The benefits of Randomization Mechanisms in Counter-Strike: Global Offensive

Timothy Schwieg

1 Research Question

In the world of video games, a market has appeared for in-game purchases. These cosmetic items affect the aesthetics of a player, but often do not influence the game-play and are sold by the designer. Recently the method of the sales has moved away from the traditional market approach of individual prices for each item, and towards the "loot box" approach. These items are sold in randomized lotteries, often given away, with a cost of opening.

Traditional economic research on randomization informs us that for the risk-neutral customer, there is no benefit to randomization, as the consumer is indifferent. So for this mechanism to be so far-reaching into the market, there must be a risk-loving nature to the consumers. This begs the question of how much money are these companies gaining by exploiting the risk-loving nature of the consumers.

Counter-Strike global offensive presents an interesting case study for these types of markets, as there is a secondary market where individuals can buy and sell these loot boxes, as well as their contents. This was one of the first games to introduce the concept of the randomized "loot box" so there is a long market history available. As important as the secondary market is the public information about the probability of obtaining the contents of the boxes, as required by Chinese Law. Because the supply of the boxes is strictly controlled, the market for these items is lively, with many items trading for hundreds of dollars, and a few entering the thousands.

These factors combine to allow for a structural estimation of demand, and risk-tolerance. I intend to combine a demand estimation BLP (1995) model with Cumulative Prospect Theory to estimate the monetary value of randomization in the market for weapon skins in Counter-Strike Global Offensive.

2 Literature Review

The literature in demand estimation is primarily focused around the seminal paper written by Berry, Levinsohn, and Pakes (1995). This paper presents a frameowrk for estimation of a heterogenous consumers in a discrete choice logit demand framework. This allows for a richer substitution framework, and an ability for the substitution affects to escape the independence of irrelevant alternatives result of traditional logit demand.

One such example of estimation is the paper by Nevo (2001). It estimates the demand in the cereal industry in order to determine the market power in the industry, and determine if the high product margins were caused by brand recognition, or by collusive behavior between the few firms in the industry. This type of counter-factual estimation is common within the literature, and is tested in many ways from both the supply and the demand side of the estimation.

I intend to take a different path from what is commonly performed with these tools, and attempt to use the estimated parameters to compute what these consumers would have been willing to pay for an item under some different policy regime (no randomness)

Estimation of these models began with the strategy first suggested by Berry et al. (1995) commonly referred to as the Nested Fixed Point Algorithm, but has recently been superseded by the Mathematical Programming under Equality Constraints suggested by Su and Judd (2012). This algorithm performs extremely well under sparse Hessian and gradients, of which my method contains many. This will allow for significantly easier estimation of the demand system.

Barberis (2013) presents a broad review of where Prospect theory has been applied, as well as its problems with its application, particularly in the choice of a reference point, which appears to be very significant, but there is little guidance on what to choose beyond possibly the expected value of the lottery. Applications of the model, originally proposed by Tversky and Kahneman (1992) exist mostly in finance and insurance. I intend to extend this body to look at the behavior of non-expert individuals in a market scenario. I believe that this area has not had many applications, likely because of the rarity of quality data outside of these fields.

3 Data

The data are market transaction history for all items sold on the *Steam Community Market* for Counter-Strike: Global Offensive.

Counter-Strike Global Offensive is a first-person shooter game where one team (terrorists) attempt to plant a bomb and defend it while the counter-terrorists attempt to defuse the bomb. Each team has specific guns that they are able to purchase at the start of every round. The in-game cost, game balance, and meta-game all contribute to the popularity of each weapon. Players may choose to purchase purely cosmetic "skins" for their weapons which change the appearance of their weapon when they buy it. These skins are sold in lotteries called weapon crates which are dropped randomly to players in-game. The drop rates are unknown, and believed to change often. Upon receiving a weapon crate, a player may elect to spend $2.50 to open it, or sell it on the community market.

These crates display which weapon skins they may contain, and the probability of obtaining each item within the crate is public knowledge. That is, the contents of the crate follow a known distribution, and can therefore be estimated under theories of risk. The contents of the crate can then be held onto, or sold at market.

3.1 Market

The market that these weapons can be sold at is the *Steam Community Market* which is run by Valve, the same company that makes Counter-Strike: Global Offensive. The market is a continuous time double-auction. Sellers may place sell orders, and buyers buy orders, and the market functions by matching the buyers and sellers, always selling at the seller's price. This is known to converge quickly to a competitive market, and will be treated as such for this project. Cripps and Swinkels (2004) There are two complications however, there is a 15% tax placed on the market by Valve, which is taken from the seller's earnings. This is complicated by the discrete nature of the selling, and the tax always rounds up in favor of Valve. That is, an item selling for $0.03 would return $.02 to Valve rather than 15%. This will not be a large factor in my model as I am primarily interested in calculating demand.

For the past 30 days, there is data on hourly median market price as well as quantity sold. For the remaining time that an item has been at market, there is data for daily median price and quantity sold. The data also contain active buy and sell orders at the time of its mining: (June 7^{th} 2018). No history for these buy and sell orders is available.

3.2 Characteristics

Since the model used will be in the characteristic space rather than the product space, I am especially interested in characteristics of the different weapons in the game. I shall ignore the characteristics that will be used to determine the market for the weapon, detailed in Assumption 1 in the model section. Unique to each weapon is a float value, between 0 and 1, which indicates the wear on the weapon. Wear does not change with use, and is determined when a weapon is un-boxed. This float is distributed uniformly, but based on its value, places the weapon into different brackets for sale. We will consider all weapons in a particular bracket as homogenous.

Float	Condition
0.00 - 0.07	Factory New
0.07 - 0.15	Minimal Wear
0.15 - 0.38	Field-Tested
0.38 - 0.45	Well-Worn
0.45 - 1.00	Battle-Scarred

Independent of wear, each item also has a 10% chance of being StatTrak™, where the gun includes a tracker that counts the number of kills a player has with this weapon. This number is reset on sale, so it can be treated simply as a binary indicator.

The contents of the crate are divided into several tiers, based on their rarity from being obtained in a box. These tiers and their probability of being obtained are given below:

Probability	Rarity
.0026	Special (Gold)
.0064	Covert (Red)
.032	Classified (Pink)
.1598	Restricted (Purple)
.7992	Mil-spec (Blue)

4 Model

I intend to estimate a structural model for the demand for the contents of the boxes, using this, we can determine the distribution of valuations for a risk-neutral consumer for the boxes, and then estimate the risk-preference of the individuals that open the loot-boxes. From there we can calculate the benefit of randomization compared to selling each item at market.

4.1 Demand Estimation

We wish to estimate the demand for this model using a discrete choice model for demand. This immediately raises the concern that it only allows for one good to be purchased, and it is common for individuals to have many weapon skins in the game. To this end, we shall split the market into several sub-markets and make a heavy identifying assumption. This assumption will allow for the discrete choice model to be applicable, and also creates price instruments for estimation.

Assumption 1. *Items are split into markets defined by the in-game role that all of the weapons in this market fulfill.*

These markets are defined by domain knowledge. For example, we treat the AK-47, the single most popular gun in the game as its own market, competing only with its own skin and condition variants. However, the M4A4 and the M4A1-S will be considered as competitors, as will the CZ75, Tec9, and Five-Seven. Weapons that fill the same role, or the same weapon slot will be considered in the same market. The assumption takes the form of claiming that one individuals do not substitute between roles, and only consider substitution between weapon skins for the same role. This ensures that consumers only purchase a single item at a time, as one could never equip multiple skins for the same role. The full power of this assumption will become clear in the instruments section.

4.1.1 BLP

To estimate the demand for the contents of the boxes, I intend to implement a standard BLP demand estimation model (1995). This is a discrete choice demand system. Consider J goods in T markets for I consumers indexed by j, t, i respectively. Assuming quasilinear utility, the utility for consumer i purchasing good j is:

$$u_{ij} = \alpha_i p_j + x_j \beta_i + \xi_j + \epsilon_{ij}$$

Where p_j is the price of good j, x_j are the observed characteristics of good j, ξ_j are the characteristics of good j observed by consumers and producers but not by the econometrician, α_i, β_i are consumer i's individual preference parameters over these characteristics, and $\epsilon_{ij} \sim T1EV(0)$ is a random shock only observed by the consumer. This is a standard logit model, but we have unobserved heterogeneity among consumers.

Consumer i then chooses the good that gives him the highest utility, the probability that that good is good j is given by:

$$\Pr(i \to j) = \frac{\exp(\alpha_i p_j + x_j'\beta_i + \xi_j)}{\sum_{k \in \mathcal{F}_t} \exp(\alpha_i p_k + x_k'\beta_i + \xi_k)}$$

Each consumer has individual logit demand. If we choose to normalize the mass of consumers to one, then the market share of good j should be equal to the expected value of this individual demand, averaged over the distribution of valuations.

$$\pi_j = \mathbb{E}\left[\Pr(i \to j)\right]$$

Let us define the observed market shares as:

$$\hat{s}_j = \frac{1}{I}\sum_{i=1}^{I} \mathbb{1}_{\{y_i = j\}}$$

From the Weak Law of Large Numbers, we believe that $\hat{s}_j \xrightarrow{p} \pi_j$. Define the distribution of (α_i, β_i) as θ. By assuming that this convergence in probability has been reached, we arrive at:

$$\hat{s}_j \approx \mathbb{E}\left[\Pr(i \to j)\right] = \int \Pr(i \to j) d\theta \approx \frac{1}{N_s}\sum_{i=1}^{N_s} \Pr(i \to j)$$

Where we approximate the integral of $\int \Pr(i \to j) d\theta$ through any numerical integration technique. This expression can then be inverted to solve for ξ_j, which is unobserved.

4.1.2 Instruments

In this specification of the model, there are two sets of endogenous variables. Price is obviously correlated with the unobserved characteristics of the model, but market share is also endogenous within the model. We shall require two sets of instruments, one for price, and one for market share.

Valid price instruments are those that are correlated with supply shocks, but are not correlated with the demand shocks in the model. It is worth defining precisely what are the supply and demand for this model.

The supply for each weapon skin is the set of people who have opened the loot box that contains that item and have elected to sell it. Shocks that will affect this are changes in consumer tastes leading to less people choosing to sell, as well as changes in the drop rates of the crates, controlling the flow of this item into the market.

Demand for this good is the individuals who elect to buy the good at the market rather than attempt to earn it through opening loot boxes. The shocks that affect these people are entrance and exit to the market as well as changes in taste. (Needs more here)

A Valid price instrument is something that is correlated with supply shocks, but not with the demand shocks. For this we will take the prices of the other contents of the box that are not in the same market as the good at hand. By Assumption 1, these prices are exogenous to the unobserved characteristics of the good at hand. They are however affected by the changes in the drop rate of the loot box that provides them, since they come from (nearly) the same supply. This is a form of the Hausman instruments used often in the literature.

For market share, we intend to use the BLP instruments, which require that the valuation of one characteristic of a good is not random across the consumers. When this is satisfied, we may use the sum of the characteristics of the competitors of the good as instruments for the market share. If necessary, following Gandhi and Houde (2017), we may construct higher order approximations of the optimal instrument for the market shares using the observed characteristics.

4.1.3 Estimation

Once a set of instruments has been computed, estimation of the model requires using the orthogonality condition of the instruments against the computed values of ξ_j. Our orthogonality condition is: $\mathbb{E}\left[\xi_j z_j\right] = 0$.

This can be estimated using the generalized method of moments. Following the method of Su and Judd (2012), we may estimate this using Mathematical Programming under Equality Constraints as follows:

$$\min_{\boldsymbol{\xi}_{j,t}, \xi_{j,t}} \boldsymbol{\xi}'_{j,t} \Omega \boldsymbol{\xi}_{j,t} \tag{1}$$

$$\text{subject to: } s_{j,t} = \frac{1}{N_s} \sum_{i=1}^{N_s} \frac{\exp(\alpha_i p_j + x'_j \beta_i + \xi_j)}{\sum_{k \in \mathcal{F}_t} \exp(\alpha_i p_k + x'_k \beta_i + \xi_k)} \tag{2}$$

$$\boldsymbol{\xi}_{j,t} = \xi_{j,t} \boldsymbol{z}_{j,t} \tag{3}$$

This method allows for the exploitation of sparseness in many commercial solvers. This is important as assumption 1 has imposed this level of sparseness on the model in part for computational ease.

4.2 Risk Preference

Following the estimation of (1) we may back out the parameters of the distribution θ. These parameters form the distribution of valuations for each of the contents of the loot box. The risk-neutral distribution of valuations can be calculated as this valuation is a convex combination of the valuations for all the contents. Of interest is how this distribution varies with the distribution of valuations observed by those purchasing loot boxes. Since these distributions have been assumed to be normally distributed, we obtain the distribution of these valuations, as the sum of normally distributed random variables is normally distributed as well.

However, it is not immediate that these two distributions are at all comparable. There must be some structural difference between individuals who choose to open the boxes, and individuals who choose to buy the goods at the secondary market, or they would be performing the same action. Either there is no heterogeneity between individuals, and the market is in equilibrium, or there is heterogeneity and these two markets exist to separate the two types.

The simpler of these two explanations is that there is no heterogeneity among risk for individuals, and that these two markets are in equilibrium.

Assumption 2. *Individuals are homogenous in risk-preferences*

In this case, the individuals that purchase the loot boxes have exactly the same distribution of valuations for the items as do the individuals that purchase the items at market. We wish to estimate this risk tolerance by the difference in the price observed at market and the risk-neutral distribution.

4.2.1 Cumulative Prospect Theory

However, these lotteries contain a mixture of small losses and large gains. Expected utility theory is relatively poor at explanatory power in this area, and for this reason I choose to estimate the primitives using Cumulative Prospect Theory.

Cumulative Prospect Theory includes several important concepts not observe in classical decision making under risk.It incorporates reference dependence, which implies that people view things in the context of losses and gains rather than changes to their overall wealth. This is attractive for computational reasons. It also utilizes loss aversion, the notion that losses are relatively more costly than gains. The model also incorporates diminishing sensitivity, the notion that valuations are concave in losses and convex in gains. The final concept is probability weighting, the notion that consumers act as if they were facing different probabilities than what they encounter.

Maintaining with the function forms first suggested by Tversky and Kahneman (1992), I will attempt to find the parameter values that maximize the likelihood of obtaining the prices observed in the market.

We note that the price of opening a case is two-fold, first the case must be bought at the market for its price, and then the price of the key, denoted p_{key} must also be paid to open the case. The time required to open the case is trivial, and will not be considered. Since gains are treated differently than losses, let the parameterization of these gains and losses be defined as:

$$v(x) = \begin{cases} x^{\alpha} & x \geq 0 \\ -\lambda(-x)^{\alpha} & x < 0 \end{cases} \tag{4}$$

In cumulative prospect theory, the cumulative mass (distribution) function is weighted such that individuals overweight the tail probabilities. This is especially important in this model, as there are many high valued rare items, that if this part of the theory is correct, heavily influence the valuation of the box, despite their extremely low probability of occurrence.

$$w(P) = \frac{P^{\delta}}{(P^{\delta} + (1-P)^{\delta})^{\frac{1}{\delta}}} \tag{5}$$

To define the decision weights π_i, we must first order the prospects of the lottery in ascending order of gains. the weight π_i then is defined by:

$$\pi_i = w\left[\sum_{j=-m}^{i} P(x_j)\right] - w\left[\sum_{j=-m}^{i-1} P(x_j)\right] \tag{6}$$

These can all be combined to form the Valuation Transform for the consumer:

$$F(V_i) = \begin{cases} \left[w\left(\sum_{j=-m}^{i} P(x_j)\right) - w\left(\sum_{j=-m}^{i-1} P(x_j)\right)\right](V_i - p_l - p_{key})^\alpha & (V_i - p_l - p_{key}) \geq 0 \\ -\lambda\left[w\left(\sum_{j=-m}^{i} P(x_j)\right) - w\left(\sum_{j=-m}^{i-1} P(x_j)\right)\right](p_l + p_{key} - V_i)^\alpha & (V_i - p_l - p_{key}) < 0 \end{cases} \tag{7}$$

Under cumulative Prospect theory the distribution of valuations for a lottery in the population is given by the distribution of:

$$\sum_{j=-m}^{n} F(V_j) \tag{8}$$

This is then the demand for the loot box as a function of the cumulative prospect parameters: α, δ, λ. It is important to note that the act of purchasing only implies that an individual has a valuation higher than the market price. Only for active buy orders can we take the valuation as non-censored.

These parameters can then be estimated using Censored Maximum likelihood Estimation. This presents its own concerns as this is a non-convex optimization problem, however since the dimension of the problem is so small, an intractable problem such as this may still be solvable. If this function is too computationally difficult to estimate, it can be simulated using Monte-Carlo methods, and the likelihood taken from the kernel-smoothed density function of the simulation.

Once these parameters have been estimated, we can then compute the monetary benefit of the randomization. This is defined as the difference between the valuation of the loot box under cumulative prospect theory, and the risk-neutral valuation of the loot box. This value integrated over the distribution of consumer types is the total value gained from the risk preferences of consumers.

$$\Pi = \int \sum_{j=-m}^{n} \left[F(V_j) - V_j\right] d\theta \tag{9}$$

References

Barberis, N. C. (2013). Thirty years of prospect theory in economics: A review and assessment. *Journal of Economic Perspectives 27*, 173–196.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica 63*(4), 841–890.

Cripps, M. W. and J. M. Swinkels (2004). Efficiency of large double auctions.

Gandhi, A. and J.-F. Houde (2017). Measuring substitution patterns in differentiated products industries.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica 69*(2), 307–342.

Su, C.-L. and K. L. Judd (2012). Constrained optimization approaches to estimation of structural models. *Econometrica 80*(5), 2213–2230.

Tversky, A. and D. Kahneman (1992). Advances in prospect theory: Cumulative reperesentation of uncertainty. *Journal of Risk and Uncertainty 5*, 297–323.