Web Appendix for

An Empirical Investigation of a Digital Platform's Release Strategy for Serialized Content

Table of Contents

Section 1. Distributions of the Percentage of Chapters Purchased across Different Groups of Consumer-Book Combinations with Different Initial Availability

We discuss the purchase patterns of consumers who start purchasing chapters of books at different points in time in the publication cycles of those books. We split all consumer–book pairs into five groups based on the proportion of chapters available when the consumer starts purchasing the book[1]: group A starts to purchase chapters of books when up to 25% of their chapters are released, group B starts to purchase chapters of books when 26% to 50% of their chapters are released, group C starts to purchase when 51% to 75% of their chapters are released, group D starts to purchase when 76% to 99% of their chapters are released, and group E starts to purchase chapters of books after all chapters are released. As we move from Group A to Group E, there is a higher availability of chapters of the focal book when the consumer makes her first purchase of chapters in that book. We first investigate whether higher availability of chapters at the time of the first purchase of a book results in the consumer purchasing a higher overall number of chapters of that book over her purchase history. In the table below, we report the size, mean and standard deviation of the proportion of the overall number of chapters in a book purchased (across the entire data period) for the entire sample and each group. Observe that the latter groups tend to read a higher percentage of books than the earlier groups.

Proportion of Chapters Purchased Across Groups of Consumer-Book Combinations

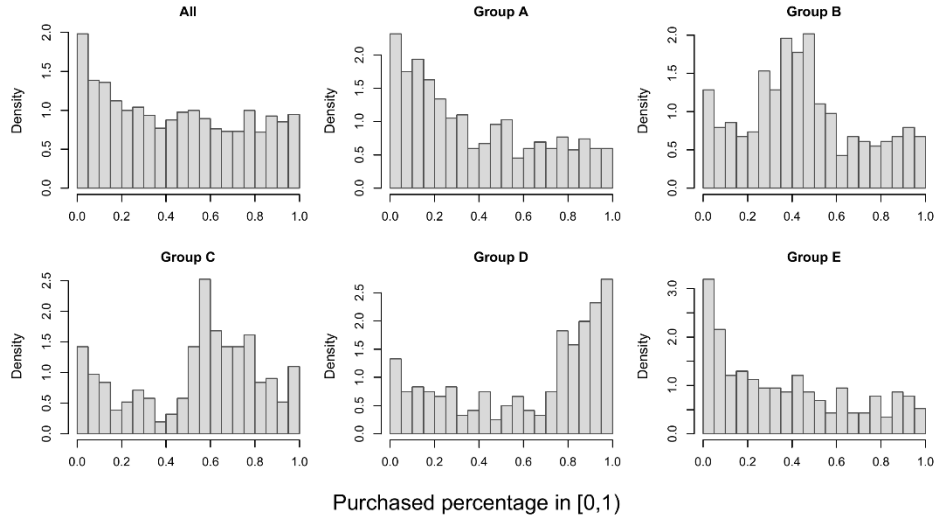
	All	Group A	Group B	Group C	Group D	Group E
Percentage of Observations	100%	40.6%	13.6%	12.4%	10.7%	22.6%
Mean	0.724	0.673	0.661	0.708	0.780	0.834
Standard Dev	0.351	0.377	0.337	0.315	0.312	0.318

In Figure A1, we plot the histogram of the percentages of chapters purchased in a book, both for the entire sample and for each group separately[2]. The following are key observations regarding how the attrition and binge consumption rates vary across the five groups. First, the figure shows that when reading ongoing books (groups A-D), once consumers catch up with a book's publication, they are more likely to stop purchasing future chapters than they were before catching up. For example, group B is more likely to stop reading between 25% and 50% (where they are likely to have caught up with the chapter releases) than between 0% and 25% (where these chapters are already released at the beginning of their purchase), and a similar pattern holds for groups C and D. Thus, consumers are less likely to stop reading when many chapters are available.

[1] Note that we have classified the 5 groups at the consumer-book level and not at the consumer level. This is because our data shows sufficient within-consumer heterogeneity in terms of when they start reading books (more on this in Section 3.4). If we instead classified the 5 groups at the consumer level, we would get very small sizes of all the groups, which would render any analysis of their purchase patterns less meaningful.

[2] In Figure A1, we intentionally omitted the highest bar value on the x-axis (purchase percentage = 100%) as its size is significantly larger than that of the other bars, which would visually shrink the other bars.

Figure A1. Distributions of the Percentage of Chapters Purchased Across Groups of Consumer-Book Combinations



Second, when reading an ongoing book, the attrition rates are highest when consumers catch up with its publication. The attrition rate after catching up decreases as we progress from groups B to group D, indicating that this attrition rate decreases as the number of chapters read before catching up increases. *Third,* when reading an ongoing book, consumers tend to purchase more chapters within a short time frame before catching up with its publication than after they have caught up. On average, they purchased 31 chapters over 2.2 purchase trips across 4.1 days before catching up, compared to 42 chapters over 32.4 trips across 47.8 days afterward. This suggests that the greater the availability of a book's chapters, the more chapters will be purchased on a given purchase trip. Moreover, if we assume that a consumer only purchases additional chapters of a book after completing all the previously purchased chapters,[3] it also suggests that the greater the number of chapters purchased of a given book on a purchase trip, the greater will be the subsequent consumption rate (and thereby the extent of binge consumption) of that book. *Fourth,* for an ongoing book, if we exclude the period in which consumers catch up with the publication, the attrition is highest when they read the first 20% of the chapters. *Finally*, for a completed book (group E), the attrition rate peaks during the first 20% of its chapters. On average, these consumers purchased 75 chapters across 1.5 purchase trips over a period of 2.5 days. This suggests that the consumption rate (and thus the extent of binge consumption) is high among readers of completed books, surpassing that of ongoing books before catching up.

Next, we demonstrate the same results by showing how a user's purchase and consumption trajectories differ over time depending on the number of chapters of a book that have already been released when the consumer started reading it. We will compare the purchase and consumption trajectories of a given consumer across two ongoing books whose chapters she purchased at different points in time. Importantly, Figure A2 refers to the case where the consumer started purchasing an ongoing book early on (when 16% of its chapters were released), and Figure A3

[3] The rationale for this assumption is explained in Section 3 of the main paper. Given this assumption, the consumer's average consumption rate of a book's chapters between two consecutive purchases of that book in periods s and t can be calculated by dividing the number of chapters of that book purchased by the consumer in period s by the number of days in the duration of that period (i.e., between periods s and t).

refers to the case where the consumer started purchasing an ongoing book later on (when 49% of its chapters were released).

Figure A2 corresponds to early adoption, illustrating the purchase and consumption trajectories of a consumer who begins reading an ongoing book early on, specifically when 16% of its chapters have been published. Before the consumer catches up (purchasing 16% of the chapters already released), the consumer first purchases a small quantity of chapters to test out. If she likes it, the consumer then purchases a large quantity of chapters within a short period, indicating a high consumption rate (or binge consumption). After the consumer has caught up, she only purchases chapters as they are released. The resulting waiting time increases the propensity for churn. The consumer stops purchasing chapters of the book after t = 71 periods, having purchased 41% of the book's chapters.

Figure A3 corresponds to the late adoption of another book by the same consumer, showing the purchase and consumption trajectories of a consumer who starts reading an ongoing book later on. The figure shows similar trajectories to those in the first figure: before the consumer catches up (with 49% of the chapters already released), the consumer first purchases a small quantity of chapters to test them out. If she likes it, then the consumer purchases a large quantity of chapters over a short period, resulting in a high consumption rate (or binge consumption). After the consumer has caught up, she only purchases chapters as they are released. This resulting waiting time increases the propensity for churn. However, the churn probability is smaller than that in the first figure – the consumer stops purchasing chapters of the book after t = 183 periods, having purchased 98% of the book's chapters.

The two figures, A2 and A3, illustrate the purchase and consumption trajectories over time for a representative user, facilitating an intuitive understanding of patterns such as binge-reading and potential attrition at specific intervals in a book's lifespan.

Figure A2. Purchases and Consumptions for Consumer 329, who started book 70 at day 55 and 16% chapters released and stopped purchase at day 71 and 41% chapters.

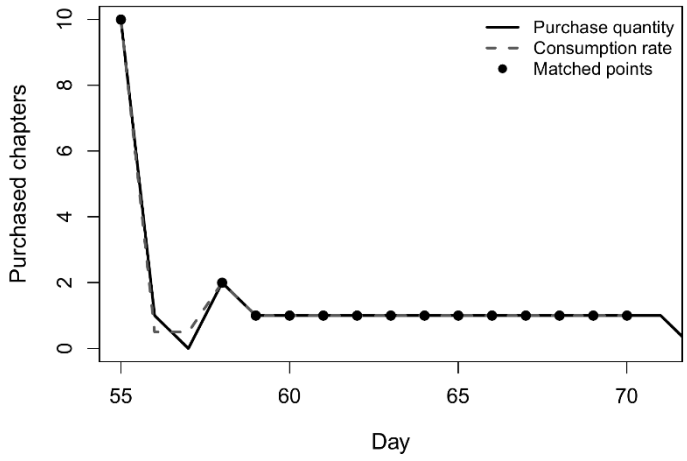
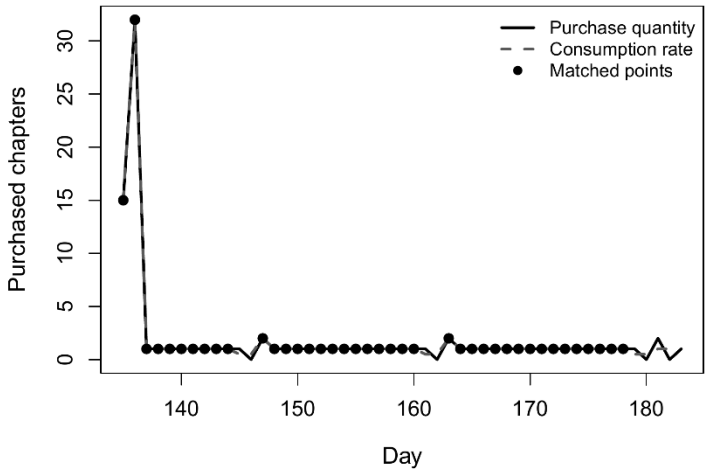


Figure A3. Purchases and Consumptions for Consumer 329, who started book 110 at day 135 with 49% chapters released and stopped purchase after day 183 and 98% chapters.



Section 2. Impact of Release Speed on Chapter Quality (Experienced vs. Inexperienced Writers)

For a given book, the release speed of chapters over time can be correlated with the quality of those chapters, for the reasons discussed in Section 3.11 of the paper. Regarding this source of endogeneity, note that our dataset consists of bestselling books by experienced writers (averaging 3 years of writing experience, with at least five books written before 2017). For such writers, we do not expect a significant correlation between a chapter's quality and the speed of its release.

To validate this, we run a regression in which the dependent variable is a chapter's *post-completion retention rate*, which is a measure of its quality. The retention rate of a chapter (k) is defined as amongst consumers who purchased the chapter (k), the percentage who purchase the next chapter $(k+1)$. We calculate the retention rate using only the purchase data after the full book is released (i.e., the *post-completion* retention rate), where consumers no longer face any availability constraint due to release progress. Therefore, the release speed of each chapter should not affect the post-completion retention rate, as long as the availability constraint on the demand side is met. The regressors are (1) the release speed associated with the past week, including the day when the focal chapter is uploaded, and (2) whether the book is from an experienced writer. The regression is conducted at the chapter level and is restricted to all books in the chosen genre released during our sample window. We control for the chapter index using second-degree polynomials, as well as whether the chapter is the last chapter of a book section, book-fixed effects, and month-fixed effects. The regression results in Table A1 show that the chapter retention rate is not significantly related to the release speed of the books in our sample. However, for other books from the same genre that are not from experienced writers, release speed has a positive and significant impact on chapter quality.

Table A1. The Effect of Release Speed on Chapter Quality

Dependent variable: Chapter retention rate (%)	
Log (No. of chapters released in the past week) * Experienced	0.11 (0.07)
Log (No. of chapters released in the past week) * Inexperienced	0.22** (0.08)
Chapter Index/100	0.68*** (0.09)
(Chapter Index/100)^2	-0.05*** (0.01)
Last chapter in a book section	-1.01 (0.63)
Book F.E.	✓
Month F.E.	✓

Note: Regressions are at the book-chapter level. Standard errors (in brackets) are clustered at the book levels. ***, **, and * indicate significance at the 0.1%, 1%, and 5% levels, respectively.

Section 3. The Effect of Release Speed on the Number of Comments Written by Readers

We investigate how a book's release speed can affect the number of comments that readers write about the book. We run regressions at the consumer-book-day level and test whether a consumer is more likely to leave comments (or an increased number of comments) when the book that she is following released chapters faster than when it is not, while controlling for consumer-FE and book specific month-FE and using the instrument to deal with potential endogeneity of release speed. The regression results in Table A2 show that neither consumers' propensity to comment nor the number of comments is significantly affected by book's release speed.

Table A2. The Effect of Release Speed on Comments

	(1)	(2)
Dependent Variable:	Whether to leave comment in a given day	Log (No. of comments in a given day + 1)
No. of chapters released in a day	0.018 (0.011)	0.018 (0.010)
Consumer F.E.	✓	✓
Book specific month F.E.	✓	✓
2SLS with instruments	✓	✓

Note: Regressions are at the consumer-book-day level. Standard errors (in brackets) are three-way clustered at the consumer, book, and day levels. ***, **, and * indicate significance at the 0.1%, 1%, and 5% levels, respectively.

Section 4. Validity of the Instrumental Variable

In this section, we discuss the validity of the instrument presented in Section 3 of the paper, specifically its relevance and whether it satisfies the exclusion conditions. We construct an instrument for the release speed, using the release speed of other books written by the same author at least 6 months earlier. We divide each book into five parts, each with an equal number of chapters, and calculate the average release speed for each quantile. To obtain the instrument variable for release speed of focal book j in week t, we first identify the quantile of the book to which the chapters released that week belong. For example, for book j in week t, if the chapters released are from the first quantile of book j, then the instrument value is the average release speed of chapters in the first quantile for all the previous (at least 6 months prior) books written by the author of book j.

The instrument is relevant because writers tend to have similar writing habits and a similar writing pace over time, and hence a similar cost structure for updating specific parts of their book chapters. The first-stage regression reports the incremental *F*-statistic for the instrument as 41.9, which suggests that we do not have a weak instrument problem.

The underlying conditions for the instrument to satisfy the exclusion restriction are that the time-varying demand shocks[4] In the nth quantile of the focal book, be uncorrelated with the corresponding time-varying demand shocks in the n^{th} quantile of the previous book(s) written by the same author. This is likely to hold because the writers in our data never release multiple books on the platform at the same time. Instead, they focus on writing one book at a time and take long breaks—at least several months—between books. Given a sufficient time gap of at least 6 months between the focal book and the previous books, the demand shocks in the n^{th} quantile of the focal book should be uncorrelated with those in the n^{th} quantile of older books by the same author.

Next, we will discuss some potential violations of the exclusion restriction. Before we discuss them, it is important to point out that the exclusion restriction will be violated if <u>all</u> the following three conditions hold true:

(i) The demand shocks in n^{th} quantile of the focal book are correlated with the corresponding demand shocks in the n^{th} quantile of previous book(s) written by the same author.
(ii) The demand shocks in the n^{th} quantile of the focal book are correlated with the release speed of that book in the n^{th} quantile.
(iii) The demand shocks under consideration are time varying (since we control for the time invariant demand shocks using book FE).

A potential violation of the exclusion restriction is as follows. There could be unobserved variation in readers' preferences over the different chapters of a book, and this unobserved variation could be similar across all books written by the same author. This could arise if a writer has a similar writing style/story formula across all her books. For instance, it could be the case that in all the books written by a writer, the stories reach their crescendos during the same n^{th} quantile of chapters of all books. This crescendo can lead to a temporary increase in readers' preferences when reading the same n^{th} quantile of chapters in all books written by the same writer (implying condition iii is satisfied). Since this temporary increase in preferences is unobserved by the econometrician, it will result in the aggregate demand shocks in the same quantiles to be correlated across all the different books written by the same writer (implying condition i is satisfied). Now if the writer takes these aggregate demand shocks into consideration when deciding the release speed of the next set of chapters, then these demand shocks will be correlated with the release speed. However, that is not so in the context of our data (implying that condition ii will not be satisfied). In our data, we only consider experienced writers, who are unlikely to make their chapter release decisions based on the popularity or quality of the most recently released chapters. As compared to inexperienced writers, experienced writers are likely to be more confident in their own skills and taste, value their creative independence more, and react less to readers' demand and comments.[5]

To validate the above, we run a regression in which the dependent variable is the release speed associated with the day that the chapter is uploaded. The regressor of interest is the previous chapter's *post-completion retention rate*, which is a measure of the previous chapter's popularity. The retention rate of a chapter (k) is defined as 'the percentage of consumers who purchase the next chapter $(k+1)$ amongst all the consumers who had purchased the chapter (k)'. We calculate

[4] The reason that we are only focusing on time varying demand shocks is because we control for book FE in our analysis.
[5] "Digital Literature: The Customer is God?" from People.cn. (http://it.people.com.cn/n/2014/0218/c1009-24386304.html)

the retention rate using only the purchase data after the full book is released (i.e., the *post-completion* retention rate), where consumers no longer face any availability constraint due to release progress. Therefore, the release speed of each chapter should not affect the post-completion retention rate, as long as the availability constraint on the demand side is met. The regression is restricted to the books in the chosen genre released in our sample window by experienced writers. We control for the book-fixed effects and month-fixed effects. The regression results in Table A3 indicate that the post-completion retention rate from the previous chapter is not significantly related to the subsequent release speed of chapters. This implies that the potential violation we discussed in the previous paragraph is not a source of concern in our data.

Table A3. The Effect of Popularity on Release Speed

Dependent variable: No. of days between two releases	
Average chapter retention rate of the previous release (%)	-0.017 (0.0098)
Book F.E.	✓
Month F.E.	✓

Note: Regressions are at the book-release day level. Standard errors (in brackets) are clustered at the book levels. ***, **, and * indicate significance at the 0.1%, 1%, and 5% levels, respectively.

Section 5. Reduced Form Evidence for the Net Effect of Notifications on New Book Purchases

We estimate the net effect of notifications on the purchase of new books. We run a regression with the DV being the number of new books purchased in each period (month). The regressors are (i) logged number of notifications received, (ii) the consumer's visit frequency in the last 7 days, (iii), average number of chapters previously purchased across all recently purchased books (averaged across books and all the days in the given period), (iv) average time since last purchase of all recently purchased books (averaged across books and all the days in the given period), and (v) the consumer FE. The results are shown in Table A4 below. Observe that the number of notifications has a positive impact on the new books.

Table A4. Net Effect of Notifications on New Books Purchased

	Dependent variable:
	Number of new books purchased in a period
	(1)
Log of No. of notifications received	0.335 *** (0.018)
Visit frequency in the last 7 days	0.116 *** (0.024)
Average No. of chapters previously purchased across all recently purchased books	-0.018 (0.028)
Average time since last purchase of all recently purchased books	0.505 *** (0.129)
Consumer F.E.	✓
Number of observations	4025
R-squared	0.30

Note: Regressions are at the consumer–monthly level. Standard errors (in brackets) are clustered at the consumer level. ***, **, and * indicate significance at the 0.1%, 1%, and 5% levels, respectively.

Section 6. Reduced Form Evidence for the Rationing Effect

We provide reduced-form evidence for the rationing effect. Unlike the product exploration effect, the rationing effect is conditional on a visit to the platform. We test the rationing effect by regressing the release speed of a book (*j*) in a day (*t*) on the number of new books tried by the consumer (*i*) in the day (*t*). We include the book and consumer FE, as well as the IV discussed earlier, to address the endogeneity of release speed. The regression analysis uses observations conditional on consumer *i* visiting the platform on a given day and conditional on consumer *i* purchasing chapters of book *j* before day *t* and catching up on the progress of book *j*'s release. This is done to ensure that the consumer faces an immediate constraint imposed by the release speed (as opposed to the scenario in which the consumer is not affected by the release speed, given that many of the available chapters have not yet been read). The results in Table A5 show that, conditional on a consumer following a book and visiting the platform, release speed has a negative and significant effect on product exploration within the same period. However, the coefficient is very small, which implies a small magnitude of the rationing effect - all else being equal, if the release speed decreases by 50%, each consumer would buy only an additional 0.96 book over the course of a year.

Table A5. Rationing Effect - The Effect of Release Speed on Purchase Variety

	Dependent variable:	
	Whether the consumer tries a new book	Log (No. of new books tried + 1)
	(1)	(2)
Average Release speed of books following: Chapters/day	-0.013 * (0.005)	-0.009 * (0.004)
Weekend	-0.002 (0.004)	-0.002 (0.003)
Consumer F.E.	✓	✓
2SLS with instruments	✓	✓
Number of observations	46,474	46,474
R-squared	0.03	0.03

Note: Regressions are at the consumer-day level. Standard errors (in brackets) are two-way clustered at the consumer and day levels. ***, **, and * indicate significance at the 0.1%, 1%, and 5% levels, respectively.

Section 7. Reduced Form Evidence for Heterogeneity in Consumers' Preferences on When to Start Reading Books

We investigate whether large and/or prolific segments of consumers exist in our data who have a strong preference for reading ongoing vs. completed books and when to start reading ongoing books. We report the fractions of consumers in the sample who have a high preference to start reading books when (i) up to 25% of their chapters have been released (Group 1), (ii) 26% to 50% of their chapters have been released (Group 2), (iii) 51% to 75% of their chapters have been released (Group 3), (iv) 75% to 99% of their chapters have been released (Group 4), and (v) 100% of their chapters have been released (Group 5).[6] For each of these groups of consumers, we also report the average number of chapters they purchased daily throughout the data period. This

6 Note that we have classified the groups 1-5 at the consumer level, which is different from the classification of the groups A-E in the Table in section 1 of the web appendix, which was at the consumer-book level.

distribution will help shed light on whether heterogeneity in consumers' preferences for reading ongoing vs. completed books and when to start reading ongoing books plays a significant role in influencing the platform's release strategy. For instance, if a sizable fraction of consumers falls into Group 3 and/or if consumers in Group 3 are avid readers, it would suggest that the platform would be better off implementing a hybrid release in which the first 51%-75% of chapters are released simultaneously and the rest are released sequentially.

To calculate the fraction of consumers in, say, Group 3, we calculate the fraction of the consumers in the sample who are such that a 'vast majority' of the books (i.e., more than 80% of the books) they started reading were those when 51% to 75% of their chapters were released. We report the relevant statistics of Groups 1-5 in Table A6 below. Observe that the sizes (i.e., fraction of consumers in each group) of groups 2-5 are small. While the size of Group 1 is larger, the average daily purchase of chapters is a lot smaller for Group 1. Moreover, the average number of chapters purchased daily is small in Groups 1 and 3. While the average number of chapters purchased is higher for Groups 2, 4, and 5, their sizes are small. This suggests that consumers' preferences for reading ongoing vs. completed books and when to start reading ongoing books *may not* play a strong role in influencing the platform's release strategies.

Table A6. Fraction of Consumers and their Average Daily Purchases in Groups 1-5

Group Number	Fraction of Consumers	Average Daily Purchase of Chapters by Consumers
1	0.112	3.30
2	0.007	11.37
3	0.003	3.37
4	0.002	8.22
5	0.024	12.31

Section 8. Justifications of Model Assumptions discussed in Section 4.1 of the paper

In section 4 of the paper, we mention the following four assumptions that we made in the second stage of the model (the utility maximization): (i) It assumes that purchase quantities (i.e., the number of chapters purchased) of the inside goods, x_{it}, are continuous variables. (ii) It assumes that the price per chapter of a book is the same across all chapters of that book. (iii) It does not have a time constraint. (iv) It assumes that consumers are myopic decision makers. We justify our assumptions as follows.

8.1 Assumption of purchase quantities of inside goods, x_{it}, being continuous.

Justification: This assumption aligns with prior literature that derives purchase conditions from a discrete-continuous utility-maximizing framework (e.g., Chiang 1991; Nair, Dube, and Chintagunta 2005; Song and Chintagunta, 2007). There are two reasons for this assumption:

i. Nair, Dube, and Chintagunta (2005) empirically compared the predictions of the continuous quantity assumption in these discrete-continuous models with those of the discrete quantity. They find that the results are similar across the two, implying that the continuous quantity assumption is robust.

ii. This assumption allows us to use Kuhn-Tucker (KT) conditions to solve the utility maximization problem. If quantity is not continuous, then we can no longer use KT conditions, and we would need to resort to comparing utilities across all possible combinations of purchases and non-purchases of books and their chapters. This is an NP-hard problem since it

will require comparing utilities across all combinations of purchase/non-purchase of 138 books and across all their possible quantity choices of chapters for each alternative (that can range from 1 to 100 chapters per book). It is important to note that prior literature has relaxed the continuous quantity assumption only for simple cases when the number of brands is less than 5 and the number of units purchased per brand is less than 3.

8.2 Assumption of Price per chapter to be the same across all chapters in the book.

Justification: There are two reasons for this assumption:

i. The platform charges a fixed price per 1,000 characters, which remains constant over time. In our data, the average within-book coefficient of variation in the number of characters per chapter is 0.15, indicating that the chapter-level price variation for a given book is relatively limited. Thus, we assume the price per chapter in a book to be the same across all chapters.

ii. This assumption allows us to use Kuhn-Tucker conditions to solve the utility maximization problem. If the price per chapter changes for each chapter, then we can no longer use the Kuhn-Tucker conditions and we would need to resort to comparing the utilities across all possible combinations of purchases and non-purchases of books and their chapters in the entire set. This is an NP-hard problem since it will require comparing utilities across all combinations of purchase/non-purchase of 138 books and across their quantity choices that can range from 1 to 100 chapters per book on a given trip. It is important to note that the prior literature has relaxed the constant per unit price assumption only for simple cases when the number of brands is less than 5 and the number of units purchased per brand is less than 3.

8.3 Assumption of Consumers as Myopic Decision Makers

Justification: We model readers as myopic decision makers for three reasons.

First, purchasing online books is a low-involvement purchase decision, and there is minimal storage cost, no price variation, and no quantity discount on the platform. Strategically altering purchase timing cannot benefit consumers, as it does in many other purchase settings.

Second, given that we use a discrete–continuous model with multiple discreteness with 138 alternatives at the purchase stage, it isn't easy to incorporate consumer's forward-looking behavior. To the best of our knowledge, no prior paper has modeled consumer's forward-looking behavior with even one alternative in a discrete continuous choice model.

Third, we provide reduced-form evidence for our assumption. Specifically, we examine the possibility that consumers strategically wait for a certain minimum number of chapters to be released before purchasing them, so that they can read or binge-consume those multiple chapters together**.** Strategic waiting can be of two types: (i) Even after catching up with the progress of an ongoing book, consumers may prefer to wait for a minimum number of its additional chapters to be released, so that they can purchase all those chapters in bulk, which would then allow them to binge-read the chapters later. (ii) Consumers prefer to purchase fully published books, which will allow them to binge-read them later. We next discuss each of these two types.

For the first type of strategic waiting behavior, consider the consumer–book combinations in which the consumer begins to purchase the chapters of a book before its publication is complete. For these consumer-book combinations, we find in the data that after a consumer catches up with an ongoing book's progress (that is, when the consumer purchases the most recently released chapter of the book), 98% of subsequent purchases happen when the next set of chapters are released. This indicates that these consumers do not engage in strategic waiting after catching up.

For the second type of strategic waiting behavior, we look at consumer–book combinations in which the consumer begins to purchase book chapters after publication is complete. For such combinations, most of the consumers start purchasing the completed book within 2 weeks of the book's finale, and they constitute approximately 16% of all consumers – book combinations. There are multiple reasons why a consumer may start reading a book after its publication is complete: strategic waiting to binge consume it later, a reduction in uncertainty, and the consumer not being aware of the book before it was fully published. To distinguish the effect of strategic waiting from other reasons, we run a regression in which we regress the number of new readers who start reading a book 2 weeks before and after its finale. The regressors are a dummy for after-finale release, book-fixed effects, and whether the book was on the platform's best-seller list. The variable "whether the book was on the best-seller list on the platform's homepage" controls for any potential decrease in uncertainty and consumer awareness about the book. The post-finale time dummy would thus capture the effect of strategic waiting on demand. The regression results are in Table A7. Based on the estimates of the dummy for the after-finale release, we find that the number of readers who start reading due to strategic waiting is 178.6. Dividing this number by the average number of readers per book, we find that the fraction of consumer–book combinations in our data that may implicate strategic waiting is 3.7%. Because this is a small fraction, our structural model assumes that consumers are myopic.

Moreover, for this small fraction, we capture their strategic waiting behavior in a reduced form fashion by including the variable 'an indicator of whether book *j* is fully published' in a book's baseline utility, and allow its coefficient to vary across different consumer segments.

Table A7. The Effect of Finale Release on New Readership

Dependent Variable:	No. of new readers who pick up book *j* in day *t*
After finale release of the book	0.3271 (0.092) **
Being on front-page ranking	0.897 (0.307) **
Book F.E.	✓

Note: Regressions are at the book-day observation level, which is restricted to 28 days for each book, including the 14 days before and after each book's final release. Standard errors (in brackets) are two-way clustered at the book-week level. ***, **, and * indicate significance at the 0.1%, 1%, and 5% levels, respectively.

8.4 Not Having a Time Constraint in the Utility Maximization

Justification: The time constraint is not needed in our utility maximization. Note that the reading speeds will enter through the weights for the quantities of books in the time constraint. Since the econometrician does not observe these, they will get subsumed into the intercept in the term α_{ijt} in the Kuhn-Tucker conditions in equations (4.6)–(4.8). Thus, including the time constraint will not affect our estimation results.

Section 9. Maximum Likelihood Estimation

We first derive the observational likelihood in the absence of the endogeneity issue. We start by deriving the second-stage likelihood and then incorporate the first-stage likelihood. Following that, we incorporate the control function into the observational likelihood to deal with endogeneity of release speed.

9.1 Specification of the Observational Likelihood without Endogeneity Correction

Second-stage Observational Likelihood. Recall from Section 4 of the main paper that any observation can be classified into three groups. The conditions for the three groups follow from equations (4.6)-(4.8) of the paper as:

Group 1: Conditions for non-purchase of books $j = 1..n_1$:

$$Inv_{ijt}>0 \text{ or } \{Inv_{ijt}=0,\ \alpha_{ijt} - p_j - \beta_i \rho s_{it} \leq 0\} \quad \text{(TA.1)}$$

Group 2: Conditions for purchase at the availability constraint for books $j = n_1 + 1..n_1 + n_2$:

$$\{Inv_{ijt}=0,\ \alpha_{ijt} - p_j - \beta_i(1-\rho)c_{ijt} - \beta_i \rho s_{it} \geq 0\} \quad \text{(TA.2)}$$

Group 3: Conditions for purchase at less than the availability constraint for books $j = n_1 + 1..n_1 + n_2$:

$$\{Inv_{ijt}=0,\ \alpha_{ijt} - p_j - \beta_i(1-\rho)x_{ijt} - \beta_i \rho s_{it} = 0\} \quad \text{(TA.3)}$$

In the above equations, the expression for α_{ijt} follows from equation (4.2) of the paper as

$$\alpha_{ijt} = \alpha_{i0} + \alpha_j + \alpha_i^T Z_{ijt} + \varepsilon_{ijt} \quad \text{(TA. 4)}$$

Recall that the econometrician's error term ε_{ijt} is assumed to have a standard normal distribution, which is IID across consumers and over time, but can be correlated across books. We assume the correlation to be the same across all pairs of books and equal to ρ_ε. We represent these equi-correlated errors as

$$\varepsilon_{ijt} = v_{ijt} + v_{i0t} \quad \text{(TA.5)}$$

where v_{ijt} is IID across all i, j, and t, v_{i0t} is IID across all i and t but the same across the books for a given consumer-period combination, and $v_{ijt} \perp v_{i0t}$. Their distributions are given by $v_{ijt} \sim N(0, 1-\rho_\varepsilon)$ and $v_{i0t} \sim N(0, \rho_\varepsilon)$. The expressions for $Inv_{ijt}>0$ and $Inv_{ijt}=0$ follow from equation (4.11) of the paper as:

$$Inv_{ijt}>0 \text{ iff } x_{j,l_j(t)} \cdot \left(1 - b \cdot \Delta t_{j,t}\right) - (a + \sigma_c \varepsilon_{cj,l_j(t)})\Delta t_{j,t} > 0 \quad \text{(TA.6)}$$

$$Inv_{ijt}=0 \text{ iff } x_{j,l_j(t)} \cdot \left(1 - b \cdot \Delta t_{j,t}\right) - (a + \sigma_c \varepsilon_{cj,l_j(t)})\Delta t_{j,t} \leq 0 \quad \text{(TA.7)}$$

Both conditions TA.7 and TA.8 are conditional on the per period consumption rate to be non negative, that is, $a + \sigma_c \varepsilon_{cj,l_j(t)} + b x_{j,l_j(t)} \geq 0$. Substituting the above expressions for the expressions for α_{ijt}, $Inv_{ijt}>0$ and $Inv_{ijt}=0$ into equations (TA.1)-(TA.3) of the paper, we obtain the following conditions for groups 1–3.

Group 1: Conditions for non-purchase of books $j = 1..n_1$:

$$\left\{\begin{matrix} x_{j,l_j(t)} \cdot \left(1 - b \cdot \Delta t_{j,t}\right) - (a + \sigma_c \varepsilon_{cj,l_j(t)})\Delta t_{j,t} \leq 0, \\ v_{ijt} \leq \beta_i \rho_i s_{it} + p_j - \alpha_{0i} - \alpha_j - \alpha_i^T Z_{ijt} - v_{i0t} \end{matrix}\right\} \cup$$
$$\left\{x_{j,l_j(t)} \cdot \left(1 - b \cdot \Delta t_{j,t}\right) - (a + \sigma_c \varepsilon_{cj,l_j(t)})\Delta t_{j,t} > 0\right\} \quad \text{(TA.8)}$$

Group 2: Conditions for purchase at the availability constraint for books $j = n_1 + 1..n_1 + n_2$:

$$\left\{\begin{matrix} x_{j,l_j(t)} \cdot \left(1 - b \cdot \Delta t_{j,t}\right) - (a + \sigma_c \varepsilon_{cj,l_j(t)})\Delta t_{j,t} \leq 0, \\ v_{ijt} \geq \beta_i(1-\rho_i)c_{ijt} + \beta_i \rho_i s_{it} + p_j - \alpha_{0i} - \alpha_j - \alpha_i^T Z_{ijt} - v_{i0t} \end{matrix}\right\} \quad \text{(TA.9)}$$

Group 3: Conditions for purchase at less than the availability constraint for books $j = n_1 + 1..n_1 + n_2$:

$$\begin{Bmatrix} x_{j,l_j(t)} \cdot \left(1 - b \cdot \Delta t_{j,t}\right) - (a + \sigma_c \varepsilon_{cj,l_j(t)})\Delta t_{j,t} \le 0, \\ v_{ijt} = \beta_i(1-\rho_i)x_{ijt} + \beta_i \rho_i s_{it} + p_j - \alpha_{0i} - \alpha_j - \alpha_i^T Z_{ijt} - v_{i0t} \end{Bmatrix} \quad \text{(TA.10)}$$

where ε_c is a standard IID normal random measurement error in the inventories. The error v_{ijt} is IID across all i, j, and t and is distributed as $v_{ijt} \sim N(0, 1-\rho_\varepsilon)$, and v_{i0t} is IID across all i and t but the same across the books for a given consumer–period combination and is distributed as $v_{i0t} \sim N(0, \rho_\varepsilon)$. The conditions in (TA.8) – (TA.10) are conditional on the per-period consumption rate being non-negative, that is, $a + \sigma_c \varepsilon_{cj,l_j(t)} + b x_{j,l_j(t)} \ge 0$. The contribution of groups 1 and 2 to the second-stage observational likelihood follows from Equations (TA.8) and (TA.9) as

$$L_{it,g1}^{stage\ 2}\left(x_{ijt} \middle| \theta_i, v_{i0t}\right) = \prod_{j=1}^{n} \left[\begin{array}{c} \dfrac{\Phi\left(\dfrac{x_{j,l_j(t)} \cdot \left(1 - b \cdot \Delta t_{j,t}\right) - a\Delta t_{j,t}}{\sigma_c \Delta t_{j,t}}\right) - \left(\dfrac{-a - b x_{j,l_j(t)}}{\sigma_c}\right)}{1 - \left(\dfrac{-a - b x_{j,l_j(t)}}{\sigma_c}\right)} + \\ \left(\dfrac{1 - \Phi\left(\dfrac{x_{j,l_j(t)} \cdot \left(1 - b \cdot \Delta t_{j,t}\right) - a\Delta t_{j,t}}{\sigma_c \Delta t_{j,t}}\right)}{1 - \left(\dfrac{-a - b x_{j,l_j(t)}}{\sigma_c}\right)} \right) \times \Phi\left(\dfrac{\beta_i \rho_i s_{it} - \alpha_{0i} - \alpha_j - \alpha_i^T Z_{ijt} - v_{i0t}}{\sqrt{1-\rho_\varepsilon}}\right) \end{array} \right] \quad \text{(TA.11)}$$

$$L_{it,g2}^{stage\ 2}\left(x_{ijt} \middle| \theta_i, v_{i0t}\right) = \prod_{j=n_1+1}^{n_1+n_2} \left[\begin{array}{c} \left(\dfrac{1 - \Phi\left(\dfrac{x_{j,l_j(t)} \cdot \left(1 - b \cdot \Delta t_{j,t}\right) - a\Delta t_{j,t}}{\sigma_c \Delta t_{j,t}}\right)}{1 - \left(\dfrac{-a - b x_{j,l_j(t)}}{\sigma_c}\right)} \right) \times \\ \left(1 - \Phi\left(\dfrac{\beta_i(1-\rho_i)c_{ijt} + \beta_i \rho_i s_{it} - \alpha_j - \alpha_{0i} - \alpha_i^T Z_{ijt} - v_{i0t}}{\sqrt{1-\rho_\varepsilon}}\right)\right) \end{array} \right] \quad (TA.12)$$

In Equations (TA.11) and (TA.12), the likelihoods are conditional on the normal random variable v_{i0t} and the parameters θ_i that vary across consumers $i = 1..I$. Also note that we remove the per-chapter price term p_j in the likelihoods in Equations (TA.11) and (TA.12) because there is no change in the price over time. Thus, the price term merely rescales the brand intercept α_j in the baseline preference. For group 3, the joint density of the observed quantities is obtained through the transformation of variables from v_{ijt} to x_{ijt} from equation (TA.10) as

$$L_{it,g3}^{stage\ 2}\left(x_{ijt} \middle| \theta_i, v_{i0t}\right) = \frac{\beta_i^{n_3}\left(1 + \rho_i(n_3 - 1)\right)(1-\rho_i)^{n_3 - 1}}{\left(\sqrt{1-\rho_\varepsilon}\right)^{n_3}}$$

$$* \prod_{j=n_1+n_2+1}^{n_1+n_2+n_3} \left[\left(\frac{1-\Phi\left(\frac{x_{j,l_j(t)} \cdot (1 - b \cdot \Delta t_{j,t}) - a\Delta t_{j,t}}{\sigma_c \Delta t_{j,t}} \right)}{1 - \left(\frac{-a - b x_{j,l_j(t)}}{\sigma_c} \right)} \right) \times \phi\left(\frac{\beta_i(1-\rho_i)x_{ijt} + \beta_i \rho_i s_{it} - \alpha_{0i} - \alpha_j - \alpha_i^T Z_{ijt} - v_{i0t}}{\sqrt{1-\rho_\varepsilon}} \right) \right] \quad (TA.13)$$

where ϕ is the standard normal probability density function, and the first term is the Jacobian of the transformation from v_{ijt} to x_{ijt}. The observational likelihood for the second stage of consumer i at period t conditional on v_{i0t} and θ_i (which is $L_{it}^{stage\ 2}(x_{ijt}|\theta_i, v_{i0t})$) is a product of the RHS of Equations (TA.11), (TA.12), and (TA.13) as

$$L_{it}^{stage\ 2}(x_{ijt}|\theta_i, v_{i0t}) = \left(L_{it,g1}^{stage\ 2}(x_{ijt}|\theta_i, v_{i0t})\right)^{d_{1it}} \left(L_{it,g2}^{stage\ 2}(x_{ijt}|\theta_i, v_{i0t})\right)^{d_{2it}} \left(L_{it,g3}^{stage\ 2}(x_{ijt}|\theta_i, v_{i0t})\right)^{d_{3it}} \quad (TA.14)$$

Where d_{1it} take a value of 1 if the observation lies in group 1 and 0 otherwise, d_{2it} take a value of 1 if the observation lies in group 2 and 0 otherwise, and d_{3it} take a value of 1 if the observation lies in group 3 and 0 otherwise. Now note that this second-stage likelihood, $L_{it}^{stage\ 2}(x_{ijt}|\theta_i, v_{i0t})$, in equation (TA.14) is conditional on the normal random error $v_{i0t} \sim N(0, \rho_\varepsilon)$.

Incorporating the First-stage Model into the Observational Likelihood. We first specify the probability of visiting the platform. Following that we will discuss how to incorporate the platform visit probability into the second-stage likelihood to get the observational likelihood. We now incorporate the first-stage platform visit decision into the observational likelihood.

$$Pr_{it}(Visit) = Pr(\gamma_0 + W_{it}\gamma_1 + \eta_{it} \geq 0) \quad (TA.15)$$

where η_{it} is a standard normal random variable. We next allow for the error η_{it} in the first stage to be correlated with the error ε_{ijt} in the second stage with a correlation of ρ_η. We allow for this correlation by representing η_{it} as

$$\eta_{it} = \omega_{it} + \frac{\rho_\eta}{\rho_\varepsilon} v_{i0t} \quad (TA.16)$$

where ω_{it} is IID across all i and t, v_{i0t} is the error from the second stage that is IID that is across all i and t but the same across the books for a given consumer-period combination, with $\omega_{it} \perp v_{i0t}$ and $\omega_{it} \perp v_{ijt}$. Note that the correlation between η_{it} in the first stage and the error ε_{ijt} in the second stage stems from the common error v_{i0t} in the expressions of both η_{it} and ε_{ijt} in equations (TA.16) and (TA.5) respectively. Recall that the distribution of v_{i0t} is $v_{i0t} \sim N(0, \rho_\varepsilon)$. Since η_{it} has a standard normal distribution, it follows from equation (TA.16) that ω_{it} will be distributed as $\omega_{it} \sim N(0, 1 - \frac{\rho_\eta^2}{\rho_\varepsilon})$. Given this, we get the probability of platform visit by consumer i at time t conditional on the error v_{i0t} as

$$Pr_{it}(Visit|\theta_i, v_{i0t}) = \Phi\left((\gamma_0 + W_{it}\gamma_1 + \frac{\rho_\eta}{\rho_\varepsilon} v_{i0t}) \Big/ \sqrt{1 - \frac{\rho_\eta^2}{\rho_\varepsilon}} \right) \quad (TA.17)$$

Given the visit probability in (TA.17) and the likelihood for the second stage in (TA.14), we can write the observational likelihood conditional on the error v_{i0t} as

$$L_{it}(x_{ijt}|\theta_i, v_{i0t}) = \begin{cases} 1 - P_{it}(Visit|\theta_i, v_{i0t}) & \text{if } visit_{it} = 0 \\ P_{it}(Visit|\theta_i, v_{i0t}) \times L_{it}^{stage\ 2}(x_{ijt}|\theta_i, v_{i0t}) & \text{if } visit_{it} = 1 \end{cases}$$

which can be written as

$L_{it}(x_{ijt}|\theta_i, v_{i0t})=$

$$\left(1 - P_{it}(Visit|\theta_i, v_{i0t})\right)^{1-Visit_{it}} \left(P_{it}(Visit|\theta_i, v_{i0t}) \times L_{it}^{stage\ 2}(x_{ijt}|\theta_i, v_{i0t})\right)^{Visit_{it}} \quad \text{(TA.18)}$$

where $Visit_{it}$=1 if consumer i makes visit to platform in period t and 0 otherwise. We integrate out v_{i0t} in the observational likelihood in equation (TA.17) by using the Gauss-Hermite quadrature. To do so, we define $\tau = \frac{v_{i0t}}{\sqrt{2\rho_\varepsilon}}$ and rewrite the likelihood above as $L_{it}(x_{ijt}|\theta_i) =$

$$\int L_{it}(x_{ijt}|\theta_i, v_{i0t}) \frac{1}{\sqrt{\rho_\varepsilon}} \phi\left(\frac{v_{i0t}}{\sqrt{\rho_\varepsilon}}\right) dv_{i0t} = \int L_{it}(x_{ijt}|\theta_i, v_{i0t} = \tau\sqrt{2\rho_\varepsilon}) \frac{1}{\sqrt{\pi}} e^{-\tau^2} d\tau\,. \quad \text{(TA.19)}$$

Next, we use the Gauss-Hermite quadrature to approximate the integral above to get the observational likelihood conditional on the consumer-specific parameters θ_i.

$$L_{it}(x_{ijt}|\theta_i) \approx \sum_{m=1}^{12} L_{it}(x_{ijt}|\theta_i, v_{i0t} = \tau_m\sqrt{2\rho_\varepsilon}) \frac{w_m}{\sqrt{\pi}}\,. \quad (TA.20)$$

where the expression for the visit probability, *P*(*Visit*), is given as where we use 12 pairs of (τ_m, w_m) of the quadrature.

9.2 Modification of the Observational Likelihood with the Control Function Approach.

Recall that the control function enters both models of the first and second stages. Let μ_{jt} be the errors in the control function which is assumed to be normally distributed with mean zero and standard deviation of σ_μ. Let $\hat{\mu}_{jt}$ be the residuals retained from the first step of the control function approach and let the estimate of the standard deviation be $\hat{\sigma}_\mu$.

We begin by discussing how we incorporate the control function into the second stage and how this modification affects the observational likelihood at that stage. As discussed above, we add the control function part into the baseline utility in the second stage if and only if the book is still ongoing and consumer i is purchasing at the availability constraint in period t, i.e., x_{ijt} belongs to group 2. Recall that in the second stage in equation (TA.5), we specified the error term ε_{ijt} as the sum of two zero-mean IID normal random variables: (1) v_{ijt} with a variance of $1 - \rho_\varepsilon$; and (2) v_{i0t} with a variance of ρ_ε. When adding the control function to the model, we need to readjust the error term variances for observations where the control function matters. We derive the modification below.

Suppose μ_{jt} and v_{ijt} are correlated with ρ_{c2}. Given that they are bivariate normally distributed, the expected value and variance of v_{ijt} conditional on $\mu_{jt} = \hat{\mu}_{jt}$ and $\sigma_\mu = \hat{\sigma}_\mu$ is

$$E(v_{ijt}|\hat{\mu}_{jt}) = \frac{\rho_{c2}\hat{\mu}_{jt}\sqrt{1-\rho_\varepsilon}}{\hat{\sigma}_\mu}, Var(v_{ijt}|\hat{\mu}_{jt}) = (1-\rho_\varepsilon)(1-\rho_{c2}^2). \quad (TA.21)$$

Therefore, we can rewrite v_{ijt} conditional on $\hat{\mu}_{jt}$ as

$$v_{ijt} = \frac{\rho_{c2}\hat{\mu}_{jt}\sqrt{1-\rho_\varepsilon}}{\hat{\sigma}_\mu} + \hat{v}_{ijt}, \quad (TA.22)$$

where $\hat{v}_{ijt}$ is a normally distributed random variable and $\hat{v}_{ijt} \sim N\left(0, (1-\rho_\varepsilon)(1-\rho_{c2}^2)\right)$. Substituting Equation (TA.22) into the condition for group 2 in Equation (TA.9), the modified condition for books $j = n_1 + 1, \ldots, n_1 + n_2$ is

Group 2: Conditions for purchase at the availability constraint for books $j = n_1 + 1..n_1 + n_2$:

$$\left\{x_{j,l_j(t)} \cdot \left(1 - b \cdot \Delta t_{j,t}\right) - a\Delta t_{j,t} + \sigma_m \varepsilon_m \leq 0,\ \hat{v}_{ijt} \geq\ \beta_i(1-\rho_i)c_{ijt} + \beta_i \rho_i s_{it} - \alpha_{0i} - \alpha_j - \alpha_i^T Z_{ijt} - \frac{\rho_{c2}\hat{\mu}_{jt}\sqrt{1-\rho_\varepsilon}}{\hat{\sigma}_\mu} - v_{i0t}\right\} \quad \text{(TA.23)}$$

Accordingly, we obtain the modified contribution to the second-stage observational likelihood for group 2 as follows:

$$L_{it,g2}^{stage\ 2}\left(x_{ijt}\middle|\theta_i, v_{i0t}\right) = \prod_{j=n_1+1}^{n_1+n_2}\left[\frac{1-\Phi\left(\frac{x_{j,l_j(t)}\cdot\left(1-b\cdot\Delta t_{j,t}\right)-a\Delta t_{j,t}}{\sigma_c \Delta t_{j,t}}\right)}{1-\left(\frac{-a-bx_{j,l_j(t)}}{\sigma_c}\right)} \times \left(1 - \Phi\left(\frac{\beta(1-\rho)c_{ijt} + \beta\rho s_{it} - \alpha_0 - \alpha^T Z_{ijt} - \frac{\rho_{c2}\hat{\mu}_{jt}\sqrt{1-\rho_\varepsilon}}{\hat{\sigma}_\mu} - v_{i0t}}{\sqrt{1-\rho_\varepsilon}\sqrt{1-\rho_{c2}^2}}\right)\right)\right] \quad \text{(TA.24)}$$

Following this, we simply replace the second-stage likelihood contribution for group 2 in Equation (TA.12) with the second-stage likelihood contribution for group 2 in the above Equation (TA.24). This completes the control function modification of the observational likelihood in the second stage. We next discuss the modification of the visiting probability in the first stage (equations TA.15 and TA.17). Suppose the residual errors in the control function, μ_{jt}, and the normal random error, ω_{ijt}, in the probit model in the first stage (equation TA.16) are correlated with ρ_{c1}. Recall that ω_{it} is distributed as $\omega_{it} \sim N(0, 1 - \frac{\rho_\eta^2}{\rho_\varepsilon})$. Given that ω_{ijt} and μ_{jt} are bivariate normally distributed, the expected value and variance of ω_{ijt} conditional on $\mu_{jt} = \hat{\mu}_{jt}$ and $\sigma_\mu = \hat{\sigma}_\mu$ are

$$E\left(\omega_{ijt}\middle|\hat{\mu}_{jt}\right) = \frac{\rho_{c1}\hat{\mu}_{jt}}{\hat{\sigma}_\mu}\sqrt{1-\frac{\rho_\eta^2}{\rho_\varepsilon}}, Var\left(\omega_{ijt}\middle|\hat{\mu}_{jt}\right) = (1-\rho_{c1}^2)\left(1-\frac{\rho_\eta^2}{\rho_\varepsilon}\right) \quad (TA.25)$$

Therefore, we can rewrite ω_{ijt} conditional on $\hat{\mu}_{jt}$ as

$$\omega_{ijt} = \frac{\rho_{c1}\hat{\mu}_{jt}}{\hat{\sigma}_\mu}\sqrt{1-\frac{\rho_\eta^2}{\rho_\varepsilon}} + \hat{\omega}_{ijt}, \quad (TA.26)$$

where $\hat{\omega}_{ijt}$ is a normally distributed random variable and $\hat{\omega}_{ijt} \sim N\left(0, (1-\rho_{c1}^2)\left(1-\frac{\rho_\eta^2}{\rho_\varepsilon}\right)\right)$. Substituting Equation (TA.26) into the visit probability into equations (TA.15) and (TA.16), we get the modified visit probability conditional on the error v_{i0t} as

$$Pr_{it}(Visit|\theta_i, v_{i0t}) = \Phi\left((\gamma_0 + W_{it}\gamma_1 + \frac{\rho_\eta}{\rho_\varepsilon}v_{i0t} + \frac{\rho_{c1}\hat{\mu}_{jt}}{\hat{\sigma}_\mu}\sqrt{1-\frac{\rho_\eta^2}{\rho_\varepsilon}})\Big/(\sqrt{1-\rho_{c1}^2}\sqrt{1-\frac{\rho_\eta^2}{\rho_\varepsilon}})\right) \quad \text{(TA.27)}$$

Following this, we simply replace the visiting probability in Equation (TA.17) with the visiting probability in Equation (TA.27). This completes the control function modification of the observational likelihood in the first stage.

The Overall Likelihood. Previously, we specified the observational likelihood in Equation (TA.20) conditional on the consumer-specific parameters θ_i, $L_{it}(x_{ijt}|\theta_i)$. Given this, we get the likelihood for consumer i across all weeks $t = 1, \ldots, T$ in the data conditional on θ_i as

$$L_i(x_i \equiv \{x_{it}\}_{t=1}^{T}|\theta_i) = \prod_{t=1}^{T} L_{it}\left(x_{ijt}|\theta_i\right) \qquad (TA.28)$$

We model heterogeneity using the latent class segmentation method, in which θ_i takes different values across different latent segments. Let there be G segments with weights as $\{w_g\}_{g=1}^{G}$. The unconditional joint likelihood of consumer i over all periods is

$$L_i(x_i) = \sum_{g=1}^{G} w_g \cdot L_i(x_i|\,\theta_g)\,. \qquad (TA.29)$$

Where for each segment g, the parameters are θ_g. The log-likelihood for the entire sample is

$$l = \sum_{i=1}^{I} \ln[L_i(x_i)] \qquad (TA.30)$$

We maximize the log-likelihood with respect to all of the parameters to obtain the model estimates.

Section 10. Justification of Assumptions in the counterfactuals regarding the writer behavior and the notification effect.

We provide justifications for our assumptions regarding counterfactuals related to writer behavior in Section 10.1 and for our assumption about the impact of notifications on platform visits in Section 10.2 below.

10.1. Assumptions in the Counterfactuals Regarding Writer Behavior

In the counterfactuals, we assume that the following aspects of writer behavior do not change with a change in the release strategy: (1) whether the writer continues writing books on the platform or switches to another platform, (2) qualities of chapters in the book, (3) the initial release time of the book, (4) number of chapters in a book, and (5) number of books written in a year.

Regarding the first aspect (whether a writer continues to write books on the platform), note that the platform is a virtual monopoly with a 70% market share in the romance genre, and all the writers have signed a long-term (5 to 10 years) exclusive contract that prevents them from switching to other platforms or multihoming. Thus, it is unlikely that the change in release strategy will stop writers from writing books on the platform or drive them to another platform.

Regarding the second aspect (namely, the impact of release strategy on the qualities of chapters of a book), the analysis in Section 2 of the Web Appendix suggests that the qualities of book chapters in our sample are not significantly affected by the release strategies. This is because the books in our sample are written by *experienced* writers, who tend to be more skilled and produce consistently high-quality content. In comparison to inexperienced writers, experienced writers are more likely to be confident in their own skills and taste, value their creative independence more, and care less about and react less to readers' comments.[7] Therefore, even if a release strategy change (such as adopting a hybrid release strategy) leads to much less reader feedback during an early stage, our sample writers are less likely to respond by changing the content of their books.

Regarding the third aspect (namely, impact of release strategy on release time of the first chapter of the book), note that our model and counterfactuals focus on release speed while taking the initial release timing as exogenous. There are two reasons why we would not expect the change in release strategy to impact the initial release time.

i. In the counterfactuals, we keep the initial release time the same as the one observed in data. We also varied the initial release time in counterfactuals and it doesn't affect the main results.
ii. The initial release time will impact the sales of the book if there is seasonality and if there is competition between the different books/writers on the platform. However, our results indicate that the extent of substitutability between books is limited, suggesting that the degree of competition between books and writers is also small. Regarding the impact of seasonality on the release of the first chapter, we have plotted the histogram of all books' initial release months in Figure A4. We see that the distribution is even across all 12 months (except for the slight blip during the Chinese New Year). This is unlike in the movie industry, in which many more movies are released during peak moviegoing seasons such as December. This suggests that seasonality does not play a strong role in the initial release time decisions.

Regarding the fourth aspect (namely, the impact of release strategy on the number of chapters in a book), note that experienced writers are less likely to change their book content based

[7] "Digital Literature: The Customer is God?" from People.cn. (http://it.people.com.cn/n/2014/0218/c1009-24386304.html)

on release strategy as compared to inexperienced writers. To validate this, we run a regression to test how the number of chapters in a book is affected by the book's average release speed, controlling for book initial release year-fixed effect, writer-fixed effect, and book sales. In this regression, we exploit the variation in the average release speed across different books written by the same author and test its impact on the number of chapters in each of the books written by that writer. The results in Table A8 show that release speed does not significantly affect the length of the books in our sample. Therefore, it is reasonable to assume that in the counterfactuals, the release strategy of a book does not change its length.

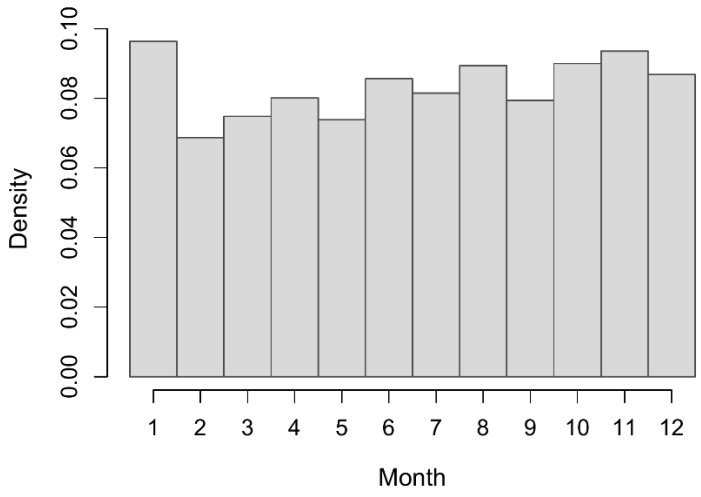
It is important to note that when it comes to the length of a serial (i.e., the number of chapters of a book or number of episodes of a TV serial), the behavior of experienced writers is different from that of the producers of a TV show. TV shows are a lot more responsive to audience ratings and feedback as compared to how responsive the experienced book writers are This is because writing involves much lower production costs than television production, which requires filming, casting, and other costly inputs. Thus, producers are willing to keep the story going if the show is doing well so there is a better return on the cost. This may not be so for individual writers (especially the experienced ones), since writing the story is an independent process, and often famous writers insist on their creative integrity since their personal style is what made them successful rather than simply following the mainstream taste.

Table A8. The Effect of Release Speed on Book Length

Dependent variable: Number of chapters in a book/100	
Release speed: No. of chapters per day	1.61 (0.86)
Log (daily sales during release period)	0.22 (0.19)
Book Initial Release Year F.E.	✓
Writer F.E.	✓

Note: Regressions are at the book level. Standard errors (in brackets) are clustered at the writer level. ***, **, and * indicate significance at the 0.1%, 1%, and 5% levels, respectively.

Figure A4. Distributions of Books' Initial Release Month



Regarding the fifth aspect (namely, the impact of release speed on the number of books written by the writer in a year), we would not expect a different release strategy to significantly change

an experienced writer's overall creative production process during the entire year. Thus, if the overall production of the writer does not change and the book content/length do not change with the change in release strategy, it follows that the number of books written by a writer will also not change with a change in release strategy.

10.2. The Effect of Notifications on Platform Visits is the same across the different Release Strategies in the Counterfactuals.

In our counterfactuals, we have assumed that the impact of notifications on platform visits will be the same across the sequential release and the hybrid release strategies[8]. This assumption raises a concern since we have modeled the impact of notifications on platform visits in a reduced form manner. Although we cannot explicitly test this assumption in our data (since our data does not include hybrid release strategies), we can implicitly test it by making use of the fact that in our data, consumers start reading an ongoing book at different points in time. For example, if a consumer starts reading an ongoing book when x% of the chapters have been released, it is similar to a scenario in which the consumer faces a hybrid release strategy in which the first x% of the chapters of that book are released simultaneously and the rest are released sequentially. Therefore, we test our assumption by checking whether the initial availability of chapters at the time when a consumer starts reading an ongoing book moderates the impact of notifications of its newly released chapters on her platform visits. If the initial availability of chapters of an ongoing book when a consumer starts reading it has an insignificant moderating impact (on the impact of notifications on her platform visits), it will suggest that the impact of notifications on platform visits will not vary across the different release strategies.

The regression specification is similar to Table 7 in section 3.3 of the main paper (in which we had provided reduced form evidence for the impact of notifications on the platform visit probability), except that we now interact the number of notifications of new chapters released with the percentage of chapters that were initially available when consumer first started purchasing an ongoing book. The results of the regression are reported in Table A9. The estimate of the interaction effect (0.016) is insignificant, which suggests that our assumption (that the impact of notifications on platform visits effect will not vary across the different release strategies in the counterfactuals) is reasonable.

Table A9. The Effect of Notifications and Initial Availability on Platform Visits

Dependent variable: Platform Visit	
Weekend	-0.010 (0.005)
No. of visits in the last 7 days	0.098 *** (0.002)
No. of notifications of new chapters released	0.117 *** (0.010)
Average initial availability of chapters of an ongoing Book when the consumer started reading it	-0.060 ** (0.021)
No. of notifications of new chapters released * Average initial availability of chapters of an ongoing Book when the consumer started reading it	0.016 (0.018)
Consumer F.E.	✓

Note: Regressions are at the consumer–day level. Standard errors (in brackets) are clustered at the consumer level. ***, **, and * indicate significance at the 0.1%, 1%, and 5% levels, respectively.

[8] As discussed in section 7 of the paper, the platform's notification policy does not apply to simultaneous release strategy.

Section 11: Details on Counterfactual Simulations Process

We provide an overview of how counterfactual simulations are conducted step by step.

1. Assign 100 individuals randomly to 4 segments based on the segment weights.
2. Pre-generate error terms from a normal distribution based on the estimated variances for all 100 individuals and all 365 days. This includes error terms for (1) first stage visit decision, (2) second stage purchase decision, (3) consumption rates.
3. Pre-simulate the chapter availability of each book on any given day based on the counterfactual assumptions – e.g., whether a book should be released in a simultaneous, sequential, or hybrid manner, the release speed for the sequential portion, the ratio of different portions under hybrid, etc.
4. For $t = 1$ and all individuals, calculate the first stage of the platform visit probability based on the estimated probit equation in equation (4.9) of the paper and pre-generated error terms. Assign $visit_i$=1 if the LHS of equation (4.9) is positive and $visit_i$=0 if it is zero or negative.
5. For $t = 1$ and all individuals, if the individual does not visit based on step 4, set their purchase amount of all books to 0. If the individual does visit based on step 4, simulate the second stage purchase decision in terms of the number of chapters to purchase for all books by solving each individual's constrained utility maximization problem using Kuhn-Tucker conditions (see more details in the paragraph to follow).
6. For $t = 1$ and all individuals, given the purchase quantity, we then update the consumer inventory level, and then calculate the consumption level of day t based on estimated consumption equation parameters, inventory level, and pre-generated consumption error terms. Using the consumption level, we then update the inventory level once more to calculate the next period's initial inventory. This updating is done using equation (4.11) in the paper.
7. Having simulated all the visits, purchases, consumption, and inventory for $t = 1$, we then update all the independent variables for the next period, such as never purchased before, whether purchased in last 7 days, log(No. of chapters purchased before), promoted on homepage, availability constraints, visit frequency in the last 7 days, etc. A book will be promoted on the home page in period t if its aggregate sales (summed across all consumers) is in the top 10 in the last 7 days. We track the notifications of all ongoing books for a given consumer, based on whether that consumer has purchased any chapters of a book before.
8. Repeat steps 4 to 7 for $t = 2, \dots, 365$.

In step 5, for each book, we calculate the purchase incidence decision and the optimal purchase quantity in the second stage as follows. For each book, the optimal solution can be in one of three cases: (a) a constrained boundary solution at 0, (b) a constrained boundary solution at the maximum availability constraint, or (c) a non-boundary solution. Recall that the estimate of the substitution parameter between books, ρ≈0. Thus, in our analysis, we assume ρ=0. This simplifies the optimal solution for book j as follows

Case 1: Condition for non-purchase of book *j*:

Inv_{ijt}>0 or $\{Inv_{ijt}$ =0, $u_{ijt}/\beta_i \leq 0\}$

Case 2: Condition for purchase at the availability constraint for book *j*:

Inv_{ijt} =0, $u_{ijt}/\beta_i \geq c_{ijt},\ x_{ijt} = c_{ijt}$

Case 3: Condition for purchase at less than the availability constraint for book *j*:

Inv_{ijt} =0, $0 < \frac{u_{ijt}}{\beta_i} < c_{ijt}$, $x_{ijt} = u_{ijt}/\beta_i$

Where u_{ijt} is consumer *i*'s sub-utility of book *j* in period t, which is given as $u_{ijt} = \alpha_{ijt} - p_j$, where p_j is the unit price of each chapter of book *j*, and the specification of α_{ijt} follows from equation 4.2 of the paper as

$$\alpha_{ijt} = \alpha_{i0} + \alpha_j + \alpha_i^T Z_{ijt} + \varepsilon_{ijt}$$

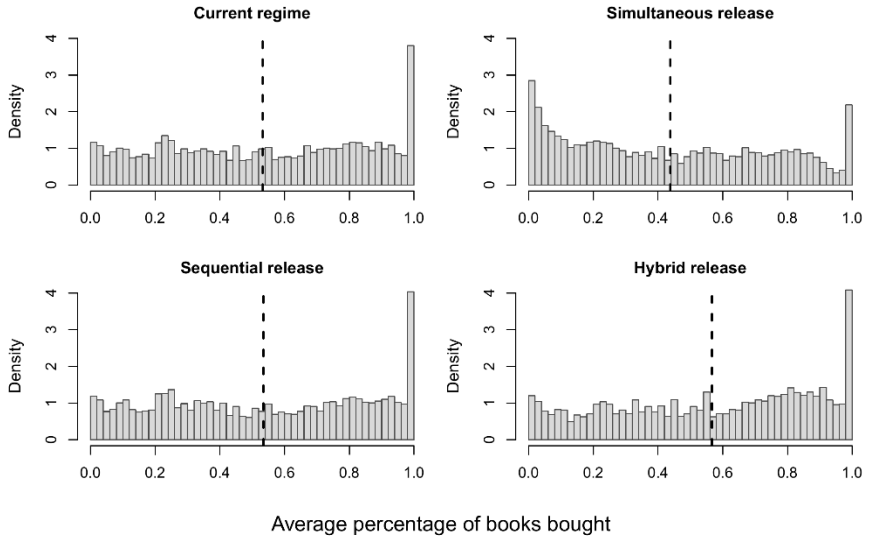
Note that by assuming the substitutability parameter ρ=0, the purchase decisions become independent across all books. The only dependency is through the contemporaneous correlations in the errors in α_{ijt} across j=1..J. Thus, given the draws of the errors ε_{ijt} from step 1, we first calculate the inventory level of consumer *i* for each book *j* in period *t*. If the inventory level is positive, then the consumer will not purchase any chapter of book *j* in period *t*. If consumer *i*'s inventory level of book *j* is zero in period *t*, then we calculate the consumer i's sub-utility for book *j* in period t, u_{ijt}. Following that, for each consumer *i* and book *j*, we check whether u_{ijt}/β_i is such that (1) $u_{ijt}/\beta_i \leq 0$, (2) $0 < \frac{u_{ijt}}{\beta_i} < c_{ijt}$, or (3) $u_{ijt}/\beta_i \geq c_{ijt}$. If (1) is true, then it means that consumer *i* will not purchase any chapters of book *j* in period t. If (3) is true, then it means that consumer *i* will purchase u_{ijt}/β_ichapters of book j in period t rounded to the nearest whole number.

Finally, note that while simulating the purchases of consumers in the counterfactuals, we do not need to account for two of the spillover effects across books that stem from the effect of (non) availability of other books on the total inventory stock, and on the satiation levels of the focal book. This is because our structural model analysis indicates that substitution between books is very small, implying that the (non)availability of one book has a minimal effect on the satiation level for other books. Moreover, since book data is in text format, storage costs are trivial. Thus, the inventory stock will play a minimal role in book purchases. The only spillover we must consider when simulating the purchases is the notification effect, as the (non) availability of chapters of the focal book can impact the sales of other books through the notifications of the focal book. In the simulations, we track the notifications of all ongoing books for a given consumer, based on whether that consumer has purchased any chapters of a book before.

Section 12. Distribution of Book Consumption and Writer Revenue across Different Release Strategies

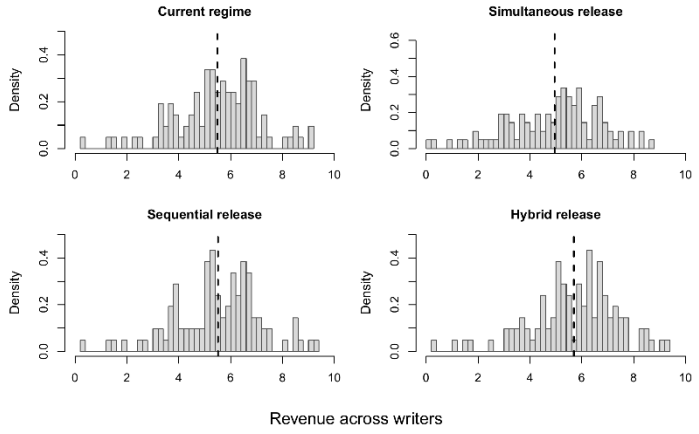
In this section, we first present the following histogram in Figure A5 to illustrate the distributions of book consumption across the 4 release strategies, as predicted by the counterfactuals.

Figure A5. Histograms of Percentage of Chapters Purchased per Book



Observe in the above figure that under simultaneous release, we see the highest bar for 100% finish rate as the binge consumption effect is the strongest. Other than that, the difference in patterns of the distribution across the 4 release strategies are more subtle. The retention patterns are similar between current and optimal sequential (1 chapter per day), with sequential release having a slightly higher finish rate. Under hybrid release, we see the percentage distribution is lower under 40%, given that the first 40% of the chapters are simultaneously released, and consumers are more likely to pass the 40% percentage. We next report the distribution of writers' revenues across the 4 release strategies, as predicted by the counterfactuals in Figure A6. Observe that there is no clear pattern of how the distribution of revenues shared by the writers changes with the release strategy.

Figure A6. Histograms of Revenues Earned by Writers



Section 13. Counterfactuals on Different Splits of Sequential and Simultaneous Release for Hybrid Release Strategy IV

Table A10. Optimal Hybrid Release Strategy

Sequential release speed	(a) Chapters purchased	(b) Number of visits	(c) Number of books explored	(d) Purchase order size
(1) Simultaneous for the first 30% of chapters + sequential for the rest at a constant speed				
1 chapter/day	506 (2.9)	61.2 (0.29)	9.9 (0.04)	11.4 (0.03)
2 chapters/day	433 (2.8)	42.9 (0.24)	9.2 (0.04)	16.6 (0.05)
(2) Simultaneous for the first 40% of chapters + sequential for the rest at a constant speed				
1 chapter/day	511 (2.9)	56.8 (0.28)	10.1 (0.05)	12.9 (0.04)
2 chapters/day	441 (2.6)	40.6 (0.21)	9.4 (0.04)	18.4 (0.05)
(3) Simultaneous for the first 50% of chapters + sequential for the rest at a constant speed				
1 chapter/day	502 (2.9)	51.3 (0.25)	9.9 (0.05)	14.8 (0.05)
2 chapters/day	434 (2.8)	37.6 (0.22)	9.3 (0.05)	20.3 (0.05)

Section 14. Counterfactual Results at the Segment Level

We provide evidence that consumers' preferences for ongoing vs. completed books, and when to start reading ongoing books, do not significantly influence the optimal release strategy for the platform. We conduct the counterfactual exercises for each segment separately to determine which release strategy is most effective: simultaneous, sequential, or hybrid? This differs from the counterfactual exercises done earlier, which assessed the impact of release strategies on the entire sample of consumers rather than a specific segment. The results of this counterfactual exercise are presented in Table A11 below.

Table A11. Breakdown of Sales from Different Counterfactuals by Segments

Segments:	(1)	(2)	(3)	(4)
Segment weights:	0.39	0.26	0.21	0.14
(I) Existing release strategy				
Chapters purchased	385 (3.0)	254 (3.1)	741 (3.6)	485 (3.5)
(II) Simultaneous release (releasing all chapters at once)				
Chapters purchased	351 (3.8)	53 (1.0)	469 (4.1)	51 (1.0)
(II) Sequential release at 1 chapter/day				
Chapters purchased	401 (3.5)	273 (2.8)	762 (3.8)	523 (3.4)
(IV) Hybrid release: Simultaneous for the first x% + sequential for the rest at 1 chapter/day				
Optimal ratio x%	60	20	60	20
Chapters purchased	442 (3.7)	293 (2.9)	941 (5.0)	524 (3.4)

The above suggests that although there are some differences across segments, we see the following similarities across all four segments: (i) platform revenues are higher for sequential release than simultaneous release, and (ii) the platform revenues are highest for hybrid release

when a certain chunk of chapters are first released simultaneously, and the rest are released sequentially. This suggests that consumers' preferences for ongoing versus completed books, and when to start reading ongoing books, do not significantly influence the optimal release strategy for the platform. This is because across all four segments, the hybrid release yields the highest platform revenue, followed by sequential release, and finally by simultaneous release.