Online Appendix of Learning when Reading: Evidence from an Online Mobile Reading Platform

EC.1. Reduced-form Evidence on In-consumption Social Listening Effects

We construct two reduced-form models to examine evidence on the impacts of in-consumption comments on consumers' purchase decisions.

EC.1.1. Modeling Consumer's Skipping vs. Non-skipping Decision

We first model consumers' decision at the end of each chapter as a binary variable, where 0 indicates the decision to continue reading the subsequent chapter, and 1 indicates the decision to skip it. We then specify this decision as a function of the in-consumption comment volume across five topics within the last chapter the consumer read, the price of the following chapter, the total amount of money the consumer has paid on the book, the informativeness of the book's chapter name list, and the consumer's current position out of the book. We also implemented the book genre-level fixed effects. The proposed model is as follows:

$$\begin{aligned} logit(Y_{im[j]t}) = \beta_0 + \beta_1 \cdot Price_{im[j]t} + \beta_2 \cdot TotalSpend_{im[j]t} + \beta_3 \cdot Informativeness_{im[j]t} + \beta_4 \cdot Position_{im[j]t} \\ + \sum_{k=1}^{5} \beta_{5k} \cdot BulCluster_{im[j]t,k} + Genre_k + \varepsilon_{im[j]t} \end{aligned} \tag{EC.1}$$

where $Y_{im[j]t}$ is 1 if the consumer i decides to skip the subsequent chapter $t+1$ at the end of chapter t of book m and 0 otherwise. $Genre-k$ denotes the genre-specific fixed effects.

As shown in Table EC.1, the directions of the in-consumption comment volume's effects across five topics align with our main model estimation results as well as our expectations based on the conceptual framework, as detailed in Section5. Specifically, consumers facing more comments making inferences based on the narrative, judging the character's behaviors, or pleading for the release of new chapters tend to continuously read the subsequent chapter. On the contrary, an increased volume of comments comprising merely happy emotional words or those depicting a desired hypothetical future scene could increase consumers' intention to skip the following chapter.

Table EC.1 Baseline Results: Skipping vs. Non-skipping[1]

Variables	Estimate	Std. error
Intercept	13.404***	0.633
Topic-1 Comment Volume[2]	4.508***	0.535
Topic-2 Comment Volume	−1.626**	0.621
Topic-3 Comment Volume	−3.138***	0.454
Topic-4 Comment Volume	2.972*	1.480
Topic-5 Comment Volume	−2.099**	0.783
Price	3.009	7.950
TotalSpend	−4.569	4.985
Informativeness	−0.791**	0.399
Position	7.255***	0.678

[1] *Significance level:* * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[2] *The comment volume is log-transformed.*

EC.1.2. Modeling Consumer's Chapter Selection Decision

To more accurately capture the consumer's decision-making process while reading a book, we further model their chapter selection as the number of chapters they choose to skip after completing each chapter. Consistent with our primary model, we assume that this behavior follows a negative binomial distribution. The proposed model, detailed below, is constructed similarly to our main model, with the key distinction being the exclusion of the consumer's learning process regarding the book's quality.

$$
\begin{aligned}
Prob(N_{im[j]t} = n) &= \frac{\Gamma(\theta+n)\theta^{\theta}\lambda_{im[j]t}^{n}}{\Gamma(n+1)\Gamma(\theta)(\lambda_{im[j]t}+\theta)^{\theta+n}} \\
\lambda_{im[j]t} &= exp(U_{im[j]t}) \\
U_{im[j]t} &= \beta_0 + \beta_1 \cdot Price_{im[j]t} + \beta_2 \cdot TotalSpend_{im[j]t} + \beta_3 \cdot Informativeness_{im[j]t} + \beta_4 \cdot Position_{im[j]t} \\
&\quad + \sum_{k=1}^{5} \beta_{5k} \cdot BulCluster_{im[j]t,k} + \varepsilon_{im[j]t}
\end{aligned}
\tag{EC.2}
$$

where $N_{im[j]t}$ denotes the number of skipped chapters by consumer i at the end of chapter t of book m.

The results are reported in Table EC.2. Our findings indicate that the volume of in-consumption comments across five topics has varying impacts on consumers' purchase decisions. The impact of these comment

topics is consistent with the estimation results of our main model. Specifically, consumers exposed to comments that infer based on the narrative, evaluate character behaviors, or request the release of new chapters tend to skip fewer chapters. In contrast, a higher volume of comments that simply express happy emotions or describe hypothetical future scenarios may lead consumers to skip more chapters.

Table EC.2 Baseline Results: Chapter Selection[1]

Variables	Estimate	Std. error
Intercept	$-1.58***$	0.14
Topic-1 Comment Volume[2]	$9.77*$	4.78
Topic-2 Comment Volume	$-9.97**$	3.56
Topic-3 Comment Volume	$-9.72*$	4.87
Topic-4 Comment Volume	$10.57*$	4.21
Topic-5 Comment Volume	$-9.66*$	4.14
Price	0.37	0.29
TotalSpend	$-0.87*$	0.41
Informativeness	$-1.36***$	0.14
Position	$4.52***$	1.70

[1] *Significance level:* * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[2] *The comment volume is log-transformed.*

However, both of the models fail to capture consumers' hidden, dynamic perceived quality of the book, which is derived from their own experience as well as in-consumption social listening. To address this gap, we develop a Bayesian learning model to unveil the underlying mechanisms of consumers' decision-making process. To further highlight the significance of incorporating consumers' learning channels, we present the fit metrics—Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC)—for the two baseline models and our proposed model in Table EC.3. Both metrics indicate that excluding the learning channels leads to a substantial reduction in model fit, underscoring the importance of modeling the consumer learning process.

EC.2. Validating the Classification of In-consumption Comments Topics

To determine the optimal number of topic clusters, we utilize the elbow method, which calculates the within-cluster sum of square (WCSS), which is the sum of the squared distance between each point and the centroid

Table EC.3 Model Fit Comparison: Baseline vs. Full

Model	AIC	BIC
Full model	9,178.7	9,808.1
Baseline model: skipping vs. non-skipping	14,870.5	15,008.2
Baseline model: chapter selection	11,951.2	12,033.4

in a cluster, and plotted the value across different numbers of clusters. As shown in Figure EC.1, WCSS would rapidly change at a point, thus creating an elbow shape. From this point, the graph starts to move almost parallel to the X-axis. This turning point denotes the optimal number of topic clusters existing among in-consumption comments.

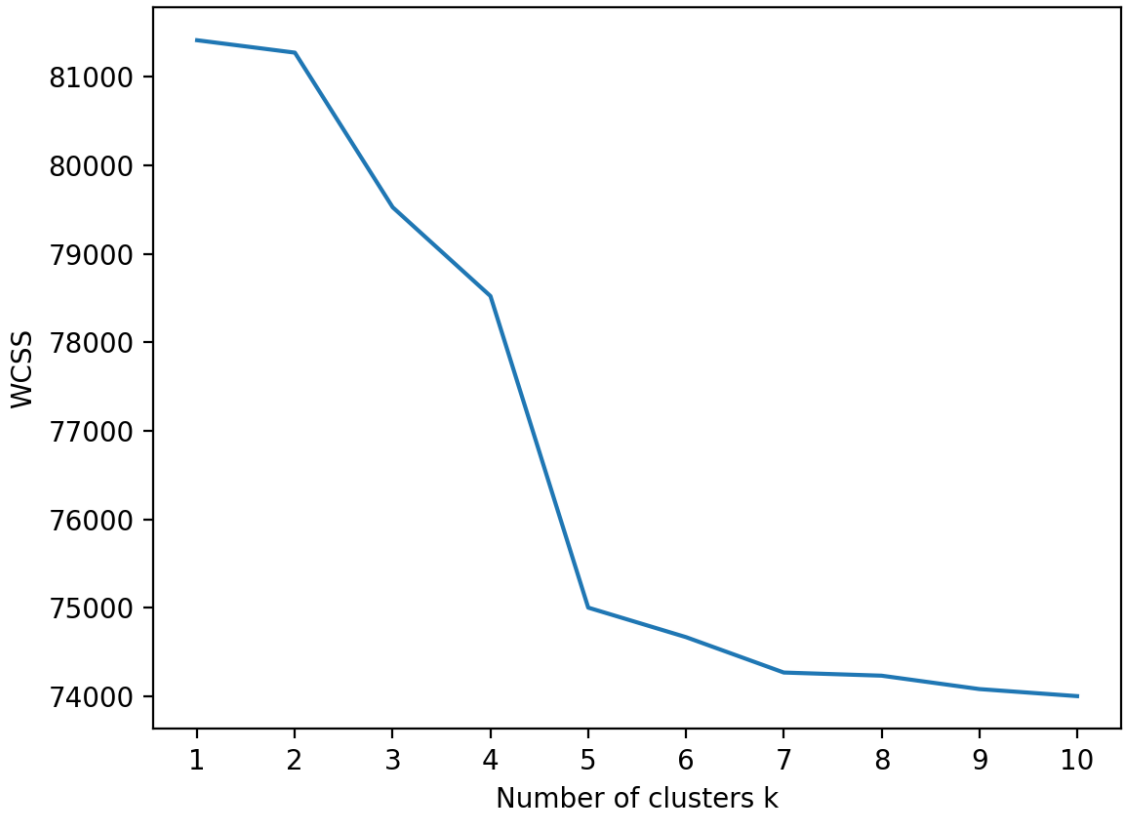

Figure EC.1 Elbow Figrue

We justify our classification analysis of the comment topics as follows. First, we choose a k-means model over a Latent Dirichlet Allocation model because the in-consumption comments are all very short (i.e., less

than 10 Chinese words). LDA model is shown to have a relatively poor performance for short text (Hong and Davison 2010). Second, we calculate the average silhouette width, which is another commonly used metric to select the optimal parameter of a k-means model (Rousseeuw 1987), to ensure the robustness of our findings. Consistently, the results reaffirmed the presence of five topics, as initially discovered. Third, we explore the potential of classifying the comments into fewer or more categories. However, these alternative classifications resulted in a decrease in interpretability, suggesting that they did not capture the nuances of the comments as effectively as our original five-topic classification.

Additionally, we conduct a robustness check by applying a state-of-the-art topic-modeling method, BERTopic, which implements the HDBSCAN (Hierarchical Density-Based Spatial Clustering) clustering algorithminstead of K-Means to identify the topics embedded within in-consumption comments (Grootendorst 2022, Campello et al. 2013). A primary advantage of HDBSCAN is that, unlike K-Means, it does not require pre-specifying the number of clusters. Instead, it automatically identifies a natural number of topics based on the underlying density of the data (i.e., how densely packed the in-consumption comments are in semantic space). Due to the constraints on our computational capacity, we randomly draw 100,000 in-consumption comments from our full sample to run the analysis. The optimal number of topics identified by HDBSCAN in our data is five. We then use the ChatGPT-4o model to generate summaries for each topic identified among the in-consumption comments, reported in below Table EC.2. For each topic, we ask ChatGPT to generate a summary both in Chinese (as presented in the second column) and in English (as presented in the third column). As shown in the table, the five topics identified among our in-consumption comments using the state-of-the-art topic modeling methods (i.e., BERTopic+HDBSCAN+ChatGPT-4o) are consistent with the five topics identified in our main analysis. Given computational constraints, we continued using the simpler K-Means clustering approach for our main analysis, instead of the more resource-intensive BERTopic method.

Based on the above comprehensive analyses, we concluded that our initial classification of in-consumption comment topics remains the most accurate and interpretable.

	Topic-CHN	**Topic-ENG**
Topic-1	正面情绪反应	Positive emotional engagement and reactions
Topic-2	剧情发展推测与人物互动反馈	Inferences and feedback on plot twists and character behavior
Topic-3	催更	Urging for updates and chapter release
Topic-4	剧情发展期待	Anticipation for plot development
Topic-5	人物负面评价	Negative reactions to characters

Figure EC.2 In-consumption Comment Topics (BERTopic+HDBSCAN+ChatGPT-4o)

EC.3. Mathematical Derivation of Consumers' Belief Updating

In this section, we gave how we derive the updating mechanism of consumers' perceived book quality as outlined in Section 4, following DeGroot (2005). Since consumers are assumed to be Bayesian learners, when receiving a new experiential signal $S_{im[j]t}$ and a set of five social listening signals $ln(1+Bul_{im[j]t,k}), k = 1,...,5$, the consumer i will update their belief of both the book quality and the mean log volume of the in-consumption comments across five topics as follows.

Update of the posterior means and variances of perceived book quality: Given Bayes Theorem, after receiving the new signals, the likelihood of the consumer's posterior perceived book quality, $A_{im[j]t}$, is formed as follows:

$$
\begin{aligned}
P(A_{im[j]t}|S_{im[j]t}, ln(1+Bul_{im[j]t,k})) &\propto P(S_{im[j]t}, ln(1+Bul_{im[j]t,k})|A_{im[j]}) \cdot P(A_{im[j]t}) \\
&\propto P(S_{im[j]t}|A_{im[j]t}) \cdot P(ln(1+Bul_{im[j]t,k})|A_{im[j]t}) \cdot P(A_{im[j]t}) \\
&\propto exp(-\frac{(S_{im[j]t}-A_{im[j]t})^2}{2\tau_j^2}) \cdot \prod_k exp(-\frac{(ln(1+Bul_{im[j]t,k}) - Bul^M_{im[j]t-1,k} - \phi_k * A_{im[j]t})^2}{2(\sigma^2_{\omega,k}+\sigma^2_{\eta,k})}) \\
&\cdot exp(-\frac{(A_{im[j]t}-\alpha_{im[j]t-1})^2}{2\sigma^2_{\alpha_{im[j]t-1}}}) \\
&\propto exp(-\frac{(S_{im[j]t}-A_{im[j]t})^2}{2\tau_j^2} - \sum_{k=1}^{5} \frac{(ln(1+Bul_{im[j]t,k}) - Bul^M_{im[j]t-1,k} - \phi_k * A_{im[j]t})^2}{2(\sigma^2_{\omega,k}+\sigma^2_{\eta,k}))} \\
&- \frac{(A_{im[j]t}-\alpha_{im[j]t-1})^2}{2\sigma^2_{\alpha_{im[j]t-1}}})
\end{aligned}
\tag{EC.3}
$$

Update of the posterior means and variances of mean log volume of the in-consumption comments: Given Bayes Theorem, after accessing the new in-consumption comments, the likelihood of the consumer's posterior perceived mean log in-consumption comment volume, $Bul^*_{im[j]t,k}$, is formed as follows:

$$\begin{aligned} P(Bul^*_{im[j]t,k}|ln(1+Bul_{im[j]t,k})) &\propto P(ln(1+Bul_{im[j]t,k})|Bul^*_{im[j]t,k}) \cdot P(Bul^*_{im[j]t,k}) \\ &\propto exp(-\frac{(ln(1+Bul_{im[j]t,k})-Bul^*_{im[j]t,k})^2}{2\sigma^2_{\omega,k}}) \cdot exp(-\frac{(Bul^*_{im[j]t,k}-Bul^M_{im[j]t-1})^2}{2\sigma^2_{Bul_{ijt-1,k}}}) \\ &\propto exp(-\frac{(ln(1+Bul_{im[j]t,k})-Bul^*_{im[j]t,k})^2}{2\sigma^2_{\omega,k}} - \frac{(Bul^*_{im[j]t,k}-Bul^M_{im[j]t-1})^2}{2\sigma^2_{Bul_{ijt-1,k}}}) \end{aligned} \tag{EC.4}$$

The posterior means and variance in Equation 3 are then derived after expanding and rearranging the above expressions.

EC.4. Mundlak's Approach

Mundlak's approach starts by including time-invariant fixed effects in a panel data model to address potential correlated omitted variables concerns. Instead of introducing common fixed effects that are free from assumptions on their dependence on observed explanatory variables, Mundlak places substantive restrictions on the distribution of the fixed effects given the explanatory variables. Specifically, considering a panel data model with time-invariant fixed effects, c_i, and the associative explanatory variables, $\mathbf{x_{it}}$, the CRE framework makes the following assumption:

$$c_i|\mathbf{x_i} \sim Normal(\psi + \bar{\mathbf{x}}_\mathbf{i}\xi, \sigma^2) \tag{EC.5}$$

where $\bar{\mathbf{x}}_\mathbf{i}$ is the average of $\mathbf{x_{it}}$ across all time periods, and σ^2 is the conditional variance of c_i, which is assumed to be independent of $\mathbf{x_i}$. Hence, the original model can be rewritten by replacing the fixed effects with the averages of potential endogenous explanatory variables over the time periods.

Following the CRE framework, we adjust our model specified in the previous section to mitigate the potential endogeneity concern. To this end, we first reform the expected utilities of a consumer i to read the next consecutive chapter and to return to the chapter list at the end of chapter t of book $m[j]$ (i.e., stage 1

utilities), respectively, as follows:

$$\begin{aligned} E[U_{im[j]t1}|I_{im[j]t}] &= \beta_{00}\cdot A_{im[j]t} + \beta_{00}\cdot r_1 \cdot E[A_{im[j]t}|I_{im[j]t}]^2 + \beta_{00}\cdot r_1 \cdot E[(A_{im[j]t}-\alpha_j)^2|I_{im[j]t}] \\ &\quad + \beta_1 \cdot Price_{im[j]t} + \beta_2 \cdot TotalSpend_{im[j]t} + \beta_{16}\cdot \bar{\mathbf{X}}_{\mathbf{im[j]}} + \varepsilon_{im[j]t1} \\ E[U_{im[j]t2}|I_{im[j]t}] &= \beta_{10}\cdot A_{im[j]t} + \beta_{10}\cdot r_2 \cdot E[A_{im[j]t}|I_{im[j]t}]^2 + \beta_{10}\cdot r_2 \cdot E[(A_{im[j]t}-\alpha_j)^2|I_{im[j]t}] \\ &\quad + \beta_3 \cdot Price_{im[j]t} + \beta_4 \cdot TotalSpend_{im[j]t} + \beta_{26}\cdot \bar{\mathbf{X}}_{\mathbf{im[j]}} + \varepsilon_{im[j]t2} \end{aligned} \tag{EC.6}$$

where $\bar{\mathbf{X}}_{\mathbf{im[j]}}$ indicates the averages of the potential endogenous variables in our models, i.e., the amount of money consumer i needs to pay for chapter $t+1$ and her current cost spent on book k.

EC.5. Constructing the Likelihood Function

Because we are unable to observe the values of the experiential signal generated from consumer learning processes, the probability of the observed reading history for consumer i on book $m[j]$, denoted by $H_{im[j]}$, is written as follows:

$$Prob(H_{im[j]}) = \int_{\delta_{im[j]t}} \prod_{t=1}^{T_{m[j]}} P(N_{im[j]t}) dF(\delta_{im[j]t}) \tag{EC.7}$$

where $T_{m[j]}$ indicates the total number of chapters in book $m[j]$. We simulate this integral using draws for $\delta_{im[j]t}$ from its distribution specified in Equation 1, applying the widely-used Monte Carlo simulation method with a simulation size of 100 (Sawilowsky 2003).

We then define the residuals in the content-quality relationship as $\omega_{im[j]t,k} = ln(1+Bul_{im[j]t,k}) - Bul_{0,k} - \phi_k * \alpha_j - \eta_{j,k}$. Let $f(\omega)$ denotes the intensities of ω, the simulated likelihood for consumer i is:

$$L_i = \prod_{m[j]} \prod_{t} \prod_{k=1}^{5} (1+Bul_{im[j]t,k})^{-1} f(\omega_{im[j]t,k}) Prob(H_{im[j]}) \tag{EC.8}$$

Notably, here the term $(1+Bul_{im[j]t,k})^{-1}$ is generated by the Jacobian.

The likelihood for the complete reading history for all consumers on all books, denoted by H, is thus specified as below:

$$L(H) = \prod_{i=1} L_i \tag{EC.9}$$

EC.6. Identification

For identification, we set the true quality levels of books across all four genres to 0 and fixed the variance of consumers' prior belief about book quality to 1. This approach allows us to identify the mean levels

of consumers' prior beliefs through their initial reading decisions across different genres, as these early choices are more strongly influenced by their initial perceived quality.

As consumers progress through their reading, the influence of their prior quality beliefs diminishes, enabling the overall skipping versus non-skipping decisions to identify the risk coefficients. Meanwhile, the degree to which consumers choose to skip chapters helps identify the signal variances, reflecting the variability in how consumers perceive and react to new information. Lastly, the variability in in-consumption comment volume is identified through the differences in observed comment volumes across various topics and genres, allowing us to capture how in-consumption social listening influences consumer behavior differently depending on the content and context.

EC.7. Policy Simulation: General Procedure

A key advantage of adopting structural models, such as our proposed framework, is their ability to conduct counterfactual policy experiments. Since the utility functions estimated in these models are invariant to changes in input data, once the model parameters are estimated, we can examine the effects of different policies by modifying the input data and feeding it back into the fitted model.

Specifically, we assume that consumers make purchase decisions according to our proposed main model, as described in Section 4. After estimating the parameters in Section 5, we can reconstruct the consumers' decision-making process. We then input data that reflects the information set available to consumers into the estimated model, excluding their actual decisions, and simulate their choices based on the likelihoods derived from the model. We then repeat this process for 1,000 times and calculate the averages as the final simulation results. his approach allows us to evaluate how changes in policies or external conditions might influence consumer behavior under different scenarios.

EC.8. Policy Simulation: Reading Experience as the Only Signal

To illustrate the importance of the in-consumption social learning channel in facilitating consumers' learning process, we conducted a study that keeps consumers' reading experience as their single learning channel. Specifically, we assume that consumers would now learn a book's true quality based only on the signal value formulated in formula 1. Upon receiving the signals, consumers would then update the mean and the variance of their belief about the book's quality as follows:

$$\alpha_{im[j]t} = \alpha_{im[j]t-1} + \beta \cdot (S_{im[j]t} - \alpha_{im[j]t-1}),$$
$$\sigma^2_{\alpha_{im[j]t}} = [\frac{1}{\sigma^2_{\alpha_{im[j]t-1}}} + \frac{1}{\tau_j^2}]^{-1}, \tag{EC.10}$$

where,

$$\beta = \frac{\frac{1}{\tau_j^2}}{\frac{1}{\sigma^2_{\alpha_{im[j]t-1}}} + \frac{1}{\tau_j^2}} \tag{EC.11}$$

EC.9. Procedures of Policy Simulation 3: Decreasing Comment Variability

We follow the following procedure to decrease the variability of the in-consumption comment volume while keeping the mean volume unchanged: (1) we find the mean comment volume for each book-genre-comment-topic pair; (2) we scale down the deviation of the volume realizations from that mean value to achieve the desired decrease in variance; (3) we determine how such a transformation affects mean and variance of log comment volume; (4) we modify the log comment volume equation accordingly to keep the mean comment volume fixed; (5) we simulate consumers' behavior given the new comment volume data and the new comment volume process.

EC.10. Robustness Check: Increasing Reading Limit

To further validate the findings of our analysis, we ran a robustness check by increasing the reading limit of the consumers in our data collection process from a maximum of 50 chapters to 60 chapters, leading to a new dataset comprising 135,296 chapter purchases made by 4,413 individual consumers of 1,166 e-books. We then re-estimate our proposed model with this new data and reported the estimation results in Table EC.6. As shown in the table, the results are qualitatively consistent with our main findings.

EC.11. Robustness Check: Book-specific Genre Qualities

We ran an additional analysis by allowing the true genre qualities to vary across books, denoted by α_{mj} for genre j and book m. Specifically, we allow $alpha_{mj} \sim N(\alpha_j, \sigma_A^2)$, where $alpha_j$ denotes the mean true quality of genre j and σ_A^2 captures the variance across books. Incorporating this into our broader model framework, the prior quality belief distributions of consumers are specified $A_{0m[j]} \sim N(\alpha_{0j}, \sigma_0^2 + \sigma_A^2)$. Similarly, the prior distribution of mean log volume of topic-k in-chapter comment are formed as $Bul^M_{0m[j],k} \sim N(Bul_{0,k} + \phi_k\alpha_{0j}, \phi_k^2(\sigma_0^2 + \sigma_A^2) + \sigma^2_{\eta,k})$ for k=1,...,5 and j=1,...,4. The remainder of the model remains unchanged, and

we maintain the same assumptions for model identification, i.e, we fix the mean true quality levels across all four genres at zero and set σ_0^2 equal to one. As shown in Table EC.7, the results are qualitatively consistent with our main findings.

We further compare the model fit against our main specification. The alternative model exhibits a deterioration in fit, with higher AIC (9301.2 vs. 9178.7) and BIC (9942.1 vs. 9808.1) values relative to the main model. Moreover, the estimated σ_A^2 is insignificant, implying an insignificant variation in the variation of true genre quality among books. Given this deterioration in model fit, along with our primary focus on modeling the consumer learning process, we retain the original model for the main analysis and interpretation.

EC.12. Tables

Table EC.4 Summary Statistics of the data

Variables	Male Sensational Fiction	Female Sensational Fiction	Gender-neutral Fiction	Teen Fiction
Distribution	0.21	0.32	0.16	0.31
Fraction of consumers skipping chapters	0.90	0.95	0.95	0.97
Average number of consumers' skipped chapters				
Mean	11.00	3.66	2.69	2.46
SD	48.22	24.03	6.56	7.98
Number of Topic 1 comments per chapter				
Mean	1.3	0.15	0.05	5.35
SD	1.25	0.14	0.38	20.36
Number of Topic 2 comments per chapter				
Mean	0.18	2.14	1.18	4.87
SD	1.31	2.67	2.53	34.19
Number of Topic 3 comments per chapter				
Mean	2.77	4.43	10.13	14.48
SD	35.27	12.25	21.33	69.74
Number of Topic 4 comments per chapter				
Mean	0.45	0.49	1.49	2.33
SD	4.56	1.75	2.82	8.36
Number of Topic 5 comments per chapter				
Mean	4.16	1.14	2.02	10.75
SD	17.55	3.89	6.29	48.53
Price per chapter				
Mean	0.02	0.03	0.22	0.05
SD	0.05	0.066	0.44	0.15
Chapter name informativeness				
Mean	0.99	0.97	0.89	0.80
SD	0.01	0.16	0.28	0.36

EC.13. Figures

Table EC.5 Parameter Estimates: Full Table[1]

Paramter	Estimate	Std. error
Consecutive Reading		
Price	0.15	0.17
TotalSpend	-1.50^{***}	0.27
Perceived Quality	3.55^{***}	0.25
	-0.02^{***}	0.00
Returning to Chapter List		
Price	0.45^{***}	0.05
TotalSpend	-2.49^{***}	0.16
Perceived Quality	2.97^{***}	0.18
	0.003^{***}	0.001
Skipping Chapter		
Informativeness	-1.95^{***}	0.02
Position	6.01^{***}	0.56
Perceived Quality	-2.13^{***}	0.02
	0.007^{***}	0.001
θ	0.01^{***}	0.00
Experiential Signaling Parameters		
α_{01}(Mean prior quality belief of Book Genre 1)	-110.22^{***}	18.66
α_{02}(Mean prior quality belief of Book Genre 2)	-113.08^{***}	9.44
α_{03}(Mean prior quality belief of Book Genre 3)	-69.04^{***}	19.44
α_{04}(Mean prior quality belief of Book Genre 4)	-122.86^{***}	19.16
τ_1^2(Experiential Signal Variance of Book Genre 1)	8.66^{***}	2.78
τ_2^2(Experiential Signal Variance of Book Genre 2)	17.99^{***}	4.69
τ_3^2(Experiential Signal Variance of Book Genre 3)	4.42^{***}	1.14
τ_4^2(Experiential Signal Variance of Book Genre 4)	20.26^{***}	5.85
Social-Listening Signaling Parameters		
$Bul_{0,1}$(Intercept of Topic-1 Comment)	0.22^{***}	0.004
$Bul_{0,2}$(Intercept of Topic-2 Comment)	0.22^{***}	0.004
$Bul_{0,3}$(Intercept of Topic-3 Comment)	1.08^{***}	0.003
$Bul_{0,4}$(Intercept of Topic-4 Comment)	0.18^{***}	0.003
$Bul_{0,5}$(Intercept of Topic-5 Comment)	0.65^{***}	0.003
ϕ_1(Slope of Topic-1 Comment Signaling Equation)	-0.28^{***}	0.03
ϕ_2(Slope of Topic-2 Comment Signaling Equation)	4.71^{***}	0.34
ϕ_3(Slope of Topic-3 Comment Signaling Equation)	1.64^{***}	0.08
ϕ_4(Slope of Topic-4 Comment Signaling Equation)	-1.86^{***}	0.14
ϕ_5(Slope of Topic-5 Comment Signaling Equation)	0.51^{***}	0.01
$\sigma_{\eta,1}^2$(Variance of Genre-specific Constants)	2.52^{***}	0.51
$\sigma_{\eta,2}^2$(Variance of Genre-specific Constants)	1.67^{***}	0.21
$\sigma_{\eta,3}^2$(Variance of Genre-specific Constants)	3.44^{***}	0.25
$\sigma_{\eta,4}^2$(Variance of Genre-specific Constants)	5.65^{***}	0.94
$\sigma_{\eta,5}^2$(Variance of Genre-specific Constants)	2.26^{***}	0.35
$\sigma_{\omega,1}^2$(Topic-1 Comment Volume Variability)	0.96^{***}	0.01
$\sigma_{\omega,2}^2$(Topic-2 Comment Volume Variability)	0.85^{***}	0.02
$\sigma_{\omega,3}^2$(Topic-3 Comment Volume Variability)	1.32^{***}	0.02
$\sigma_{\omega,4}^2$(Topic-4 Comment Volume Variability)	0.74^{***}	0.01
$\sigma_{\omega,5}^2$(Topic-5 Comment Volume Variability)	1.19^{***}	0.02
$\eta_{1,1}$	-0.18^{***}	0.01
$\eta_{1,2}$	-0.15^{***}	0.01
$\eta_{1,3}$	-0.23^{***}	0.01
$\eta_{1,4}$	-0.11^{***}	0.01
$\eta_{1,5}$	-0.19^{***}	0.01
$\eta_{2,1}$	-0.20^{***}	0.05
$\eta_{2,2}$	-0.14^{***}	0.02
$\eta_{2,3}$	-0.26^{***}	0.01
$\eta_{2,4}$	-0.16^{***}	0.04
$\eta_{2,5}$	-0.29^{***}	0.03
$\eta_{3,1}$	-0.18^{***}	0.02
$\eta_{3,2}$	-0.11^{***}	0.01
$\eta_{3,3}$	0.14^{***}	0.01
$\eta_{3,4}$	-0.15^{***}	0.02
$\eta_{3,5}$	-0.15^{***}	0.02
$\eta_{4,1}$	0.57^{***}	0.05
$\eta_{4,2}$	0.40^{***}	0.03
$\eta_{4,3}$	0.36^{***}	0.01
$\eta_{4,4}$	0.42^{***}	0.03
$\eta_{4,5}$	0.62^{***}	0.04

[1] *Significance level:* * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table EC.6 Parameter Estimates: Reading Limit[1]

Paramter	Estimate	Std. error
Consecutive Reading		
Price	0.26	0.20
TotalSpend	-1.34^{***}	0.25
Perceived Quality	4.01^{***}	0.35
r_1(Risk Coefficien)	-0.02^{***}	0.00
Returning to Chapter List		
Price	0.38^{***}	0.02
TotalSpend	-3.01^{***}	0.17
Perceived Quality	2.89^{***}	0.14
r_2(Risk Coefficient)	0.002^{***}	0.000
Skipping Chapter		
Informativeness	-1.87^{***}	0.01
Position	7.12^{***}	0.45
Perceived Quality	-3.07^{***}	0.02
r_3(Risk Coefficient)	0.005^{***}	0.001
θ	0.01^{***}	0.00
Experiential Signaling Parameters		
α_{01}(Mean prior quality belief of Book Genre 1)	-98.55^{***}	20.16
α_{02}(Mean prior quality belief of Book Genre 2)	-102.22^{***}	13.21
α_{03}(Mean prior quality belief of Book Genre 3)	-63.02^{***}	15.31
α_{04}(Mean prior quality belief of Book Genre 4)	-142.80^{***}	23.37
τ_1^2(Experiential Signal Variance of Book Genre 1)	7.52^{***}	1.96
τ_2^2(Experiential Signal Variance of Book Genre 2)	17.71^{***}	3.82
τ_3^2(Experiential Signal Variance of Book Genre 3)	3.12^{***}	1.01
τ_4^2(Experiential Signal Variance of Book Genre 4)	21.25^{***}	3.67
Social-Listening Signaling Parameters		
$Bul_{0,1}$(Intercept of Topic-1 Comment)	0.21^{***}	0.003
$Bul_{0,2}$(Intercept of Topic-2 Comment)	0.22^{***}	0.003
$Bul_{0,3}$(Intercept of Topic-3 Comment)	1.07^{***}	0.001
$Bul_{0,4}$(Intercept of Topic-4 Comment)	0.18^{***}	0.003
$Bul_{0,5}$(Intercept of Topic-5 Comment)	0.65^{***}	0.003
ϕ_1(Slope of Topic-1 Comment Signaling Equation)	-0.39^{***}	0.08
ϕ_2(Slope of Topic-2 Comment Signaling Equation)	4.13^{***}	0.25
ϕ_3(Slope of Topic-3 Comment Signaling Equation)	1.43^{***}	0.02
ϕ_4(Slope of Topic-4 Comment Signaling Equation)	-2.10^{***}	0.11
ϕ_5(Slope of Topic-5 Comment Signaling Equation)	0.52^{***}	0.10
$\sigma_{\eta,1}^2$(Variance of Genre-specific Constants)	2.31^{***}	0.41
$\sigma_{\eta,2}^2$(Variance of Genre-specific Constants)	1.65^{***}	0.25
$\sigma_{\eta,3}^2$(Variance of Genre-specific Constants)	3.01^{***}	0.19
$\sigma_{\eta,4}^2$(Variance of Genre-specific Constants)	4.98^{***}	0.25
$\sigma_{\eta,5}^2$(Variance of Genre-specific Constants)	2.11^{***}	0.23
$\sigma_{\omega,1}^2$(Topic-1 Comment Volume Variability)	0.94^{***}	0.01
$\sigma_{\omega,2}^2$(Topic-2 Comment Volume Variability)	0.79^{***}	0.01
$\sigma_{\omega,3}^2$(Topic-3 Comment Volume Variability)	1.35^{***}	0.02
$\sigma_{\omega,4}^2$(Topic-4 Comment Volume Variability)	0.73^{***}	0.01
$\sigma_{\omega,5}^2$(Topic-5 Comment Volume Variability)	1.19^{***}	0.02
$\eta_{1,1}$	-0.18^{***}	0.01
$\eta_{1,2}$	-0.13^{***}	0.01
$\eta_{1,3}$	-0.21^{***}	0.01
$\eta_{1,4}$	-0.09^{***}	0.01
$\eta_{1,5}$	-0.17^{***}	0.02
$\eta_{2,1}$	-0.24^{***}	0.03
$\eta_{2,2}$	-0.14^{***}	0.02
$\eta_{2,3}$	-0.24^{***}	0.02
$\eta_{2,4}$	-0.15^{***}	0.01
$\eta_{2,5}$	-0.27^{***}	0.02
$\eta_{3,1}$	-0.20^{***}	0.02
$\eta_{3,2}$	-0.11^{***}	0.01
$\eta_{3,3}$	0.17^{***}	0.02
$\eta_{3,4}$	-0.17^{***}	0.01
$\eta_{3,5}$	-0.16^{***}	0.02
$\eta_{4,1}$	0.58^{***}	0.07
$\eta_{4,2}$	0.42^{***}	0.01
$\eta_{4,3}$	0.37^{***}	0.01
$\eta_{4,4}$	0.45^{***}	0.02
$\eta_{4,5}$	0.67^{***}	0.10

[1] *Significance level:* * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table EC.7 Parameter Estimates: Varied Qualities[1]

Paramter	Estimate	Std. error
Consecutive Reading		
Price	0.31	0.19
TotalSpend	-1.56^{***}	0.11
Perceived Quality	3.89^{***}	0.27
r_1(Risk Coefficien)	-0.02^{***}	0.00
Returning to Chapter List		
Price	0.41^{***}	0.01
TotalSpend	-3.41^{***}	0.22
Perceived Quality	2.37^{***}	0.12
r_2(Risk Coefficient)	0.001^{***}	0.000
Skipping Chapter		
Informativeness	-2.01^{***}	0.03
Position	7.01^{***}	0.37
Perceived Quality	-3.35^{***}	0.01
r_3(Risk Coefficient)	0.004^{***}	0.001
θ	0.01^{***}	0.00
Experiential Signaling Parameters		
α_{01}(Mean prior quality belief of Book Genre 1)	-99.01^{***}	18.26
α_{02}(Mean prior quality belief of Book Genre 2)	-100.10^{***}	12.29
α_{03}(Mean prior quality belief of Book Genre 3)	-70.13^{***}	10.27
α_{04}(Mean prior quality belief of Book Genre 4)	-139.75^{***}	26.12
σ_A^2(Consumer-specific Variance in Quality)	2.11	1.56
τ_1^2(Experiential Signal Variance of Book Genre 1)	7.26^{***}	1.43
τ_2^2(Experiential Signal Variance of Book Genre 2)	18.25^{***}	4.01
τ_3^2(Experiential Signal Variance of Book Genre 3)	2.76^{***}	0.89
τ_4^2(Experiential Signal Variance of Book Genre 4)	23.01^{***}	5.46
Social-Listening Signaling Parameters		
$Bul_{0,1}$(Intercept of Topic-1 Comment)	0.19^{***}	0.003
$Bul_{0,2}$(Intercept of Topic-2 Comment)	0.20^{***}	0.002
$Bul_{0,3}$(Intercept of Topic-3 Comment)	0.98^{***}	0.001
$Bul_{0,4}$(Intercept of Topic-4 Comment)	0.20^{***}	0.002
$Bul_{0,5}$(Intercept of Topic-5 Comment)	0.54^{***}	0.003
ϕ_1(Slope of Topic-1 Comment Signaling Equation)	-0.26^{***}	0.04
ϕ_2(Slope of Topic-2 Comment Signaling Equation)	4.69^{***}	0.21
ϕ_3(Slope of Topic-3 Comment Signaling Equation)	1.21^{***}	0.01
ϕ_4(Slope of Topic-4 Comment Signaling Equation)	-1.98^{***}	0.08
ϕ_5(Slope of Topic-5 Comment Signaling Equation)	0.56^{***}	0.08
$\sigma_{\eta,1}^2$(Variance of Genre-specific Constants)	2.55^{***}	0.28
$\sigma_{\eta,2}^2$(Variance of Genre-specific Constants)	1.98^{***}	0.21
$\sigma_{\eta,3}^2$(Variance of Genre-specific Constants)	2.88^{***}	0.12
$\sigma_{\eta,4}^2$(Variance of Genre-specific Constants)	5.04^{***}	0.31
$\sigma_{\eta,5}^2$(Variance of Genre-specific Constants)	2.23^{***}	0.17
$\sigma_{\omega,1}^2$(Topic-1 Comment Volume Variability)	0.89^{***}	0.02
$\sigma_{\omega,2}^2$(Topic-2 Comment Volume Variability)	0.74^{***}	0.01
$\sigma_{\omega,3}^2$(Topic-3 Comment Volume Variability)	1.66^{***}	0.03
$\sigma_{\omega,4}^2$(Topic-4 Comment Volume Variability)	0.69^{***}	0.01
$\sigma_{\omega,5}^2$(Topic-5 Comment Volume Variability)	1.21^{***}	0.02
$\eta_{1,1}$	-0.15^{***}	0.01
$\eta_{1,2}$	-0.16^{***}	0.02
$\eta_{1,3}$	-0.19^{***}	0.01
$\eta_{1,4}$	-0.11^{***}	0.01
$\eta_{1,5}$	-0.14^{***}	0.01
$\eta_{2,1}$	-0.20^{***}	0.02
$\eta_{2,2}$	-0.13^{***}	0.02
$\eta_{2,3}$	-0.22^{***}	0.03
$\eta_{2,4}$	-0.12^{***}	0.02
$\eta_{2,5}$	-0.30^{***}	0.01
$\eta_{3,1}$	-0.23^{***}	0.01
$\eta_{3,2}$	-0.09^{***}	0.02
$\eta_{3,3}$	0.15^{***}	0.02
$\eta_{3,4}$	-0.14^{***}	0.01
$\eta_{3,5}$	-0.12^{***}	0.03
$\eta_{4,1}$	0.72^{***}	0.05
$\eta_{4,2}$	0.38^{***}	0.02
$\eta_{4,3}$	0.40^{***}	0.01
$\eta_{4,4}$	0.41^{***}	0.01
$\eta_{4,5}$	0.72^{***}	0.10

[1] *Significance level:* * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(a) Chapter page (b) In-chapter comments page

Figure EC.3 A sample chapter page and bullet page.

(a) screenshot of a sample book chapter. The number appearing below each sentence indicates the number of in-chapter comments of this sentence; the number shown at the bottom of the page indicates the number of in-chapter comments of the current page. (b) screenshot of sample in-chapter comments' content page. The page shows text content of all the comments to the current sentence or page.

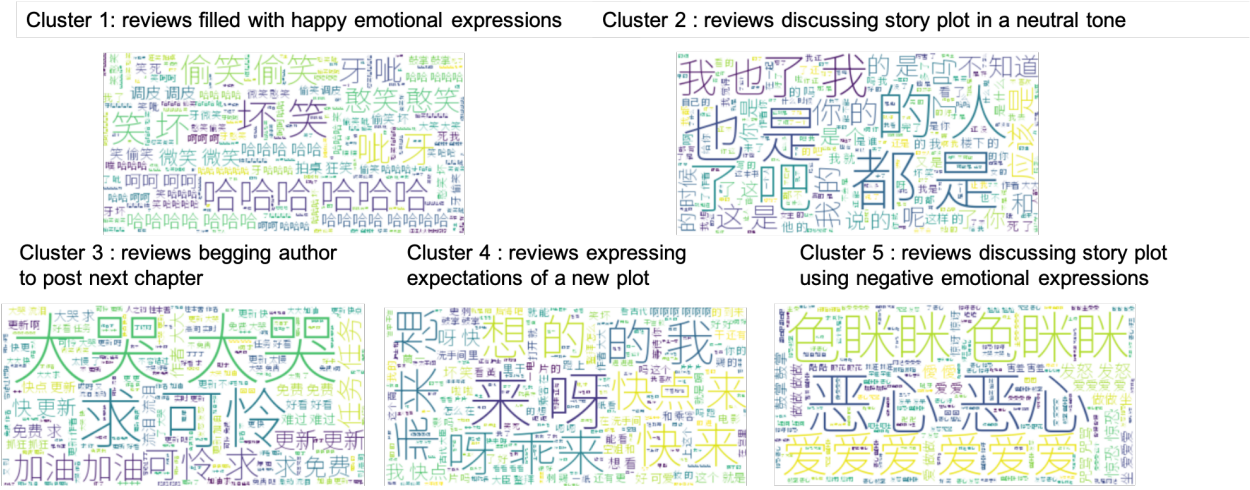

Figure EC.4 In-consumption comment topics in Chinese.

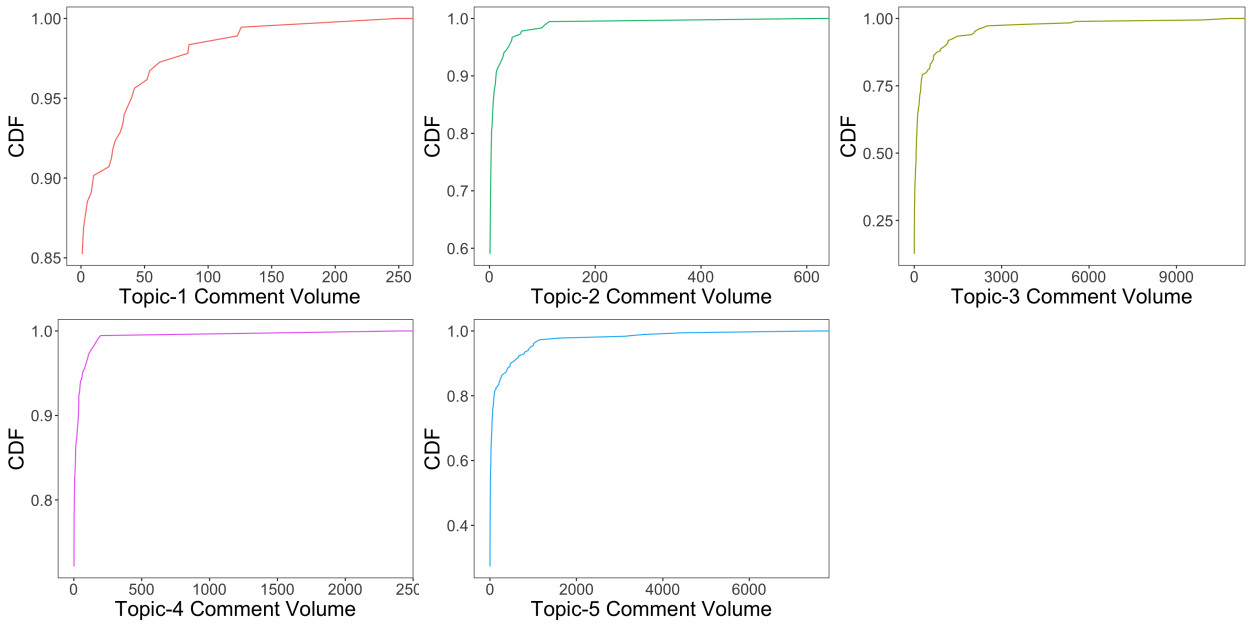

Figure EC.5 **CDF of the in-consumption comment volume across five topics within male sensational fiction**

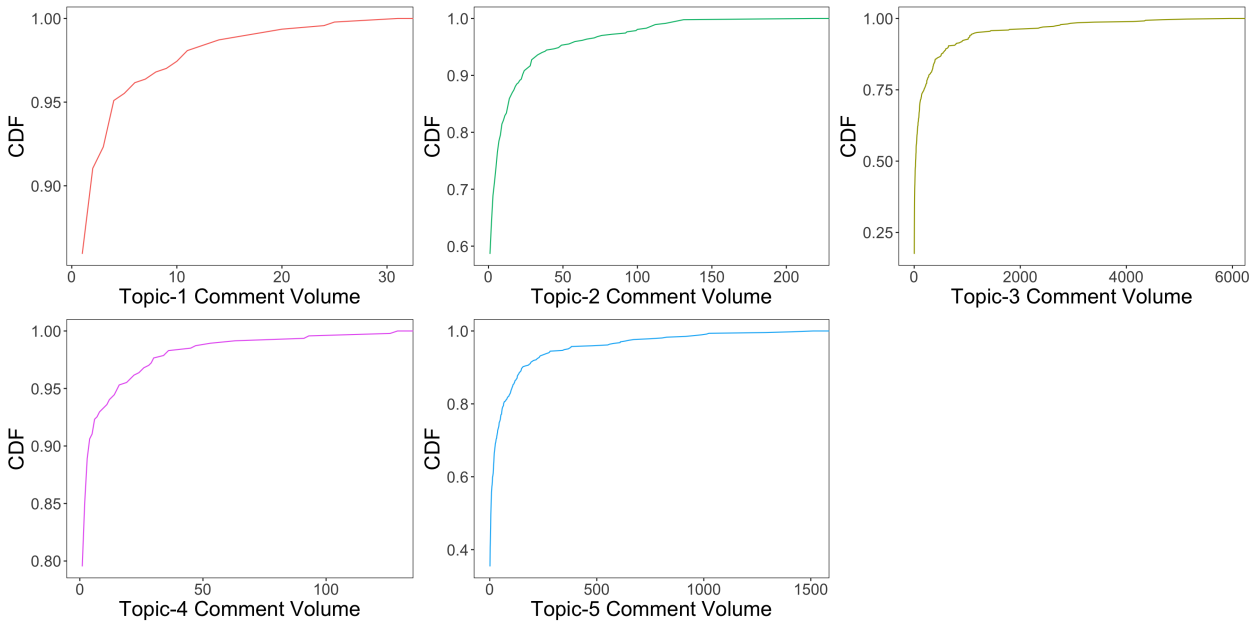

Figure EC.6 **CDF of the in-consumption comment volume across five topics within female sensational fiction**

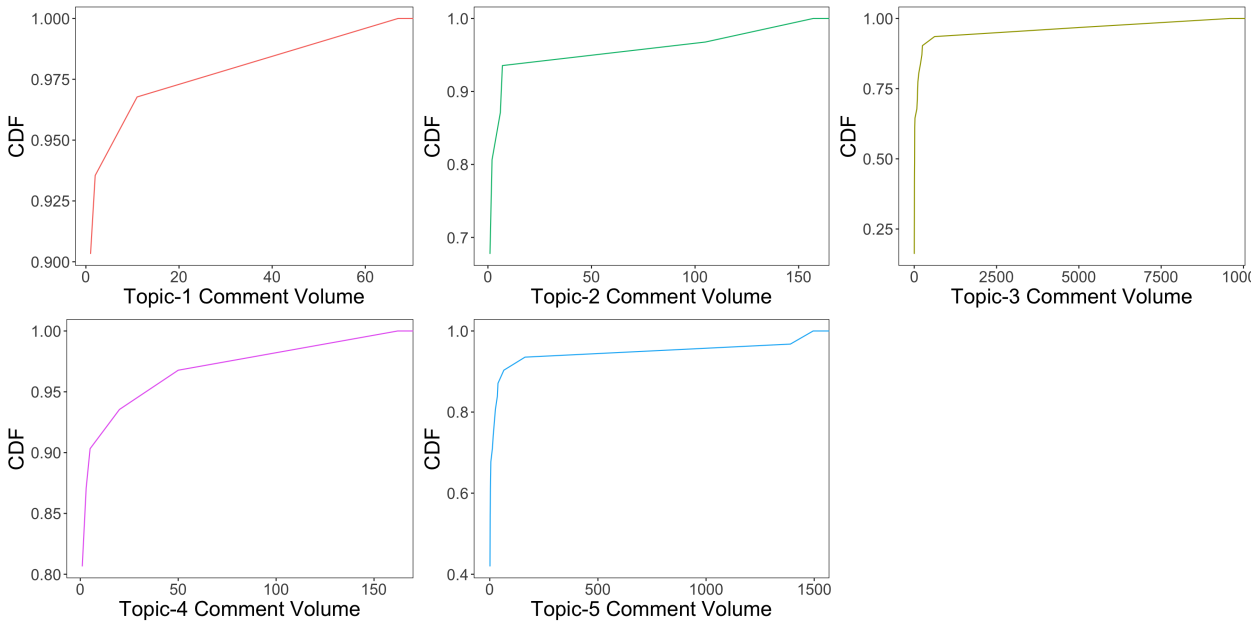

Figure EC.7 **CDF of the in-consumption comment volume across five topics within gender-neutral fiction**

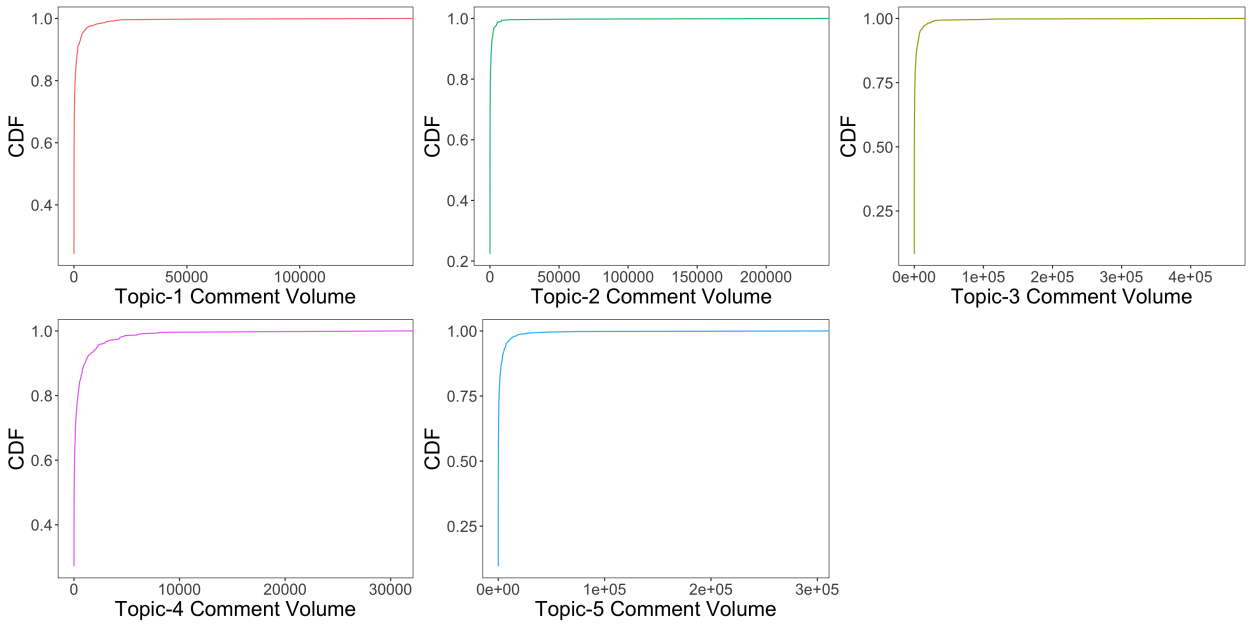

Figure EC.8 **CDF of the in-consumption comment volume across five topics within teen fiction**

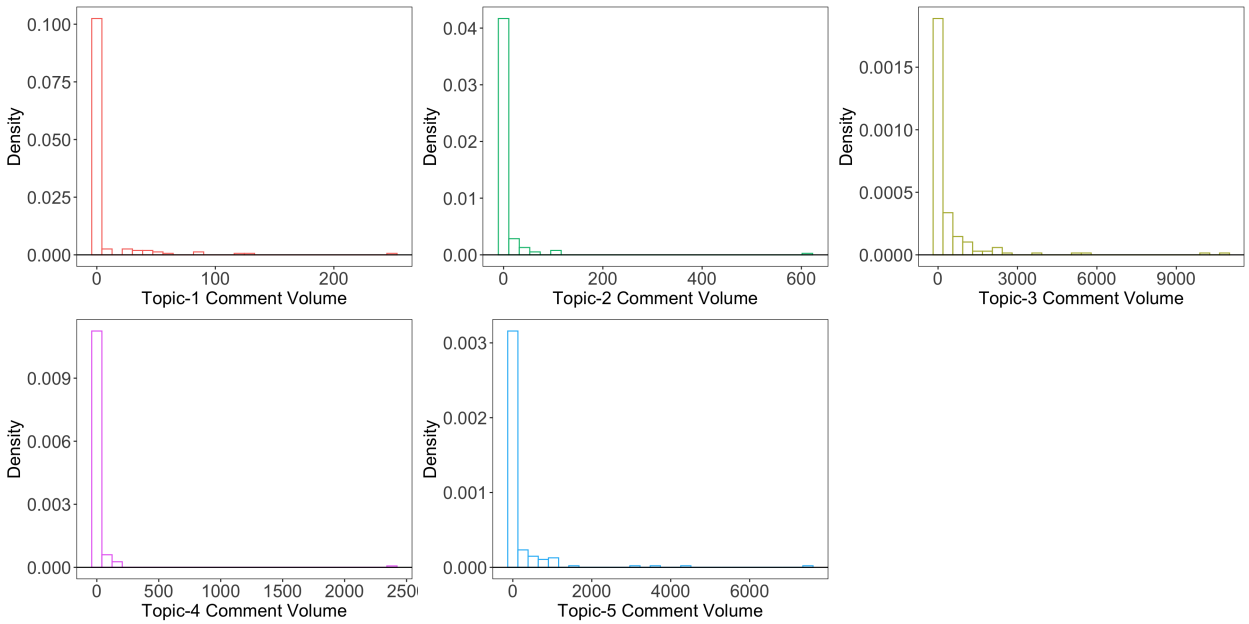

Figure EC.9 **Histogram of the in-consumption comment volume across five topics within male sensational fiction**

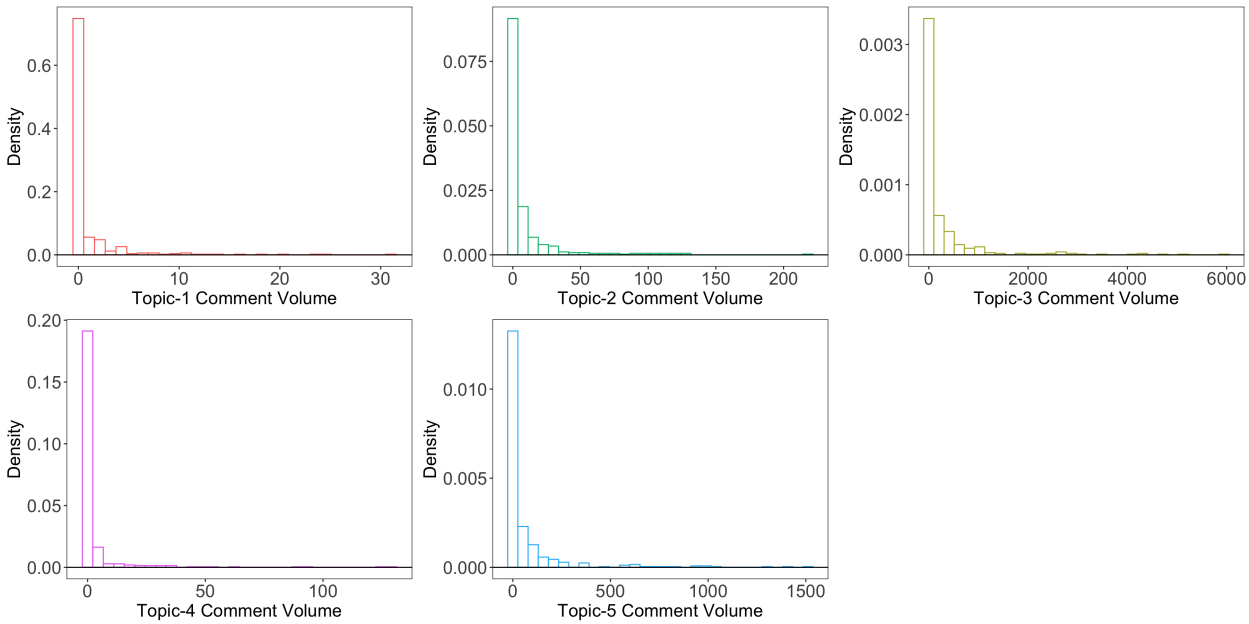

Figure EC.10 **Histogram of the in-consumption comment volume across five topics within female sensational fiction**

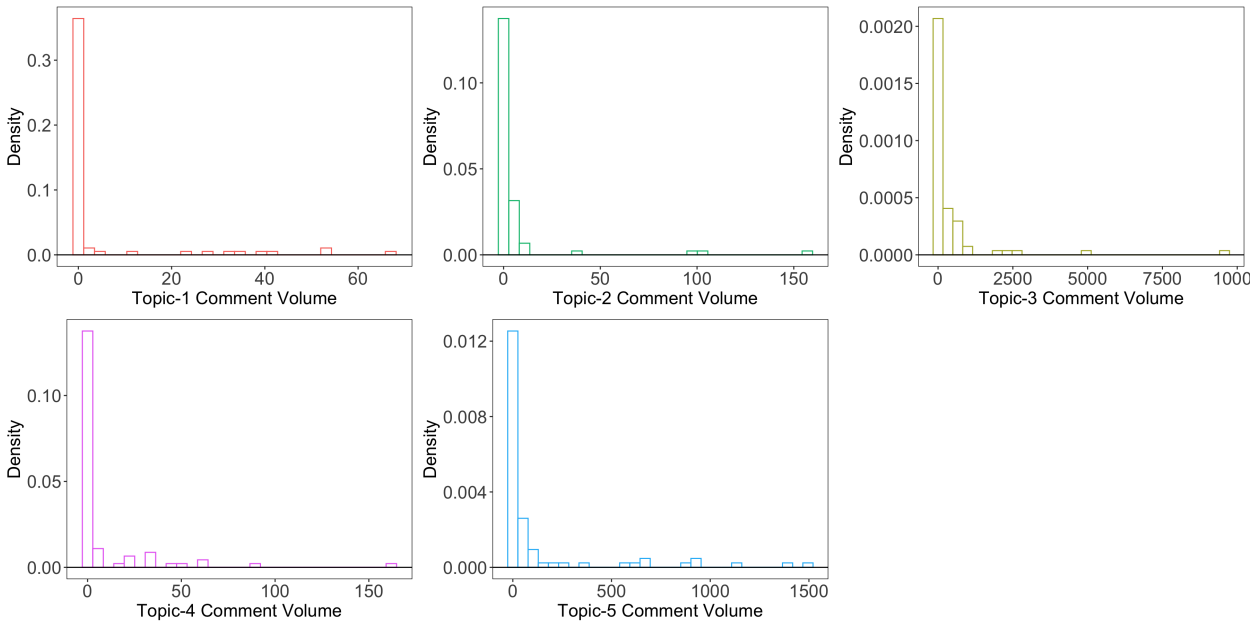

Figure EC.11 **Histogram of the in-consumption comment volume across five topics within gender-neutral fiction**

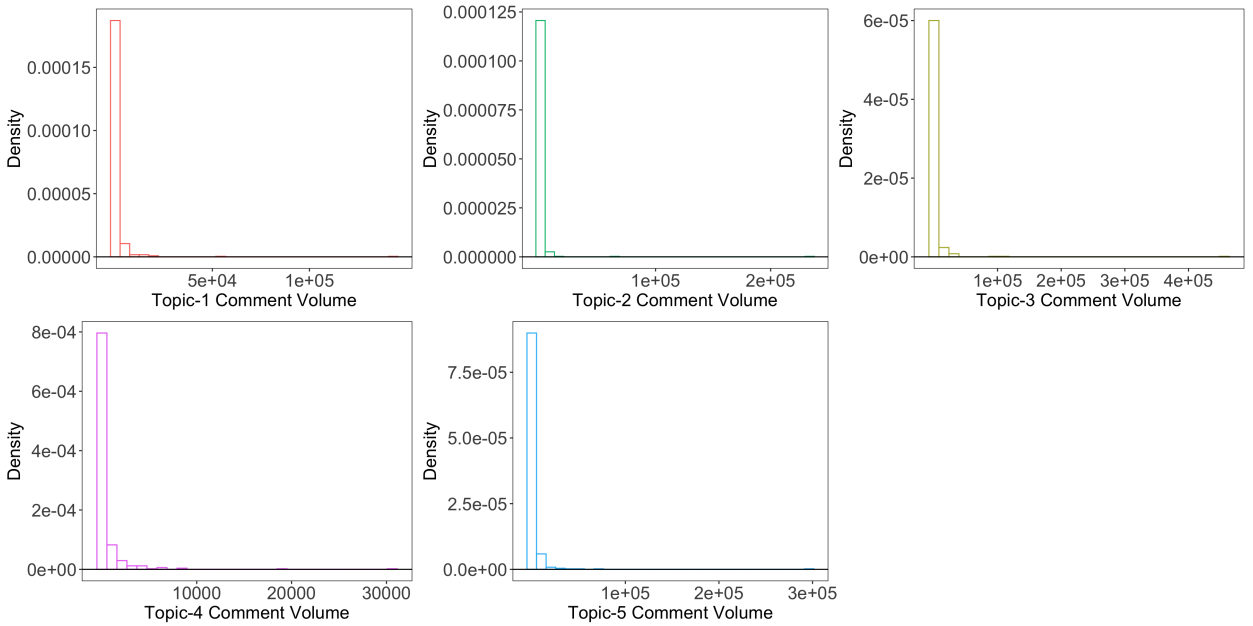

Figure EC.12 **Histogram of the in-consumption comment volume across five topics within teen fiction**