Appendix A: Literature Review on Modeling Consumer Behavioral Dynamics

Table EC.1 Literature Review on Modeling Consumer Behavioral Dynamics

Paper	Empirical Context	Model	Model Flexibility	Representation of the Event History	Representation Capacity	Possibility of Modeling the Concurrency of Multiple Behaviors
Xu et al. (2014)	Ad click and purchase behaviors in online advertising	Bayesian multivariate Hawkes process	Medium	N/A	N/A	The original form cannot directly model the concurrency; need to explicitly enumerate all possible behavior combinations
Schweidel and Moe (2016)	Binge-watching behavior in video streaming	Hazard and generalized linear model	Low	A number of covariates constructed from the event history	Low	Can possibly use a separate generalized linear model for each distinct behavior
Du et al. (2016); Mei and Eisner (2017)	Badge acquisition behaviors in online platforms	RNN-based neural point process	High	A single high-dimensional hidden state vector in RNN	Medium	The original form cannot directly model the concurrency; need to explicitly enumerate all possible behavior combinations
Dew and Ansari (2018)	Purchase behavior in video games	Gaussian process	High	A number of covariates constructed from the event history	Low	N/A
Huang et al. (2019)	Game-play behavior in video games	Discrete-time HMM	Medium	A single discrete latent state in HMM	Low	Can possibly use a separate latent-state-dependent hurdle model for each distinct behavior
Rutz et al. (2019)	Game-play behavior in mobile games	Generalized linear model	Low	A number of covariates constructed from the event history	Low	N/A
Jacobs et al. (2021)	Large-scale purchase behavior in retailing	Probabilistic hierarchical model	Medium	A single continuous latent vector in the hierarchical model	Medium	N/A
Liu et al. (2021)	Web search behavior in TV show search	Probabilistic hierarchical model	Medium	A single continuous latent vector in the hierarchical model	Medium	N/A
Dhillon and Aral (2021)	Consumption behavior of online news	RNN-based matrix factorization	High	A single high-dimensional hidden state vector in RNN	Medium	N/A
This paper	Game-play and purchase behaviors in video games	Attention-based neural point process	High	Multiple continuous-time high-dimensional representations	High	Can efficiently handle the concurrency by leveraging rich behavior-specific representations of the event history in hurdle models

Appendix B: Recurrent Neural Network (RNN)

RNN is a class of deep neural networks specifically designed for modeling and processing sequential data (Goodfellow et al. 2016). By recursively applying a nonlinear operation to update the state of hidden units with the current input and the previous hidden state, RNN is able to model the complex dynamics and thus has been successfully employed in many time-series-related tasks, such as speech recognition, machine translation, and financial forecasting. In most tasks, RNN is trained to encode certain task-relevant information from the past sequence of inputs in the hidden state, which then can be used to make future predictions.

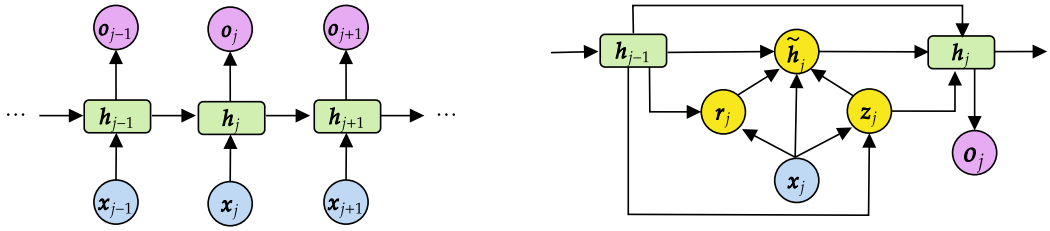
Figure EC.1a shows the network structure of a standard RNN. At each time step j, the network first applies a nonlinear transformation on the previous hidden state $\boldsymbol{h}_{j-1}$ and the current input $\boldsymbol{x}_j$ to update the hidden state $\boldsymbol{h}_j$, and then performs another transformation on $\boldsymbol{h}_j$ to give the output $\boldsymbol{o}_j$. The network structure is recurrent in the sense that the same nonlinear function is applied to update the hidden state at each time step and the parameters of the transformations are shared across all time steps. Formally, the hidden state $\boldsymbol{h}_j$ is updated as follows:

$$\boldsymbol{h}_j = \tanh(\boldsymbol{W}_l[\boldsymbol{h}_{j-1}, \boldsymbol{x}_j] + \boldsymbol{b}_l), \tag{EC.1}$$

where the vector $[\boldsymbol{h}_{j-1}, \boldsymbol{x}_j]$ is the concatenation of the previous hidden state $\boldsymbol{h}_{j-1}$ and the current input $\boldsymbol{x}_j$, $\boldsymbol{W}_l$ and $\boldsymbol{b}_l$ are the network parameters to learn (referred to as the weight matrix and bias vector respectively), and $\tanh(x) = (e^{2x}-1)/(e^{2x}+1)$ is the hyperbolic tangent activation function. Compared with other sequential models such as hidden Markov models, the nonlinearity of RNN makes the models more powerful and adaptive at learning from sequential data. Moreover, the hidden units in RNN are typically represented by continuous high-dimensional vectors, instead of a few discrete states as in most hidden Markov models, so they can store richer information of the underlying dynamics.

Training RNN requires computing the gradient of a certain loss function with respect to the network parameters $\boldsymbol{W}_l$ and $\boldsymbol{b}_l$. Because these parameters are shared in the recurrent structure, the gradient tends to quickly explode or vanish for long sequences since the same weight matrix $\boldsymbol{W}_l$ is repeatedly multiplied over many time steps in the backpropagation training algorithm. This is known as the "vanishing and exploding gradient problem": the vanishing gradient makes it difficult to effectively update the parameters, whereas the exploding gradient can cause the training procedure to be unstable (Goodfellow et al. 2016). Many variants of RNN, including the notable long short-term memory (LSTM, Hochreiter and Schmidhuber 1997) and gated recurrent unit (GRU, Cho et al. 2014), have been proposed to address this issue. The two variants often achieve comparable performance in various tasks of sequence

Figure EC.1 Illustration of the Standard Recurrent Neural Network and Gated Recurrent Unit



(a) Standard RNN (b) GRU

Note. (a) x_j, h_j, and o_j are the input, hidden state, and output, respectively. (b) z_j and r_j are the update and reset gates, and $\tilde{\boldsymbol{h}}_j$ is the "candidate" hidden state.

modeling (Chung et al. 2014), but GRU is a more streamlined variant that is significantly faster to train because it has fewer parameters. Figure EC.1b illustrates the network structure of a GRU. Compared with the standard RNN, GRU consists of two additional gating units, an update gate $\boldsymbol{z}_j$ and a reset gate $\boldsymbol{r}_j$, which are two vectors containing values between 0 and 1 responsible for modulating the flow of information across consecutive hidden units. Similar to Equation (EC.1), they are computed using the previous hidden state $\boldsymbol{h}_{j-1}$ and the current input $\boldsymbol{x}_j$:

$$\boldsymbol{z}_j = \sigma_S(\boldsymbol{W}_z[\boldsymbol{h}_{j-1}, \boldsymbol{x}_j] + \boldsymbol{b}_z), \qquad \boldsymbol{r}_j = \sigma_S(\boldsymbol{W}_r[\boldsymbol{h}_{j-1}, \boldsymbol{x}_j] + \boldsymbol{b}_r), \tag{EC.2}$$

where $\sigma_S(x) = e^x/(1+e^x)$ is the sigmoid function that ranges from 0 to 1, and $\boldsymbol{W}_z, \boldsymbol{b}_z, \boldsymbol{W}_r, \boldsymbol{b}_r$ are the gate parameters. The GRU also introduces a "candidate" hidden state $\tilde{\boldsymbol{h}}_j$, such that the hidden state $\boldsymbol{h}_j$ at time step j is a linear interpolation between the previous hidden state $\boldsymbol{h}_{j-1}$ and the "candidate" hidden state $\tilde{\boldsymbol{h}}_j$:

$$\boldsymbol{h}_j = (1 - \boldsymbol{z}_j) \circ \boldsymbol{h}_{j-1} + \boldsymbol{z}_j \circ \tilde{\boldsymbol{h}}_j. \tag{EC.3}$$

Here $\circ$ denotes the element-wise multiplication between two vectors, and the update gate $\boldsymbol{z}_j$ determines the portion of $\boldsymbol{h}_{j-1}$ to be retained and the portion of $\tilde{\boldsymbol{h}}_j$ to be used for forming $\boldsymbol{h}_j$. The "candidate" hidden state $\tilde{\boldsymbol{h}}_j$ is also computed based on the previous hidden state $\boldsymbol{h}_{j-1}$ and the current input $\boldsymbol{x}_j$, with the reset gate $\boldsymbol{r}_j$ deciding the amount of the past information to be used:

$$\tilde{\boldsymbol{h}}_j = \tanh\left(\boldsymbol{W}_h[\boldsymbol{r}_j \circ \boldsymbol{h}_{j-1}, \boldsymbol{x}_j] + \boldsymbol{b}_h\right). \tag{EC.4}$$

With this specification, the two gating units in GRU can be trained to selectively keep the relevant information from the past and add new information on top of it, which prevents the gradient from exploding or vanishing because of the addition operation in Equation (EC.3).

Appendix C: Classic Marked Point Processes

A marked point process is a stochastic process of events $\{(t_j, \boldsymbol{m}_j)\}_{j=1}^N$, where t_j is the occurrence time of the j-th event and $\boldsymbol{m}_j$ is the associated vector of marks. Most marked point processes consist of a ground process, the ordinary point processes of event time $\{t_j\}_{j=1}^N$ characterized by its conditional density function (Daley and Vere-Jones 2003)

$$\lambda^*(t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}(N(t+\Delta t) - N(t) > 0 \mid \mathcal{H}_t)}{\Delta t}, \tag{EC.5}$$

where $\mathcal{H}_t = \{(t_j, \boldsymbol{m}_j) \mid \forall j : t_j \leq t\}$ is the event history up to time t and $N(t)$ is the associated counting process that counts the number of events up to time t, i.e., $N(t) = \sum_{j=1}^N \mathbb{1}(t_j \leq t)$. The *ground intensity function* $\lambda^*(t) \triangleq \lambda(t \mid \mathcal{H}_t)$ specifies the instantaneous occurrence rate of events conditioned on the event history $\mathcal{H}_t$.

Compared with ordinary point processes (e.g., Poisson processes), which are only able to model the occurrence time of events, marked point processes allow for the modeling of additional characteristics (known as marks) of an event. To achieve such flexibility, marked point processes define the *mark density function* $f^*(\boldsymbol{m} \mid t) \triangleq f(\boldsymbol{m} \mid t, \mathcal{H}_t)$ to characterize the generative mechanism of observed event marks at time t given the event history $\mathcal{H}_t$. A special case is the multivariate point process with categorical marks, in which each category often represents a distinct event type, such as the type of online advertisement being clicked (Xu et al. 2014). Event marks can also be ordinal variables, e.g., the magnitude of an earthquake (Ogata 1988) and the follower count of a retweeting account (Chen and Tan 2018). In our empirical context of video games, marks of each in-game activity correspond to the game-play duration and purchase count (Figure 1).

Most classic marked point processes require strong parametric specifications about the functional forms of $\lambda^*(t)$. For example, multivariate Hawkes process (MHP, Hawkes 1971a,b), or mutually exciting point process, is a multivariate point process composed of K marginal self-exciting point processes $\{N_k(t)\}_{k=1}^K$. The ground intensity function of the MHP model is specified as the sum of K marginal intensity functions, i.e., $\lambda^*(t) = \sum_{k=1}^K \lambda_k^*(t)$, and the k-th marginal intensity is defined as

$$\lambda_k^*(t) = \mu_k + \sum_{k'=1}^K \sum_{j=1}^{N_{k'}(t)} \alpha_{k'k} e^{-\beta_{k'k}(t - t_{k'j})}, \tag{EC.6}$$

where $\mu_k > 0$ is the baseline intensity of event type k, $\alpha_{k'k} > 0$ is the degree of exciting effect from an event of type k' to an event of type k, and $\beta_{k'k} > 0$ is the decaying rate of this effect. With this specification, the MHP model assumes that all the past events have an *additive* and *stimulating* effect on the occurrence of future events, and such influence is *exponentially decaying* over time. Such assumption might be reasonable in certain application scenarios, such as the

dynamic interactions among various types of advertisement clicks and purchases in the context of e-commerce (Xu et al. 2014) and the influence of past idea acquisition on future idea generation (Aggarwal et al. 2021). However, we believe that it can lead to model misspecification in the context of digital content consumption, e.g., video game players' past in-game experiences are not guaranteed to increase their future engagement level, as discussed in Section 2.2.

Moreover, many classic marked point processes often assume the independence of event marks and event history (Rasmussen 2018), i.e., $f(\boldsymbol{m} \mid t, \mathcal{H}_t) = f(\boldsymbol{m} \mid t)$. Though it can significantly reduce the model complexity, such a simplification renders marked point processes unable to predict random marks of interest based on past information, making the model much less realistic and useful. Some marked point processes may not assume the independence of event marks and event history, but they adopt a strong parametric form of the mark density function $f^*(\boldsymbol{m} \mid t)$. For example, the mark density function in the MHP model assumes that, given the occurrence time of the next event, the probability of the next event type is proportional to its corresponding marginal intensity, i.e.,

$$f^*(\boldsymbol{m} = k \mid t) = \frac{\lambda_k^*(t)}{\sum_{k'=1}^{K} \lambda_{k'}^*(t)} \tag{EC.7}$$

Appendix D: Monte Carlo Algorithm for Evaluating the Likelihood Function

Recall from Section 4.5 that the likelihood function in our proposed model can be written as:

$$L = \prod_{i=1}^{I} \left[\prod_{j:c_j^i=1}^{N_i} \lambda_i^*(t_j^i) \prod_{k \in \{\text{Play, Pur}\}} f_{i,k}^*(m_{j,k}^i \mid t_j^i) \right] \cdot \exp\left(- \int_0^T \lambda_i^*(\tau) d\tau \right), \tag{EC.8}$$

where $\lambda_i^*(t)$ is the ground intensity function of the player i. The stochastic integration term $\wedge_i = \int_0^T \lambda_i^*(\tau) d\tau$ in the likelihood can be estimated by a Monte Carlo unbiased estimator

$$\hat{\wedge}_i = \frac{T}{N} \sum_{j=1}^{N} \lambda_i^*(t_j), \tag{EC.9}$$

where N is the number of Monte Carlo samples and $t_j \sim \text{Unif}(0, T), j = 1, \ldots, N$. In practice, it is preferable to choose a relatively large N to reduce the variance of the estimator. Although selecting a large N would lead to a higher computational cost, we only need to conduct the Monte Carlo evaluation of $\hat{\wedge}_i$ for each player i once in an optimization step. Therefore, the final computational cost would not increase much if the number of Monte Carlo samples is chosen to be much smaller than the total number of players, i.e., $N \ll I$. We choose $N = 100$ for our dataset with $I = 2,020$ players.

Appendix E: Simulation Algorithm for Our Attention-Based Neural Point Process

We extend Ogata's thinning algorithm (Ogata 1981) and develop a simulation algorithm to generate data from our attention-based neural point process model. Applying it to our empirical context of video games can simulate the occurrence time, behavior combination (i.e., login only, game-play only, purchase only, and concurrent game-play and purchase), and consumption quantity (i.e., game-play duration and purchase count) of in-game activities under the influence of real-world sports matches. Below we provide the detailed algorithm to simulate player i's in-game activities between the activation time t_0^i and the ending time T.

1. **Input**: Activation time t_0^i, ending time T, occurrence time $\{t_j\}_{j=1}^{N_E}$ and marks $\{\boldsymbol{m}_j\}_{j=1}^{N_E}$ of real-world sports matches (ordered in time), and the model parameters $\boldsymbol{\Theta}$.
2. Initialize $n=0, j_E=\arg\min\{j\in\{1,\cdots,N_E\}: t_{j_E}>t_0^i\}, t=t_0^i, \mathcal{S}_i=\varnothing, \bar{\lambda}_i=\sigma_{\text{SP}}(b_{\text{Time}}^i)$.
3. Repeat until $t>T$:
   (a) Simulate $\tau\sim\text{Exponential}(\bar{\lambda}_i)$. Update $t=t+\tau$.
   (b) If $t<T$, calculate $\lambda_i^*(t)=\sigma_{\text{SP}}\left(\boldsymbol{v}_{\text{Time}}^\top \boldsymbol{s}_{\text{Time}}^i(t)+b_{\text{Time}}^i\right)$ where $\boldsymbol{s}_{\text{Time}}^i(t)$ is obtained from Equations (1) to (4) with $\mathcal{S}_i$ as the input. Simulate $u\sim\text{Uniform}(0,1)$. If $u\le\lambda_i^*(t)/\bar{\lambda}_i$:
      (i) If $t\ge t_{j_E}$ and $j_E\le N_E$, then repeat the following steps until $t<t_{j_E}$ or $j_E>N_E$ or $t_{j_E}\ge T$:
         (A) Let $t=t_{j_E}, c=2, \boldsymbol{m}=\boldsymbol{m}_{j_E}$.
         (B) Update $n=n+1, j_E=j_E+1, \mathcal{S}_i=\mathcal{S}_i\cup\{(t,c,\boldsymbol{m})\}$.
         (C) Calculate $\boldsymbol{h}_n^i$ according to Equations (1) and (2) with $\mathcal{S}_i$ as the input and update $\bar{\lambda}_i=\max\{\bar{\lambda}_i,\sigma_{\text{SP}}\left(\boldsymbol{v}_{\text{Time}}^\top\boldsymbol{h}_n^i+b_{\text{Time}}^i\right)\}$.
         (D) Simulate $\tau\sim\text{Exponential}(\bar{\lambda}_i)$. Update $t=t+\tau$. Simulate $u\sim\text{Uniform}(0,1)$. Repeat these steps until $t\ge T$ or $u\le\lambda_i^*(t)/\bar{\lambda}_i$. Here $\lambda_i^*(t)=\sigma_{\text{SP}}\left(\boldsymbol{v}_{\text{Time}}^\top\boldsymbol{s}_{\text{Time}}^i(t)+b_{\text{Time}}^i\right)$ where $\boldsymbol{s}_{\text{Time}}^i(t)$ is obtained from Equations (1) to (4) with $\mathcal{S}_i$ as the input.
      (ii) If $t<t_{j_E}$ or $j_E>N_E$ or $t_{j_E}\ge T$, and if $t<T$:
         (A) Let $c=1$. Simulate consumption quantity $\boldsymbol{m}$ at time t according to Equations (7) to (12).
         (B) Update $n=n+1, \mathcal{S}_i=\mathcal{S}_i\cup\{(t,c,\boldsymbol{m})\}$.
         (C) Calculate $\boldsymbol{h}_n^i$ according to Equations (1) and (2) with $\mathcal{S}_i$ as the input and update $\bar{\lambda}_i=\max\{\bar{\lambda}_i,\sigma_{\text{SP}}\left(\boldsymbol{v}_{\text{Time}}^\top\boldsymbol{h}_n^i+b_{\text{Time}}^i\right)\}$.
4. **Output**: $\mathcal{S}_i=\{(t_j,c_j,\boldsymbol{m}_j)\}_{j=1}^n$

To demonstrate the effectiveness of our simulation algorithm, we first train our model on the original sports video game dataset introduced in Section 3. Then, using the estimated model parameters $\widehat{\boldsymbol{\Theta}}$ and the same real-world sports matches, we simulate the entire sequences of in-game activities for $I=2,020$ players (the same sample size as the

Table EC.2 Comparison of Summary Statistics in Simulated and Original In-Game Activities

Aggregate Level	Variable	Simulated Mean	Original Mean	Simulated SD	Original SD
Event	Game-Play duration (in hours)	1.90	1.87	2.13	2.28
	Purchase count	0.49	0.44	1.25	1.21
Sequence	Total number of in-game activities	78.49	71.04	41.35	42.45
	Total game-play duration (in hours)	167.90	133.20	131.65	116.28
	Total purchase count	36.04	31.70	59.24	43.2

Note. Number of sequences/players $I = 2,020$ for both the simulated and original sports video game dataset.

original dataset), with each simulated sequence starting from the same activation time as in its original sequence. Table EC.2 compares the summary statistics of the simulated and original in-game activities at both event and sequence levels (see also Table 1 and Table 2 in Section 3). At the event level, the simulated and original in-game activities show significant similarity in terms of game-play duration and purchase count. At the sequence level, the total number of in-game activities, the cumulative game-play duration, and the total purchase count, are also similar in the simulated and original sequences.

Appendix F: Evaluating the Estimation Procedure

Because of the sheer amount of parameters in our attention-based neural point process model, it is impractical to validate the estimation accuracy of all parameters as in a traditional parametric model. Therefore, instead of examining each parameter separately, we conduct simulation studies to evaluate, both qualitatively and quantitatively, whether our estimation procedure in Section 4.5 can accurately recover the ground intensity functions.

F.1. Qualitative Evidence: Intensity Visualization

We first use the original sports video game dataset to obtain the estimated model parameters $\widehat{\mathbf{\Theta}}$. Next, we apply the simulation algorithm in Appendix E to simulate $I = 2,020$ sequences of in-game activities (the same sample size as the original dataset) from our attention-based neural point process model with the parameters $\widehat{\mathbf{\Theta}}$. We then re-estimate the model parameters, denoted as $\widetilde{\mathbf{\Theta}}$, based on the simulated sequences. If our estimation procedure is accurate, the re-estimated intensities $\lambda(t; \widetilde{\mathbf{\Theta}})$ evaluated with $\widetilde{\mathbf{\Theta}}$ are expected to be close to the target ground intensities $\lambda(t; \widehat{\mathbf{\Theta}})$. Figure EC.2 shows a visual comparison of the target and re-estimated intensity functions for 16 randomly selected players in a representative run of the simulation, which provides qualitative evidence that our estimation procedure can successfully recover the target ground intensities for those players, despite a significant difference in their trends.

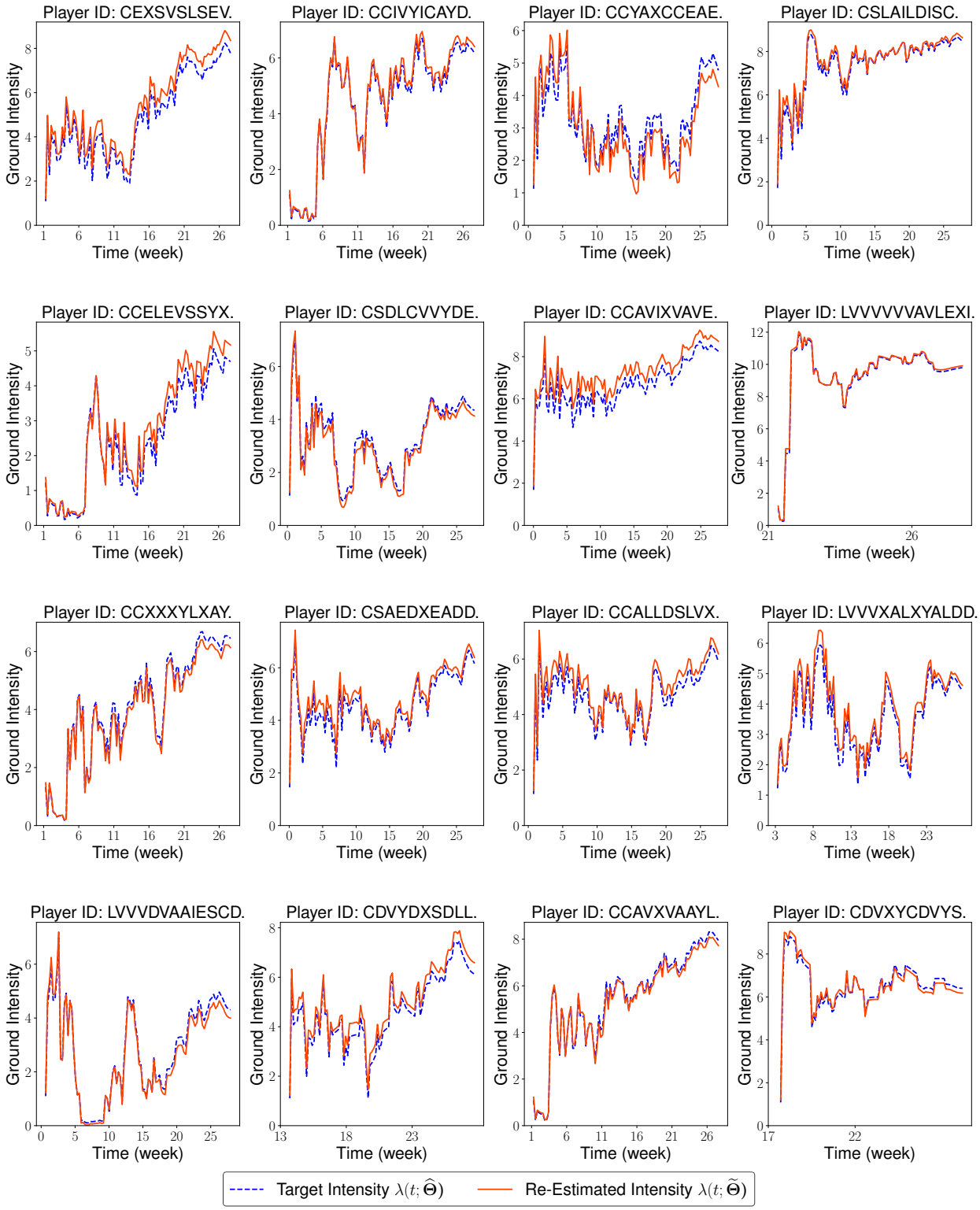
F.2. Quantitative Evidence: Residual Analysis

To further quantitatively validate the accuracy of our estimation procedure, we repeat the aforementioned simulation and re-estimation procedure in Appendix F.1 for 50 independent runs and obtain 50 different re-estimated parameter values $\{\widetilde{\boldsymbol{\Theta}}_1, \widetilde{\boldsymbol{\Theta}}_2, \ldots, \widetilde{\boldsymbol{\Theta}}_{50}\}$. We then apply the residual analysis approach proposed by Ogata (1988) to examine whether the 50 re-estimated ground intensities $\{\lambda(t; \widetilde{\boldsymbol{\Theta}}_l), l = 1, \ldots, 50\}$ are quantitatively close to the target intensities $\lambda(t; \widehat{\boldsymbol{\Theta}})$. Specifically, according to Papangelou (1972), after we use the target intensities $\lambda(t; \widehat{\boldsymbol{\Theta}})$ to simulate the occurrence time $\{t_j\}$ of in-game activities, the transformed time $\{\wedge(t_j; \widehat{\boldsymbol{\Theta}}) := \int_0^{t_j} \lambda(\tau; \widehat{\boldsymbol{\Theta}}) d\tau\}$ follow the distribution of a stationary Poisson process with intensity 1. Therefore, if the re-estimated intensities $\lambda(t; \widetilde{\boldsymbol{\Theta}}_l)$ are a good approximation to the target intensities $\lambda(t; \widehat{\boldsymbol{\Theta}})$, it is expected that the transformation of $\{t_j\}$ under the re-estimated intensities $\{\tau_{jl} = \wedge(t_j; \widetilde{\boldsymbol{\Theta}}_l) := \int_0^{t_j} \lambda(\tau; \widetilde{\boldsymbol{\Theta}}_l) d\tau\}$ would also behave like a stationary Poisson process and the inter-activity time $\{\tau_{(j+1)l} - \tau_{jl}\}$ are i.i.d. distributed as Exponential(1). Following this rationale behind the residual analysis, we use each re-estimated ground intensities $\lambda(t; \widetilde{\boldsymbol{\Theta}}_l)$ to obtain the transformed time $\{\tau_{jl}\}$, and then apply the Kolmogorov-Smirnov goodness-of-fit test to examine the distribution of the inter-activity time in each run of the 50 independent simulations. The test results show no statistical evidence that the distribution of the inter-activity time deviates from an exponential distribution of rate 1 (all p-values > 0.3). As a result, we conclude that the re-estimated intensities $\{\lambda(t; \widetilde{\boldsymbol{\Theta}}_l), l = 1, \ldots, 50\}$ accurately approximate the target intensities $\lambda(t; \widehat{\boldsymbol{\Theta}})$ and that our estimation procedure works very effectively in practice.

Appendix G: More Evaluation Results

To demonstrate the performance of our model in predicting the behaviors of out-of-sample players, we split the sports video game dataset introduced in Section 3 by individuals so that the training, validation, and test sets include 60%, 20%, and 20% of total players. Then, we train our model on the training set, use the validation set to select the optimal values of hyperparameters, and evaluate our model on the test set. To apply the trained model to players in the validation and test sets, we use the median values of player-specific parameters (see Equations (6), (7), (9), and (11)) calculated from players in the training set as their player-specific parameters. Table EC.3 shows that the two different model evaluation strategies, i.e., splitting the data by time (see Section 5 for more details) and splitting the data by players, lead to qualitatively similar conclusions about the overall predictive performance of our model.

Figure EC.2 Comparison of the Target and Re-Estimated Ground Intensity Functions



Note. Blue dotted lines: the target ground intensity functions $\lambda(t;\widehat{\boldsymbol{\Theta}})$ are evaluated with the model parameters $\widehat{\boldsymbol{\Theta}}$ estimated from the original sports video game dataset. Red solid lines: the re-estimated ground intensity functions $\lambda(t;\widetilde{\boldsymbol{\Theta}})$ are evaluated with the model parameters $\widetilde{\boldsymbol{\Theta}}$ estimated from the simulated datasets. The simulated datasets are generated from our point process with the model parameters $\widehat{\boldsymbol{\Theta}}$, using the simulation algorithm in Appendix E.

Table EC.3 Model Evaluation Based on Splitting the Data by Time vs. Splitting Data by Players

Splitting Strategy	Occurrence Time		Behavior Combination		Game-Play Duration		Purchase Count	
	RMSE	MAE	Micro-F1	Macro-F1	RMSE	MAE	RMSE	MAE
By Time	0.42	0.35	0.88	0.62	2.05	1.30	1.25	0.45
By Players	0.45	0.37	0.87	0.58	2.12	1.35	1.27	0.48

We also examine how the performance of our model would evolve as we have more observations of in-game activities. To this end, we vary the proportion of data used to train our model from 60% to 80%, and the proportions of data used to validate and evaluate our model from 20% to 5%, respectively. Table EC.4 indicates that the out-of-sample predictive performance of our model improves with increasing number of in-game activities in the training data. This result is consistent with the literature showing that RNN-based models tend to work better with more training data.

Table EC.4 Out-of-Sample Predictive Performance for Different Proportions of Training Data

Proportion of	Occurrence Time		Behavior Combination		Game-Play Duration		Purchase Count	
Training Data	RMSE	MAE	Micro-F1	Macro-F1	RMSE	MAE	RMSE	MAE
60%	0.42	0.35	0.88	0.62	2.05	1.30	1.25	0.45
70%	0.40	0.33	0.88	0.63	2.00	1.27	1.20	0.43
80%	0.39	0.32	0.89	0.68	1.97	1.25	1.18	0.40

Appendix H: Effects of Event Occurrence on Players' Future Engagement

For each event $(t_j^i, c_j^i, \boldsymbol{m}_j^i)$ in player i's event sequence, we consider two scenarios of his event history up to time t_j^i: (1) the "actual" scenario $\mathcal{H}_{t_j^i}^i = \{(t_j^i, c_j^i, \boldsymbol{m}_j^i) \mid \forall j : t_j \le t_j^i\}$, and (2) the "hypothetical" scenario $\widetilde{\mathcal{H}}_{t_j^i}^i = \{(t_j^i, c_j^i, \boldsymbol{m}_j^i) \mid \forall j : t_j < t_j^i\}$ that includes all the events *before* time t_j^i and thus assumes the j-th event did not occur in the history. We examine *the effects of this event occurrence on the player's future engagement* by measuring the differences in the player's subsequent (a) login frequency (ΔLoginFreq), (b) game-play duration (ΔGameDur), and (c) purchase count (ΔPurCnt) within a period of Δt between the two scenarios:

$$
\begin{aligned}
\Delta\text{LoginFreq}(t_j^i, c_j^i, \boldsymbol{m}_j^i) &= \mathbb{E}\left[N_i(t_j^i + \Delta t) \middle| \mathcal{H}_{t_j^i}^i \right] - \mathbb{E}\left[\widetilde{N}_i(t_j^i + \Delta t) \middle| \widetilde{\mathcal{H}}_{t_j^i}^i \right], \\
\Delta\text{GameDur}(t_j^i, c_j^i, \boldsymbol{m}_j^i) &= \mathbb{E}\left[\sum_{l=N_i(t_j^i)+1}^{N_i(t_j^i+\Delta t)} m_{l,\text{Play}}^i \middle| \mathcal{H}_{t_j^i}^i \right] - \mathbb{E}\left[\sum_{l=\widetilde{N}_i(t_j^i)+1}^{\widetilde{N}_i(t_j^i+\Delta t)} m_{l,\text{Play}}^i \middle| \widetilde{\mathcal{H}}_{t_j^i}^i \right], \\
\Delta\text{PurCnt}(t_j^i, c_j^i, \boldsymbol{m}_j^i) &= \mathbb{E}\left[\sum_{l=N_i(t_j^i)+1}^{N_i(t_j^i+\Delta t)} m_{l,\text{Pur}}^i \middle| \mathcal{H}_{t_j^i}^i \right] - \mathbb{E}\left[\sum_{l=\widetilde{N}_i(t_j^i)+1}^{\widetilde{N}_i(t_j^i+\Delta t)} m_{l,\text{Pur}}^i \middle| \widetilde{\mathcal{H}}_{t_j^i}^i \right].
\end{aligned}
\tag{EC.10}
$$

Here, $N_i(t)$ and $\widetilde{N}_i(t)$ are the corresponding counting processes for player i's in-game activities under the "actual" and "hypothetical" scenarios, respectively. As such, the expressions in Equation (EC.10) directly measure the changes in player i's login frequency, game-play duration, and purchase count within a period of Δt, incurred by the occurrence of his j-th event in the history. We choose Δt to be a week to match the unit of time used in our model training and evaluation (see Section 4).

Because of the complex nature of our attention-based neural point processes, the exact computation of expectations in Equation (EC.10) is intractable. To address this challenge, we first use the simulation algorithm in Appendix E to simulate $R = 10,000$ sequences of in-game activities under each scenario and then apply the Monte Carlo method to numerically approximate the expected differences in Equation (EC.10):

$$\begin{aligned}
\Delta\text{LoginFreq}(t_j^i, c_j^i, \boldsymbol{m}_j^i) &\approx \frac{1}{R}\sum_{r=1}^{R}\left[N_i^{(r)}(t_j^i+\Delta t)\right] - \frac{1}{R}\sum_{r=1}^{R}\left[\widetilde{N}_i^{(r)}(t_j^i+\Delta t)\right], \\
\Delta\text{GameDur}(t_j^i, c_j^i, \boldsymbol{m}_j^i) &\approx \frac{1}{R}\sum_{r=1}^{R}\sum_{l=N_i^{(r)}(t_j^i)+1}^{N_i^{(r)}(t_j^i+\Delta t)} m_{l,\text{Play}}^{i(r)} - \frac{1}{R}\sum_{r=1}^{R}\sum_{l=\widetilde{N}_i^{(r)}(t_j^i)+1}^{\widetilde{N}_i^{(r)}(t_j^i+\Delta t)} m_{l,\text{Play}}^{i(r)}, \\
\Delta\text{PurCnt}(t_j^i, c_j^i, \boldsymbol{m}_j^i) &\approx \frac{1}{R}\sum_{r=1}^{R}\sum_{l=N_i^{(r)}(t_j^i)+1}^{N_i^{(r)}(t_j^i+\Delta t)} m_{l,\text{Pur}}^{i(r)} - \frac{1}{R}\sum_{r=1}^{R}\sum_{l=\widetilde{N}_i^{(r)}(t_j^i)+1}^{\widetilde{N}_i^{(r)}(t_j^i+\Delta t)} m_{l,\text{Pur}}^{i(r)}.
\end{aligned} \tag{EC.11}$$

Appendix I: Effects of Event Marks on Players' Future Engagement

We conduct regression analysis to estimate the effects of event marks (i.e., the game-play duration $m_{j,\text{Play}}^i$ and purchase count $m_{j,\text{Pur}}^i$ for in-game activities; the number of highlights $m_{j,\text{Highlight}}^i$ and absolute score difference $m_{j,\text{Score}}^i$ for sports matches) on players' future engagement. The dependent variables are the outcomes generated by the simulations in Section 6.2 (see also Appendix H). Considering that players' unobserved characteristics may drive their behaviors over time, we apply fixed-effects models to account for player heterogeneity. The model specifications are shown below:

$$\Delta Y_j^i = \begin{cases} \delta_i + \eta_1^\top m_{j,\text{Play}}^i + \eta_2^\top m_{j,\text{Pur}}^i + \boldsymbol{\zeta}^\top \boldsymbol{\kappa}_j^i + \epsilon_j^i, & c_j^i = 1; \\ \tilde{\delta}_i + \tilde{\eta}_1^\top m_{j,\text{Highlight}}^i + \tilde{\eta}_2^\top m_{j,\text{Score}}^i + \tilde{\boldsymbol{\zeta}}^\top \boldsymbol{\kappa}_j^i + \tilde{\epsilon}_j^i, & c_j^i = 2, \end{cases} \tag{EC.12}$$

Here ΔY_j^i corresponds to the estimated changes in player i's login frequency ($\Delta\text{LoginFreq}$), or game-play duration ($\Delta\text{GameDur}$), or purchase count (ΔPurCnt), incurred by the occurrence of his j-th event in the history. $\boldsymbol{\kappa}_j^i$ denotes the control variables, including player experience measured by the number of weeks since the player's first-time login to the game, and monthly and day-of-the-week (DoW) dummies. Since in-game activities ($c_j^i = 1$) and sports matches ($c_j^i = 2$)) are characterized by different event marks, we perform two separate regressions to estimate their effects or coefficients: (η_1, η_2) for the game-play duration and purchase count, and $(\tilde{\eta}_1, \tilde{\eta}_2)$ for the number of highlights and

absolute score difference. All the event marks are standardized, so the estimated coefficients corresponding to the same dependent variable are comparable.

Columns (1)-(3) in Table EC.5 present the estimated effects of the game-play duration and purchase count. All the coefficients are significant and positive, and their magnitudes exhibit several patterns. First, players' future behaviors are more affected by past consumption behaviors of the same type, suggesting the existence of behavioral inertia over time. For instance, the effect of past game-play duration is significantly stronger than past purchase count on future game-play (0.442 vs. 0.190, F-stat. = 11,391, p-value < 0.001). Similarly, the impact of past purchase count is significantly larger than past game-play duration on subsequent purchases (0.233 vs. 0.009, F-stat. = 16,819, p-value < 0.001). Second, purchase count, on average, has a greater influence than game-play duration on subsequent login frequency (0.051 vs. 0.043, F-stat. = 21.239, p-value < 0.001). This is probably because in-game purchases are primarily for enhancing the game-play experience, so players who have made purchases are more likely to log in again to use the enhancement packs. Moreover, the results show that the effect of past purchases on future game-play tends to be larger than that of past game-play on future purchases (0.190 vs. 0.009).

Table EC.5 Effects of Event Marks on Players' Future Engagement

	In-game Activities			Sports Matches		
	(1)ΔLoginFreq	(2)ΔGameDur	(3)ΔPurCnt	(4)ΔLoginFreq	(5)ΔGameDur	(6)ΔPurCnt
Game-play duration	0.043^{***}	0.442^{***}	0.009^{***}			
	(0.002)	(0.014)	(0.001)			
In-game purchase count	0.051^{***}	0.190^{***}	0.233^{***}			
	(0.002)	(0.009)	(0.005)			
Number of highlights				0.020^{***}	0.053^{***}	0.031^{***}
				(0.001)	(0.003)	(0.001)
Abs. score difference				0.007^{***}	0.041^{***}	0.010^{***}
				(0.001)	(0.004)	(0.001)
Player fixed	✓	✓	✓	✓	✓	✓
Player experience controlled	✓	✓	✓	✓	✓	✓
Month and DoW dummies	✓	✓	✓	✓	✓	✓
Adjusted R^2	0.117	0.104	0.225	0.325	0.142	0.099

Note. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. Robust clustered error terms are reported in the parentheses. Shown in the columns labeled ΔLoginFreq, ΔGameDur, and ΔPurCnt are the estimated effects of event marks on future login frequency, game-play duration, and purchase count, respectively.

Columns (4)-(6) in Table EC.5 report the estimated effects of the number of highlights and absolute score difference. As discussed in Section 2.2, real-world sports matches can act as external stimuli to "prime" players' affective feelings and cognitive impressions, and in turn, to affect in-game activities. The results show that both event marks

of sports matches have significantly positive effects on players' future engagement (all p-values < 0.001). Moreover, in comparison to the effect of the absolute score difference, the effect of the number of highlights is 185.7% larger on login frequency (0.020 vs. 0.007, F-stat. = 106, p-value < 0.001), 29.3% larger on game-play (0.053 vs. 0.041, F-stat. = 3.77, p-value < 0.05), and 210.0% larger on purchase count (0.031 vs. 0.010, F-stat. = 181, p-value < 0.001). Therefore, the video game firm can motivate greater player engagement by making good use of the external stimuli and, in particular, highlighting the exciting moments therein. For example, to attract more potential players, the firm can advertise in athletic fields and stadiums. The audience who have experienced exciting moments during the matches are likely to seek to revisit their feelings through the virtual video game. The firm can also advertise on social media by posting content of exciting moments in sports matches to activate potential players' desire to play the linked video game. To motivate existing players, especially experienced ones who are more susceptible to the influence of external sports matches, the firm can send notifications or newsletters reminding them of exciting moments from recent matches.

We also examine how the effects of event marks on players' future engagement vary with the four player segments. Two dummy variables, *HighPlay* and *HighPurchase*, are created to indicate whether a player's game-play and purchase propensities are above the median level. We then include the interaction terms between these two dummy variables and event marks in the regression models. All the other model specifications are the same as those in the aforementioned main effect analysis.

Columns (1)-(3) in Table EC.6 present the estimated effects of the game-play duration and purchase count across player segments. Column (1) shows that the interaction effects of the propensity dummies and event marks (the game-play duration and purchase count) on future login frequency are all negative, with p-values lower than 0.05. This suggests that past in-game activities have the highest effects on subsequent logins for the players in the reference group, i.e., the "casual" segment whose game-play and purchase propensities are below the median level. By contrast, the login decisions of players in the other segments are less responsive to past in-game activities. On the other side, Column (2) shows that the positive impact of past game-play duration on future game-play engagement is greater among players with a high game-play propensity (Coef. = 0.054, p-value < 0.05). These players might be more addicted to game-play and thus exhibit stronger susceptibility to previous playing activities. Similarly, Column (3) shows that the positive influence of past purchase count on future purchase behavior is enhanced among players with a high purchase propensity (Coef. = 0.078, p-value < 0.001). Probably due to their lower price sensitivity and greater willingness to pay, these players are more likely to continue purchasing enhancement packs to enrich their game experience. In

addition, the result in Column (2) shows that the positive effects of past purchase count on future game-play are strengthened among players with either a high game-play or a high purchase propensity. The interactions between purchase count and the two propensity dummies are both significant (Coef. = 0.055 and 0.075, p-values< 0.001). This result suggests that past purchases are effective in encouraging future game-play for players who intrinsically like playing or purchasing in the video game.

Columns (4)-(6) in Table EC.6 report the estimated effects of event marks in sports matches across player segments. The result shows that the effects of the number of highlights and absolute score difference are relatively stable and would not vary across player segments in most cases. In Column (4), the interaction effects of both event marks and propensity dummies on subsequent logins appear insignificant. In Columns (5) and (6), the effect of the number of highlights on future game-play is stronger among players with a high game-play propensity (Coef. = 0.027, p-value < 0.001), and the effect of the number of highlights on future purchases increases when players have a high purchase propensity (Coef. = 0.035, p-value < 0.001). However, we cannot find significant interaction effects of the score difference and propensity dummies.

Given the varying effects of event marks across player segments, marketing practitioners can customize their strategies to better engage players in video games. Overall, the results suggest that past consumption behaviors could stimulate the same type of behavior in the future, and such effects are particularly stronger among players who intrinsically like the behavior. As such, "casual" players should be the main target group for marketing practitioners to improve their engagement, because their future game-play and purchases are relatively less susceptible to past in-game activities. The video game firm can adjust the difficulty level of the game so that "casual" players can win more easily to increase their feelings of efficacy and pleasure. The firm can also introduce more functions to raise "casual" players' curiosity to keep playing the game. To encourage them to make purchases, the firm can offer larger product discounts and promotions. In comparison, for players with high behavioral propensities, the video game firm can focus on improving their game satisfaction and customizing their in-game experiences. For example, the firm can recommend more enhancement packs to "hardcore" and "buyer" groups, or create more challenges for "hardcore" and "gamer" segments, which may also increase their need to make in-game purchases for enhancing their ability to level up. In addition to players' in-game activities, real-world sports matches are a practical resource that can be leveraged to motivate player engagement in related video games. Compared with the score difference, exciting moments in sports matches show more effectiveness in triggering players' reactions and inspiring those who intrinsically like playing or purchasing to do more in the future.

Table EC.6 Effects of Event Marks on Players' Future Engagement Across Player Segments

	In-game Activities				Sports Matches		
	(1)ΔLoginFreq	(2)ΔGameDur	(3)ΔPurCnt		(4)ΔLoginFreq	(5)ΔGameDur	(6)ΔPurCnt
Game-play duration	0.066^{***}	0.390^{***}	0.011^{***}	Number of highlights	0.019^{***}	0.035^{***}	0.013^{***}
	(0.004)	(0.022)	(0.003)		(0.001)	(0.004)	(0.001)
× HighPlay	-0.025^{***}	0.054^{*}	-0.005	× HighPlay	0.000	0.027^{***}	0.004
	(0.004)	(0.025)	(0.003)		(0.001)	(0.006)	(0.002)
× HighPurchase	-0.009^{**}	0.018	0.003	× HighPurchase	0.000	0.010	0.035^{***}
	(0.004)	(0.027)	(0.003)		(0.001)	(0.006)	(0.002)
Purchase count	0.082^{***}	0.218^{***}	0.176^{***}	Abs. score difference	0.007^{***}	0.037^{***}	0.011^{***}
	(0.006)	(0.019)	(0.011)		(0.001)	(0.005)	(0.001)
× HighPlay	-0.012^{*}	0.055^{***}	-0.007	× HighPlay	0.002	0.006	0.000
	(0.005)	(0.016)	(0.010)		(0.002)	(0.007)	(0.002)
× HighPurchase	-0.031^{***}	0.075^{***}	0.078^{***}	× HighPurchase	-0.003	0.002	-0.001
	(0.006)	(0.021)	(0.011)		(0.002)	(0.007)	(0.002)
Player fixed	✓	✓	✓		✓	✓	✓
Player experience controlled	✓	✓	✓		✓	✓	✓
Month and DoW dummies	✓	✓	✓		✓	✓	✓
Adjusted R^2	0.119	0.108	0.233		0.326	0.154	0.116

Note. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. Robust clustered error terms are reported in the parentheses. Shown in the columns labeled ΔLoginFreq, ΔGameDur, and ΔPurCnt are the estimated effects of event marks (and their interaction effects with propensity dummies) on future login frequency, game-play duration, and purchase count, respectively.

References

Aggarwal V, Hwang EH, Tan Y (2021) Learning to be creative: A mutually exciting spatiotemporal point process model for idea generation in open innovation. *Information Systems Research* 32(4):1214—-1235.

Chen F, Tan WH (2018) Marked self-exciting point process modelling of information diffusion on twitter. *The Annals of Applied Statistics* 12(4):2175–2196.

Cho K, van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y (2014) Learning phrase representations using RNN encoder–decoder for statistical machine translation. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 1724–1734.

Chung J, Gulcehre C, Cho K, Bengio Y (2014) Empirical evaluation of gated recurrent neural networks on sequence modeling. *NIPS 2014 Deep Learning and Representation Learning Workshop*, URL http://arxiv.org/abs/1412.3555.

Daley DJ, Vere-Jones D (2003) *An Introduction to the Theory of Point Processes. Vol. I. Probability and Its Applications* (Springer).

Dew R, Ansari A (2018) Bayesian nonparametric customer base analysis with model-based visualizations. *Marketing Science* 37(2):216–235.

Dhillon PS, Aral S (2021) Modeling dynamic user interests: A neural matrix factorization approach. *Marketing science* 40(6):1059–1080.

Du N, Dai H, Trivedi R, Upadhyay U, Gomez-Rodriguez M, Song L (2016) Recurrent marked temporal point processes: Embedding event history to vector. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1555–1564.

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

Hawkes AG (1971a) Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society: Series B (Methodological)* 33(3):438–443.

Hawkes AG (1971b) Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58(1):83–90.

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

Huang Y, Jasin S, Manchanda P (2019) "Level up": Leveraging skill and engagement to maximize player game-play in online video games. *Information Systems Research* 30(3):927–947.

Jacobs B, Fok D, Donkers B (2021) Understanding large-scale dynamic purchase behavior. *Marketing Science* .

Liu J, Toubia O, Hill S (2021) Content-based model of web search behavior: An application to TV show search. *Management Science* 67(10):6378–6398.

Mei H, Eisner JM (2017) The neural Hawkes process: A neurally self-modulating multivariate point process. *Advances in Neural Information Processing Systems*, 6754–6764.

Ogata Y (1981) On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory* 27(1):23–31.

Ogata Y (1988) Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association* 83(401):9–27.

Papangelou F (1972) Integrability of expected increments of point processes and a related random change of scale. *Transactions of the American Mathematical Society* 165:483–506.

Rasmussen JG (2018) Lecture notes: Temporal point processes and the conditional intensity function.

Rutz O, Aravindakshan A, Rubel O (2019) Measuring and forecasting mobile game app engagement. *International Journal of Research in Marketing* 36(2):185–199.

Schweidel DA, Moe WW (2016) Binge watching and advertising. *Journal of Marketing* 80(5):1–19.

Xu L, Duan JA, Whinston A (2014) Path to purchase: A mutually exciting point process model for online advertising and conversion. *Management Science* 60(6):1392–1412.