbf{U}_c, \mathbf{b}_c\}$ are the network parameters of the forget, input, output gates and the candidate memory, respectively. $\{\mathbf{W}_d, \mathbf{b}_d\}$ are the network parameters of the subspace decomposition. Δt is the elapsed time between $\mathbf{x}_{t-1}$ and $\mathbf{x}_t$. $g(\Delta t) = 1/\log(e + \Delta t)$ can be chosen as the decay function.

Appendix E.3. MA-GRU Network

Multirepresentational Attention GRU (Yin et al., 2024) introduces a novel adaptation of the traditional GRU model to better handle continuous-time data within consumer digital footprints. This adaptation is crucial because standard GRU models, which process sequences using the following formula:

$$\mathbf{h}_t = \text{GRU}(\mathbf{h}_{t-1}, \mathbf{x}_t)$$

are typically structured to handle discrete-time data and therefore do not incorporate the continuous-time dynamics naturally present in consumer behavior data.

To address this limitation, multi-representational attention mechanism is designed to layer on top of the GRU to capture the elapsed time between events effectively. This mechanism computes an influence score for each hidden state produced by the GRU, using the formula:

$$e(t) = \tanh(\mathbf{u}^\top \mathbf{h}_t \gamma(t - t_j))$$

where $\gamma(t - t_j) = e^{-w(t-t_j)}$ represents a time-decaying function that diminishes the impact of past states over time. The parameters $\mathbf{u}$ and w are learned through the model training process, enabling the system to adaptively determine the relevance of each historical hidden state $\mathbf{h}_t$ in predicting the outcome. Higher influence scores indicate greater relevance.

Finally, the attention mechanism aggregates these influence scores to construct a continuous-time representation of the consumer's past behaviors. This is done by computing a weighted average of all previous hidden states, where the weights are the normalized influence scores:

$$\mathbf{s}(t) = \sum_{t_j<t} a(t_j)\mathbf{h}_{t_j}, \quad \text{where } a(t_j) = \frac{\exp(e(t_j))}{\sum_{t_j<t} \exp(e(t_j))}$$

This weighted summary, $s(t)$, effectively captures the significant historical influences on the current state, enhancing the model's ability to make more accurate predictions based on a comprehensive view of the consumer's past interactions. In the experiment, we compute the context vector at the last time point of each month, i.e., $\mathbf{s}(T)$, for classification.

Appendix E.4. GRU-Time2Vec

Time2Vec (Kazemi et al., 2019) is a temporal encoding technique designed to transform scalar time data into a high-dimensional vector representation, enhancing the ability of deep learning models to capture and leverage temporal dependencies in sequential data. Unlike traditional methods that treat time as a simple scalar or positional index, Time2Vec represents time as a combination of a linear component and multiple sinusoidal components, allowing the model to encode both linear trends and periodic patterns.

Mathematically, for a given time step t, Time2Vec generates an embedding $\mathbf{v}(t) \in \mathbb{R}^k$, where k is the dimension of the embedding. The first component of this embedding is a linear transformation:

$$v_0(t) = w_0 \cdot t + b_0$$

where w_0 and b_0 are learnable parameters that allow the model to capture linear temporal trends.

The remaining components are sinusoidal functions that capture periodic patterns:

$$v_i(t) = \sin(w_i \cdot t + b_i), \quad i = 1, 2, \ldots, k-1$$

where w_i and b_i are learnable parameters that determine the frequency and phase of each sinusoidal component, enabling the model to learn complex cyclical patterns in the data.

The final Time2Vec embedding is formed by concatenating the linear component and the sinusoidal components:

$$\mathbf{v}(t) = [v_0(t), v_1(t), \ldots, v_{k-1}(t)]$$

This embedding $\mathbf{v}(t)$ is then used as input to models such as GRU, LSTM, or Transformers, providing them with a rich temporal representation that enhances their ability to recognize and leverage the timing and periodicity of events in the data.

Appendix F. Computation Infrastructure

All studies and experiments are run on Dell Precision 7920 Workstations with Intel(R) Xeon(R) Gold 6256 CPU at 3.60GHz and three sets of NVIDIA Quadro GV100 GPUs. All models are implemented in Python 3.8. The versions of the main packages of our code are Pytorch 1.8.1+cu102, torchdiffeq: 0.2.2, Sklearn: 0.23.2, Numpy: 1.19.2, Pandas: 1.1.3, Matplotlib: 3.3.2.