The Longshot bias in market data: Evidence from Counter-Strike: Global Offensive

Timothy Schwieg

March 14, 2019

Loot Boxes

- Many video games have chosen to sell cosmetic alterations to their games using randomization mechanisms called “loot boxes”
- Economic Literature tells us that there is no benefit to randomization for risk-neutral consumers, so the benefit must come from risk-loving consumers.
- What aspect of these lotteries is generating the revenue for the companies selling them?
- How much more revenue-generating is this compared to traditional selling mechanisms?

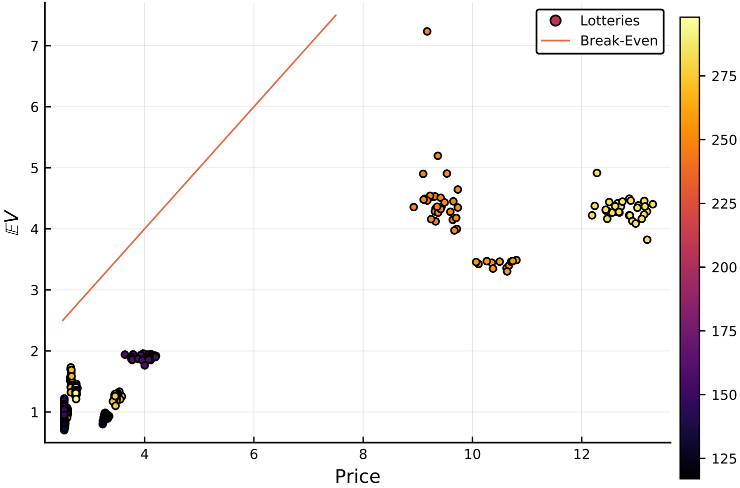
Why do we care?

- We are interested in discovering what drives this market to feature randomization mechanisms.
- Are consumers inherently more risk-loving when they play video games?
- Is this driven by consumers over-weighting tiny probabilities as cumulative prospect theory suggests?
- Are consumers weighing benefits and losses differently?
- What is the magnitude of these gains from randomization?

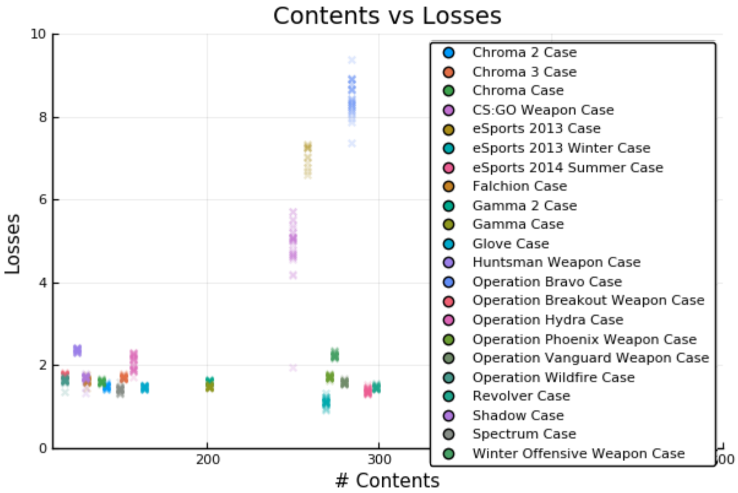
The Data

- Contains complete market history for all items sold in the Steam Community Market for *Counter-Strike: Global Offensive*
- Market history is specific to the hour for the last 30 days, specific to the day for the remaining time the item has existed.
- Contains all active buy and sell orders for each of these items as of March 31st 2018.
- Number of active players per day and unique twitch viewers per day

Pictures


Lotteries
Break-Even
$\mathbb{E}V$
Price
7
6
5
4
3
2
1
4
6
8
10
12
275
250
225
200
175
150
125

Does Size Matter?



Lotteries

	Values		Number of Contents				
Case	$\mathbb{E}[V]$	Price	#Blue	#Purple	#Pink	#Red	#Gold
Operation Wildfire	0.89891	2.5307	26	18	14	9	50
Operation Breakout	0.77011	2.5305	24	15	12	10	56
Falchion Case	0.95072	2.5323	27	24	11	9	59
Shadow Case	0.85299	2.5349	29	17	14	10	59
Huntsman Weapon Case	0.95531	3.3181	25	17	12	8	62
Spectrum Case	0.98146	2.53	34	23	15	9	68
Chroma 2 Case	1.0058	2.53	25	13	13	9	81
Chroma 3 Case	0.66099	2.53	30	19	11	10	81
Chroma Case	0.83215	2.55	23	20	10	4	81
Glove Case	0.84301	2.53	27	26	9	12	89
Operation Hydra	1.5465	4.0827	25	20	14	9	89
Gamma 2 Case	0.68335	2.53	31	22	13	7	128
Gamma Case	0.80717	2.53	31	21	11	10	128

High Content Lotteries

	Values		Number of Contents				
Case	$\mathbb{E}[V]$	Price	#Blue	#Purple	#Pink	#Red	#Gold
CS:GO Weapon	4.4611	9.3248	7	6	7	2	228
eSports 2013 Case	3.2708	10.354	8	13	7	2	228
eSports 2013 Winter	1.5687	2.6441	18	9	11	3	228
eSports 2014 Summer	1.4136	2.7414	21	19	16	9	228
Operation Bravo	4.3567	12.628	26	15	9	6	228
Operation Phoenix	0.85507	2.5416	15	12	9	7	228
Operation Vanguard	1.038	2.5928	17	13	12	10	228
Revolver Case	1.1045	2.53	24	25	12	9	228
Winter Offensive	1.299	3.5079	14	14	12	6	228

Discrete Choice - Berry (1994)

Utility for these lotteries is quasi-linear

$$u_{ijt} = V(x_{jt}, p_{jt}; \theta) + \xi_{jt} + \epsilon_{ij} \quad \epsilon_{ij} \sim Gumbel$$

Consumers choose the lottery that has the highest utility for them:

$$\Pr(i \rightarrow j) = \frac{\exp(V(x_{jt}, p_{jt}; \theta) + \xi_{jt})}{\sum_{k \in \mathcal{F}} \exp(V(x_{jt}, p_{jt}; \theta) + \xi_{kt})}$$

Using an outside option that is normalized so that it has zero utility:

$$\log s_{jt} - \log s_{0t} = V(x_{jt}, p_{jt}; \theta) + \xi_{jt}$$

Implications

- Differentiated Goods
- Prediction based on market shares
- Homogeneous Consumers - Is this reasonable?
- No structure placed on ξ

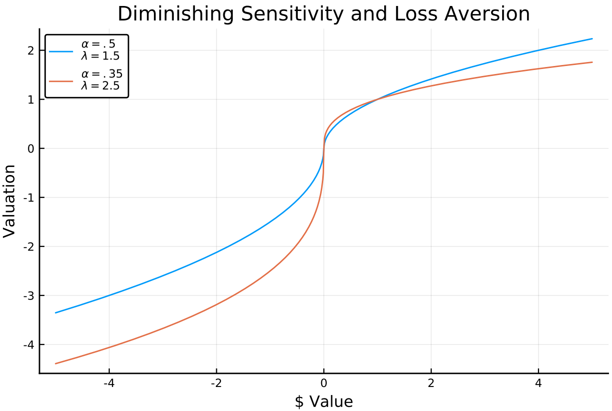
Cumulative Prospect Theory

- Four main components: Reference dependence, loss aversion, diminishing sensitivity, and probability weighting
- Diminishing sensitivity and loss aversion are summarized by the valuation function for each content of the lottery.
- x is not the content of the lottery, but the value of the gain or loss of that content relative to some reference point.

$$v(x) = \begin{cases} x^{\alpha} & x \geq 0 \\ -\lambda(-x)^{\alpha} & x < 0 \end{cases}$$

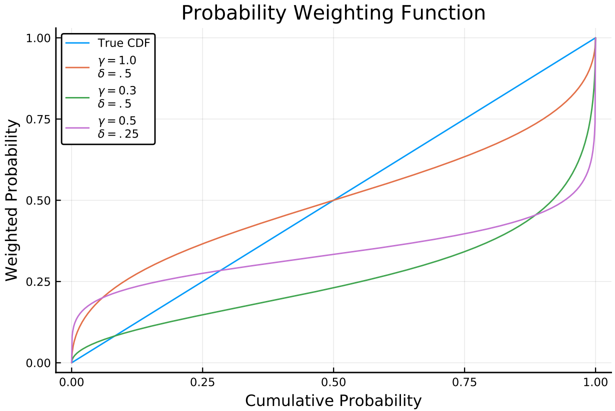
Reference dependence and Loss Aversion

- What is the proper reference point?
- Can it be estimated?
- How is loss aversion tied to the reference point?



Probability Weighting Function

$$w(p) = \frac{\gamma p^{\delta}}{\gamma p^{\delta} + (1-p)^{\delta}}$$



Valuation of a Lottery

- We compute the "as-if" probability taking differences of the weighted-CDF function.

$$\Pi_{s_i} = \sum_{j=1}^{s_i} \pi_{s_j}$$

$$p_i = w(\Pi_{s_i}) - w(\Pi_{s_{i-1}})$$

$$F(x_i) = [w(\Pi_{s_i}) - w(\Pi_{s_i-1})]\, v(x_i - R)$$

$$V = \sum_{i=1} F(x_i)$$

Constant Term

- To normalize the utility to an outside good, need a constant term
- There is no interpretation for this constant term.
- Combines mis-specification of outside good, expected value of ξ and the normalizing utility of the outside good.

Estimation

- Price is determined by intersection of supply and demand and is therefore endogenous
- Instrument with the changes in daily player base from the average number of players

$$\xi_{jt} = \log s_{jt} - \log s_{0t} - \beta - V(x_{jt}, p_{jt}; \theta)$$

Using the orthogonality of ξ_{jt} to the instruments and exogenous parameters:

$$\min_{\boldsymbol{\xi}_{j,t}, \xi_{j,t}} \sum_{j,t} \boldsymbol{\xi}'_{j,t} \Omega \boldsymbol{\xi}_{j,t}$$

$$\text{subject to: } \xi_{j,t} = \log s_{jt} - \log s_{0t} - \beta - V(x_{jt}, p_{jt}; \theta)$$

$$\boldsymbol{\xi}_{j,t} = \xi_{j,t} \boldsymbol{Z}_{j,t}$$

Computation

- Estimated using KNITRO
- RMSE is computed both in sample and for an out-of-sample test to determine over-fitting
- $\bar{R}^2$ is computed as $1 - \frac{\mathbb{V}(\xi)}{\mathbb{V}(Y)}$
- J-Statistic Critical Values:

– Fixed Effects: 5% 314.6784, 1% 332.4796
– No Fixed Effects: 5% 337.1254, 1% 355.5251

Results

$\mathbb{E}[V]$ + Price Reference Point			
α	0.56534 (2.03484)	λ	1.36844 (10.8477)
γ	1.0 (6.36280)	δ	1.0 (9.47887)
In Sample RMSE	1.23649	Out Sample RMSE	1.4337
$\bar{R}^2$	0.18880	J-Statistic	825.185
$\mathbb{E}[V]$ + Price Reference Point and Fixed Effects			
α	0.79549 (4.6084)	λ	0.60091 (9.85376)
γ	1.0 (9.79014)	δ	1.0 (21.5814)
In Sample RMSE	1.08121	Out Sample RMSE	1.10642
$\bar{R}^2$	0.51688	J-Statistic	558.41
Price Reference Point			
α	0.47457 (5.4068)	λ	0.54667 (14.56967)
γ	1.0 (10.97014)	δ	1.0 (10.85583)
In Sample RMSE	1.52584	Out Sample RMSE	1.5258
$\bar{R}^2$	0.08117	J-Statistic	860.261
Price Reference Point and Fixed Effects			
α	0.8215 (7.6682)	λ	0.3152 (7.3252)
γ	1.0 (7.6753)	δ	1.0 (15.7306)
In Sample RMSE	1.01432	Out Sample RMSE	1.07900
$\bar{R}^2$	0.54053	J-Statistic	351.73
Rational - CRRA			
α	0.17411 (6093.657)	β	-0.32488 (0.28937)
In Sample RMSE	0.98513	Out Sample RMSE	1.10097
$\bar{R}^2$	0.52163	J-Statistic	570.394

What stories does this tell?

- Rational CRRA story is one of risk aversion
- Poor fit without fixed effects means that individuals may not be sensitive to price changes
- Cumulative Prospect Models do not tell a story of probability weighting.
- Low fit means that there is more driving this effect than a single representative agent

Where to go from here?

- Belief Heterogeneity
- Preference Heterogeneity
- Non-parametric fit
- Explore other alternatives for views of price
- Larger amount of data used