The open banking era

Citation for published version:
Liu, J, Huang, S, Fu, Q, Luo, Y, Qin, S, Cao, Y, Zhai, J & Yang, S 2024, 'The open banking era: An optimal model for the emergency fund', *Expert Systems with Applications*, vol. 244, 122915.
https://doi.org/10.1016/j.eswa.2023.122915

Digital Object Identifier (DOI):
10.1016/j.eswa.2023.122915

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Expert Systems with Applications

The open banking era: an optimal model for the emergency fund

Junxiu Liu[1], Shaodong Huang[1], Qiang Fu[1*], Yuling Luo[1*], Sheng Qin[1],
Yi Cao[2], Jia Zhai[3], Su Yang[4]
[1]Guangxi Key Lab of Brain-inspired Computing and Intelligent Chips, School of Electronic and Information Engineering, Guangxi Normal University, Guilin, 541004, China
[2]Management Science and Business Economics Group, Business School, University of Edinburgh, Edinburgh, UK
[3]Business School, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu, China
[4]Department of Computer Science, Swansea University, Swansea, UK
j.liu@ieee.org, 371483169@qq.com, *{qiangfu, yuling0616}@gxnu.edu.cn,
qinsheng@gxnu.edu.cn, yi.cao@ed.ac.uk, Jia.Zhai@xjtlu.edu.cn, su.yang@swansea.ac.uk

Abstract: The COVID-19 outbreak has negatively impacted the income of many bank users. Many users without emergency funds had difficulty coping with this unexpected event and had to use credit or apply to the government for bailout funds. Therefore, it is necessary to develop spending plans and deposit plans based on transaction data of users to assist them in saving sufficient emergency funds to cope with unexpected events. In this paper, an emergency fund model is proposed, and two optimization algorithms are applied to solve the optimal solution of the model. Secondly, an early warning mechanism is proposed, i.e. an unexpected prevention index and a consumption index are proposed to measure the ability of users to cope with unexpected events and the reasonableness of their expenditure respectively, which provides early warning to users. Finally, the model is experimented with real bank users and the performance of the model is analysed. The experiments show that compared to the no-planning scenario, the model helps users to save more emergency funds to cope with unexpected events, furthermore, the proposed model is real-time and sensitive.

Keywords: Emergency fund; Optimization algorithms; Open Banking;

1. Introduction

At the end of 2019, COVID-19 exploded worldwide and a series of work stoppages, production shutdowns and quarantines led to a sharp drop in corporate profits, which resulted in a significant impact on people's the income. Many people struggled without emergency funds, resorting to credit or government bailout funds. It is reported that one in five British workers had no savings before the COVID-19 crisis. As a result of the COVID-19, people are aware of the importance of saving for emergency funds when unexpected events happen. The emergency fund is a type of savings to cover future unexpected events, which usually require a high level of expenditure and may be one-off or ongoing, such as a COVID-19 crisis, unemployment, major illness, etc. (Kumajas & Wuryaningrat, 2021). Usually people simply save 3 to 6 months of living expenditures as an emergency fund, which makes it difficult to help users cope with longer-cycle unexpected events.

The anonymous transaction data reveals that even higher income groups, often without an emergency fund, struggle with financial management. This is primarily due to a lack of rational expenditure planning and emergency fund allocation. Developing consumption plans tailored for different income groups could foster better consumption habits and resilience against future unforeseen events. This research has significant implications for both users and banks, as it can inform real-time alerts and future planning. The core challenge is twofold: helping users prepare for future unexpected events and extending the coping time in case of event failure to minimize reliance on credit or government bailouts. The model must consider users' varied expenditure and income levels, and for those with adequate emergency funds, optimize spending behaviour to improve expenditure quality. Essentially, the optimization problem is: how can the model advise on consumption while helping users prepare for or endure unexpected events for a sufficient duration?

The global spread of Covid-19 has led to numerous social challenges including economic losses, mental health issues, social isolation, educational disruptions, and medical resource shortages (Tisdell, 2020). In response, researchers have been exploring optimal decision-making strategies to balance epidemic readiness against these social issues. A variety of mathematical models have been developed to predict the trajectory and impact of the Covid-19 outbreak (Das et al., 2023; Taylor & Taylor, 2023), models based on infectious disease dynamics (Chang & Kaplan, 2023), including models based on infectious disease dynamics, machine learning, and deep learning (Science et al., 2020). The assist policy makers in predicting outbreak spread and impacts, facilitating appropriate response strategies. With the advent of Covid-19 vaccines, optimal vaccine distribution and vaccination strategies have been proposed to maximize public health and economic development (Carlo et al., 2023; Thul & Powell, 2023). The Covid-19 outbreak led to a strain on healthcare resources such as hospital beds, ventilators and medical staff, and decision-makers need to develop optimal healthcare resource allocation strategies to minimise the healthcare system. Many studies have proposed models and algorithms based on decision optimization to help decision-makers make the best decisions (K. Liu et al., 2023; Zhang et al., 2023). The socio-economic impact of the Covid-19 epidemic has been severe. Many approaches have developed policies and measures from an optimization perspective to help policymakers assess the socioeconomic impact of the epidemic (Chen et al., 2023; Zhang et al., 2023).

Decision optimization involves selecting the best decision to maximise or minimize an objective function within given constraints. It's essential in modern management and engineering, aiding decision-makers in understanding complex systems and developing optimal solutions. Decision optimization is widely applied in production, logistics, finance, energy sectors (Du et al., 2019; L, 2021; Nadizadeh et al., 2022; Zhu et al., 2021) , and in supply chain management. It can enhance operational efficiency and reduce costs by optimizing production planning, inventory management, and distribution routes (Yao et al., 2020). In the financial field, decision optimization can help investors to improve returns and reduce risks through portfolio optimization and risk management (Jalota et al., 2023; Steuer & Utz, 2023; Wu et al., 2022). Decision optimization in the energy sector allows companies to enhance energy production and dispatching efficiency, thereby reducing costs (L, 2021). Building and solving appropriate mathematical models is crucial in these problems. Various models, such as mixed integer linear programming (Er-rahmadi & Ma, 2022; Warwicker & Rebennack, 2023), dynamic programming (Cervellera, 2023; Lyu & Huang, 2023), and others like stochastic, non-linear, and multi-objective programming have been developed. Numerous effective solution algorithms have also emerged, including genetic, ant colony, and particle swarm algorithms, among others. Deep learning-based models and algorithms, such as deep reinforcement learning and deep neural networks, have also been applied in decision optimization (Herrera-viedma, 2020; Jang, 2019; Nour et al., 2020; J. J. Q. Yu et al., 2019). Multi-objective optimization, which focuses on balancing multiple objectives to achieve an optimal solution, is a significant aspect of decision optimization (Cao et al., 2020; He et al., 2021; Huang et al., 2020; Y. Liu et al., 2020). In recent years, many algorithms have emerged for multi-objective optimization problems, such as multi-objective genetic algorithms (Labani et al., 2020), multi-objective particle swarm optimization algorithms (Hu et al., 2021; Labani et al., 2020), and dominant ranking genetic algorithms. In addition, decision makers are often faced with uncertain environments. As a result, uncertainty decision-making has become an important direction in the study of decision optimization problems. It is mainly concerned with how to make decisions in an uncertain environment, including risky/fuzzy decision-makings, and robust optimization (Eko et al., 2023). In recent years, many effective algorithms and models have also emerged for uncertain decision problems, such as stochastic programming (Cristina et al., 2023; Fusco et al., 2023), robust optimization (Goerigk & Kurtz, 2023), and fuzzy programming (G. Yu & Li, 2022). With the advent of the era of big data, decision optimization problems are also facing new challenges and opportunities. Big data decision-making is mainly concerned with how to mine effective information in massive amounts of data to support decision-making. This requires research into new models and algorithms, such as machine learning, deep learning, data mining and other techniques based on big data. Big data decision-making research has been widely used in many fields, such as finance, healthcare and transportation (Jain et al., 2023). Therefore, the use of rich big data information can be effective in optimizing decisions. The aim of this paper is to optimise the expenditure of users by analysing their transaction data and building up an emergency fund, which helps users to save more emergency funds for unexpected events.

The COVID-19 pandemic has highlighted the importance of emergency funds in coping with unexpected events. However, many individuals, even those with higher incomes, struggle with financial management and lack adequate emergency funds due to improper expenditure planning and allocation. The research problem, therefore, lies in developing a model that can help users prepare for future unexpected events, extend their coping time in case of event failure, and optimize their spending behaviour to improve expenditure quality. The model should be able to advise on consumption while helping users prepare for or endure unexpected events for a sufficient duration, considering varied expenditure and income levels. The objective of this research is to optimize the expenditure of users by analyzing their transaction data and building up an emergency fund. The study aims to maximize the coverage time of the emergency fund for unexpected events, improve users' spending quality when they can handle unexpected events, and maximize the duration that the emergency fund covers the event when the model is deployed after the unexpected event has occurred. The study also seeks to develop an alert mechanism for real-time monitoring and early alerts to guide user consumption behaviour and assess their preparedness for unexpected events. The model's effectiveness will be validated using both real-world banking transaction data and simulation data. The ultimate goal is to assist users in effectively saving for emergency funds and provide a new metric system to measure the capacity of users to handle unexpected events and the reasonableness of their expenditure.

This paper makes three assumptions for the problem at hand. First, it is assumed that the user's emergency fund cannot fully cover the duration of the unexpected event, and the model's optimization goal is to maximize this coverage time. Second, it is assumed that the planned emergency fund can cover the entire event, or even longer, with the model's optimization goal being expenditure optimization, i.e. it aims to enable users to spend more rationally and improve spending quality when they can handle unexpected events. Third, it is assumed that the model is deployed after the unexpected event has occurred, with the model's optimization goal being to maximize the duration that the emergency fund covers the event. An objective function is formulated for the model to solve this problem. Genetic Algorithms (GA) and Particle Swarm Optimization algorithms (PSO) are employed to find the optimal value. To provide real-time monitoring and early alerts, an alert mechanism is proposed to guide user consumption behaviour and assess their preparedness for unexpected events. Specifically, the unexpected prevention capability index is proposed to measure users' ability to handle unexpected events and a consumption index is proposed to measure real-time user spending, gauging the extent of actual expenditure deviation from the plan.

To validate the model, firstly, several typical user types extracted from anonymous banking transaction data provided by our FinTech partners. Secondly, the data from the Ornstein-Uhlenbeck (OU) process simulations are used for further validation. The model's feasibility is tested across the three scenarios mentioned earlier and compared with situations where the model is not utilized. This comparison helps verify the model's validity and flexibility. Finally, to test the early warning performance of the proposed indices, a scenario where a user does not follow the model's suggestions is simulated. The resulting changes in the index values over the months demonstrate the model's ability to provide real-time early warnings for users.

The main contributions and the motivations of this research are summarized as follows.

1) It is the first time that this real-life problem raised by COVID-19 is considered and described from an optimization perspective, i.e., assuming that an unexpected event occurs in the future, how the users plan their savings and spending to cope with the event.
2) A model is formulated for this realistic problem, in addition, two optimization algorithms are applied to solve the problem and the corresponding solution performance is analysed.

3) A solution to the problem is proposed, i.e. banks should be proactive and give certain advice and warnings about the spending habits and deposit situations of their users.
4) It is the first time that the unexpected prevention capability index is proposed to measure the ability of users to cope with unexpected events. It is the first time that a consumption index is proposed to measure the reasonableness of the expenditure of the user. In addition, a real-time monitoring and alerting mechanism are provided for users based on these two indexes.

In summary, this is a novel approach that can assist users building a more effectively saving for emergency funds. Additionally, the proposed early warning mechanism, including the unexpected prevention index and consumption index, offers a new metric system to measure the capacity of users to handle unexpected events and the reasonableness of their expenditure. These are elements not present in existing technology, thus filling a technological gap.

The remainder of the paper is organised as follows. Section 2 describes and analyses the emergency fund problem and introduces the proposed model. Section 3 analyses and verifies the model based on real user transaction data. Section 4 concludes the paper.

2. Methodology

The objective of the model is to assist users in saving sufficient emergency funds to cover their living expenses under unexpected events by optimizing their spending structure. In this section, the notation of the model is defined and the emergency fund is described as an optimization problem. To define the notation, we assume the current time is T_n and the event occurs at the time T_s and ends at T_e, the user's emergency fund can sustain the event until T_m, where we usually have $T_s \leq T_e$ and $T_n \leq T_m$ but other orders are unclear.

Assuming we start from the time T_n with a situation where an individual user has income, daily spending, certain debts and an emergency fund, all of which can be non-negative and stochastic. An unexpected event that lasts for t_2 months will happen at T_s time. The income of the user is impacted very negatively by the event and reduces at different rates. If the user keeps spending the same amount, they risk running out of money. This paper aims to answer the question: After the outbreak of the event, as users' income declines, how should users adjust their consumption to tide over the difficulties?

As illustrated in Figure 1, we describe three scenarios of T_s and T_e as well as T_n and T_m, based on an overall consideration of emergency funds saving throughout those four time periods. The details of modelling the user's future financial situation are discussed in the following sections.

(1) Scenario 1 in Figure 1 considers a sequence of $T_n \leq T_s \leq T_e \leq T_m$, where the unexpected event occurs after the current time T_n, and the users can cover their living expenses with some loans. Hence, during the period from T_n to T_e, users' total income (including the loan) can be higher than their total expenditure. The optimization objective of our model in this scenario is to make the user's spending more reasonable and reduce the loan while being able to cope with events.
(2) Scenario 2 in Figure 1 considers a sequence of $T_n \leq T_s \leq T_m \leq T_e$, where the

unexpected event occurs after the current time T_n, during the period from T_n to T_e, the user's total income (including loans) is lower than their total expenditure and cannot cope with the event. The optimization objective of the model in this scenario is to maximise the response time t_3 to reduce the bailout and debit requested by the user.

(3) Scenario 3 in Figure 1 considers a sequence of $T_s \leq T_n \leq T_m \leq T_e$, where the unexpected event occurs before the current time point T_n and lasts for the time t_4. Hence, during the period from T_n to T_e, the total income of user (including the loan) can be higher or lower than their total expenditure. In practice, because of the impact of unexpected events, the total income of most users is lower than their total expenditure. In this scenario, the objective of the model optimization is to maximise the duration t_3 and provide an early warning and a new plan.

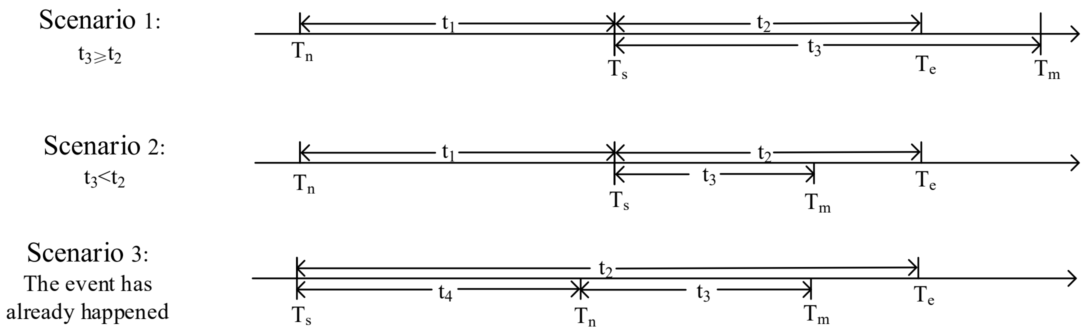


Figure 1. Model consideration of the situation.

Figure 2 shows the workflow of our proposed methodology. The objective optimization process includes the following steps:

(1) Data Extraction: Income and various categories of expenditure data are extracted from the anonymous user transaction data for 2017 and 2018. This data serves as the basis for calibrating the parameters (θ, *mu*, *k*) of the Ornstein-Uhlenbeck (OU) process.

(2) Parameter calibration and future estimation: Firstly, by using the extracted data, we calibrate the OU process parameters (θ, *mu*, *k*). The purpose of calibrating these parameters is to ensure that the results estimated by the OU process match the user's characteristics. Second, the OU process is used to estimate the user's future income and expenditure. These steps are described in detail in Section 3.2.1.

(3) Optimization: The calibrated data is input into the model's objective function, which is then optimized using GA and PSO algorithms to determine the maximum coping time (t_{max}) and the corresponding advised expenditure values ($\overline{eb}_i, \overline{er}_i, \overline{el}_i$).

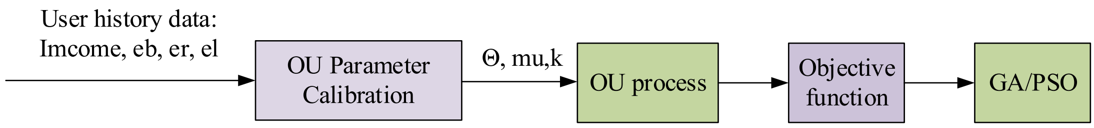


Figure 2. Objective parameter optimization process.

2.1. Formulation of optimization model of the emergency fund

2.1.1. Emergency fund model objective function

The model notation is defined in Table I. Superscript $\sim$ indicates the actual value and superscript $-$ indicates the suggested value. T represents the time point and t represents the length of time.

Table I. Model notation definitions.

$\sim$	Actual value.
$-$	Suggested value.
T	Point of time.
t	Length of time.
i	Time index.
T_n	Current time, the time point of using the model.
T_s	The point in time when the event happened.
T_e	The time point at when the event ends.
T_m	The maximum point in time that an emergency fund can cover an event.
t_1	The length of time from T_n to T_s, that is, how long after the event occurs.
t_2	The length of time the event lasts, from T_s to T_e.
t_3	The maximum time that the emergency fund can cover the event, from T_s to T_m.
s	An additional one-time expense due to the event, i.e., medical expense during Covid-19.
$\widetilde{income}$	Actual income amount for this month.
$\overline{income}$	Estimated income amount for this month.
$\widetilde{AVG}_{income_i}$	The monthly average income of the ith month is an actual value.
$\widetilde{eb}_i$	The actual basic expenditure amount in the ith month.
$\widetilde{er}_i$	The actual amount of current expenditure in the ith month.
$\widetilde{el}_i$	The actual expenditure amount of luxury goods in the ith month.
$\widetilde{AVG}_{eb_i}$	The monthly average basic expenditure amount of the ith month.
$\widetilde{AVG}_{er_i}$	The monthly average recurrent expenditure value of the ith month.
$\widetilde{AVG}_{el_i}$	The Monthly average luxury expenditure in the ith month.
$\widetilde{m}_i$	The actual amount of emergency funds in the ith month.
$\widetilde{m}_{s_i}$	A total emergency fund of the ith month.
$\overline{m}_i$	A suggested emergency fund for ith month.
$\overline{eb}_i$	Suggested basic expenditure amount in the ith month.

$\overline{er}_i$	Suggested recurrent expenditure amount in the ith month.
$\overline{el}_i$	Suggested luxury expenditure amount in the ith month.
σ	Income decay factor.

The model aims to maximize the time t_3 that the user may live with adequate living expenses when an unexpected event occurs. To achieve this objective, the model adjusts the living expenses with the reference to user's historical data. The model output the adjusted living expenses as the suggested expenditure values $\overline{eb}_i$、$\overline{er}_i$ and $\overline{el}_i$ and the optimized t_3 is the maximum duration that the user may live. The optimization is determined by the following factors:

The actual amount of emergency funds $\widetilde{\boldsymbol{m}}_{s_i}$ saved after i months of using the model. It is worth noting that at any time T_n of using our model, the user's income and expenditure after the occurrence of the event are unknown, so the decisions made by the model at any time point are based on estimates of the user's future income and expenditure. The model expects the emergency fund to grow monotonically before an unexpected event happens, but in practice, users may not implement the model's suggestions and expenditure may exceed the model's suggested value or even exhaust the existing emergency fund.

User income. Since the model is planned before unexpected events, the income should contain two components. The first component is the income before the unexpected event, which is the income during the period t_1. The income during this period is not affected by the unexpected event, so the income should be normal and we use the historical income data as a reference. We assume that at the normal time, the income of most people is a stable stochastic process with some variation. The OU process, which is explained in the next section, is a frequently used stochastic process in financial mathematics and is used in this study to model the user's income during the normal time. The second component is the income after the occurrence of the event in t_3 period. Due to the negative impact of the event, income is often reduced or even zero during the event period. For instance, the majority of people experience varying degrees of income loss during the COVID-19 pandemic, with some even earning nothing at all. This effect continued throughout the epidemic period. Our model describes this effect using an exponential decay factor $\sigma = e^{-\lambda t}$ on the user's normal income, with λ determining the magnitude to which it is affected.

The users' actual living spending. In this paper, we use the anonymous transaction data provided by the collaborated fintech company. The daily transactional data is pre-processed and the user expenditure is summarised into three categories: basic expenditure $\widetilde{eb}_i$, recurrent expenditure $\widetilde{er}_i$, and luxury expenditure $\widetilde{el}_i$. The basic expenditure, $\widetilde{eb}_i$, is usually the fundamental expenses for living, such as utility bills, clothing, food, housing and transport. The recurrent expenditure, $\widetilde{er}_i$, includes entertainment and buying household goods. The luxury expenditure, $\widetilde{el}_i$, however, is used for holidays and buying luxury goods.

Spending on unexpected events. The model assumes a total expenditure on unexpected events s addition to the living expenses. For example, the medical expenditure during the Covid-19 pandemic period.

The objective function according to the three scenarios can be established in Figure 1. Because users' living expenses and users' income are typically updated monthly from the transactional data we use in this work, all time variables, t_1, t_2, and t_3 are expressed in months. We can simply denote the total income before and after the occurrence of the event by $\overline{income} * t_1$ and $\overline{income} * \sigma * t_3$ respectively, where the decay factor σ is defined as $\sigma = e^{-\lambda t}$ to model the income decrease. Hence, the total amount of cash that users can spend on living expenses over t_1 and t_3 can be denoted as the sum of

$$\widetilde{m}_{s_i} + \overline{income} \times t_1 + \overline{income} \times \sigma \times t_3 - s. \tag{1}$$

Furthermore, we consider the total amount of cash during t_3 only and calculate the time t_3 as

$$t_3 = \frac{\widetilde{m}_{s_i} + \left[\overline{income} - \left(\widetilde{eb}_i + \widetilde{er}_i + \widetilde{el}_i\right)\right] \times t_1 + \overline{income} \times \sigma \times t_3 - s}{\overline{eb}_i + \overline{er}_i + \overline{el}_i}, \tag{2}$$

where $\left[\overline{income} - \left(\widetilde{eb}_i + \widetilde{er}_i + \widetilde{el}_i\right)\right] \times t_1$ denote the cash that the user may have at the time point of the occurrence of the event T_s. To simplify this equation, we denote the $\overline{income} - \left(\widetilde{eb}_i + \widetilde{er}_i + \widetilde{el}_i\right)$ as $\overline{income}'$, and calculate the t_3 as

$$t_3 = \frac{\widetilde{m}_{s_i} + \overline{income}' \times t_1 - s}{\left(\overline{eb}_i + \overline{er}_i + \overline{el}_i\right) - \overline{income} \times \sigma}. \tag{3}$$

In scenario 1, we have $t_3 \geq t_2$, where the users can cover their living expenses with some loans. To reduce the amount of the loan, the objective of the model is to minimize the living expenses while maintaining a proper living conditions and satisfying the constraint $t_3 \geq t_2$, which is the time user can cover their living expenses with some loan that is longer than or equal to the time of the unexpected event. Therefore, the objective function in scenario 1 can be expressed as

$$\underset{\overline{eb}_i, \overline{er}_i, \overline{el}_i}{\operatorname{argmax}} \left[\frac{\widetilde{m}_{s_i} + \overline{income}' \times t_1 - s}{\left(\overline{eb}_i + \overline{er}_i + \overline{el}_i\right) - \overline{income} \times \sigma}\right], w.r.t\ t_3 \geq t_2. \tag{4}$$

In scenario 2, the model objective is identical while the constraint is changed to $t_2 \geq t_3$, which means that the users cannot cover their living expenses. The goal is to make the users' self-sufficiency time as close as feasible to the length of the unexpected event. As a result, the objective function of scenario 2 can be written as

$$\underset{\overline{eb}_i,\overline{er}_i,\overline{el}_i}{\operatorname{argmax}}\left[\frac{\widetilde{m}_{s_i}+\overline{\imath ncome}'\times t_1-s}{\left(\overline{eb}_i+\overline{er}_i+\overline{el}_i\right)-\overline{\imath ncome}\times\sigma}\right], w.r.t\ t_2\geq t_3. \tag{5}$$

In scenario 3, the unexpected event occurs before the current time point T_n and lasts for a very long time that the user is unable to endure. The aim is to increase the user endurance t_3 as much as feasible to alleviate the pressure of government for financial aid. We can denote the endurance t_3 as

$$t_3=\frac{\widetilde{m}_{s_i}+\overline{\imath ncome}'\times\sigma\times t_4+\overline{\imath ncome}\times\sigma\times t_3-s}{\overline{eb}_i+\overline{er}_i+\overline{el}_i}. \tag{6}$$

After re-organizing the equation, the objective function of scenario 3 can be denoted as

$$\underset{\overline{eb}_i,\overline{er}_i,\overline{el}_i}{\operatorname{argmax}}\left[\frac{\widetilde{m}_{s_i}+\overline{\imath ncome}'\times\sigma\times t_4-s}{\left(\overline{eb}_i+\overline{er}_i+\overline{el}_i\right)-\overline{\imath ncome}\times\sigma}\right], w.r.t\ t_2>t_3. \tag{7}$$

$\widetilde{AVG}_{eb_i}$, $\widetilde{AVG}_{er_i}$, $\widetilde{AVG}_{el_i}$ denote the level of the expenditure of the user in the short term, and the level of the spending of the user in the short term is relatively stable. Assuming that users continue to maintain such spending, the emergency fund can cover the event t_m time. t_m can be considered as an early warning value, indicating that the emergency fund can cover the event t_m time at most. t_m can be written as

$$t_m=\frac{\widetilde{m}_{s_i}+\overline{\imath ncome}\times\sigma\times t_3-s}{\widetilde{AVG}_{eb_i}+\widetilde{AVG}_{er_i}+\widetilde{AVG}_{el_i}}. \tag{8}$$

2.1.2. Constraints

Based on the discussion and the objective functions, we summarize all constraints in our model. Our initial constraints include that the length of time, income, consumption expenditure, emergency fund and event spent in the objective function should all be non-negative, i.e.

$$t_1,t_2,t_3,\widetilde{\imath ncome},\widetilde{eb}_i,\widetilde{er}_i,\widetilde{el}_i,\widetilde{m}_{s_i},s\geq 0. \tag{9}$$

Considering that sometimes the income is not enough to cover the expenditure, the user can only use the current emergency fund $\widetilde{m}_{s_i}$. As a result, the emergency fund $\widetilde{m}_i$ stored each month may be a negative amount. The total emergency fund is the sum of the actual emergency fund storage amounts for each month from the moment T_n. The total emergency fund $\widetilde{m}_{s_i}$ can be calculated by

$$\widetilde{m}_{s_i}=\sum_{0}^{i}\widetilde{m}_i. \tag{10}$$

Since users' income and expenditure often have certain randomness, fluctuating up and down in a long-term equilibrium value, we model those stochastic variables using the OU process (Uhlenbeck G.E. & Ornstein L.S., 1930). The OU process is often applied to modelling stochastic asset prices in financial mathematics. In our work, the incomes of users and expenditures exhibit the same characteristics as asset prices: they are variables that fluctuate around their long-term mean over time.

Denoting the rate of mean reversion as θ, volatility as k, the Wiener process as W_i, and the user's average monthly income over the last n months as AVG_{income_i}, we can write the OU process for the income of the user as

$$\overline{income}_{i+1} = \overline{income}_i + \theta_{income} \times \left(AVG_{income_{i-1}} - \overline{income}_{i-1}\right) + k_{income} \times dW_i, \tag{11}$$

where the average monthly income of the user is calculated using their monthly income over the previous n months, $AVG_{income_i} = \sum_{i-n}^{i} \widehat{income}_i, n > 0$. In this OU process, the Wiener process describes the randomness of the user's income and the volatility k denotes the amplitude of the uncertainty: a higher value of k suggests a greater degree of uncertainty.

Experiments are conducted to identify the proper parameters to make the income and spending values described by the OU process more compatible with the actual financial condition of the user. It's important to maintain an emergency fund so that they can cover essential living expenses. Consequently, spending constraints must be established. In this work, we analyse anonymous transactional data of users provided by a collaborative FinTech firm and divide user expenditures into three groups: basic, recurrent, and luxury expenditures.

We use historical data to calibrate the OU process to determine the appropriate parameters so that the income and expense values described by the OU process are more in line with the actual financial situation of the user. Thus, we establish the OU process for three types of user expenditure: basic ($\widetilde{eb}_{i+1}$), recurrent ($\widetilde{er}_{i+1}$), and luxury ($\widetilde{el}_{i+1}$), i.e.

$$\widetilde{eb}_{i+1} = \widetilde{eb}_i + \theta_{eb} \times \left(\widetilde{AVG}_{eb_i} - \widetilde{eb}_i\right) + k_{eb} \times dW_i, \tag{12}$$

$$\widetilde{er}_{i+1} = \widetilde{er}_i + \theta_{er} \times \left(\widetilde{AVG}_{er_i} - \widetilde{er}_i\right) + k_{er} \times dW_i, \tag{13}$$

$$\widetilde{el}_{i+1} = \widetilde{el}_i + \theta_{el} \times \left(\widetilde{AVG}_{el_i} - \widetilde{el}_i\right) + k_{el} \times dW_i. \tag{14}$$

Specifically, the Kalman filter is used to calculate the optimal estimate, then the maximum likelihood function is calculated and the unknown model parameters are estimated. We implement the Kalman filter using the Python library Pykalman, and the maximum likelihood optimization is implemented by the Scipy library. This part of the work is easy to implement in Python.

2.2. Model solution and evaluation index

2.2.1. Model solution

The objective function and constraints are established in (2) - (9) in Section 2.1.2. Genetic Algorithms (GA) and Particle Swarm Optimization (PSO) algorithms are global optimization algorithms. The algorithms search in the global solution space and focus their search on the optimal region, resulting in faster convergence (Amer & Namaane, 2013). Therefore, in this paper, GA and PSO are used to solve for the optimal value t_3 and its corresponding optimal inputs $\overline{eb}_i, \overline{er}_i$ and $\overline{el}_i$. In the methodology we use, constraints are managed as shown by (9) which delineates the constraints intrinsic to our model, demarcating the permissible range of values for the variables within our model, i.e. a range that is predicated on the real-world scenario under investigation. Subsequently, the optimization algorithm embarks on a search for the optimal solution within this explicitly defined value range. This approach guarantees that the solutions generated by the proposed model are mathematically optimal, while concurrently ensuring their practicality and applicability to the real-world scenarios.

The GA algorithm is a method for searching for optimal solutions by simulating the natural evolutionary process. The following is a description of some of the terminology of the GA algorithm, where the population is a set of possible solutions, the individual is one of the solutions, the chromosome can be understood as the number of variables, and the probability of variation indicates the probability of a change in the chromosome. The GA algorithm calculates the fitness of an individual by simulating the mechanisms of selection and variation of the evolutionary process. Fitness can be understood as the value of the objective function corresponding to the individual. In each iteration, a set of candidate individuals is selected based on fitness. Repeating these steps, the population evolves over several generations and the best population is obtained, which is the optimal solution to the objective function. Specifically, the population is initialised within the constrained search range respectively, each individual in the population is substituted into the objective function of the model, the fitness of each individual is calculated, and some individuals are selected to evolve. The iterations are repeated until the optimal fitness is found, which is the optimal value of the model t_3. The specific parameters of the genetic algorithm are as follows: 100 individuals of the population are initialised, the number of chromosomes is 3, the number of iterations is 100, the mutation probability is set to 0.2 and the crossover probability is 0.2. Figure 3 shows the iterative process of the GA, and the model converges at an iteration number approximately equal to 10.

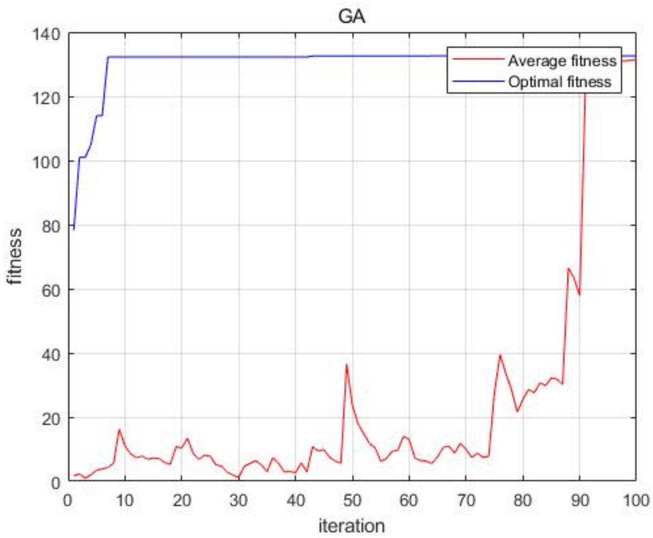


Figure 3. The iterative process of the GA.

Similar to the GA algorithm, the PSO algorithm searches each particle individually in space, the fitness being used to decide the solution to the objective function. In contrast to GA, PSO can memorise the current optimal solution and continue the search in this direction. The inertia weights, self-learning coefficients and population learning coefficients are important parameters of PSO. The inertia weight is the speed of particle movement and it determines the speed of convergence. The self-learning coefficient and the group-learning coefficient can be understood as the step size of the search. The specific parameters of the particle swarm algorithm are as follows: 100 individuals of the population are initialised, the number of iterations is 200, the inertia weight is 0.8, the self-learning factor is 0.5 and the population learning factor is 0.5. Figure 4 illustrates the iterative process of the PSO, where the model converges at several iterations approximately equal to 10.

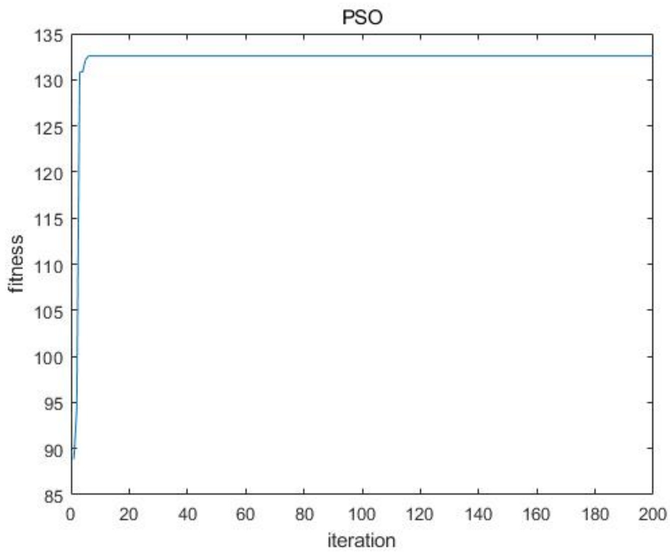


Figure 4. The iterative process of the PSO.

The optimal solution t_3 corresponds to the variables $\overline{eb_i}, \overline{er_i}$ and $\overline{el_i}$ which are the expenditures suggested by the model, and the emergency fund for the month is calculated by

$$\widetilde{m}_i = \widetilde{AVG}_{income_i} - \overline{eb}_i - \overline{er}_i - \overline{el}_i. \quad (15)$$

(10) calculates the total amount of emergency funds at each point in time. $\widetilde{m}_{s_i} \geq 0$ shows that the current user's emergency fund can still cover the event, $\widetilde{m}_{s_i} < 0$ shows that the user is already in debt and cannot cover the unexpected event.

For the genetic algorithm, the time complexity is O(*T***M** *N*), where *T* represents the number of iterations, *M* is the population size, and *N* is the length of the individual chromosome. The space complexity is O(*M***N*), i.e. the memory required to store the population. For the particle swarm optimization algorithm, the time complexity is O(*T***P***N*), where *P* is the size of the particle swarm. The space complexity is O(*P***N*), i.e. the memory needed to store all particles and their positions and velocity information.

2.2.2. Prevention capability index

To evaluate the ability of users to cover unexpected events, the Prevention capability index (PCI) is proposed to measure the current emergency funds and the ability to plan for unexpected events. It can be calculated by

$$PCI = \frac{t_3}{t_3 + t_2}, \quad (16)$$

where t_3 is the max time that the user can cover for an unexpected event and t_2 is the duration of the unexpected event. 0.5 is the threshold for the user's ability to cope with unexpected events. $PCI < 0.5$ means that the current emergency fund and the plan cannot cover the contingency to the end. $PCI \geq 0.5$ means that the currently planned emergency fund covers the entire unexpected event. PCI is directly proportional to the user's ability to cover unexpected events. For example, in scenario 1, $PCI < 0.5$, which indicates that the user is better able to cope with the event. In scenario 2, $PCI \geq 0.5$, which indicates that the user is unable to cope with the unexpected event to the end of the event. In scenario 3, the user's PCI may be less than or greater than 0.5.

2.2.3. Consumer Index

To guide users to spend wisely, a consumption index (CI) is proposed in this paper to measure the reasonableness of users' various expenditures. The CI considers several factors, the first being the degree of deviation between the actual and suggested values of expenditure. The second is the degree of balance between the various types of expenditure. CI is calculated from (17), $CI \in [0,1)$. A smaller CI indicates that the actual value deviates less from the suggested value, and when $CI = 0$, it indicates that the user has fully complied with the model's suggestions. The actual and suggested values of expenditure are different at each point in time. Therefore, the CI changes dynamically over time. It can be calculated by

$$CI = \frac{1}{3} \times \left(\left| \frac{\widetilde{eb}_i - \overline{eb}_i}{\widetilde{eb}_i + \overline{eb}_i} \right| + \left| \frac{\widetilde{er}_i - \overline{er}_i}{\widetilde{er}_i + \overline{er}_i} \right| + \left| \frac{\widetilde{el}_i - \overline{el}_i}{\widetilde{el}_i + \overline{el}_i} \right| \right). \quad (17)$$

2.2.4. Real-time monitoring and alerts

Because PCI and CI are dynamic over time, the ability of users to cover unexpected events and the reasonableness of expenditure can be measured in real-time by PCI and CI respectively. Therefore, PCI and CI can provide real-time monitoring and guidance. The model sets a threshold value for PCI and CI to alert the user. For example, when $PCI < 0.5$, users are reminded that their current consumption situation and plans cannot cover future unexpected events and that it would be dangerous to continue to overspend. When $CI = 0.9$, the user is alerted to the fact that the current risk is low and the user can have more spending. Similarly, when the $CI > 0.6$, the user is alerted that the current expenditure has deviated significantly from the plan and that continuation will increase the risk. Of course, the PCI and CI thresholds can be set by the user to meet the needs of users with different risk mentalities. PCI and CI can provide monitoring and alerts to users, hence reducing the risk to them. In addition, if users implement the model's plan, they will only need to apply for a small or no bailout from the government when the unexpected happens, which greatly reduces the financial pressure on the government and lowers the risk to society.

3. Experiments and analysis

3.1. Datasets

In this paper, our work is based entirely on the anonymous bank transaction data provided by our collaborating FinTech companies, which are regulated under the Open Banking Agreement and comply with the General Data Protection Regulation (GDPR) (*General Data Protection Regulation*, n.d.). This dataset contains transaction data from bank customers for 2017 and 2018 and contains 10 million transactions from nearly 20,000 customers. Each transaction record contains 15 attributes, including category, sub-category, transaction type, account ID, account provider, account type, amount, company ID, debit, merchant line of business, description, provider category, transaction date, and transaction. The saving of the emergency fund is only relevant to the financial situation of the user, therefore, only attributes relevant to the financial situation are selected from the raw data. Table 2 summarises the attributes that are relevant to this study, including category, sub-category, transaction type, account ID, amount and date of transaction. We analysed the raw data and found that the types of transactions included income and expenditure. The income categories are divided into non-recurrent income and recurrent income. The expenditure categories are divided into three types of expenditure: basic, discretionary and luxury. These income and expenditure categories are also divided into several sub-categories, e.g. non-recurrent income includes expense refunds and others, and basic expenditure includes bank repayment, bank fees, etc. Each transaction is given a specific sub-category and amount, and the features required by the model (income, base expenditure, discretionary expenditure and luxury expenditure) are extracted from these attributes. The income of the user is obtained by summing all the sub-categories of the income of the customer. Similarly, basic, discretionary and luxury expenditures are obtained by summing the amounts of their corresponding sub-categories.

Table II. Raw transaction data.

<table>
<tr><th>Transaction Type</th><th>Category</th><th>Subcategory</th><th>Amount</th><th>Transaction Date</th><th>User id</th></tr>
<tr><td rowspan="6">Income</td><td rowspan="2">Non-Recurrent Income</td><td>Expense Refund</td><td rowspan="27">Amount of the transaction</td><td rowspan="27">Date of transaction: (Year-Month-Day)</td><td rowspan="27">User ID, a number</td></tr>
<tr><td>Other</td></tr>
<tr><td rowspan="4">Recurrent Income</td><td>Salary</td></tr>
<tr><td>Rental Income</td></tr>
<tr><td>Benefits</td></tr>
<tr><td>Interest Income</td></tr>
<tr><td rowspan="21">Expenditure</td><td rowspan="12">Basic Expenditure</td><td>Bank Repayment</td></tr>
<tr><td>Bank Fees</td></tr>
<tr><td>Tax</td></tr>
<tr><td>Education</td></tr>
<tr><td>Housing</td></tr>
<tr><td>Other</td></tr>
<tr><td>Bank Interest</td></tr>
<tr><td>Transport</td></tr>
<tr><td>Utilities</td></tr>
<tr><td>Groceries</td></tr>
<tr><td>Insurance</td></tr>
<tr><td>Health</td></tr>
<tr><td rowspan="6">Discretionary Expenditure</td><td>Products</td></tr>
<tr><td>Services</td></tr>
<tr><td>Entertainment</td></tr>
<tr><td>Cash</td></tr>
<tr><td>Other</td></tr>
<tr><td>Food Drink</td></tr>
<tr><td rowspan="3">Luxury Expenditure</td><td>Holidays</td></tr>
<tr><td>Luxury Services</td></tr>
<tr><td>Luxury Products</td></tr>
</table>

As shown in Figure 5, the majority of users are in debt during the two years 2017 and 2018, indicating that these customers are out-spending their income affordability. In addition, although many customers have an emergency fund in some months, they do not have an emergency fund in most months, which indicates a lack of proper planning for their spending. Only a very small number of customers maintained an emergency fund during these two years. The results in Table 3 indicate that it is very meaningful to build models to help customers optimise their spending behaviour and to store emergency funds.

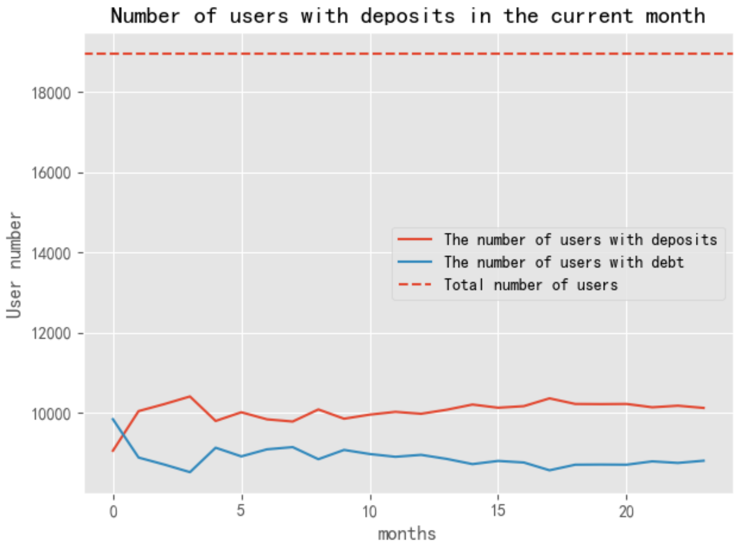


Figure 5. Emergency fund deposits of the surveyed customers.

3.2. Model validation

3.2.1. Event impact & OU process

To verify the model, four typical categories of users are selected from the selection data, and one user from each category is randomly chosen for analysis. As shown in Table III, these users have distinct characteristics. The income and expenditure of these four categories of customers in 2017 and 2018 are shown in Figure 6. In the experiments that follow, these four categories of users are used to validate the proposed model.

Table III. Features of the four users.

User	User 1	User 2	User 3	User 4
Average monthly income	2686	2122	32497	72561
Average monthly expenditure	2848	1350	13178	39093
Total income in 2017 and 2018	64474	50933	779925	1741484
Total expenditure in 2017 and 2018	68366	32421	316281	938246

Feature	Low income, high expenditure, no savings	Low income and low expenditure but have savings	High income, high expenses and no savings	High income, low expenditure and have savings

The purpose of the model is to plan for the future spending of the user, however, for the current point in time, the future income and expenditure of the user are unknown. Therefore, the model needs to simulate the future income and expenditure of the user and make plans based on this data. To enable the simulated data to reflect the real financial situation of the users, the experiments to simulate the future data of the users are based on user data from 2017 and 2018. The income and expenditure of users are values that fluctuate around the mean in the near term, so the OU process is used to simulate the future income and expenditure of the four users shown in Table III. Because the OU process is based on historical data, the results can be considered as an approximation of the future income and expenditure of the user.

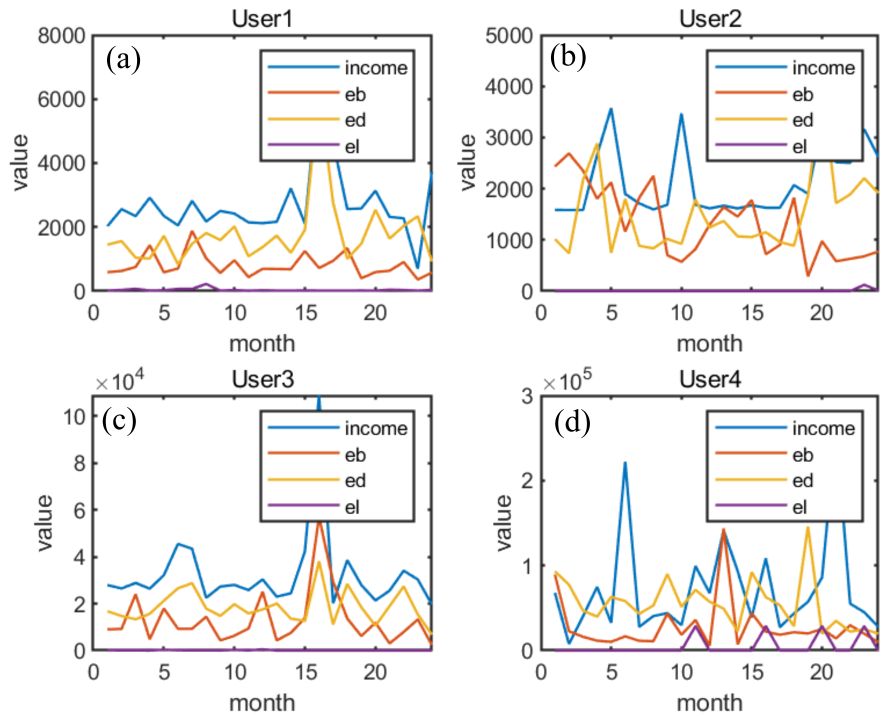


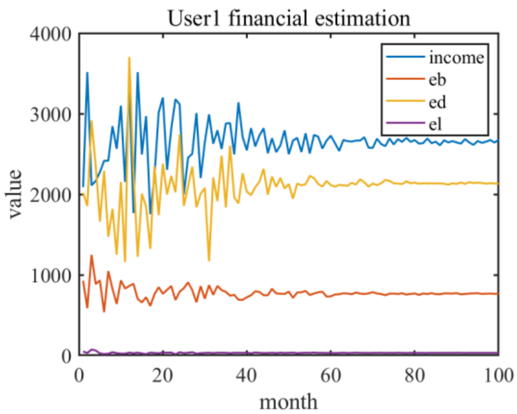
Figure 6. Income and expenditure of the four users for 2017 and 2018. (a) Income and expenditure of user 1. (b) Income and expenditure of user 2. (c) Income and expenditure of user 3. (d) Income and expenditure of user 4.

The choice of the four sets of parameters in Table IV is based on several considerations: Firstly, the efficacy of the model across different types of users is intended to be validated. Thus four categories of users from Table III are selected, where each of them exhibits distinct characteristics in terms of income and expenditure. Consequently, the income and expenditure data for each of these categories are estimated separately, and they are provided in Table IV. Secondly, the model is designed such that users' future emergency funds can be planned and the duration they can withstand unexpected events can be assessed. These evaluations are dependent on the user's income ($\overline{income}$) and various types of expenditures ($\overline{eb}_i$, $\overline{er}_i$ and $\overline{el}_i$). Thus, it is necessary for these four variables to be projected for each user, the values of which are also included in Table IV. Finally, as demonstrated by (11) through (14), the historical data (2017 and 2018) are utilized to estimate each user's income ($\overline{income}$) and various types of expenditures ($\overline{eb}_i$, $\overline{er}_i$ and $\overline{el}_i$). These estimates exhibit distinct characteristics in the existing historical data. To ensure that the estimated income and expenditures align more closely with the user's characteristics, each value is calibrated based on the user's historical data. This calibration involved the OU parameters θ, mu, k being adjusted for each variable of each user. These reasons primarily informed the choice of the four sets of parameters, which are believed to sufficiently reflect the functionality and effectiveness of the proposed model. To set the appropriate parameters to make the simulated income, basic expenditure, recurrent expenditure and luxury expenditure similar to the real situation of the user, the data from the user for 2017 and 2018 are used to calibrate the OU process. Specifically, the Kalman filter is used to calculate the best estimate and then to calculate the maximum likelihood function and estimate the parameters of the unknown model. We use the Python library Pykalman to implement the Kalman filter, while the maximum likelihood optimization is implemented by the Scipy library. Table IV shows the parameters obtained from the calibration of the data of the users for 2017 and 2018, which are used to simulate the future income and expenditure of the four users. The OU process simulates the data according to the parameters in Table IV. As shown in Figure 7, income and expenditure data for the four types of users are simulated, where the x-axis indicates the number of months simulated. The OU process simulates the financial situation of users as shown in Figure 7, which shows that these income and expenditure values fluctuate around the mean value.

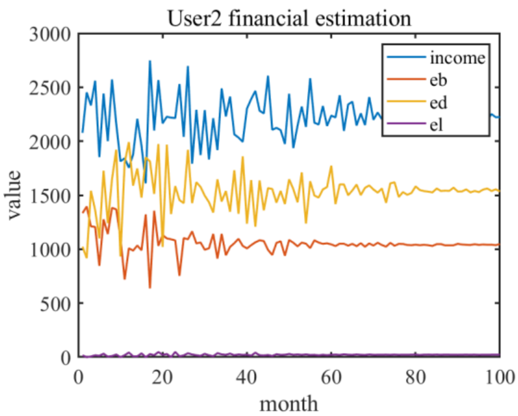
Table IV. Parameters of the OU process.

User	Estimated variable	Parameters		
		θ	mu	k
User1	*income*	0.03142	0.048619	0.054328
	eb	9.8096e-05	0.44428	0.36873
	er	0.00028496	0.12106	0.5751
	el	7.8839e-06	0.48973	0.14333
User2	*income*	0.009052	0.088093	0.059729

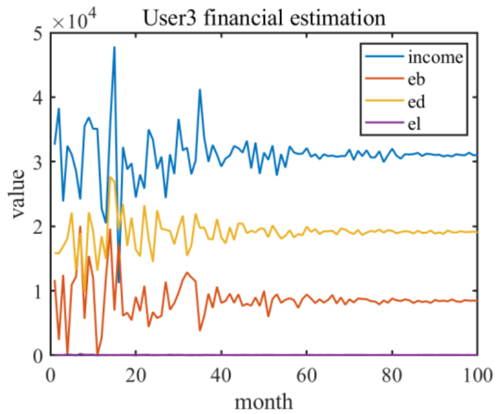
	eb	4.2392e-05	0.30815	0.71694
	er	0.0045556	0.088854	0.10088
	el	2.978e-06	0.48972	0.14338
User3	*income*	4.3641	0.32692	-0.17673
	eb	1.5831	0.13159	0.11661
	er	15.0852	0.18574	0.07012
	el	8.2157e-05	0.48972	0.14338
User4	*income*	9.2852	0.72752	0.6098
	eb	10.0026	0.25212	-0.27209
	er	10.2014	0.53425	0.2858
	el	12.4256	0.050116	0.10922

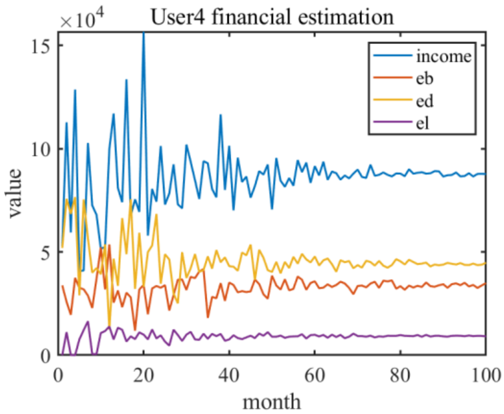


(a)



(b)

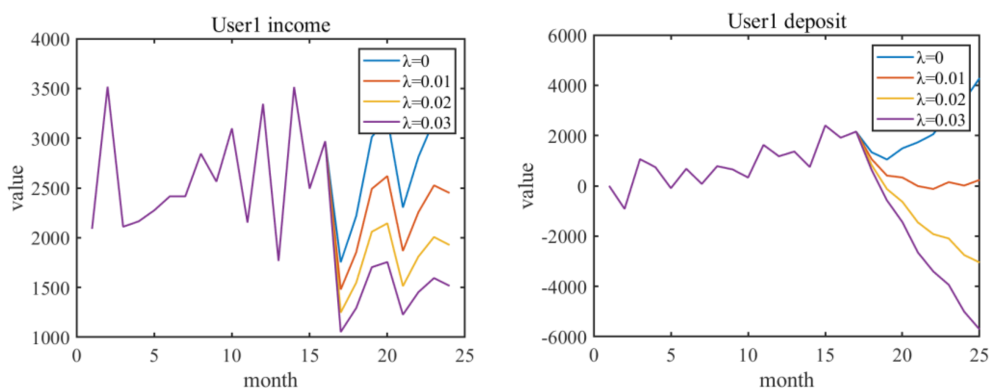




(c) (d)

Figure 7. Simulation of the financial situation of four types of users. (a) The financial estimation of user 1. (b) The financial estimation of user 2. (c) The financial estimation of user 3. (d) The financial estimation of user 4.

In real-life scenarios, income is affected to varying degrees by unexpected events, but the degree of the impact of these events on income is unknown in advance. The decay factor λ is taken as different values to indicate the degree of effect of the unexpected event on income, respectively. To explore the emergency fund situation of users when events impact differently on income, the following assumptions are made. Assume an unexpected event occurs after t_1 months that maintains t_2 months. Meanwhile, it is assumed that the unexpected event may impact the income of the user to different degrees. The emergency fund situation when the user does not take any action is explored (Figure 8), where $t_1 = 16, t_2 = 8, \lambda = 0,\ 0.01,\ 0.02,\ 0.03$. Figure 8 (a), (c), (e) and (g) represent the income of four user types when the decay factor λ takes different values. A larger λ indicates a greater impact of the event on income, $\lambda = 0$ indicates that the event does not happen or the event does not impact on income. Figure 8 (b), (d), (f) and (h) represents the emergency funds for the four types of users when the decay factor λ takes different values. Figure 8 (a), (c), (e) and (g) indicates that after the event ($t1 = 16$), the income of the each customer begins to decline and the decline is continuous and exponential. As λ increases, user income declines at an even faster rate. Figure 8 (b), (d), (f) and (h) show that before the event ($month \leq 16$), the income of each user is normal. During this period, users have emergency funds in some months, or the debt is relatively low. However, after the event ($month > 16$), the emergency funds of the users become less due to the impact of the event ($\lambda = 0.01, \lambda = 0.02, \lambda = 0.03$). In addition, the debt of the user is very high and the user is without emergency funds for most of the months. Furthermore, the debt becomes more severe as λ increases. For example, before the event, User 1 and User 2 have emergency funds for a few months, but most of the time after the event they have no emergency funds and the debt becomes severe as the event continues. Figure 8 shows that even for users in a better financial situation (e.g. user 1 and user 2), the emergency fund can only cover the event for a short time if they do not plan for it before the event occurs.



(a) (b)

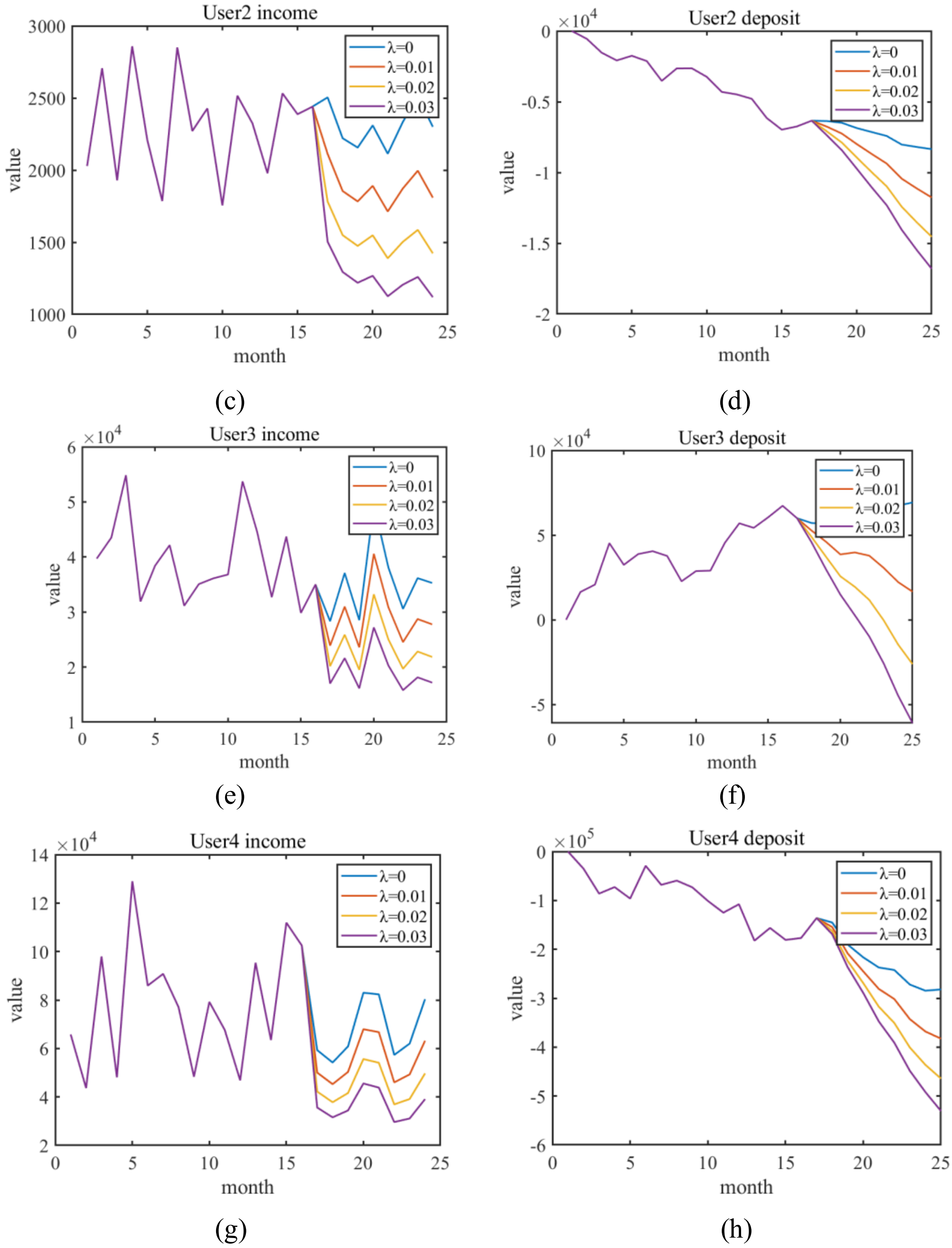
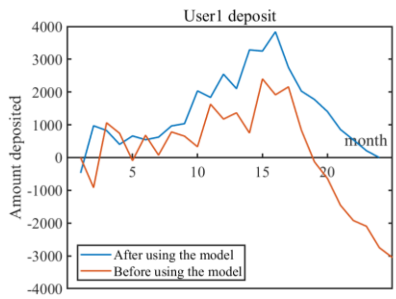


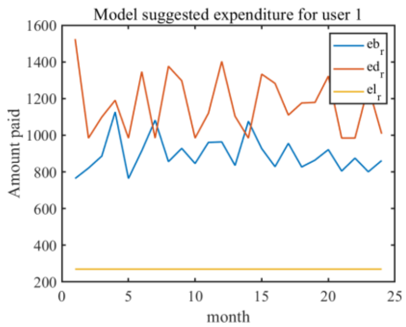
Figure 8. Effect of different λ on user income and emergency funds. (a) The income of user 1 when the decay factor λ takes different values. (b) The emergency funds of user 1 when the decay factor λ takes different values. (c) The income of user 2 when the decay factor λ takes different values. (d) The emergency funds of user 2 when the decay factor λ takes different values. (e) The income of user 3 when the decay factor λ takes different values. (f) The emergency funds of user 3 when the decay factor λ takes different values. (g) The income of user 4 when the decay factor λ takes different values. (h) The emergency funds of user 4 when the decay factor λ takes different values.

3.2.2. Scenario 1

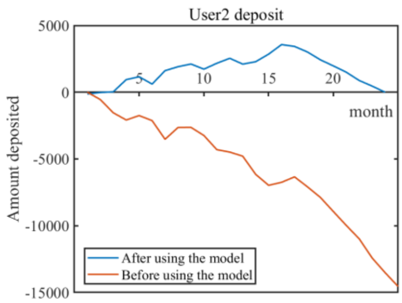
To verify the validity of the model, the proposed model is used to plan. The model is optimised to obtain optimal values for income, base expenditure, recurrent expenditure and luxury expenditure. Meanwhile, the emergency fund for the current month $\widetilde{m}_i$ is added to the emergency fund already deposited $\widetilde{m}_{s_i}$ to obtain the current emergency fund. Figure 9 (a), (c), (e) and (g) shows the emergency fund deposits for the four types of users. In the absence of the model, the emergency funds are difficult to maintain until the end of the event, even for users in a better financial situation (e.g. user 4). With the use of the proposed model, the emergency fund of the user can be maintained until the end of time. Before the event ($month < 16$), the emergency fund tends to rise, which means that users deposit emergency funds before events to prepare for the event. After the event ($month \geq 16$), the emergency fund is decreased, which is consonant with reality. After the event, the income of the user is affected and the income cannot cover normal expenses. Therefore, user can only spend their existing emergency fund, which reduces the emergency fund from month to month. If the user follows the suggestions of the model to spend, then the emergency fund can cover the event until the end. Figure 9 (b), (d), (f) and (h) show the various types of expenditure suggested by the model. As shown in Figure 9 (a), (c), (e) and (g), the emergency fund of the user can cover the event to the end, which means that the proposed model can help users cope better with unexpected events compared to no measure at all.



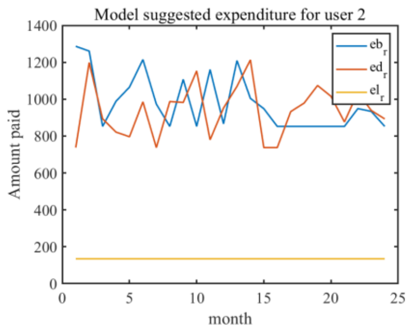
(a)



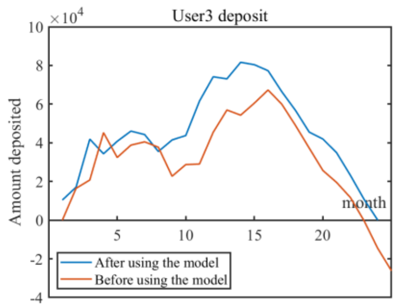
(b)

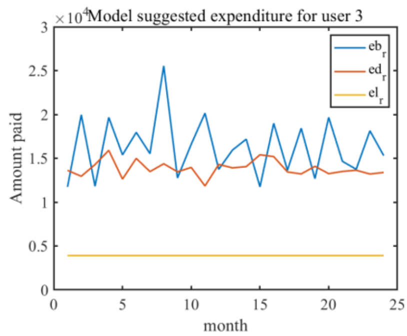


(c)

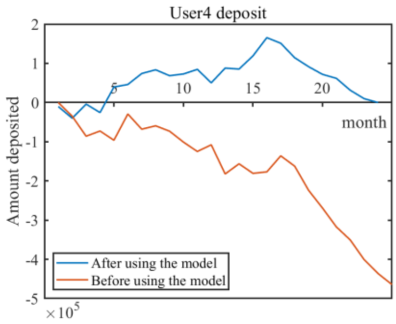


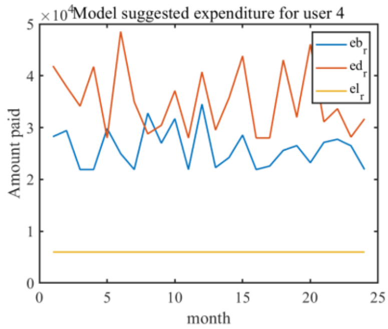
(d)





(e) (f)





(g) (h)

Figure 9. The suggestions of the proposed model to the users. (a) The emergency fund deposits for the four types of user 1. (b) The model suggests various types of expenditure for user 1. (c) The emergency fund deposits for the four types of user 2. (d) The model suggests various types of expenditure for user 2. (e) The emergency fund deposits for the four types of user 3. (f) The model suggests various types of expenditure for user 3. (g) The emergency fund deposits for the four types of user 4. (h) The model suggests various types of expenditure for user 4.

3.2.3. Scenario 2

Assuming the event duration t_2 is infinite, the user is bound to be unable to cope with the event at some point, regardless of how the model is planned. It corresponds to scenario 2 described in section 2. To analyse the maximum time to cope with an event for the model, this work assumes $t_1 = 16$ and then increasing t_2. The model first estimates the normal future income of the user based on data from 2017 and 2018 respectively. The income is decayed by $\lambda = 0.01, \lambda = 0.02, \lambda = 0.03$ at $t_1 = 16$, which indicates that the user is affected by an unexpected event. The model optimises the base, recurrent and luxury expenditures respectively based on the data obtained from the simulation. The current emergency fund is calculated using (3). As shown in Figure 10, Figure 11 and Figure 12, the model is unable to cope with the event when t_2 increases to a certain value, at which point the value of t_2 taken as $\mathrm{t}_{2\max}$ is the limiting time value for the model. If $t_2 > t_{2max}$, it indicates that the model cannot help the user to cope with the event until the end, but only to maximise the time to cope with the event. For different users, t_{2max} is different, which is related to the financial situation of the users. Figure 10, Figure 11 and Figure 12 show that in comparison to the case without the model, the use of the proposed model enabled the emergency fund to cover the event for a longer time. In addition, the larger the λ, i.e. the greater the impact of the event on the income of the user, the shorter the maximum time for the user to cope with the unexpected event, whether the model is used or not. In general, the use of models can make the maximum time to cope with the event longer.

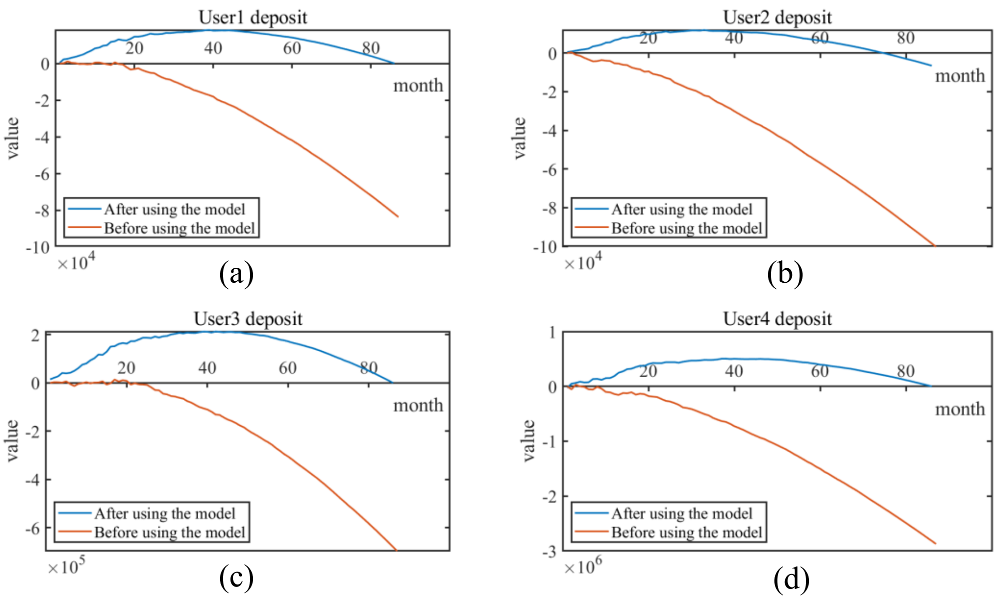


Figure 10. The maximum duration of emergency fund maintenance events before and after the use of the model ($\lambda = 0.01$). (a) The deposit of user 1. (b) The deposit of user 2. (c) The deposit of user 3. (d) The deposit of user 4.

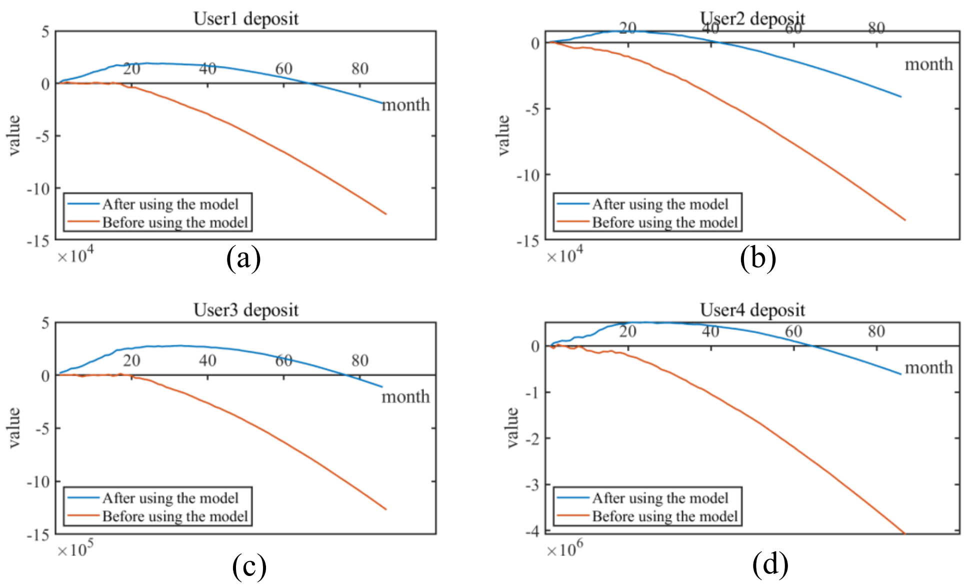


Figure 11. The maximum duration of emergency fund maintenance events before and after the use of the model ($\lambda = 0.02$). (a) The deposit of user 1. (b) The deposit of user 2. (c) The deposit of user 3. (d) The deposit of user 4.

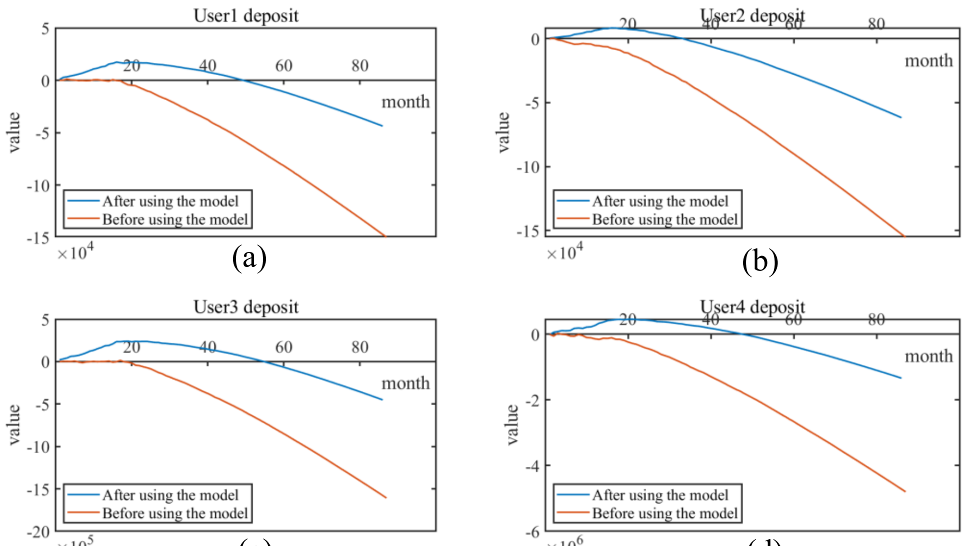
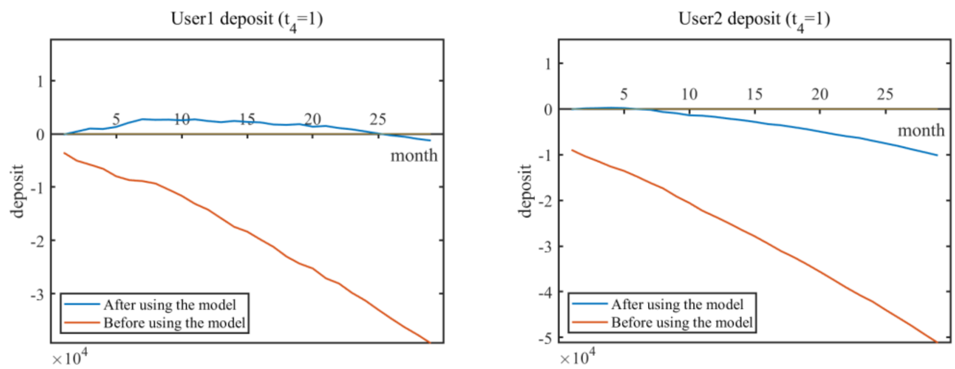


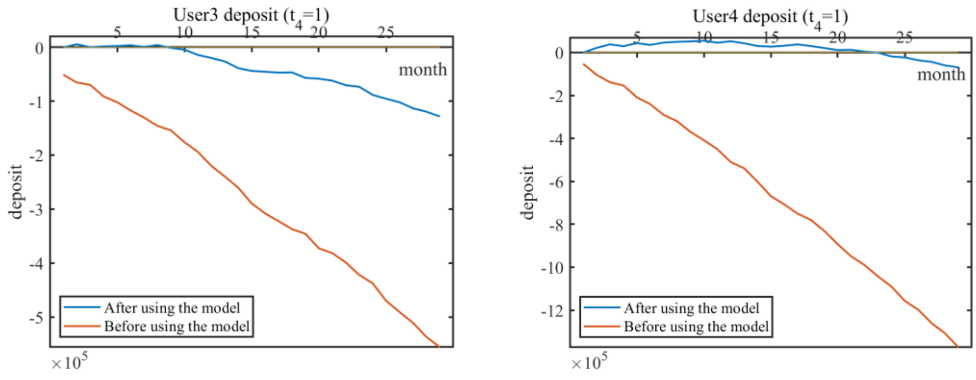
Figure 12. The maximum duration of emergency fund maintenance events before and after the use of the model ($\lambda = 0.03$). (a) The deposit of user 1. (b) The deposit of user 2. (c) The deposit of user 3. (d) The deposit of user 4.

3.2.4. Scenario 3

A scenario is also considered in this paper in which it is assumed that the unexpected event has occurred at the time t_4 when the model is used. This corresponds to scenario 3 described in section 2. To verify the limits of the model response to the event, assume $\lambda = 0.02, t_1 = 0$ and $t_2 = 30$. The model still uses the data estimated by OU in the previous experiments. As shown in Figure 13, t_4 indicates the length from the time of the event to the current time, where t_4 take different values ($t_4 = 1$, $t_4 = 5$). As shown in Figure 13 and Figure 14, the time to cover the event t_3 is inversely proportional to t_4. In addition, some users (e.g. user 2 and user 3) are unable to cope with events after the event occurs t_4 time. The result is compatible with realistic scenarios, as most users in a poor financial situation are hard-pressed to deposit emergency funds under normal circumstances, and even harder after an unexpected event. In addition, in comparison to not using the model, the use of the proposed model can help users to cope with unexpected events for a longer time.

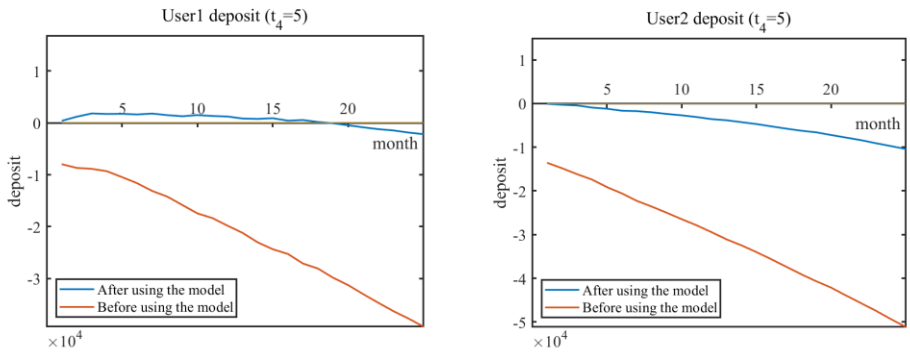


(a) (b)

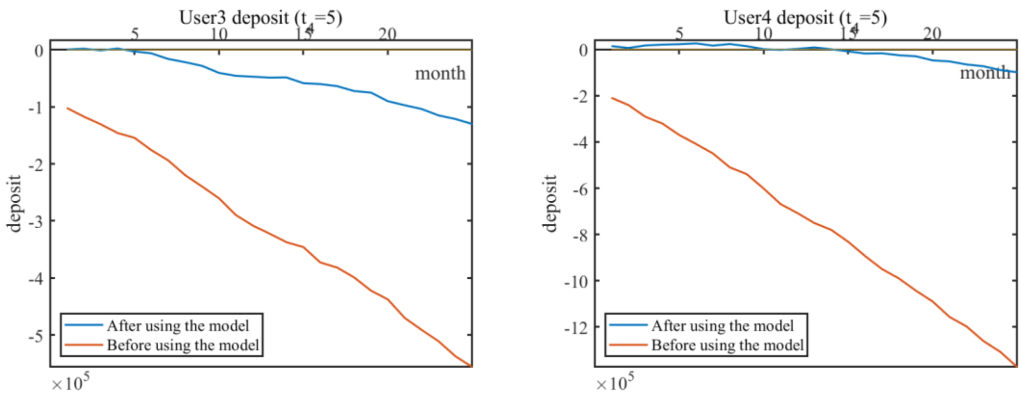


(c) (d)

Figure 13. The effect of using the model on the emergency fund after an event (t_4 = 1). (a) The deposit of user 1. (b) The deposit of user 2. (c) The deposit of user 3. (d) The deposit of user 4.



(a) (b)



(c) (d)

Figure 14. The effect of using the model on the emergency fund after an event t_4 = 5. (a) The deposit of user 1. (b) The deposit of user 2. (c) The deposit of user 3. (d) The deposit of user 4.

3.2.5. Validity of model

To verify the validity of the model, the proposed model is used at different times. It is assumed that the model is used at t_i = 1,8,16,24 respectively, and the maximum time for the user to cope with the event is compared. Assume λ = 0.02 and t_1 = 16.

The model still uses the data estimated by OU in the previous experiment. For different values of t_i, the base expenditure, recurrent expenditure and luxury expenditure are optimised and an emergency fund is calculated according to (3). The emergency fund $\widetilde{m}_{s_i} = 0$ indicates that the user is unable to cope with the event. Figure 15 shows the situation of the four types of users who cope with unexpected events. In each subplot, each curve represents the change in the emergency fund when the user uses the model at different times. The point where the curve intersects the $x = 0$ axis indicates the emergency fund $\widetilde{m}_{s_i} = 0$. As shown in Figure 15, t_i and cope time are inversely proportional, which visually demonstrates the validity of the model.

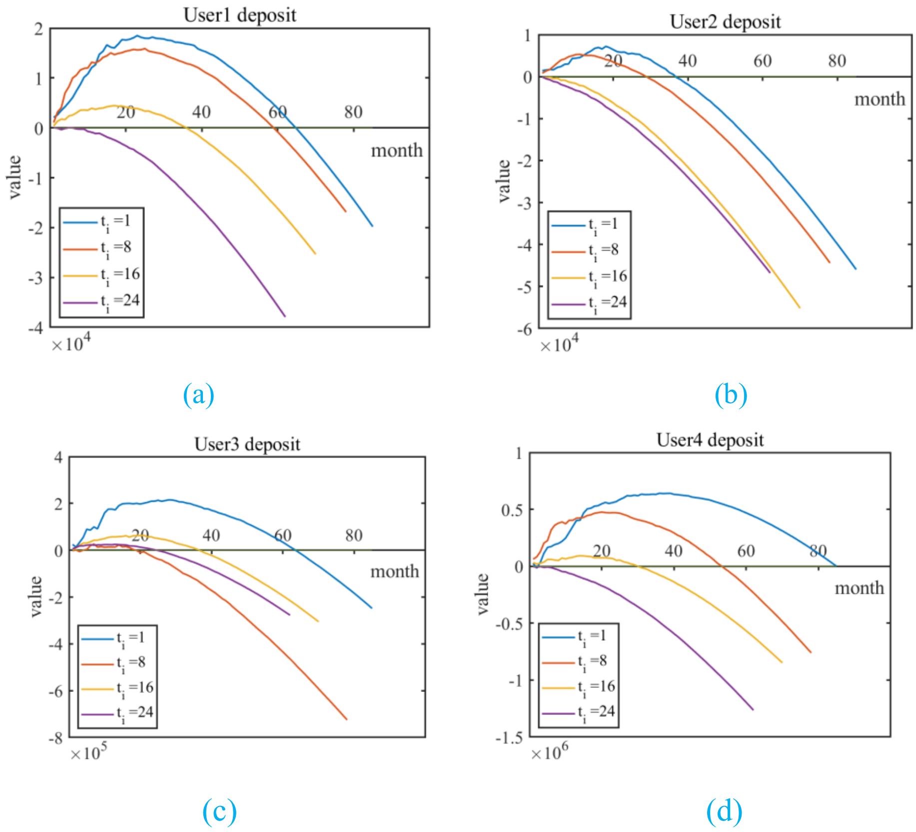
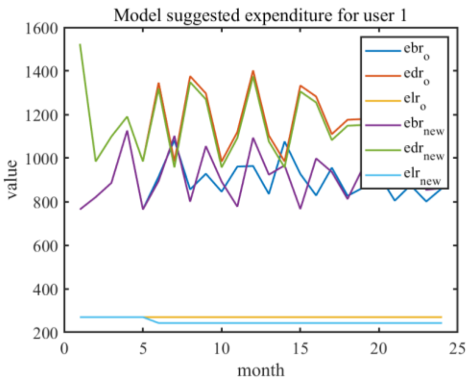


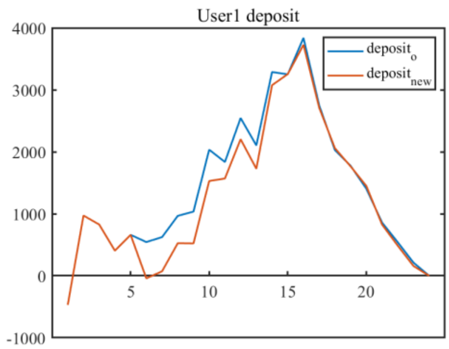
Figure 15. Compare the emergency fund and cope time in using the model at different times. (a) The deposit of user 1. (b) The deposit of user 2. (c) The deposit of user 3. (d) The deposit of user 4.

3.2.6. Dynamic update of the model

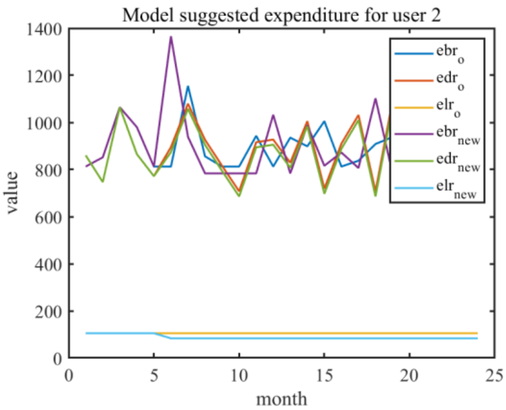
In real scenarios, users may not comply with the suggestions of the model, and this is an issue well worth considering for the model. To verify the real-time performance of the model, the following realistic assumptions are made. Assume that at a certain month, a category of expenditure is not spent as originally suggested and is overspent. Specifically, it is assumed that $t_6 = 6$, the actual base expenditure of the user is 1.2 times that suggested by the model. Furthermore, the same as in the previous experiment, assume $\lambda = 0.02$, $t_1 = 16$ and $t_2 = 8$. The model still uses the data estimated by OU in the previous experiment. The current emergency fund $\tilde{m}_{s_i}$ is calculated by (3). As shown in Figure 16, during $t < t_6$, the actual and suggested expenditures are equal, because the user complies with the suggestions of the model. However, the model gives a new suggestion for expenditure at $month > 6$ as the actual base expenditure of the user increases at $t = 6$. Figure 16 (b), (d), (f) and (h) shows that for $t > 6$, the emergency fund plan for the user also changes compared to the original plan, but it enables the customer to cope with unexpected events in the end. Figure 16 shows that the model can update the plan according to the actual expenditure of the user.



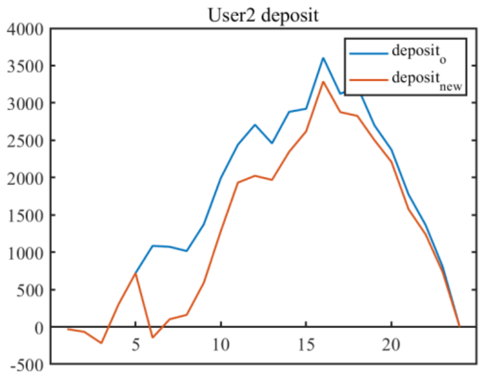
(a)



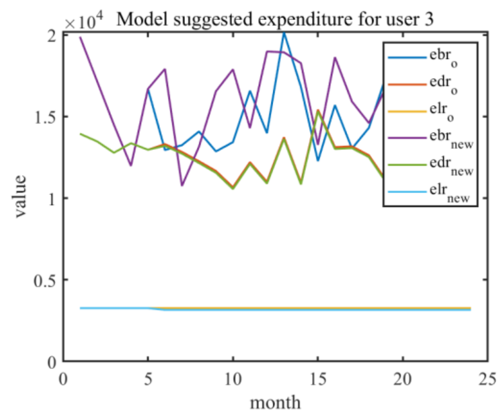
(b)



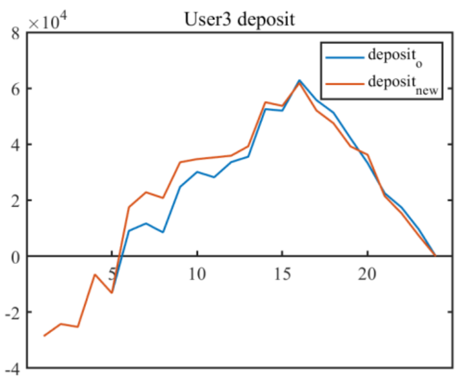
(c)



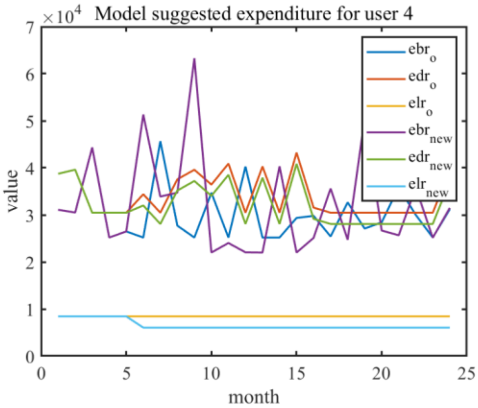
(d)



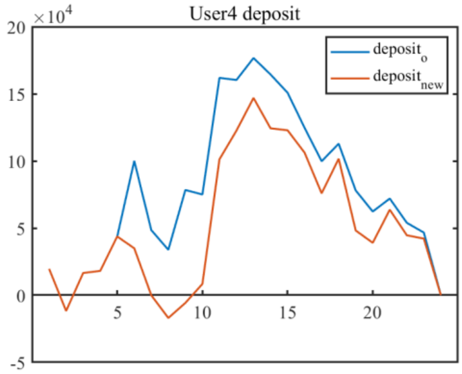
(e)



(f)



(g)



(h)

Figure 16. The actual expenditure of the user changes and the model updates the suggestion in real-time. (a) The model suggested expenditure for user 1. (b) The deposit of user 1. (c) The model suggested expenditure for user 2. (d) The deposit of user 2. (e) The model suggested expenditure for user 3. (f) The deposit of user 3. (g) The model suggested expenditure for user 4. (h) The deposit of user 4.

3.3. Monitoring and alerts

To provide the user with visual monitoring and alerts, the proposed CI index is used to measure the ability of users to currently cope with risk and the degree of deviation from consumption. CI is used to measure several categories of users separately, as shown in Figure 17. The $CI = 0$ when the user fully complies with the plan and $CI > 0$ when the actual and suggested expenditures is equal. A higher CI indicates a higher degree of deviation from the plan in actual expenditure. As in section 2.2.4, it is assumed that in a certain month, a certain type of expenditure is not spent in the way initially suggested and is overspent. Specifically, it is assumed that in $t_6 = 6$, the actual basic expenditure of the user is 1.2 times the model suggestion. Further, as in the previous experiment, set λ = 0.02, $t_1 = 16$ and $t_2 = 8$. Figure shows two scenarios, one assuming the user complies with the plan throughout, and the other assuming that the user does not comply with the plan in t_6, when $CI > 0$ in theory. Figure 17 shows four users, where $CI > 0$ when $t_6 = 6$, indicating that in this month the user deviated from the plan and the magnitude of the CI can indicate the degree of deviation. Since the actual base expenditure for all four users is set to 1.2 times the model suggestion, therefore $CI = 0.03$, which indicates that although the actual expenditure of users deviates from the plan, the degree of deviation is acceptable. The CI value should be zero if the user follows the plan exactly. The CI value gives the user a clear idea of whether their current consumption is reasonable or not.

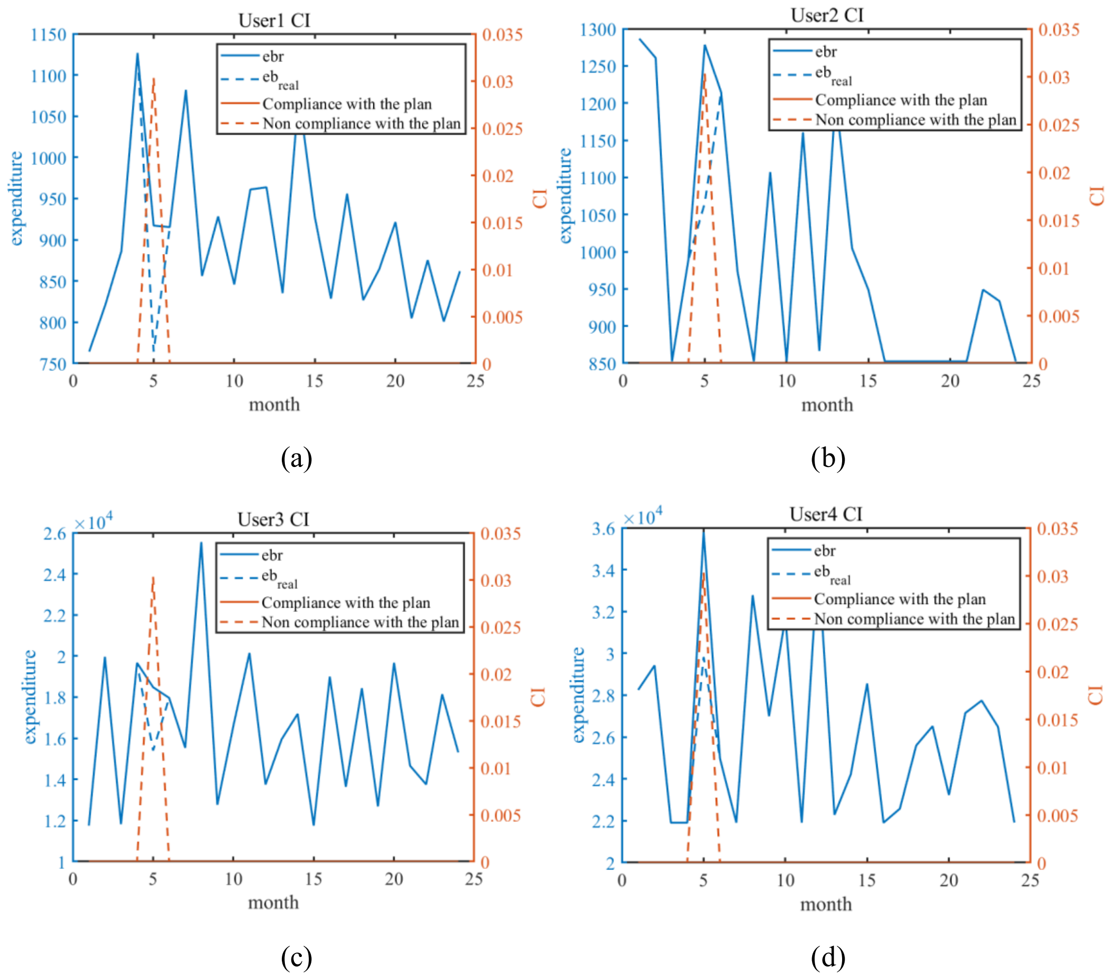


Figure 17. The CI of the four users. (a) The CI of user 1. (b) The CI of user 2. (c) The CI of user 3. (d) The CI of user 4.

4. Conclusion

In this paper, a novel emergency fund saving model is proposed to help users to save emergency funds to better cope with future unexpected events. A contribution to the research literature is proposed in this paper in the following ways. (1) It is the first time that the real-life social issues raised by COVID-19 are considered and described from an optimization perspective, i.e. how to plan to better cope with future contingencies, assuming that they will occur. (2) A model is formulated for solving the emergency fund problem and an optimization algorithm is applied to solve this problem, providing a targeted solution for users. (3) An index of prevention and an index of consumption is proposed to evaluate the reasonableness of consumption and the ability of users to cope with events and to provide a real-time monitoring and early warning solution. (4) The feasibility and validity of the proposed models are verified based on real Open Banking customer data. The experiments compared the emergency fund saved by customers and their ability to respond to incidents before and after the use of the model respectively. The experiments shown that it can be effective to use the model to plan consumption early, before unexpected events occur, to cope with future unexpected events. The proposed model can help customers cope with unexpected events or extend the time to cope with unexpected events. In future research, the model can be extended to a family emergency fund model, which is also very realistic with family members working together to cope with unexpected events. In addition, refining time to a day-by-day basis provides customers with real-time planning and monitoring of spending. This allows for better cope with unexpected events. The limitations of this study are mainly reflected in the following two aspects: (1) Data source: The study uses anonymous transaction data provided by FinTech partners and data from the OU process simulations for model validation. The model's effectiveness in real-world scenarios may vary depending on the accuracy and representativeness of this data. (2) Financial behaviour: The model is designed to optimize user expenditure and build up an emergency fund. However, it may not fully account for individual financial behaviours, preferences, and unexpected changes in income or expenditure.

Acknowledgments

This research is supported by the National Natural Science Foundation of China under Grant 61976063, the Guangxi Natural Science Foundation under Grant 2022GXNSFFA035028, research fund of Guangxi Normal University under Grant 2021JC006, the AI+Education research project of Guangxi Humanities Society Science Development Research Center under Grant ZXZJ202205.

References

Amer, M., & Namaane, A. (2013). Optimization of hybrid renewable energy systems (HRES) using PSO for cost reduction. *Energy Procedia*, *42*(3), 318–327.

Cao, B., Dong, W., Lv, Z., & et al. (2020). Hybrid Microgrid Many-Objective Sizing Optimization With Fuzzy Decision. *IEEE Transactions on Fuzzy Systems*, *28*(11), 2702–2710.

Carlo, G., Parino, F., Zino, L., & et al. (2023). Dynamic planning of a two-dose vaccination campaign with uncertain supplies. *European Journal of Operational Research*, *304*(3), 1269–1278.

Cervellera, C. (2023). Optimized ensemble value function approximation for dynamic programming. *European Journal of Operational Research*, *309*(2), 719–730.

Chang, J. T., & Kaplan, E. H. (2023). Modeling local coronavirus outbreaks. *European Journal of Operational Research*, *304*(1), 57–68.

Chen, K., Seng, C., & Ying, H. (2023). Efficient social distancing during the COVID-19 pandemic : Integrating economic and public health considerations. *European Journal of Operational Research*, *304*(1), 84–98.

Cristina, A., Henrique, L., Brito, B., Fernando, J., & Antónia, M. (2023). Discrete Optimization A stochastic programming approach to the cutting stock problem with usable leftovers. *European Journal of Operational Research*, *308*(1), 38–53.

Das, S., Bose, I., & Kumar, U. (2023). Predicting the outbreak of epidemics using a network-based approach. *European Journal of Operational Research*, *309*(2),

819–831.
Du, J., Yu, F. R., Chu, X., & et al. (2019). Computation Offloading and Resource Allocation in Vehicular Networks Based on Dual-Side Cost Minimization. *IEEE Transactions on Vehicular Technology*, *68*(2), 1079–1092.
Eko, T., Figueira, G., & Almada-lobo, B. (2023). Hybrid MCDM and simulation-optimization for strategic supplier selection. *Expert Systems With Applications*, *219*(1), 119624.
Er-rahmadi, B., & Ma, T. (2022). Data-driven mixed-Integer linear programming-based optimization for efficient failure detection in large-scale distributed systems R. *European Journal of Operational Research*, *303*(1), 337–353.
Fusco, A., Gioffr, D., Francesco, A., Bovo, C., & Martelli, E. (2023). *A multi-stage stochastic programming model for the unit commitment of conventional and virtual power plants bidding in the day-ahead and ancillary services markets Expected Value of Perfect Information*. *336*(4), 120739.
General Data Protection Regulation. (n.d.). Https://Gdpr-Info.Eu/; IEEE.
Goerigk, M., & Kurtz, J. (2023). Computers and Operations Research Data-driven robust optimization using deep neural networks. *Computers and Operations Research*, *151*(5), 106087.
He, Z., Yen, G. G., & Ding, D. (2021). Knee-Based Decision Making and Visualization in Many-Objective Optimization. *IEEE Transactions on Evolutionary Computation*, *25*(2), 292–306.
Herrera-viedma, E. (2020). Composite Monte Carlo decision making under high uncertainty of novel coronavirus epidemic using hybridized deep learning and fuzzy. *Applied Soft Computing Journal*, *93*(12), 106282.
Hu, Y., Zhang, Y., & Gong, D. (2021). Multiobjective Particle Swarm Optimization for Feature Selection With Fuzzy Cost. *IEEE Transactions on Cybernetics*, *51*(2), 874–888.
Huang, Y., Jin, Y., & Hao, K. (2020). Decision-making and multi-objectivization for cost sensitive robust optimization over time. *Knowledge-Based Systems*, *199*(1), 105857.
Jain, A., Chandra, A., Rao, S., & et al. (2023). Optimized levy flight model for heart disease prediction using CNN framework in big data application. *Expert Systems With Applications*, *223*(1), 119859.
Jalota, H., Kumar, P., Thakur, M., & Mittal, G. (2023). A novel approach to incorporate investor ' s preference in fuzzy multi-objective portfolio selection problem using credibility measure. *Expert Systems With Applications*, *212*(8), 118583.
Jang, H. (2019). A decision support framework for robust R & D budget allocation using machine learning and optimization. *Decision Support Systems*, *121*(6), 1–12.
Kumajas, L. I., & Wuryaningrat, N. F. (2021). Dana darurat di masa pandemi Covid-19. *Modus*, *33*(1), 1–17.
L, A. J. L. (2021). A robust decision-support method based on optimization and simulation for wildfire resilience in highly renewable power systems. *European Journal of Operational Research*, *294*(2), 723–733.

Labani, M., Moradi, P., & Jalili, M. (2020). A multi-objective genetic algorithm for text feature selection using the relative discriminative criterion. *Expert Systems With Applications*, *149*(7), 113276.

Liu, K., Liu, C., Xiang, X., & et al. (2023). Testing facility location and dynamic capacity planning for pandemics with demand uncertainty. *European Journal of Operational Research*, *304*(1), 150–168.

Liu, Y., Ishibuchi, H., & Yen, G. G. (2020). Handling Imbalance Between Convergence and Diversity in the Decision Space in Evolutionary Multimodal Multiobjective Optimization. *IEEE Transactions on Evolutionary Computation*, *24*(3), 551–565.

Lyu, Z., & Huang, G. Q. (2023). Cross-docking based factory logistics unitisation process : An approximate dynamic programming approach. *European Journal of Operational Research*, *1*(4), 1.

Nadizadeh, S., Neysani, N., Minaei, F., & et al. (2022). A decision model based on decision tree and particle swarm optimization algorithms to identify optimal locations for solar power plants construction in Iran. *Renewable Energy*, *187*(1), 56–67.

Nour, M., Cömert, Z., & Polat, K. (2020). A Novel Medical Diagnosis model for COVID-19 infection detection based on Deep Features and Bayesian Optimization. *Applied Soft Computing Journal*, *97*(8), 106580.

Science, N., Phenomena, C., Lalmuanawma, S., & et al. (2020). Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic : A review. *European Journal of Operational Research*, *139*(1), 110059.

Steuer, R. E., & Utz, S. (2023). Non-contour efficient fronts for identifying most preferred portfolios in sustainability investing. *European Journal of Operational Research*, *306*(2), 742–753.

Taylor, J. W., & Taylor, K. S. (2023). Combining probabilistic forecasts of COVID-19 mortality in the United States. *European Journal of Operational Research*, *304*(1), 25–41.

Thul, L., & Powell, W. (2023). Stochastic optimization for vaccine and testing kit allocation for the COVID-19 pandemic. *European Journal of Operational Research*, *304*(1), 325–338.

Tisdell, C. A. (2020). Economic , social and political issues raised by the COVID-19 pandemic. *Economic Analysis and Policy*, *68*(1), 17–28.

Uhlenbeck G.E., & Ornstein L.S. (1930). On the theory of the Brownian motion. *Physical Review*, *36*(5), 823.

Warwicker, J. A., & Rebennack, S. (2023). A unified framework for bivariate clustering and regression problems via mixed-integer linear programming. *Discrete Applied Mathematics*, *336*, 15–36.

Wu, Q., Liu, X., Qin, J., Zhou, L., & Mardani, A. (2022). Knowledge-Based Systems An integrated generalized TODIM model for portfolio selection based on financial performance of firms. *Knowledge-Based Systems*, *249*(8), 108794.

Yao, J., Shi, H., & Liu, C. (2020). Optimizing the configuration of green supply chains under mass personalisation. *International Journal of Production Research*, *58*(24), 7420–7438.

Yu, G., & Li, D. (2022). Computers & Industrial Engineering A novel intuitionistic fuzzy goal programming method for heterogeneous MADM with application to regional green manufacturing level evaluation under multi-source information. *Computers & Industrial Engineering*, *174*(2006), 108796.

Yu, J. J. Q., Yu, W., & Gu, J. (2019). Online Vehicle Routing With Neural Combinatorial Optimization and Deep Reinforcement Learning. *IEEE Transactions on Vehicular Technology*, *20*(10), 3806–3817.

Zhang, W., Shi, X., Huang, A., Hua, G., & Teunter, R. H. (2023). Optimal stock and capital reserve policies for emergency medical supplies against epidemic outbreaks. *European Journal of Operational Research*, *304*(1), 183–191.

Zhu, P., Saadati, H., & Khayatnezhad, M. (2021). Application of probability decision system and particle swarm optimization for improving soil moisture content. *Water Supply*, *21*(8), 4145–4152.