CARF CENTER FOR ADVANCED RESEARCH IN FINANCE

CARF Working Paper

CARF-F-550

Big data applications with theoretical models and social media in financial management

Taiga Saito
Graduate School of Economics, The University of Tokyo

Shivam Gupta
Department of Information Systems, Supply Chain Management
& Decision Support, NEOMA Business School

December 8, 2022

CARF is presently supported by Nomura Holdings, Inc., Sumitomo Mitsui Banking Corporation, The Dai-ichi Life Insurance Company, Limited, The Norinchukin Bank, MUFG Bank, Ltd. and Ernst & Young ShinNihon LLC. This financial support enables us to issue CARF Working Papers.

Big data applications with theoretical models and social media in financial management

Taiga Saito[1] and Shivam Gupta[2*]

[1]Graduate School of Economics, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, 1130033, Tokyo, Japan.
[2*]Department of Information Systems, Supply Chain Management & Decision Support, NEOMA Business School, France.

*Corresponding author(s). E-mail(s): shivam.gupta@neoma-bs.fr;
Contributing authors: staiga@e.u-tokyo.ac.jp;

Abstract

This study presents big data applications with quantitative theoretical models in financial management and investigates possible incorporation of social media factors into the models. Specifically, we examine three models, a revenue management model, an interest rate model with market sentiments, and a high-frequency trading equity market model, and consider possible extensions of those models to include social media. Since social media plays a substantial role in promoting products and services, engaging with customers, and sharing sentiments among market participants, it is important to include social media factors in the stochastic optimization models for financial management. Moreover, we compare the three models from a qualitative and quantitative point of view and provide managerial implications on how these models are synthetically used along with social media in financial management with a concrete case of a hotel REIT. The contribution of this research is that we investigate the possible incorporation of social media factors into the three models whose objectives are revenue management and debt and equity financing, essential areas in financial management, which helps to estimate the effect and the impact of social media quantitatively if internal data necessary for parameter estimation are available, and provide managerial implications for the synthetic use of the three models from a higher viewpoint. The numerical experiment along with the proposition indicates that the model can be used in the revenue management of hotels, and by improving the social media factor, the hotel can work on maximizing its sales.

Keywords: Big data applications, Revenue management, Social media, Financial management

1 Introduction

Social media is important for business since social media includes valuable information that the other data that companies own normally do not contain. From social media such as Twitter, Instagram, and Facebook, companies can collect customers' impressions and reviews and even make advertisements to the customers. Also, people's views on financial markets and economics are observed on social media. The revenue management model by Saito et al. (2016), (2019), the interest rate model with sentiments by Nishimura et al. (2019) and Nakatani et al. (2020), and the high-frequency trading equity market model by Saito and Takahashi (2019), are quantitative theoretical models that express behaviors of customers, investors, and market participants. Moreover, these models incorporate big data, such as booking data, financial data with text data, and order and execution data into modeling. Since social media includes information that these data do not have, thus, the models become more valuable if they take into account the social media data that reflects people's sentiments. In this study, we review these three theoretical models from an integrated perspective and consider possible extensions of the models to include social media factors. Also, we compare the three models to investigate what common features are and what the differences are by considering managerial implications about in which scenario each model works.

As big data are used in our daily lives, it is becoming increasingly important to utilize big data and theoretical models that describe mechanisms of phenomenon. Kar and Dwivedi (2020) also point out the need for studies of theoretical modeling incorporating big data, which explains phenomena caused by the interaction of people. For example, the customers on the online hotel booking website in Saito et al. (2016), (2019) choose the hotel to book from a group of hotels in the same area by comparing the room charges and other characteristics of the hotels. In the interest rate model in Nishimura et al. (2019) and Nakatani et al. (2020), the investors maximize the expected utility of their wealth by choosing portfolio allocations under sentiments. In the high-frequency trading equity market model in Saito and Takahashi (2019), the market participants in a high-frequency trading market choose their trading strategies to maximize their expected profits. Those quantitative theoretical models describe human behaviors in social circumstances and explain patterns of the outcomes observed in reality. With applications of big data, the models become remarkably useful such that they can be utilized for determining optimal strategies in financial management.

In detail, Saito et al. (2016), (2019) presented big data applications to revenue management for hotels, utilizing big data, which is online booking data collected from a hotel booking website. With the big data, they estimated the quantitative revenue management models that described booking behaviors of the customers in response to the room charges of hotels and showed how those models could be used in hotel revenue management. Nishimura et al. (2019) and Nakatani et al. (2020) proposed applications of big data, which is financial news, by text mining and estimated market sentiment factors in interest rate models. Nishimura et al. (2019) and Nakatani et al. (2020) assessed the sentiment factors by text mining in economic news, which is big data, and proposed interest rate models incorporating those factors, which can be used by financial managers in companies as well as financial institutions.

In this study, we consider what further information helps describe the phenomena in detail and how it should be incorporated into the modeling using additional big data. Social media are essential in the utilization of big data since they contain information from customers, which can be utilized in marketing, and information on the market participants' sentiments. We investigate how social media data can be added to the three quantitative theoretical models with big data for practical use. Accordingly, we set the first research question as *(i) How can we add social media factors into the quantitative theoretical models incorporating big data?*

In the revenue management model in Saito et al. (2016), (2019), the importance of social media was indicated, but the social media factor was not incorporated into the modeling. Data from social media is helpful since it reflects the customers' sentiments against the hotels for instance, which affect the customers' choice behavior. We incorporate the social media factor into the revenue management model in Saito et al. (2016), (2019), which could be useful for revenue managers in hotels to maximize their expected sales. Also, sentiments of the investors on the financial market, which are also observed on social media, are important to the interest rate model with sentiments in Nishimura et al. (2019) and Nakatani et al. (2020) and the high-frequency trading equity market model in Saito and Takahashi (2019), which could be helpful for financial managers companies to predict when the best timing for financing with corporate bonds and equities, for central banks to control term structures of interest rates, and for financial authorities to regulate the market. We consider possible utilization of the social media information into the interest rate model and the high-frequency trading equity market model.

Moreover, we investigate the three applications of big data with the quantitative theoretical optimization models in financial management from an integrated point of view. Particularly, we compare these three models and consider how these three models are utilized in financial management in different roles and scenarios. Thus, we set the second research question as *(ii) How are those three theoretical models different from an integrated point of view and how can a company synthetically use the three models depending on a scenario?*

This study is new in discussing the applications of big data with theoretical models in financial management from an integrated point of view and their extensions to include social media from the stakeholders' viewpoints. The contribution is the extension of the three quantitative theoretical models incorporating big data to include further social media factors, a comparison of the three models from an integrated point of view, and the managerial implication for the synthetic use of the models by companies. Also, this study connects social media with the quantitative theoretical models and provides how those models are utilized in financial management for companies, investors, and financial authorities.

The organization of the paper is as follows. Section 2 proposes an extension of the revenue management model to include social media factors. Sections 3 and 4 investigate the interest rate model with market sentiments and the high-frequency trading equity market model by discussing possible extensions to include social media information. Section 5 compares the three theoretical models, and Section 6 investigates the managerial implications of the models and discusses the limitation of this study. Finally, Section 7 concludes.

2 Revenue management model

In this section, as an application of big data with a theoretical model in financial management, we extend the revenue management model for hotels in Saito et al. (2016), (2019), whose mathematical details are explained in Appendix A.1, to include social media information. In the hotel industry, quantitative revenue management models are used for day-to-day revenue control in online booking. The customers booking behavior is estimated, and the hotel revenue manager sets a price for booking to maximize the profit. By effectively utilizing the revenue management model, hotels can maximize their revenue by setting the room charge appropriately.

Specifically, we propose a quantitative revenue management model for hotels incorporating social media factors estimated from internal information such as reviews on online booking websites, Twitter, and Facebook, which is an extension of the revenue management model in Saito et al. (2019). Since social media has valuable information for marketing and enhancement for the services of hotels, hotels can better use social media so that it affects the probability of booking more effectively.

2.1 Data

Along with the model described in Appendix A.1, Saito et al. (2016)(2019) use the online hotel booking data crawled from a Japanese hotel online booking website for the estimation of the parameters. In Saito et al. (2016)(2019), the collected data includes room charges and the number of available rooms for booking for a certain booking period for collective check-in dates for four major hotels in front of Kyoto station and two major luxurious hotels in the Shinjuku area, respectively. In equations (A1)-(A3) in Appendix A.1, the big data that include the room charges and numbers of rooms booked, and the availability for booking of hotels in the same area are used for estimation. Particularly, the data are used in estimating the model for the sensitivity of the hotel's score

β and the characteristic part α_i in equation (A3) through the expression of booking probabilities p_i^γ in (A2).

Specifically, in Saito et al. (2016), optimal room charges for the four hotels in front of Kyoto station are calculated, particularly the price competition among the four hotels when each hotel chooses its optimal room charge given the room charges of the other hotels is investigated.

2.2 Extension to include social media information

Although publicly available big data, online booking data crawled from a Japanese online booking website, are used in Saito et al. (2016), (2019), utilization of internally available information or information from social media, such as reviews of the hotels, Twitter, and Facebook, which reflect sentiments of the customers, is not considered.

In this subsection, we extend the model to include the social media effect to V_i, the score of hotel i. In the constant term α_i in V_i in (A3), which represents static characteristics of hotel i, some social media information of hotel i, such as the reputation of the hotel, is reflected. Thus, we decompose α_i as

$$\alpha_i = \bar{\alpha}_i + \nu v_i, \ \nu > 0, \tag{1}$$

where $v_i \in \mathbf{R}$ represents the social media factor of hotel i and ν expresses the sensitivity of the factor on α_i, the static characteristics of hotel i. We remark that v_i can take not only a positive value but also a negative value, which affects positively (when it is positive)/negatively (when it is negative) the hotel i's score V_i compared with the case where $\alpha_i = \bar{\alpha}_i$.

This implies that if we identify the social factor v_i and estimate the sensitivity of the social media factor ν in α_i in (1), we can observe how the expected sales change if the social media factor v_i shifts and how the optimal room charge and overbooking level vary. The social media factor v_i, which is a real value, positively affects the booking probability p_i^γ and the expected sales of the hotel in (A4) through V_i, the score of the hotel i, in (A3), and p_i^γ, the booking probability, in (A2) if it increases. This is an extension of the static score inherent to hotel i, α_i, to incorporate the time-varying effect on α_i, the characteristic of the hotel, by social media.

One way to identify the social media factor v_i and estimate the sensitivity ν incorporated in α_i is by utilizing social media data, such as frequencies of some specific words in online reviews or other social media, along with detailed results of the hotel's revenue from online booking, which the hotel owns as private information.

In detail, a hotel may collect and utilize customer reviews in TripAdvisor, where perceived review credibility, review usefulness, and ease of use predict customer satisfaction (Filieri et al. (2020)), such as the scores rated by the customers and words in the customer reviews that affected the scores. (For other researches on online reviews, see Saumya et al. (2019) and Ismagilova et al. (2020) for instance). Particularly, a hotel may identify influential reviews from loyal customers who repeatedly book the hotel by those methods and utilize the reviews together with privately available detailed customer information.

Moreover, other social media components, such as frequencies of positive or negative words in social media such as Twitter, Facebook, and Instagram, could be used. As output data, detailed information on sales through online bookings, attributes of customers who booked, and the hitting ratio in the booking website will help estimate the effect of the social media factor more precisely.

2.3 Impact of social media factor changes

In this subsection, using the estimation result for the model in Saito et al. (2019) and assuming the sensitivity of the social media factor ν in (1), we observe how the expected sales, optimal room charge, and overbooking level change, if the social media factor of hotel shifts.

With the model parameters in Saito et al. (2019), whose attributes are described in Appendix A.1, assuming the social media factor, we consider the maximization of the expected sales with respect to the optimal room charge and the overbooking level. In short, the hotel aims to set the room charge and the overbooking level optimally to maximize its expected revenue, where there are trade-offs such that if the hotel sets the room charge high, the booking probability decreases, and if the hotel accepts more overbooking for last-minute cancellations, the hotel needs to repay the cost to decline the overbooked customers when the cancellation is less than expected. Specifically, we use the parameters originally estimated from the data collected from a Japanese online booking website for check-in dates ranging from March 1st to April 30th 2017 for standard nonsmoking twin rooms of two major luxurious hotels in Shinjuku area in Tokyo, Japan. We assume the two hotels' case $L = 2$ and name the hotels as hotel 1 and hotel 2. The parameters are as follows: $T = 14$, $\lambda = 2.1429$, $\mu = -8.161672$, $\sigma = 1.053053$, $\bar{\alpha}_1 = 0$, $\alpha_2 = -2.094104$, $\delta_2 = 0.849365$, $q_1 = 20$, and $x_2 = 42{,}292$. We remark that α_2 was originally estimated with $\alpha_1 = 0$ in the estimation in Saito et al. (2019) due to the degree of freedom of the parameters since only the difference $\alpha_2 - \alpha_1$ affects the booking probabilities in (A2).

Then, we can calculate the expected sales, the optimal room charge, and the optimal overbooking. Particularly, the social media factor positively affects the expected sales when it increases. Here, we assume the high cancellation rate with the distribution $P(r_1^H = 100\%) = P(r_1^H = 50\%) = P(r_1^H = 30\%) = \frac{1}{3}$, the high over-sale cost $c_1^H = 100,000$ for hotel 1, and the sensitivity of the social media factor $\nu = 1.0$.

Firstly, the following proposition holds.

Proposition 1 *The booking probability of hotel i by a customer, p_i^γ in (A2), increases for any availability of the hotels, γ, when the social media factor of hotel i, v_i, increases. Also, the optimal expected sales* $\max_{0<x_i<\infty} E[x_i R_T^i]$ *in the expected* *sales maximization of $E[x_i R_T^i]$ in (A4) increases as the social media factor v_i increases.*

Proof. Since by (1), the hotel i's characteristic factor α_i is increasing with respect to the social media factor v_i. Then, the hotel i's score V_i in (A3) is increasing and thus the booking probability by a random customer p_i^γ in (A2) is also increasing with respect to v_i. Since the booking probability by a customer p_i^γ increases in any availability of the hotels γ, the expected number of rooms booked $E[R_T^i]$ increases with respect to v_i for any $0 < x_i < \infty$, thus the expected revenue $E[x_i R_T]$ also increases with respect to v_i, for any $0 < x_i < \infty$. Therefore, the maximized value of $E[x_i R_T]$ with respect to $0 < x_i < \infty$ is also increasing with respect to the social media factor v_i. $\square$

In the overbooking case, the maximization of the expected revenue in the optimal overbooking problem $E[\min(q_i, (1 - r_i)R_T^i)x^{(i)} - \max((1 - r_i)R_T^i - q_i, 0)c_i]$ in (A5), when the number of rooms booked R_T^i increases as the social media factor v_i increases, the total oversale cost $\max((1 - r_i)R_T^i - q_i, 0)c_i$ also increases. Thus, when the overbooking level $L_i^{(ob)}$ is large, where the hotel accepts exceedingly more customers than the actual capacity, if the number of rooms booked R_T^i becomes large as the social media factor v_i increases, the total revenue decreases due to the effect of the over-sale cost $\max((1 - r_i)R_T^i - q_i, 0)c_i$. In such a case, an increase in the social media factor v_i does not necessarily mean an increase in the expected revenue for fixed overbooking level and room charge. However, in the optimization, the hotel i can set the overbooking level $L_i^{(ob)}$ lower so that the over-sale cost effect does not exceed the revenue increase. $\square$

Remark 1 The proposition indicates that the way the social media factor is incorporated into the revenue management model is plausible since the increase in the social media factor results in a rise in the booking probability and the expected sales of the hotel, which is compliant with the practical situation. Thus, if internal data necessary for parameter estimation are available, and the model is estimated, the quantitative model with the social media factor can be used to measure the sensitivity and the effect of the factor on the booking probability and expected sales quantitatively, which is the novelty of this proposition. As a managerial implication, hotels can first work on identifying what effort could increase the social media factor, then both the social media marketing team and the revenue management team cooperate to increase the revenue. Specifically, the proposition implies that the social media marketing team first works on identifying the social media factor and increasing the social media factor effect, and then the revenue management team conducts revenue maximization by optimally setting the room charge.

Remark 2 The proposition partially answers the research questions. For the first research question, by showing that a social media factor is suitably incorporated in the revenue management model, this proposition indicates that the social media factor in the model can be estimated from booking probability, and with the obtained estimation result, the company can work on revenue management to maximize its sales by using the revenue management model. For the second research question, the proposition suggests a managerial implication that the hotels can set up a social media marketing team in the financial department, which works on the social media advertising strategy that can increase the booking probability of the hotel and monitors social media impact in the financial market. By setting up the team, the finance department could conduct the revenue management and financing strategies with debt and equity better in an integrated way.

In this numerical example, the optimal overbooking level and room charge are solved, and we observe that the optimal expected revenue also increases as the social media factor increases.

In this example, the optimal expected sales of hotel 1 is JPY 447,155 when the overbooking level and the room charge are $(L_1^{(ob)}, x_1) = (29, 43{,}000)$ in the case of $v_1 = 1.0$, while it is JPY 428,082 when $(L_1^{(ob)}, x_1) = (29, 42{,}000)$ in the case of $v_1 = 0$. Therefore, the expected revenue increases as the social media factor increase from $v_1 = 0$ to $v_1 = 1.0$ in this case.

Figures 1, 2 describe the expected sales of hotel 1 when the optimal room charge and the overbooking level vary, when the social media factor v_1 is 0 and 1.0, respectively.

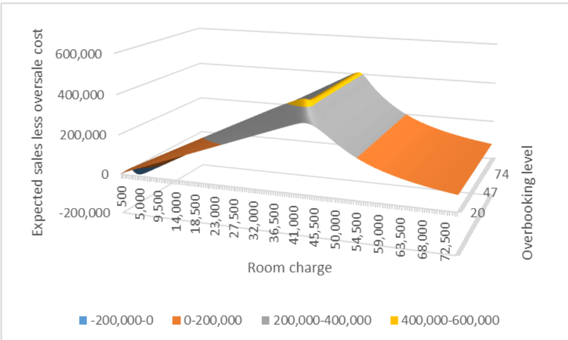

Fig. 1 The expected sales in the case of the high cancellation rate and the high oversale cost per room when $v_1 = 0$. The maximized expected sales is JPY 428,082 when $(L_1^{(ob)}, x_1) = (29, 42{,}000)$

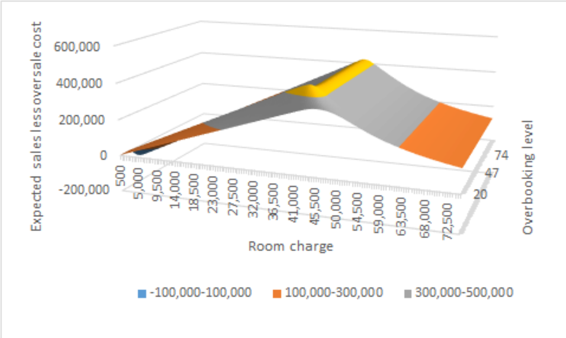

Fig. 2 The expected sales in the case of the high cancellation rate and the high over-sale cost per room when $v_1 = 1.0$. The maximized expected sales is JPY 447,155 when $(L_1^{(ob)}, x_1) = (29, 43{,}000)$

Moreover, Figure 3 shows the change in the expected sales when the social media factor v_1 is increased from 0 to 1.0 for fixed overbooking level $L_1^{(ob)} = 29$, which is optimal in both cases. We observe that by increasing the social media factor v_1 from 0 to 1.0, the graph of the expected sales curve shifts to the right above, and the optimal room charge changes from 42,000 to 43,000, where the corresponding expected sales increase from JPY 428,082 to JPY 447,155.

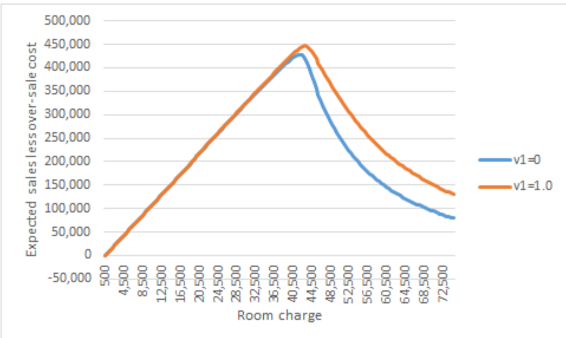

Fig. 3 The expected sales in the case of the high cancellation rate and the high over-sale cost per room when the room charge varies for $v_1 = 0$ and 1.0

3 Interest rate model with market sentiments

Next, as the second example of big data applications with models in financial management, we introduce an interest rate model with sentiment factors in Nishimura et al. (2019) and Nakatani et al. (2020). (For other studies on applications of interest rate models, see Menkveld et al. (2000), Morelli (2021), and Bali (2007), for instance.)

After the global financial crisis, we have observed global monetary easing and the resultant low-interest-rate environment. In such an environment, market sentiments mainly affect asset prices, particularly in the government bond markets, and thus it is essential to incorporate the sentiment factors in financial modeling.

In this section, we review the model in Nishimura et al. (2019) and Nakatani et al. (2020), whose mathematical details are explained in Appendix A.2, as an example in which the sentiment factors are estimated with big data, which are financial news, by text mining. The model exhibits impacts of the sentiments on the interest rate, which can be used by central banks that aim to control the market better to improve the economy. The model can also be useful for large traders, such as institutional investors and hedge funds, who invest money from pension funds in the financial markets. Moreover, we propose an extension of the model in Nishimura et al. (2019) and Nakatani et al. (2020), in which the sentiment factors are estimated not only by financial news but also by social media information.

3.1 Data

In Nishimura et al. (2019) and Nakatani et al. (2020), the three-factor model in Appendix A.2 is estimated by a stochastic filtering method, in which the frequencies of words closely related to the steepening (pessimistic) or the flattening (optimistic) factor are used in the observation equations in the filtering. In detail, those words relevant to the steepening and flattening factors are specified with financial news text data from Reuter. The estimated three-factor interest model helps predict how the sentiment-related words affect the yield curve shape, which can be used in trading by hedge funds or the yield curve control by central banks.

3.2 Possible extension to include social media information and estimation

We extend the observation equations in the stochastic filtering in Nishimura et al. (2019) and Nakatani et al. (2020) to include social media information. In detail, we find words closely correlated with the steepening or flattening factor and incorporate the frequencies of the words in the observation equations. The estimated three-factor interest rate model implies how social media information, in addition to financial news, affects the yield curve shape. For instance, Grover et al. (2019) investigated social media impacts on voting behavior during an election through acculturation of ideologies and polarization of voter preferences. As social media information, we may consider specific words closely related to the steepening or flattening factor in Twitter for political or economic events, for example.

Firstly, we assume the stochastic dynamics for x_1, x_2 and x_3 in the system equations (A7). In Nishimura et al. (2019) and Nakatani et al. (2020), along with the observation equations for the bond yields and frequencies of specified words by text mining in financial news, the parameters in the system equations are estimated.

Let $Y_t(n)$ be the yield of the n-year bond at time t and $F(A_i), i = 1, \dots, I_A, F(B_i), i = 1, \dots, I_B$, be the frequencies of the steepening related words and the flattening related words, respectively, specified by text mining in financial news.

Then, the discretized system equations and the observation equations are as follows.

System equations (discrete):

$$\begin{aligned} x_{j,t} &= e^{-\kappa_j^P \Delta t} x_{j,t-\Delta t} + \frac{\sigma_j}{\sqrt{2}} \sqrt{\frac{1 - e^{-2\kappa_j^P \Delta t}}{2\kappa_j^P}} \epsilon_{j,t}, \ \ j = 1, 2, \\ x_{3,t} &= x_{3,t-\Delta t} + \sigma_3 \sqrt{\Delta t} \epsilon_{3,t}, \end{aligned} \tag{2}$$

where $\kappa_j^P = \kappa_j^Q - \sigma_{x,j}\sigma_{c,j}$, $\lambda_3^Q = -\sigma_{x,3}\sigma_{c,3}$, $\Delta t = \frac{1}{250}$, $\epsilon_{j,t} \sim i.i.d.N(0, 1)$.

Observation equations:

$$\begin{aligned} Y_t(10) - Y_t(2) &= \sum_{l=1}^{3} \{X_{l,t}(10) - X_{l,t}(2)\} + e_{t,10-2y}, \\ Y_t(20) - Y_t(10) &= \sum_{l=1}^{3} \{X_{l,t}(20) - X_{l,t}(10)\} + e_{t,20-10y}, \\ Y_t(20) &= \sum_{l=1}^{3} X_{l,t}(20) + e_{t,20y}, \\ Y_t(30) &= \sum_{l=1}^{3} X_{l,t}(30) + e_{t,30y}, \end{aligned} \tag{3}$$

$$\begin{aligned} \log(\sum_{i=1}^{I_A} F(A_i) + 1) &= \xi_{1,c} + \xi_1 x_{1,t}^2 + e_{t,w_1}, \\ \log(\sum_{i=1}^{I_B} F(B_i) + 1) &= \xi_{2,c} + \xi_2 x_{2,t}^2 + e_{t,w_2}, \end{aligned} \tag{4}$$

where $X_j, j = 1, 2, 3$ are defined as in (A11), $e_{t,j} \sim i.i.d.N(0, \gamma_j^2)$, $j =$ 10-2y, 20-10y, 20y, 30y, w_1, w_2.

This indicates that as observable data in the observation equations in (3) and (4), we use the bond yield spreads for 10 year - 2 year, 20 year - 10 year, the yields of 20 year and 30 year bonds, frequencies of the steepening related words and the flattening related words and estimate the system equation parameters by the Monte Carlo filtering method. Here, $e_{t,j}$ are called the observation noise, and ϵ_j is the system noise. For details of the Monte Carlo filtering method, see Nakatani et al. (2019).

Since the investors monitor social media, such as online polls and Twitter, to observe the views of other investors and announcements from governors, social media affect the sentiments and views of investors. Thus, it is important to incorporate the effect of social media in addition to the financial news into the modeling.

In addition to the observation equations, we further include

$$\log(\sum_{i=1}^{I_C} F(C_i) + 1) = \eta_{1,c} + \eta_1 x_{1,t}^2 + e_{t,s_1},$$
$$\log(\sum_{i=1}^{I_D} F(D_i) + 1) = \eta_{2,c} + \eta_2 x_{2,t}^2 + e_{t,s_2}, \tag{5}$$

in the observation equations. Here, $C_i, i = 1, \ldots, I_C$, and $D_i, i = 1, \ldots, I_D$, stand for the frequencies of the steepening related words and the flattening related words from social media, respectively, and $e_{t,j} \sim i.i.d.N(0, \gamma_j^2),\ j = s1, s2$.

4 High-frequency trading equity market model

Furthermore, as the third model, we review the high-frequency trading equity market model in Saito and Takahashi (2019), whose mathematical details are explained in Appendix A.3, and consider the possible estimation with big data. Saito and Takahashi (2019) analyzed how the parameter shifts affect the stock price movement and equilibrium trading strategies of the three types of players with the model described in Appendix A.3, which expresses the interactions among the trading of three types of players. Although the estimation of the parameters with data has not been done in Saito and Takahashi (2019), if the financial authorities or stock exchanges utilize their internal data, we can observe how changes in regulations affect the behaviors of the participants and the asset price movements.

Specifically, the following internally available big data for financial authorities investigated in Saito et al. (2018) for Tokyo Stock Exchange, for example, could be used. To observe the transaction information, one needs high-frequency trading data, such as the trader ID, the order amount, the price, and the type of the order (the market order, the limit order, or the cancellation). If one has such data, one first analyzes the transaction data and classifies the traders into types by their trading patterns. Then, one estimates the parameters in the stochastic differential equations of the model by a stochastic filtering method and will be able to use the model to investigate the impacts of regulatory changes.

4.1 Incorporation of social media in estimation

Moreover, market sentiment-related words in social media could also be used in the estimation. Not only financial news but also views of the investors and announcements of politicians on social media can change sentiments in the market, which could affect the trading behaviors of the market participants. Particularly, since social media can affect elections and government policies, for example, monitoring social media and incorporating the sentiments are also important. By incorporating social media in estimating the model, we may estimate how the sentiments or announcements observed in social media affect the asset price.

5 Comparison of the three theoretical models

In this section, we compare the three quantitative theoretical models. The three models are common in that they are stochastic models for optimization, which utilize big data. In detail, the revenue management model maximizes the expected revenue of hotels, the high-frequency trading equity market model deals with the maximization of the expected revenue of the different types of traders, and the interest rate model is an equilibrium of agents who solve an optimal investment problem with sentiments, where the big data are online booking data, which include room charge and sales, interest data with sentiment-related words, and order and execution data of market participants, respectively.

These three models are described and compared from a quantitative perspective as follows. Firstly, the hotel revenue management model is for optimization with a random choice model, the interest rate model is a model whose parameters are estimated with stochastic filtering, and the high-frequency market model is a stochastic differential game. Secondly, these models are common in that they are stochastic models, which aim to capture the random choice behavior of customers, interest rate movement, and trading behaviors of players in the high-frequency stock market. Specifically, the revenue management model deals with the score of each hotel as a function of the room charge and the hotel's characteristic term, which includes the social media factor. The interest rate model expresses the short rate model with the steepening, flattening, and level factor, where the steepening and flattening factors could be estimated with financial news and social media-related information. The high-frequency stock market model deals with the stochastic trading behavior of the algorithmic trader, the market maker, and the general trader, and the stock price process where the expected return could be possibly estimated with a social media factor. Thus, they are common in dealing with optimization, maximization of expected revenue, minimization of error for estimation, and expected profit maximization.

Moreover, the three models are qualitatively described as follows. The three models are used for financial management. The revenue management model is for the business department of hotels which aims to increase sales. The interest rate model and the high-frequency trading equity market model can be used in the finance department to plan the timing for debt financing and equity financing. Also, the three models describe human transactions in the environment with randomness. In the revenue management model, the customers choose a hotel to book depending on the room charge and the hotel-specific factor. In the interest rate model, the agents solve optimal investment problems, and the high-frequency trading equity market model deals with the investment activity of the algorithm traders, the market makers, and general traders.

On the other hand, the three models are distinct in the following points. Firstly, they are different in usage. The revenue management model is for the maximization of the expected revenue. The interest rate model is for estimation of the model that describes the movement of the term structure of interest rates with sentiments, and the high-frequency trading equity market model solves for the equilibrium stock price under different types of traders. Also, the optimizing agents are dissimilar. In the revenue management model, optimization by a hotel revenue manager given the behavior of the customers is considered. In contrast, in the interest rate model and the high-frequency trading equity market model, the agents in the market maximize their objective functions. Furthermore, the big data types are different. The revenue management model utilizes online booking data, the interest rate model deals with interest rate data with the text news, and the high-frequency trading equity market model needs the order and execution data, respectively.

Remark 3 We consider three models in financial management, particularly for the finance department of a company, which deals with revenue and expenditure from the business and access to the financial market for financing by issuing corporate bonds and stocks. These three areas are important in terms of revenue management, debt financing, and equity financing. Therefore, we use the three models from the important three areas, revenue management in business, financing by debt and equity, for the finance department. Also, the three quantitative models reflect the practical aspects of each market, the mechanism of online booking, the government bond market with different maturity and the financial news, and the interaction of market participants in a high-frequency market. In addition, we note that the basis of the three models is a mixed logit model for random choice behavior and its application to revenue management for hotels, a yield curve model with quadratic Gaussian factor processes, which reflect the steepening and the flattening effect of a yield curve, with stochastic filtering incorporating text mining, and a stochastic differential game that solves the Nash equilibrium of the strategies of three different types of players.

6 Discussion

6.1 Theoretical implications

In this study, we have investigated three quantitative theoretical models for financial management utilizing big data and considered possible extensions to include social media factors. The contribution of this paper is the possible extensions of the quantitative modeling in financial management to include social media factors and the qualitative and quantitative comparison of the three models. Specifically, we have proposed the incorporation of the social media factor in a customer's score on a hotel, the steepening and the flattening factor for the interest rate movement, and a stock price movement.

There has been a vast of research on the utilization of big data and social media in management. Particularly, some research incorporates social media factors into quantitative modeling for optimization. For instance, Kumar et al.(2021) proposed a dynamic transmission model to investigate the impact of social media on the number of influenza and COVID-19 cases. Nilsang et al. (2019) investigated a model that considers real-time data from a social media application to minimize the response time and cost during emergencies and disasters. Zhu et al. (2021) developed the two-sided platform's scalable decisions on when to cooperate and how to optimize the pricing and investment decisions. (For other utilization of big data in various aspects of information systems and information management, see Gupta et al. (2018),(2019),(2020), Kamboj and Gupta (2020), Kamboj et al. (2018), Modgil et al. (2021), Duan et al. (2019), Dwivedi et al. (2019), Kumar et al. (2021). For utilization of social media, see Giannakis et al. (2022), Rad et al. (2018), Grover et al. (2022), Wamba et al. (2019), Bogaert et al. (2018)).

To the best of our knowledge, this study is the first attempt to investigate the comparison of the quantitative model using big data and social media in the field of financial management, which enables us to estimate the effect and the impact of social media quantitatively by introducing the social media factor as a new variable, if internal data necessary for parameter estimation are available, and consider the synthetic use of the three models in financial management with managerial implications from an integrated and higher viewpoint.

Specifically, for the three models investigated in this study, Saito et al.(2019) worked on revenue management with the online booking data for two luxurious hotels in Shinjuku area in Tokyo, considering cancellation and overbooking strategies. Nakatani et al.(2020) estimated a yield curve model with the steepening and flattening sentiment factors using stochastic filtering with text mining for financial news. Saito and Takahashi(2019) considered a theoretical model that describes the different types of players in a high-frequency stock market where the trading by algorithmic traders, market makers, and general traders interact with each other and affect the stock price movement. Although these models implement practical aspects of financial management, the social media factor is not incorporated. Therefore, we have investigated possible social media extensions of the three models for revenue management, interest rate, and the stock market, which could be used in financial management in the measurement of the effect and the impact of social media quantitatively if internal data necessary for parameter estimation are available.

6.2 Managerial implications

Big data applications with theoretical models in financial management positively affect a wide range of stakeholders such as companies, investors, financial institutions, financial authorities, and the economy in the country as follows.

6.2.1 Revenue management model

The model incorporates the social media factor, and its increase affects the expected sales positively through the increase in the choice probability of the hotel. This revenue management model with social media has the following managerial implications. First of all, revenue managers in hotels, who set room charges of their hotel on online booking websites, can decide the room charges and take the overbooking strategy optimally so that the hotel can maximize its expected revenue. Hotels can analyze how social media affect the customers and utilize social media effectively by improving their services so that it can affect the booking from online customers positively. Moreover, as stakeholders, customers can benefit from the improved services and obtain valuable sales information through social media and book through online booking. Hotel investors can make the hotel introduce the revenue management model for better profitability, which leads to improvement of the investment performance.

Also, in the revenue management model in Appendix A.1, hotel revenue managers can maximize the revenue of the hotel by optimally setting the room charge and managing the social media so that it can affect sales positively. Customers can enjoy the merit of the marketing efforts by the hotel and obtain valuable information through social media. Furthermore, the city and the companies in the area can earn revenue from tourism, where hotels make efforts to attract more visitors and transmit information through social media.

6.2.2 Interest rate sentiment model

The interest rate model with market sentiment could be estimated more precisely if we incorporate social media information in addition to the interest rate data and financial news. This interest rate model with sentiments and social media has the following managerial implications.

Firstly, central banks can monitor how the words in financial news and social media affect the interest rate market through the sentiment model and effectively conduct monetary policies by making announcements strategically. Moreover, investors such as hedge funds and insurance companies trade effectively, monitoring how the words related to sentiments observed in the financial news and social media affect the bond prices. Furthermore, government liaises with central banks to make announcements, takes fiscal policies, and successfully controls the economy in the country.

Secondly, in the interest rate model with sentiments in Section 3, investors such as institutional investors, hedge funds, and pension funds can trade, observing what financial news and words on social media would affect the yield curves. Moreover, central banks can effectively conduct monetary policies by monitoring the sentiment factors estimated from the financial news and social media in the model, which leads to a better economy in the country.

6.2.3 High-frequency trading equity market model

The high-frequency trading model incorporates the interactions among the different types of players. Moreover, with detailed internal transaction data and social media, the model becomes useful for financial authorities and large investors. If such internal data, all order and execution data with server IDs that the stock exchanges own and are shared with financial authorities by concluding a confidential agreement for research purposes, are available and the parameters are estimated with such data and social media as discussed in Section 4, by the estimated parameters in the model with the data that include the order and transaction data of algorithmic traders, market makers, and general traders, and data from social media, the market participants, which include the listed companies whose stocks are traded in the market, investors, and the financial authorities who regulate the high-frequency trading market, could predict how the stability of the current market by investigating the estimated parameters. In detail, this high-frequency trading market model has the following managerial implications.

Firstly, the financial managers of a company, who needs to decide when to issue new stocks for equity financing, can predict how the algorithmic traders' trading activities affect the price and predict the best timing for equity financing by conducting public offering for the new issuance.

Secondly, financial authorities who need to regulate the rules in the financial markets can figure out how the trading of the algorithmic traders affects the stability of the market and set regulations so that the market becomes fair and stable. Stock exchanges can suitably maintain fees from all the participants for providing them a fair market, while they also satisfy the algorithmic traders by keeping them as good customers. Also, a central bank can consider the best timing for releasing the announcement, considering how the announcement affects the market.

Moreover, although the data accessibility is limited compared to the financial authorities, institutional investors who care about price impacts caused by trading a large volume of stocks, with their internally available data, they can optimally execute their orders taking into account the algorithmic traders' accelerations in the trades. Private investors can also predict when the market is unstable and about to crash and trade along with the large price moves.

Furthermore, algorithmic traders can predict when the market becomes unstable by the parameters estimated with social media information and trade immediately when unexpected market news occurs.

Furthermore, in the high-frequency trading market model in Appendix A.3, institutional investors can optimally execute large sizes of trades without causing price impacts. Private investors can follow the rapid price changes predicting the signals of instability. Additionally, financial authorities can set regulations for the stability of the market by observing what trading activities could affect the stability. Stock exchanges can set the trading fees for algorithmic traders properly to maintain a stable and fair trading environment for institutional and private investors.

6.3 Synthetic use of the models by a company (Hotel REIT's case)

The finance department of a hotel conducts revenue management to increase sales and manage interest rate risk and the share price of the hotel for debt and equity financing. In addition to the hotel's booking data, interest rate data, and stock market data, by incorporating social media factors, which affect the sentiment of the customers and the markets, the models become more useful in optimizing the whole business of the hotels.

These three models can be used in the finance department of companies. For instance, hotel REITs (Real Estate Investment Trusts) can utilize the models to increase their financial results. A hotel REIT is a company that purchases hotel buildings and lends the buildings to hotels to earn rent fees as revenue. The REIT pays dividends from the revenue to REIT's shareholders. The rent fee can be fixed or linked to the performance of the hotels. Then, the REIT can use the models as follows. Firstly, REITs make hotels that pay the performance-linked rent fee use the revenue management model to increase their sales, which results in an increase in the rent fee. Secondly, the REIT that issues corporate bonds for debt financing manages the risk of the interest rate hikes by the interest rate model with sentiments. Thirdly, the REIT, whose share is traded in the high-frequency trading equity market and subject to the trading of HFT traders, can increase its capital by public offering at the best timing by utilizing the high-frequency trading equity market model.

After the outbreak of COVID-19, it has become more important for the hotel industry to conduct financial management optimally. According to Japan hotel REIT (2021), amid the COVID-19 pandemic, the domestic demand for hotels in Japan decreased, and the hotel management's important metrics deteriorated. The hotel REIT introduced the variable rent fee scheme and also conducted third-party allocation of shares and debt financing with long maturities. As the COVID-19 situation calms down, it is expected that domestic and inbound demand will recover, and due to the changes in the market environment, hotels need to increase revenues by understanding customer needs. In such a situation, the hotel REIT can utilize the revenue management model to increase their revenue and predict long-term interest rates by the sentiment model and the best timing to conduct public offering with the high-frequency trading model.

6.4 Limitations and future research direction

There are some limitations to the extension of the revenue management model. Firstly, it is currently unable to collect new hotel booking data in this COVID-19 situation in which people are unable to travel. (For a collective insight on the impact of COVID-19 on information management research, see Dwivedi et al. (2020)).

Secondly, even with the past booking data, as in Saito et al. (2019), it is still challenging to find a clear connection between the online reviews and the publicly available customers' booking data since the data period is one month, which is relatively short compared with the gradual effect of the social media. To find a clear connection between social media and online booking, we need some additional internal information that hotels own, such as detailed sales data, attributes of the customers who booked rooms through the booking website, and answers to the questionnaires from customers. Thus, in this research, we limit ourselves to assuming parameters on the social media factors, conduct numerical experiments, and discuss possible methodologies to estimate the social media factors.

Filieri et al. (2020) show that perceived review credibility is one of the most crucial determinants of travelers' satisfaction and continued use of user-generated content (UGC) platforms. Particularly, Saumya et al. (2019) proposed a method to predict the most helpful online review, and Ismagilova et al. (2020) examined relations between emotions in the reviews and their perceived helpfulness. Identification of the social factor using influential reviews by those methods with internally available information and estimation for the sensitivity of the social media factor will be one of our future research topics.

7 Concluding remarks

In this study, we have investigated big data applications with theoretical models in financial management. Firstly, we have explored the hotel revenue management model, interest rate model with market sentiments, and high-frequency trading equity market model from an integrated perspective and discussed possible extensions to include social media factors. Moreover, in the extended revenue management model, we have conducted numerical experiments to observe how the revenue from online bookings changes and where the optimal room charges and overbooking levels are determined when the social media factor shifts. Finally, we have compared the models and discussed the managerial implications of those applications for stakeholders in financial management.

Declarations

- Funding: This research was supported by CARF (Center for Advanced Research in Finance).
- Conflict of interest/Competing interests: The authors have no relevant financial or non-financial interests to disclose.
- Availability of data and materials: Data sharing is not applicable to this article as no datasets were generated or analyzed.

References

[1] Bali, T. G. (2007). Modeling the dynamics of interest rate volatility with skewed fat-tailed distributions. *Annals of Operations Research*, 151(1), 151-178.

[2] Bogaert, M., Ballings, M., & Van den Poel, D. (2018). Evaluating the importance of different communication types in romantic tie prediction on social media. *Annals of Operations Research*, 263(1), 501-527.

[3] Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data–evolution, challenges and research agenda. *International Journal of Information Management*, 48, 63-71.

[4] Dwivedi, Y. K., Hughes, D. L., Coombs, C., Constantiou, I., Duan, Y., Edwards, J. S.,& Upadhyay, N. (2020). Impact of COVID-19 pandemic on information management research and practice: Transforming education, work and life. *International Journal of Information Management*, 55, 102211.

[5] Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., & Williams, M. D. (2019). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 101994.

[6] Filieri, R., Acikgoz, F., Ndou, V., & Dwivedi, Y. (2020). Is TripAdvisor still relevant? The influence of review credibility, review usefulness, and ease of use on consumers' continuance intention. *International Journal of Contemporary Hospitality Management*, 33 (1), 199-223.

[7] Giannakis, M., Dubey, R., Yan, S., Spanaki, K., & Papadopoulos, T. (2022). Social media and sensemaking patterns in new product development: demystifying the customer sentiment. *Annals of Operations Research*, 308, 145-175.

[8] Grover, P., Kar, A. K., Dwivedi, Y. K., & Janssen, M. (2019). Polarization and acculturation in US Election 2016 outcomes: Can twitter analytics predict changes in voting preferences ? Technological Forecasting and Social Change, 145, 438-460.

[9] Grover, P., Kar, A. K., & Dwivedi, Y. K. (2022). Understanding artificial intelligence adoption in operations management: insights from the review of academic literature and social media discussions. *Annals of Operations Research*, 308, 177-213.

[10] Gupta, S., Altay, N., & Luo, Z. (2019). Big data in humanitarian supply chain management: A review and further research directions. *Annals of Operations Research*, 283(1), 1153-1173.

[11] Gupta, S., Kar, A. K., Baabdullah, A., & Al-Khowaiter, W. A. (2018). Big data with cognitive computing: A review for the future. *International Journal of Information Management*, 42, 78-89.

[12] Ismagilova, E., Dwivedi, Y. K., & Slade, E. (2020). Perceived helpfulness of eWOM: Emotions, fairness and rationality. *Journal of Retailing and Consumer Services*, 53.

[13] Japan hotel REIT (2021). Kessan Tanshin (Summary of financial results.) *https://www.jhrth.co.jp/file/term-089c9a0a7367506473473ba4a56136fa0a0693ef.pdf*

[14] Kar, A. K., & Dwivedi, Y. K. (2020). Theory building with big data-driven research–Moving away from the "What" towards the "Why". *International Journal of Information Management*, 54, 102205.

[15] Kamboj, S., & Gupta, S. (2020). Use of smart phone apps in co-creative hotel service innovation: an evidence from India. *Current Issues in Tourism*, 23(3), 323-344.

[16] Kamboj, S., Sarmah, B., Gupta, S., & Dwivedi, Y. (2018). Examining branding co-creation in brand communities on social media: Applying the paradigm of Stimulus-Organism-Response. *International Journal of Information Management*, 39, 169-185.

[17] Kumar, P., Polonsky, M., Dwivedi, Y. K., & Kar, A. (2021). Green information quality and green brand evaluation: the moderating effects of eco-label credibility and consumer knowledge. *European Journal of Marketing*, 55(7), 2037-2071.

[18] Kumar, S., Xu, C., Ghildayal, N., Chandra, C., & Yang, M. (2021). Social media effectiveness as a humanitarian response to mitigate influenza epidemic and COVID-19 pandemic. *Annals of Operations Research*,

https://doi.org/10.1007/s10479-021-03955-y.

[19] McGroarty, F., Booth, A., Gerding, E., & Chinthalapati, V. R. (2019). High frequency trading strategies, market fragility and price spikes: an agent based model perspective. *Annals of Operations Research*, 282(1), 217-244.

[20] Menkveld, A. J., & Vorst, T. (2000). A pricing model for American options with Gaussian interest rates. *Annals of Operations Research*, 100(1), 211-226.

[21] Modgil, S., Gupta, S., Sivarajah, U., & Bhushan, B. (2021). Big data-enabled large-scale group decision making for circular economy: An emerging market context. *Technological Forecasting and Social Change*, 166, 120607.

[22] Morelli, G. (2021). Fair prices under a unified lattice approach for interest rate derivatives. *Annals of Operations Research*, 299, 429-441.

[23] Nakatani, S., Nishimura, K. G., Saito, T., & Takahashi, A. (2020). Interest rate model with investor attitude and text mining. *IEEE Access*, 8, 86870-86885.

[24] Nakatani, S., Nishimura, K. G., Saito, T., & Takahashi, A. (2019). Online appendix for Interest rate model with investor attitude and text mining (No. CIRJE-F-1136). *CIRJE, Faculty of Economics, University of Tokyo.*

[25] Nilsang, S., Yuangyai, C., Cheng, C. Y., & Janjarassuk, U. (2019). Locating an ambulance base by using social media: A case study in Bangkok. *Annals of Operations Research*, 283(1), 497-516.

[26] Nishimura, K. G., Sato, S., & Takahashi, A. (2019). Term structure models during the global financial crisis: A parsimonious text mining approach. *Asia-Pacific Financial Markets*, 26(3), 297-337.

[27] Rad, A. A., Jalali, M. S., & Rahmandad, H. (2018). How exposure to different opinions impacts the life cycle of social media. *Annals of Operations Research*, 268(1), 63-91.

[28] Saito, T., & Takahashi, A. (2019). Stochastic differential game in high frequency market. *Automatica*, 104, 111-125.

[29] Saito, T., Takahashi, A., & Tsuda, H. (2016). Optimal room charge and expected sales under discrete choice models with Limited capacity. *International Journal of Hospitality Management*, 57, 116-131.

[30] Saito, T., Takahashi, A., Koide, N., & Ichifuji, Y. (2019). Application of online booking data to hotel revenue management. *International Journal of Information Management*, 46, 37-53.

[31] Saito, T., Adachi, T., Nakatsuma, T., Takahashi, A., Tsuda, H., & Yoshino, N. (2018). Trading and Ordering Patterns of Market Participants in High Frequency Trading Environment: Empirical Study in the Japanese Stock Market. *Asia-Pacific Financial Markets*, 25(3), 179-220.

[32] Saumya, S., Singh, J. P., & Dwivedi, Y. K. (2019). Predicting the helpfulness score of online reviews using convolutional neural network. *Soft Computing*, 1-17.

[33] Sun, E. W., Kruse, T., & Yu, M. T. (2014). High frequency trading, liquidity, and execution cost. *Annals of Operations Research*, 223(1), 403-432.

[34] Wamba, S. F., Edwards, A., & Akter, S. (2019). Social media adoption and use for improved emergency services operations: The case of the NSW SES. *Annals of Operations Research*, 283(1), 225-245.

[35] Zhu, X., Yang, C., Liu, K., Zhang, R., & Jiang, Q. (2021). Cooperation and decision making in a two-sided market motivated by the externality of a third-party social media platform. *Annals of Operations Research*, https://doi.org/10.1007/s10479-021-04109-w.

Appendix A Mathematical details of the models

A.1 The revenue management model

Firstly, we present the revenue management model in Saito et al. (2016), (2019). The model is described as follows.

We assume that there are L hotels named hotels $1, \ldots, L$, in the same area with the same grade. We consider the revenue management model where customers, who visit an online booking website to book a room in the area, choose a hotel among those L hotels. We fix the room type that the customers are aiming to book and the check-in date. Let $[0, T]$ be the booking period where time 0 is the first date of the checking period, and T is the last date of the check-in period, which is the same as the check-in date.

Suppose that the customers visit the website at a frequency following a Poisson process $\{N_t\}_{0\leq t\leq T}$ with the intensity λ and choose a hotel to book among the L hotels. Hotel i, $i = 1, \ldots, L$ has a limited capacity of q_i rooms, but accept overbooking from the customers up to $L_i^{(ob)}$ rooms, where $q_i \leq L_i^{ob}$, for the last minute cancellation at the check-in date T.

Moreover, let R_t^i $(0 \leq R_t^i \leq L_i^{(ob)})$ be the number of rooms booked for hotel i by time $t \in [0, T]$, where $R^i,\ i = 1, \ldots, L$ satisfy

$$\sum_{i=1}^{L} R_t^i = N_t. \tag{A1}$$

Furthermore, we suppose that a customer, who visits the website aiming to book a room among those L hotels at time $t \in [0, T]$ at the random frequency following the Poisson process N, chooses hotel i, $i = 1, \ldots, L$ with the probability p_i given by

$$p_i^\gamma = \int_{-\infty}^{\infty} \frac{\exp(V_i) 1_{\{\gamma_i = 1\}}}{\sum_{j=1}^{L} \exp(V_j) 1_{\{\gamma_j = 1\}}} h(z) dz, \tag{A2}$$

where $\gamma = (\gamma_1, \ldots, \gamma_L)$ stands for the availability of the hotels, in which $\gamma_i = 1$ indicates that hotel i is available for booking ($R_{t-}^i < L_i^{(ob)}$, where $t-$ indicates the time just before t), and $\gamma_i = 0$ expresses that hotel i is fully booked ($R_{t-}^i = L_i^{(ob)}$) and not available for booking, and each V_i expresses hotel i's score, a linear combination of factors, namely, the room charge $x_i \in (0, \infty)$, a holiday dummy variable y ($y = 0$ if time t is a week day and $y = 1$ if it is a day before a holiday), and a constant $\alpha_i \in \mathbf{R}$ that represents the static score inherent to hotel i as

$$\begin{gathered} V_i = \alpha_i - \beta x_i + \delta_i y,\ i = 1, \ldots, L,\ \beta > 0,\ \delta_i \in \mathbf{R}, \\ \beta = e^{\mu + \sigma z},\ h(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}. \end{gathered} \tag{A3}$$

Here $-\beta < 0$ indicates that as the room charge x_i increases, the score of hotel i decreases, and the hotel is less likely to be chosen by the customers. The equation (A3) describes that the coefficient follows a log-normal distribution which reflects the randomness of the sensitivity of the room charge for the customers.

Then, Saito et al. (2016) consider the case where there is no overbooking, i.e. $L_i^{(ob)} = q_i,\ i = 1, \ldots, L$ is considered and the hotel i's objective is set to maximize the expected revenue given by

$$E[x_i R_T^i], \tag{A4}$$

by optimally choosing the room charge x_i, where $E[\cdot]$ stands for the expectation of the random variable.

Moreover, Saito et al. (2019) investigate an overbooking strategy for last minute cancellations, i.e., hotel i maximizes

$$E[\min(q_i, (1 - r_i) R_T^i) x^{(i)} - \max((1 - r_i) R_T^i - q_i, 0) c_i], \tag{A5}$$

by optimally choosing the room charge x_i and the overbooking level $L_i^{(ob)}$, where the rooms are booked up to $L_i^{(ob)} \in \mathbf{N}$ rooms, instead of q_i, q_i is the actual capacity, $r_i \in [0,1]$ is the last minute cancellation ratio, which we assume to be a random variable independent of N and the hotel choices by the customers, namely, $r_i R_T^i$ customers do not appear at the check-in data. $(1 - r_i)R_T^i$ customers show up at the check-in date, however, if it exceeds the actual capacity q_i, hotel i has to decline $(1 - r_i)R_T^i - q_i$ customers by compensating $c_i \geq 0$ per room. In other words, this objective function describes that hotel i maximizes the expected revenue minus the overbooking cost by optimally setting the room charge and the overbooking level.

A.2 The interest rate model with market sentiments

Next, in this subsection, we introduce the interest rate model with market sentiments in Nishimura et al. (2019) and Nakatani et al. (2020). In the model, a steepening factor and a flattening factor for an interest rate curve, representing pessimistic and optimistic sentiment, respectively, are estimated by financial news text data along with interest rate data. The model is expressed as follows.

Let $\{r_t\}_{0 \leq t < \infty}$ be a short rate process expressed as a linear combination of three factors x_1^2, x_2^2, and x_3,

$$r_t = c_1 x_{1,t}^2 + c_2 x_{2,t}^2 + x_{3,t}, \tag{A6}$$

where $c_1 < 0$, $c_2 > 0$, and x_j, $j = 1,2,3$, satisfy stochastic differential equations (SDEs)

$$\begin{aligned} dx_{j,t} &= -\kappa_j^Q x_{j,t} dt + \sigma_{x,j} dB_{j,t}^Q, \ j = 1,2, \\ dx_{3,t} &= \lambda_3^Q dt + \sigma_{x,3} dB_{3,t}^Q, \end{aligned} \tag{A7}$$

where $\kappa_j^Q > 0, j = 1,2, \sigma_{x,j} > 0, j = 1,2,3$, and $\lambda_3^Q \in \mathbf{R}$. Here, B^Q is a three-dimensional Brownian motion under the risk neutral probability measure Q.

Here, x_1^2 and x_2^2 represent the steepening factor and the flattening factor, respectively. $c_1 x_{1,t}^2 < 0$ $(c_2 x_{2,t}^2 > 0)$ affects the short rate r_t negatively (positively), which fades away as time passes, since $x_{1,t}^2$ $(x_{2,t}^2)$ decreases to 0 due to the mean reversion of x_1 (x_2) in SDE (A7). This implies that $c_1 x_1^2 < 0$ $(c_2 x_2^2 > 0)$ pushes down (up) the short rate r_t first, but it fades away, which makes the yield curve shape at time 0 steepen (flatten), and thus we call it a bull-steepening (bull-flattening) effect.

Thus, x_1^2 and x_2^2 control the slope of the yield curve, a collection of yields for bonds with different maturities. x_1^2 and x_2^2 also correspond to the pessimistic and the optimistic factor, respectively, since when the market is pessimistic (optimistic), the yields for the near future go lower (higher), which makes the upward sloping curve steepen (flatten). x_3 is the level factor, a Gaussian process that controls the absolute level of the short rate r.

Then, $P_t(\tau)$, the zero coupon bond price with time to maturity τ at time t and the zero coupon bond yield $Y_t(\tau)$ for the time to maturity τ at time t are calculated as

$$P_t(\tau) = E_t^Q[e^{-\int_t^{t+\tau} r_u du}] \tag{A8}$$
$$= \exp(-\tau(X_{1,t}(\tau) + X_{2,t}(\tau) + X_{3,t}(\tau))), \tag{A9}$$

and

$$Y_t(\tau) = X_{1,t}(\tau) + X_{2,t}(\tau) + X_{3,t}(\tau), \tag{A10}$$

where

$$\begin{aligned} X_{j,t}(\tau) &= \frac{-1}{\tau}[A_j(\tau) + C_j(\tau)x_{j,t}^2],\ j = 1, 2, \\ X_{3,t}(\tau) &= x_{3,t} + \frac{\lambda_3^Q}{2}\tau - \frac{\sigma_3^2}{6}\tau^2, \end{aligned} \tag{A11}$$

with

$$\begin{aligned} C_j(\tau) &= C_{0,j} + \frac{1}{z_j(\tau)}, \\ A_j(\tau) &= \frac{\sigma_j^2}{2}\int_0^\tau C_j(s)ds, \end{aligned} \tag{A12}$$

where

$$\begin{aligned} C_{0,j} &= \frac{\kappa_j^Q + \sqrt{(\kappa_j^Q)^2 + c_j\sigma_j^2}}{\sigma_j^2}, \\ z_j(\tau) &= \frac{\sigma_j^2}{\alpha_j} - e^{\alpha_j\tau}(\frac{1}{C_{0,j}} + \frac{\sigma_j^2}{\alpha_j}), \\ \alpha_j &= 2(\kappa_j^Q - \sigma_j^2 C_{0,j}), \\ \sigma_j &= \sqrt{2}\sigma_{x,j},\ j = 1, 2,\ \sigma_3 = \sigma_{x,3}. \end{aligned} \tag{A13}$$

Here, $E_t^Q[\cdot]$ denotes the conditional expectation under Q with respect to the augmented filtration generated by B^Q at time t.

A.3 High-frequency trading equity market model

Finally, as the third example of the application of big data along with theoretical models in financial management, we explain the high-frequency trading market model introduced by Saito and Takahashi (2019). The model describes the trading behaviors of three types of players and those interactions in high-frequency trading markets, where the players can trade in a millisecond interval, and the resultant price actions. This is important in the current financial markets, where algorithmic traders play a central role, and the price actions cause large economic effects. (For other studies on high-frequency trading markets, McGroarty et al. (2019) present an agent-based simulation for investigating algorithmic trading strategies, and Sun (2014) et al. propose a discrete optimization framework to describe how high-frequency trading can be utilized to supply liquidity and reduce execution cost, for example.)

As a model to describe high-frequency trading markets, in which trading patterns of different types of players and the stock price dynamics are observed, Saito and Takahashi (2019) proposed a stochastic differential game model. Specifically, there are three types of players in the model, algorithmic traders, market makers, and general traders, who aim to maximize their objective functions, and equilibrium strategies of the three types of players, in which each type maximizes its objective given the others' strategies, are obtained. The model indicates how the stock price moves, depending on the parameters in the model, which is useful in observing how the rapid price fall occurs and how the financial authorities should set appropriate regulations to prevent excessive price movements in the high-frequency trading environment.

The model is described as follows. Firstly, suppose that there are three types of players, algorithmic traders (player 1), market makers (player 2), and general traders (player 3), who optimally trade the asset in the high-frequency trading market. Let $[0, T]$ be the trading period, X_t^0 be the mid-price of the asset at time t, and $X_t^j,\ j = 1, 2, 3$ be the positions of player j. Also, let $\alpha_t^j dt,\ j = 1, 2$ be the units of the asset bought by player j in $[t, t + dt]$ (if $\alpha_t^j < 0$, it indicates $|\alpha_t^j|dt$ units are sold by player j), and α_t^3 be the spread from the mid-price set by the market makers, player 3. Moreover, we assume the following dynamics for X^0 and $X^j, j = 1, 2, 3$, and objective functions for players 1,2,3. Let W be one-dimensional Brownian motion and $\chi_t = 1_{\{t \leq t_0\}}$ is an indicator function, which takes a value 1 until $t = t_0$ and 0 thereafter. Here $t_0 \in [0, T]$ is a time lag after which the general traders can respond and start trading.

- Mid-price process:

$$dX_t^0 = (\mu + (\gamma_1 \alpha_t^1 + \gamma_2 \chi_t \alpha_t^2 + \delta \alpha_t^3))dt + \sigma_t dW_t,$$
$$X_0^0 = x, \tag{A14}$$

  where $\mu \in \mathbf{R}, \gamma_1, \gamma_2, \delta, \sigma_t > 0$.
  dX_t^0 stands for the change in the mid-price process of the asset in $[t, t + dt)$. In addition to the drift term μdt and the diffusion term $\sigma_t dW_t$, there is a market impact term $(\gamma_1 \alpha_t^1 + \gamma_2 \chi_t \alpha_t^2 + \delta \alpha_t^3)dt$ caused by the trading of the three types of players. In detail, when the algorithmic traders or the general traders sell the asset ($\alpha_t^1 < 0$ or $\alpha_t^2 < 0$), it pushes down the mid-price by the amount proportional to the units of assets sold by them, and if the market makers set the spread from the mid-price negatively ($\alpha_t^3 < 0$) to buy the

amount sold by the algorithmic traders and the general traders, the price moves down by the amount proportional to the spread.

- The position of the algorithmic traders is

$$dX_t^1 = \alpha_t^1 dt,\ X_0^1 = 0, \tag{A15}$$

and the objective function of the algorithmic traders for maximization is

$$\begin{aligned} &J^1(\alpha_1, \alpha_2, \alpha_3) \\ &= E\left[-\int_0^T \alpha_t^1 (X_t^0 + \alpha_t^3 + \lambda(\alpha_t^1 + \chi_t \alpha_t^2))dt + X_T^1 X_T^0 - \frac{1}{2}c_1 (X_T^1)^2\right], \end{aligned} \tag{A16}$$

where $\lambda, c_1 > 0$.

Here, dX_t^1 expresses the change in the position of the algorithmic traders (player 1), where X^1 starts from 0, the flat position, at time 0. The algorithmic traders (player 1) start trading from time 0 and buy $\alpha_t^1 dt$ units of the asset in $[t, t+dt)$. The algorithmic traders can trade the asset at the price $X_t^0 + \alpha_t^3 + \lambda(\alpha_t^1 + \chi_t \alpha_t^2)$, in which α_t^3 is the spread from the mid-price (e.g. $\alpha_t^3 > 0$) set by the market makers (player 3), and $\lambda(\alpha_t^1 + \chi_t \alpha_t^2)$ is the price slippage caused by the tradings from the algorithmic traders and the general traders (e.g. $\lambda(\alpha_t^1 + \chi_t \alpha_t^2) > 0$, when $\alpha_t^1, \alpha_t^2 > 0$). Thus, $-\int_0^T \alpha_t^1 (X_t^0 + \alpha_t^3 + \lambda(\alpha_t^1 + \chi_t \alpha_t^2))dt$ indicates the cash paid for the trading in the period, $X_T^1 X_T^0$ is the mark-to-market value of the asset position at time T, and $-\frac{1}{2}c_1 (X_T^1)^2$ describes the liquidation cost paid for X_T^1 units of the asset at time T. Thus, the algorithmic traders aim to maximize $J^1(\alpha_1, \alpha_2, \alpha_3)$, the expectation of the mark-to-market value of the position at time T, by choosing α_1 when α_2, α_3 are given.

- The position of the general traders is

$$dX_t^2 = \chi_t \alpha_t^2 dt,\ X_0^2 = x_2 > 0, \tag{A17}$$

and the objective function of the general traders for maximization is

$$\begin{aligned} &J^2(\alpha_1, \alpha_2, \alpha_3) \\ &= E\Bigg[-\int_0^T \chi_t \alpha_t^2 (X_t^0 + \alpha_t^3 + \lambda(\alpha_t^1 + \chi_t \alpha_t^2))dt \\ &-\frac{\eta}{2}\int_0^T \chi_t (X_t^2)^2 \sigma_t^2 dt + (X_T^2 X_T^0 - X_0^2 X_0^0) - \frac{1}{2}c_2 (X_T^2)^2\Bigg], \end{aligned} \tag{A18}$$

where $\eta, c_2 > 0$.

dX_t^2 describes the change in the position of the general traders (player 2), where X^2 starts from a long position $x_2 > 0$ at time 0. In contrast to the algorithmic traders, the general traders can start trading from time t_0, the lag of the trading speed of the algorithmic traders. As in the algorithmic traders' case, the general traders aim to maximize the trading profit with the risk aversion term $-\frac{\eta}{2}\int_0^T \chi_t (X_t^2)^2 \sigma_t^2 dt$, which implies that the general traders prefer to reduce the position size.

- The position of the market makers is

$$dX_t^3 = (-(\alpha_t^1 + \chi_t \alpha_t^2) + k\alpha_t^3)dt, \; X_0^3 = 0, \tag{A19}$$

and the objective function of the market makers is

$$J^3(\alpha_1, \alpha_2, \alpha_3) = E\Bigg[-\int_0^T (-(\alpha_t^1 + \chi_t \alpha_t^2) + k\alpha_t^3)(X_t^0 + \alpha_t^3 + \lambda(\alpha_t^1 + \chi_t \alpha_t^2))dt + X_T^3 X_T^0 - \frac{1}{2}c_3(X_T^3)^2\Bigg], \tag{A20}$$

where $k, c_3 > 0$.
dX_t^3 represents the change in the position of the market makers (player 3). The market makers set the spread α_t^3 from the mid-price X_t^0.
For example, if $\alpha_t^1, \alpha_t^2 < 0$ and $\alpha_t^3 < 0$, $-(\alpha_t^1 + \chi_t \alpha_t^2)dt$ describes the amount the market makers accept against the selling orders from the algorithmic traders and the general traders. The market makers aim to buy at a lower level $X_t^0 + \alpha_t^3 + \lambda(\alpha_t^1 + \chi_t \alpha_t^2)$, by setting the spread $\alpha_t^3 < 0$. However, $k\alpha_t^3 dt < 0$ indicates that if the market makers set the spread $\alpha_t^3 < 0$ wide at a large negative value, they also miss the orders from the other two types by the amount proportional to the spread, where the fourth type of traders (others), who do not have any objective function, take the rest of the orders the market makers missed. Thus, the market makers aim to maximize the trading profit from the market making by setting the spread optimally.