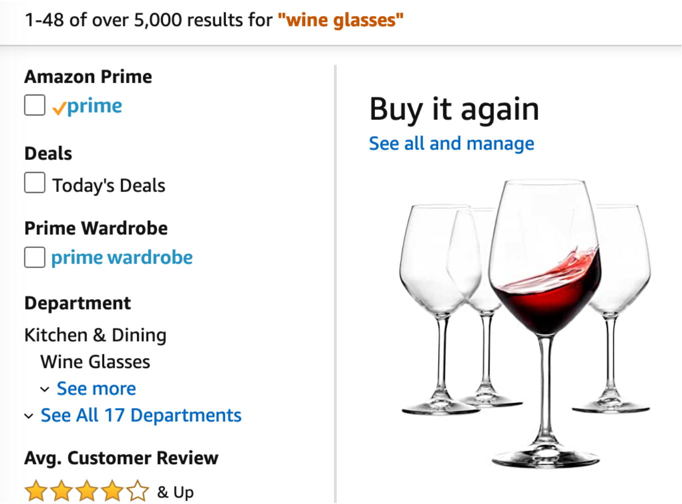
Reaching for the Stars: Discounts and Review Tier Transitions in the Video Games Market

Dmitry Sorokin (NYU)

January 22, 2021

Economics of Attention



Economics of Attention

- Online consumer reviews help to simplify choice in many markets
  - → Individual consumers leave reviews
  - → Reviews are often aggregated into **review tiers** (★★★★☆)
  - → Future customers rely on reviews
- Firms try to manage their reviews
  - → Soliciting reviews from their buyers
  - → Responding to the reviews they receive
  - → Leaving fake reviews
- What firms can do to affect their reviews is important for market design
  - → "Zero Reviews Trap" as a barrier to entry
- I study a new form of review management
  - → Can firms use price promotions to upgrade their review tiers?
  - → A link between discounts and review tier **transitions**

Why Discount?

- Consider a product that is a few reviews short of obtaining an extra star
- Would it be profitable for this product to launch a price promotion?
- The **selection effect**:
  - Customers who buy during a discount are potentially different
  - They could leave better or worse reviews
  - The sign of the selection effect is an empirical question
- The **variance effect**:
  - Randomness in the outcome of the discount
  - A good outcome catapults the product to the next review tier
  - A bad outcome might not affect the review tier
  - The opposite effect for a product close to a review tier downgrade

Stylized Model

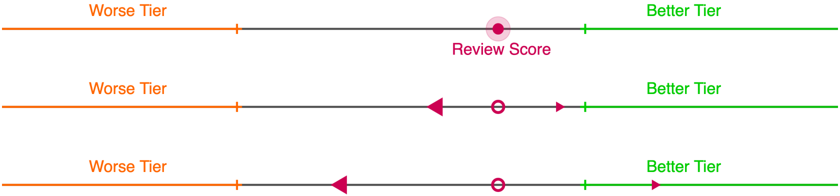
▸ With Formulas

Stylized Model

▸ With Formulas

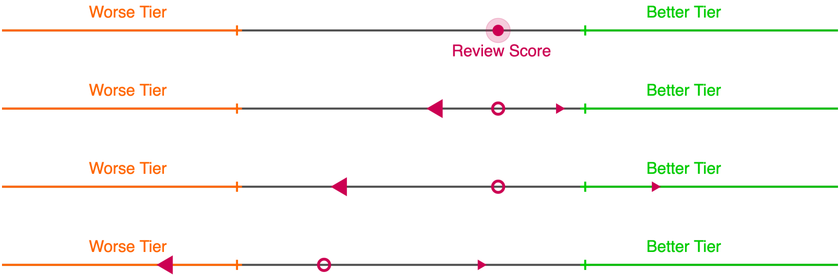
Stylized Model

▸ With Formulas



Stylized Model

▸ With Formulas



What I do

1. Model of demand and reviewing on Steam
   - The best review tier increases sales by 23% compared the worst tier
   - Discounts are followed by a 38% spike in sales
2. Discounting around transitions
   - Products that are close to upgrading their review tier are 4-9% more likely to sell at a discount
   - Evidence that proximity to a downward transition decreases the probability to sell at a discount by up to 4%
3. Selection vs. Variance
   - The selection effect is small
   - Controlling for the selection effect, proximity to a negative transition decreases the probability of a discount by 6%
   - Consumers are more likely to leave reviews when they buy during a discount

Literature

- Regression discontinuity with review tiers: Anderson and Magruder (2012), Luca (2016), Sorokin and Stevens (2020)
  - How random is assignment around the cutoffs?
- Promotions: is the selection effect negative?
  - Ifrach et al. (2017), Acemoglu et al. (2019)
  - Byers et al. (2012), Li (2016), Zhu et al. (2019), Cabral and Li (2015)
- A negative selection effect poses a puzzle:
  - Zegners (2017): free pricing leads to bad reviews
- Design of recommendation systems in digital markets
  - Vellodi (2020): reviews as barriers to entry and pricing as a way to solve it

Roadmap for Today

1. Institutional Setting
2. Data
3. Empirical Model
4. Discounts Around Transitions
5. Selection vs. Variance

Roadmap for Today

Steam

- In 2018 the U.S. video games market revenue matched the one of the U.S. film industry: $43.4 billion (Minotti 2019)
- *Steam* is the major online marketplace for selling video games
  - Think "App Store" for games on PC and Mac
  - 2013: 75% of all PC games sold online (Cliff 2013)
  - January 2018: over 47 million daily active users
- Until recently, virtually no competition
  - Developers: Steam is a powerful player that you can't ignore

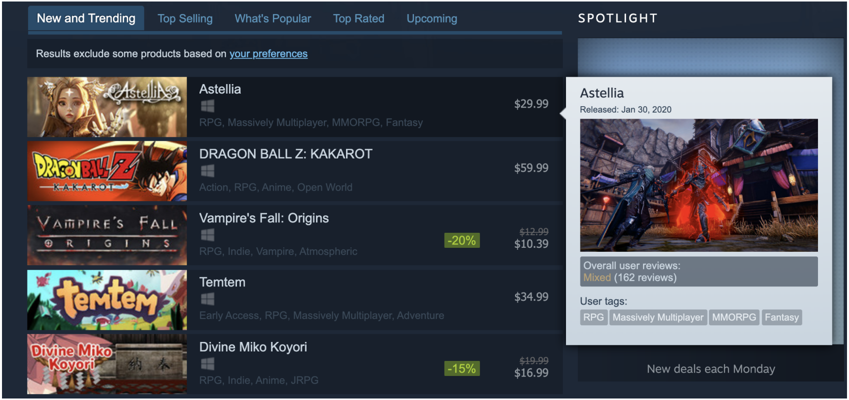
Reviews on Steam

- Transparency and simplicity
  - Users give "thumbs up" or "thumbs down"
- A **Review Score** is the fraction of positive reviews among all reviews
  - A game with 6 good reviews out of 10 total reviews has a score of 60%
- **Review Bins** based on the score:

Review Bin	N. of Reviews	Score
No Score	[0, 10)	-
Negative	Any	[0, 40)
Mixed	Any	[40, 70)
Mostly Positive	Any	[70, 80)
Positive	[10, 50)	[80, 100)
Very Positive	More than 50	[80, 100)
Overwhelmingly Positive	More than 500	[95, 100]

Visibility and Reviews

- Labels better than "Negative" have a similar contribution to visibility
- The effect is through the information that reviews convey

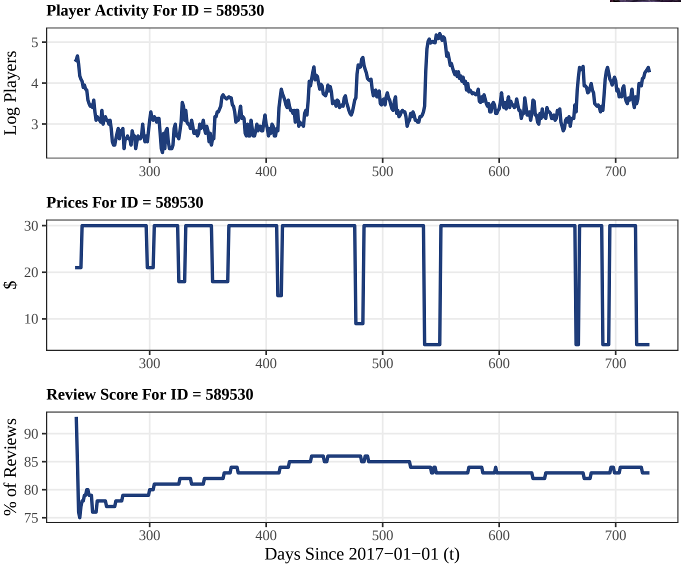


Roadmap for Today

Data

- Main source: SteamDB.com, a third party website
- Player count: the maximum number of concurrent players on any given day
- Price history: daily observations of prices
- Review history:
  - Cumulative numbers of positive and negative reviews on any day
  - Review score and review labels are reverse-engineered

Example Game



Sample

▸ Descriptive Stats

- Single player, not free-to-play games
  - → Games with stable quality: no more than 5 small patches
- Released in 2017 or later
  - → Stable review system
- At least 4 concurrent players on the median day
- This leaves us with 906 games, about 319,000 observations

Transitions

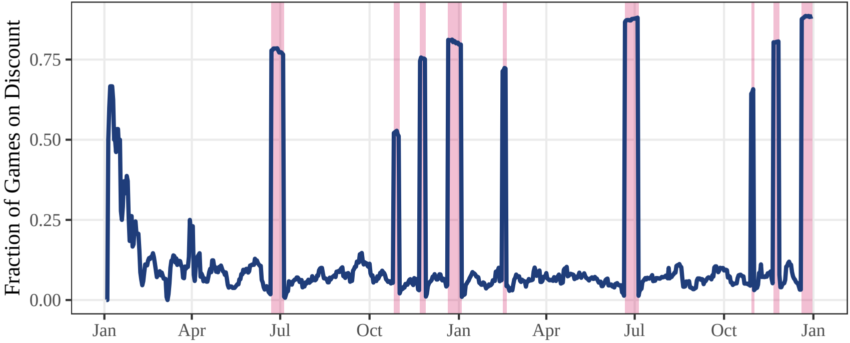
Transitions Stats

- 596 games have transitioned, 1688 unique transitions
  - 1225 are followed by at least 7 days in the new bin
- Every tier is well-represented in the data Transitions Matrix

Discounts

- Nominal prices rarely change, but instead products go on discounts
  - 6331 discounts vs. 43 price changes
  - Consumers can set up discount notifications

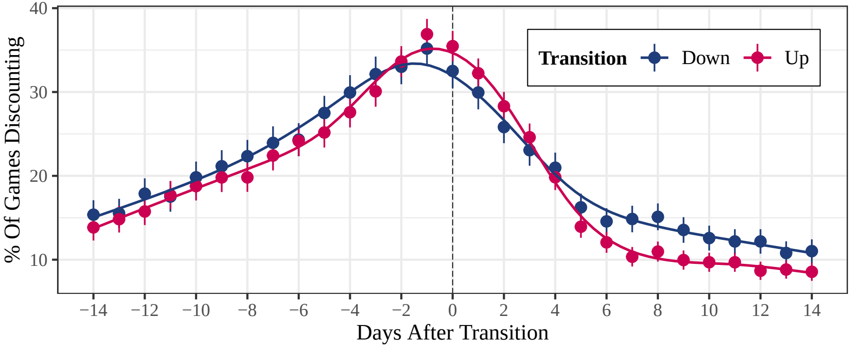
- Custom and Curated discounts
  - Seasonal sale: 64% of the sample



Discounts Around Transitions

▸ Regression

- Transitions between review bins are preceded by discounts
- Extra day closer to the transition $\iff$ the probability of discount $\uparrow$3%
- Extra day after the transition $\iff$ the probability of discount $\downarrow$7%



Roadmap for Today

1. Institutional Setting
2. Data
3. Empirical Model
4. Discounts Around Transitions
5. Selection vs. Variance

Empirical Model of Demand and Gaming

- How impactful are discounts on Steam?
- How much do various review labels matter for sales?
- Data limitation: no sales data
  - → I use the model to infer demand parameters from product usage
- Prerequisite for a model of reviewing

Arrival of Customers

- Game i on day t sells to B_{it} new buyers
  - $B_{it} \sim P(\lambda_{it})$, $\lambda_{it} = \lambda_i(1 + x_{it}'\beta)$
  - x_{it}: price, review information, age, seasonality
  - $Q_{it}^D(x_{it}) = \mathbb{E}[B_{it}] = \lambda_i(1 + x_{it}'\beta)$ is the demand function

- Active players, A_{it}: bought the game and keep playing

- Game i on day t forever loses E_{it} active players
  - $E_{it} \sim Binomial(A_{it-1}; 1 - \psi_i)$

- The evolution of A_{it}: $A_{it} = A_{it-1} + B_{it} - E_{it}$
  - $\mathbb{E}[A_{it} \mid A_{it-1}, x_{it}] = \psi_i A_{it-1} + \lambda_i(1 + x_{it}'\beta)$

- Parameters: $\theta = (\{\psi_i, \lambda_i\}_{i=1}^n, \beta)$, $\dim(\theta) = 2n + \dim(x_{it})$

Identification Challenges: Unobserved Buyers

- Both the number of buyers and the number of exiters are unobserved

$$A_{it} - A_{it-1} = B_{it} - E_{it}$$

- Key: entry is in absolute numbers, but exit is proportional to player count
  - Exit rates are identified from the spikes in the player count Show me!
  - Entry rates are identified from the level at which the player count settles
  - Crucial heterogeneity in ψ_i

Identification Challenges: Causal Effects

- Time varying unobserved demand shifters correlated with the observables
  - Sample selection $\implies$ fixed quality

- Identifying assumption: conditional on the controls, the unobserved factors are not correlated with the review tier
  - Rely on the panel structure of the data
  - "Compare" sales between the periods spent in different review tiers

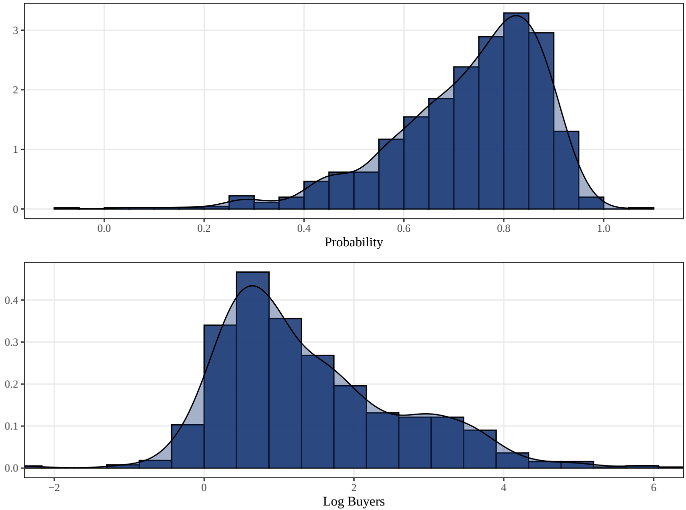
Estimation

- The model implies the following regression equation:

$$A_{it} = \psi_i A_{it-1} + \lambda_i (1 + x_{it}' \beta) + u_{it} \quad (1)$$
$$0 = \mathbb{E}\left[u_i \mid A_{it-1}, x_{it}\right] \quad (2)$$

- Parameters: $\theta = (\{\psi_i, \lambda_i\}_{i=1}^n, \beta)$, $\dim(\theta) = 2n + k$
- Conditional on β, ψ_i and λ_i could be obtained by OLS
- Minimize the concentrated sum of squares with respect to β Details

Results: Distributions of $\hat{\lambda}_i(1 + \bar{x}_i\hat{\beta})$ and $\hat{\psi}_i$



Results: Demand Parameters $\hat{\beta}$

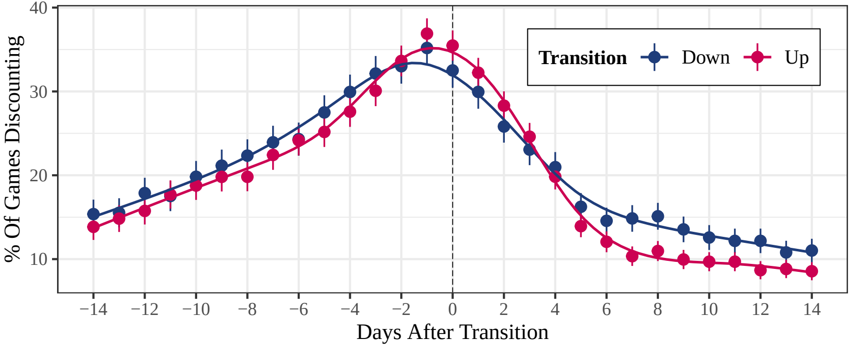
Price	-0.342^{***}
	(0.010)
New Discount	0.230^{***}
	(0.013)
Seasonal Sale	-0.006
	(0.004)
No Score	-0.128^{***}
	(0.019)
Negative	0.005
	(0.020)
Mostly Positive	0.019^{**}
	(0.008)
Positive	0.054^{***}
	(0.010)
Very Positive	0.052^{***}
	(0.010)
Overwhelmingly Positive	0.100^{***}
	(0.013)
Observations	355900
R^2	0.850

Roadmap for Today

1. Institutional Setting
2. Data
3. Empirical Model
4. Discounts Around Transitions
5. Selection vs. Variance

Discounts and Realized Transitions

▸ Regression



Discounts and Realized Transitions

- Realized transitions are preceded by discounts
- It is merely a correlation
  - Discounts could be causing transitions
  - In fact, they better be, for my story to hold up!
- Firms that give a discount, but do not transition, are not accounted for
  - Firms avoiding a bad transition and not making a good one
- Solution: **potential transitions**
  - Remember the stylized model!
  - How close the game is to a transition is a state variable
  - Discount today can not cause the potential of transition

Potential Transitions

- The goal is to see if proximity to a transition increases the probability of discounting
- Define a measure of proximity to a transition
- For each game:
  - How many positive (negative) reviews are needed for a positive (negative) transition
  - Calculate the speed of positive (negative) review arrival
  - Days to potential transition if only receiving one type of reviews
  - Example

Empirical Strategy

- One has to study proximities to positive and negative thresholds together
  - A firm close to an upgrade might be willing to give a discount
  - A firm close to a downgrade might be willing to stay put

Empirical Strategy

- One has to study proximities to positive and negative thresholds together
  - → A firm close to an upgrade might be willing to give a discount
  - → A firm close to a downgrade might be willing to stay put

- Let $T_{it}^{+} \in \{ 0, 1 \}$ indicate if proximity to a $+$ transition is less than 14 days

- Let $T_{it}^{-} \in \{ 0, 1 \}$ indicate if proximity to a $-$ transition is less than 14 days

Empirical Strategy

- One has to study proximities to positive and negative thresholds together
  - A firm close to an upgrade might be willing to give a discount
  - A firm close to a downgrade might be willing to stay put
- Let $T^+_{it} \in \{ 0, 1 \}$ indicate if proximity to a $+$ transition is less than 14 days
- Let $T^-_{it} \in \{ 0, 1 \}$ indicate if proximity to a $-$ transition is less than 14 days
- Estimate

$$disc_{it} = \beta^+ T^+_{it} + \beta^- T^-_{it} + X_{it}\beta + f_i + \tau_t + \varepsilon_{it}, \tag{3}$$

  - $disc_{it} = \mathbb{1}\{discount_{it} > 0\}$
  - X_{it}: log proximities to transitions, log review count, score, review bin dummies, time without a discount, age
  - f_i: game effect
  - τ_t: day of the week and week effects

Results

- Close to an upgrade: 4-9% increase in the daily probability of a discount
- Close to a downgrade: 4% decrease in the daily probability of a discount

	Dependent variable:	
	Discount Probability	
	14 days	7 days
Close to Pos. Transition	0.015^{***}	0.007^{**}
	(0.003)	(0.003)
Close to Neg. Transition	-0.006^{**}	0.001
	(0.003)	(0.004)
Observations	295,228	295,228
R^2	0.428	0.428
Note:	*p<0.1; **p<0.05; ***p<0.01	

Roadmap for Today

1. Institutional Setting
2. Data
3. Empirical Model
4. Discounts Around Transitions
5. Selection vs. Variance

Selection vs. Variance

▸ Model

- Attenuation of the effect of the proximity to a negative transition on the probability of the discount
  - → Positive selection effect vs. negative variance effect

- Estimate the difference in review sentiment on and off a discount

$$\Delta score_i = \begin{bmatrix} \text{Expected Score} \\ \text{on a Discount} \end{bmatrix} - \begin{bmatrix} \text{Expected Score} \\ \text{off a Discount} \end{bmatrix}$$

- Using the estimates of the selection effect, check for the variance effect:
  - → Products on the verge of a review slump should discount less

The Variance Effect

- Estimate

$$disc_{it} = (\beta_0^+ + \Delta score_i \beta_1^+) T_{it}^+ + (\boldsymbol{\beta_0^-} + \Delta score_i \beta_1^-) T_{it}^- + X_{it}\beta + f_i + \tau_t + \varepsilon_{it}$$

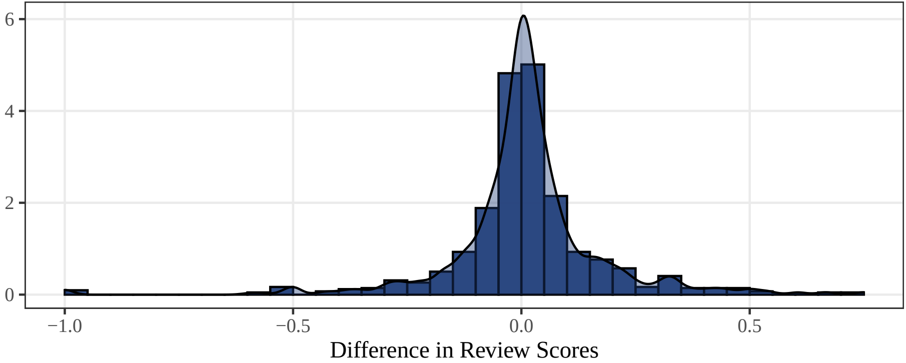
- $\boldsymbol{\beta_0^-}$ measures the effect of proximity to a negative transition when the selection effect $\Delta score_i$ is zero

- Additional test: do consumers who buy during a discount leave reviews with a higher probability?

Results: $\Delta score_i$

▸ Estimation

- Can not reject the hypothesis that the selection effect is negative ($p = 0.14$)
- Reject the hypothesis that users are less likely to leave a review during a discount ($t = 9.16$)
  - → They are 24-40% more likely to do so



	Dependent variable:	
	Discount Probability	
	Controlling for $\Delta score_i$	Not controlling for $\Delta score_i$
Pos. Transition	0.014^{***}	0.015^{***}
	(0.003)	(0.003)
Neg. Transition	-0.009^{***}	-0.006^{**}
	(0.003)	(0.003)
Selection × Pos. Tr.	0.009	-
	(0.013)	
Selection × Neg. Tr.	0.019	-
	(0.015)	
Observations	290,628	295,228
R^2	0.434	0.428
Note:		*p<0.1; **p<0.05; ***p<0.01

Conclusion

- I estimate a model of demand and reviewing behavior
  - → The best review tier increases sales by 23% compared to the worst tier
  - → Discounts are followed by a 38% spike in sales
- Proximity to a review transition affects discounting
  - → Review upgrade: 4-9% more likely to sell at a discount
  - → Review downgrade: 4% more likely to sell at a discount
- Two major effects at play: the selection and the variance effects
  - → The selection effect is small on Steam
  - → The variance effect decreases the probability of a discount by 6% for products close to a negative transition

Game Sizes at 180 Days

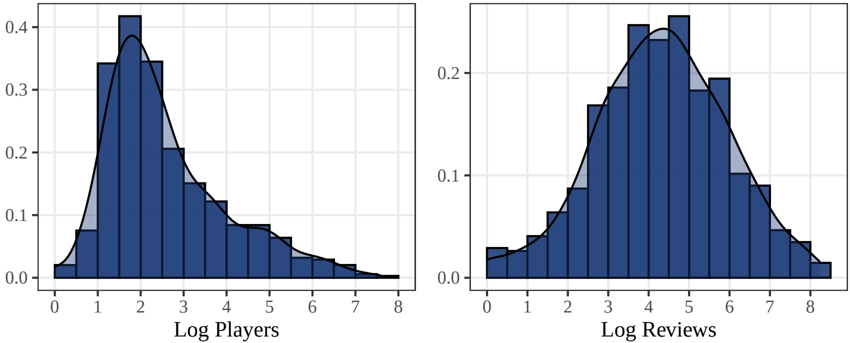


Figure: Distribution Of the Number of Reviews and Players at 180 Days.

Reviews

◂ Back

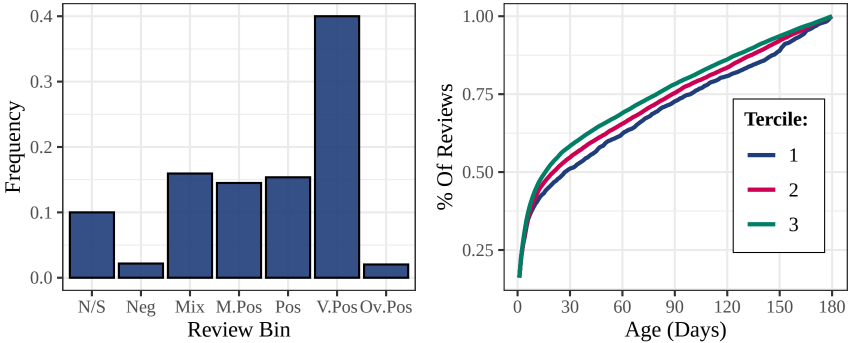


Figure: Distribution Of the Number of Reviews and Players at 180 Days.

Stats Before Transitions

Back

Table: Average Discount, Age, and Review Count Two Weeks Before A Transition

Statistic	N	Mean	St. Dev.	Pctl(25)	Pctl(75)
Mean Discount (%)	1,225	10	14	0	15.9
Age	1,225	142	157	8	226
Reviews	1,225	130	343	26	81

Transition Matrix

◂ Back

Probabilities							Counts					
Neg	Mix	M. Pos	Pos	V. Pos	Ov. Pos		Neg	Mix	M. Pos	Pos	V. Pos	Ov. Pos
0	100	0	0	0	0	Negative	0	58	0	0	0	0
24	0	76	0	0	0	Mixed	60	0	189	0	0	0
0	43	0	25	33	0	M. Positive	0	245	0	141	188	0
0	0	40	0	60	0	Positive	0	2	207	0	310	0
0	0	84	0	0	16	V. Positive	0	0	220	0	0	43
0	0	0	0	100	0	Ov. Positive	0	0	0	0	25	0

Discounts Around Transitions

◂ Back

Days to Transition	0.005***
	(0.0005)
Days After Transition	−0.011***
	(0.0005)
Negative	−0.007
	(0.011)
Mostly Positive	0.005
	(0.006)
Positive	0.044***
	(0.009)
Very Positive	0.014
	(0.009)
Ov. Positive	0.020
	(0.016)
Score	−0.0005*
	(0.0003)
Log Reviews	−0.010***
	(0.002)
Age	0.0003***
	(0.00001)
Const	0.903***
	(0.021)
Observations	32,112
R^2	0.396
Time Effects	Weekdays, Week
Game Effects	×
Polynomial(t W/O Discount)	$d = 2$

Proximity to a Transition: Example

Back

- Consider a game with 19 positive and 6 negative reviews
  - Score = 76%, "Mostly Positive"
  - Needs 5 positive reviews to transition to "Positive": $24/30 = 80\%$
  - Needs 3 negative reviews to transition to "Mixed": $19/28 \approx 67\%$
- If at age of 200 the game has 80 positive and 20 negative reviews, then
  - Average of 0.4 positive and 0.1 negative reviews per day
  - Needs $5/0.4 = 12.5$ days of only positive reviews
  - Needs $3/0.1 = 30$ days of only negative reviews

Reviews

Back

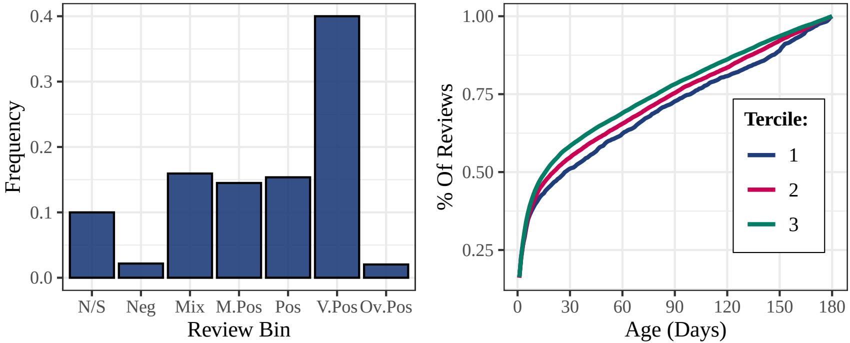


Figure: Distribution Of the Number of Reviews and Players at 180 Days.

Demand: Identification of β

- Goals: prediction of customer arrival and the effect of reviews on sales
- Unobserved factors $\tilde{x}_{it}$ that are correlated with the observed ones can lead to the OVB

$$A_{it} = \psi_i A_{it-1} + \lambda_i(1 + x_{it}'\beta) + \lambda_i \tilde{x}_{it}'\beta + \tilde{u}_{it}$$

- Solution:
  - Sample selection ruling out changes in quality
  - λ_i controls for time invariant features
  - Time effects to account for day of the week seasonality and big seasonal sales
  - Other controls: age, price, number of reviews, review score
- Review labels don't exhibit enough variation within games
  - Extrapolate changes in customer inflow from games that switch labels

Demand: Estimation

- The NLLS estimator $\hat{\theta}$ is the solution to

$$\min_{\theta} \sum_{i=1}^{n} \sum_{t=2}^{T_i} w_i (A_{it} - \psi_i A_{it-1} - \lambda_i (1 + x_{it}' \beta))^2 \tag{4}$$

- Conditional on β, the problem is OLS
  - Solve for $\hat{\lambda}_i(\beta)$ and $\hat{\psi}_i(\beta)$
  - Concentrate λ_i and ψ_i out

- Solve the concentrated out minimization problem

F.O.C. for the NLLS Problem

Back

$$\psi_i : \sum_{t=2}^{T_i} (y_{it} - \psi_i y_{it-1} - \lambda_i(1 + x_{it}'\beta)) y_{it-1} = 0 \quad (5)$$

$$\lambda_i : \sum_{t=2}^{T_i} (y_{it} - \psi_i y_{it-1} - \lambda_i(1 + x_{it}'\beta))(1 + x_{it}'\beta) = 0 \quad (6)$$

$$\beta : \sum_{i=1}^{n} \sum_{t=2}^{T_i} w_i (y_{it} - \psi_i y_{it-1} - \lambda_i(1 + x_{it}'\beta)) \lambda_i x_{it} = 0 \quad (7)$$

Stylized Model

Back

- A product has of a review status $s \in (0,1)$ can give a discount at cost c
- Status goes up by x with prob. p, down by x with prob. $1-p$
  - Upgrade: s exceeds 1. Downgrade: s goes below 0.
- Payoffs from downgrade, no change, or upgrade: $u^L < 0 < u^H$
- Give the discount if $\mathbb{E}\left[U(s+X)\right] = pU(s+x) + (1-p)U(s-x) - c > 0$
- If $s > 1/2$, then

$$\mathbb{E}\left[U(s+X)\right] = \begin{cases} -c & x < 1-s \\ pu^H - c & x \in [1-s, s) \\ pu^H + (1-p)u^L - c & x \geqslant s \end{cases} \tag{8}$$

Arrival of Reviews

Back

- A buyer at $t-k$ leaves a positive (negative) review with probability r^+_{it} (r^-_{it})
- Good reviews $G_{it} \sim P(\cdot\,; r^+_{it}\lambda_{it-k})$, bad reviews $B_{it} \sim P(\cdot\,; r^-_{it}\lambda_{it-k})$
Review Arrival
- The **like rate** and the **dislike rate** are linear functions of observables

$$r^+_{it} = w'_{it}\rho^+_i \qquad r^-_{it} = w'_{it}\rho^-_i \tag{9}$$

- $w_{it} = [1, disc_{it-k}, \ldots]'$, $\rho^+_i = [\rho^+_{i0}, \rho^+_{i1}, \ldots]$

The Selection Effect

- Probabilities of a like and a dislike off a discount: ρ_{i0}^{+} and ρ_{i0}^{-}
- During a discount: $\rho_{i0}^{+} + \rho_{i1}^{+}$ and $\rho_{i0}^{-} + \rho_{i1}^{-}$
- Review score: probability of a like conditional on a review
- Measure the selection effect by the difference between the expected scores on and off a discount

$$\Delta score_i := \frac{\rho_{0i}^{+} + \rho_{1i}^{+}}{\rho_{0i}^{+} + \rho_{1i}^{+} + \rho_{0i}^{-} + \rho_{1i}^{-}} - \frac{\rho_{0i}^{+}}{\rho_{0i}^{+} + \rho_{0i}^{-}}$$

- Test

$$H_0 : \frac{\rho_0^{+} + \rho_1^{+}}{\rho_0^{+} + \rho_1^{+} + \rho_0^{-} + \rho_1^{-}} - \frac{\rho_0^{+}}{\rho_0^{+} + \rho_0^{-}} \leqslant 0$$
$$H_1 : \frac{\rho_0^{+} + \rho_1^{+}}{\rho_0^{+} + \rho_1^{+} + \rho_0^{-} + \rho_1^{-}} - \frac{\rho_0^{+}}{\rho_0^{+} + \rho_0^{-}} > 0$$

Reviews: Identification

Back

- Turns out, G_{it} and B_{it} are independent
  - → We can estimate the like and dislike rates separately
- The lag between a purchase and review k is set to be the max of the impulse response function of review count to discounts Regression
- Estimation is by maximum likelihood, using the estimates $\hat{\lambda}_{it}$ in place of λ_{it}

$$\mathbb{P}(G_{it} = g) = \frac{(r_{it}^{+}\hat{\lambda}_{it-k})^{g}}{g!} e^{-r_{it}^{+}\hat{\lambda}_{it-k}}$$

- If discount is a dummy, a closed-form solution can be derived

$$\hat{\rho}_{0i}^{+} + \hat{\rho}_{1i}^{+} = \frac{\sum_t disc_{it-k} g_{it}}{\sum_t disc_{it-k} \hat{\lambda}_{it-k}} \tag{10}$$

$$\hat{\rho}_{0i}^{+} = \frac{\sum_t (1 - disc_{it-k}) g_{it}}{\sum_t (1 - disc_{it-k}) \hat{\lambda}_{it-k}} \tag{11}$$

Reviews: Estimates

	(1)	(2)
Constant (Like)	0.074***	0.067***
	(0.002)	(0.001)
Discount (Like)	0.032***	0.017***
	(0.003)	(0.001)
Constant (Dislike)	0.015***	0.013***
	(0.001)	(0.000)
Discount (Dislike)	0.004***	0.002***
	(0.001)	(0.000)
Young (Like)		0.296***
		(0.015)
Old (Like)		−0.045
		(0.001)
Young (Dislike)		0.067***
		(0.006)
Old (Dislike)		−0.008
		(0.000)
Observations	355077	355077

Identification of k

Back

	Dependent variable:		
	New Reviews	New Reviews(%)	Score (0-100)
Discount 0 Days Ago	1.094^{***}	1.801^{***}	-0.159
	(0.122)	(0.135)	(0.152)
Discount 1 Day Ago	3.254^{***}	4.282^{***}	-0.557^{***}
	(0.276)	(0.228)	(0.184)
Discount 2 Days Ago	1.377^{***}	1.969^{***}	-0.470^{***}
	(0.139)	(0.100)	(0.176)
Discount 3 Days Ago	0.743^{***}	1.197^{***}	-0.370^{**}
	(0.110)	(0.072)	(0.170)
Discount 4 Days Ago	0.622^{***}	0.847^{***}	-0.401^{**}
	(0.096)	(0.066)	(0.166)
Discount 5 Days Ago	0.418^{***}	1.023^{***}	-0.395^{**}
	(0.104)	(0.125)	(0.161)
Discount 6 Days Ago	0.071	0.396^{***}	-0.368^{**}
	(0.086)	(0.063)	(0.159)
Discount 7 Days Ago	-0.120^{*}	0.254^{***}	-0.336^{**}
	(0.072)	(0.076)	(0.155)
Weekdays + Week Effects	✓	✓	✓
Game Effects	✓	✓	✓
Observations	350,819	350,819	350,819
R^2	0.059	0.105	0.154
Note:			$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$