

# Outline

Timothy Schwieg

## 1 The Data

The data is market data taken from the steam community market on most of the items that can be bought and sold at market. Almost all items in counter-strike are trade-able, and while there are some slight restrictions such as no-trade periods on certain items after purchase and other things, it is relatively well modeled by a continuous double auction. Each item sold has a quality attached to it that is distributed uniformly on  $[0, 1]$ ; these qualities are broken down into several classes and each of those are sold separately on the market.

One important characteristic of the auction is that while there are buy and sell orders placed, and these are used to facilitate the exchanges between the participants, the seller's price is always paid. Without considerations of dynamics, this would mean that the buyer faces a dominant strategy of revealing his valuation when placing his bid. However the effects of the seller shading will likely be minimal since this is a very large online double auction, and is converging to the competitive equilibrium very quickly.

The data collected take three forms: Most of the data is price and quantity history data taken from market transactions. Within the last 30 days there is hourly data on the median price and quantity of items sold at the market. There is no specific buy and sell order data about these transactions. This complication makes estimating the model as a double auction difficult, as there is no data on the specific buy and sell orders made, even the winning ones. All that is recovered is the price and the quantity. The structure of the auction, where the seller's price is paid means that we do observe the seller's bids, but not the buyer's bids. However, only the median price is observed, not the actual price of all transactions occurring within that hour. This means that there is some error, but since it is pretty negligible it will just have to be ignored.

The other forms the data take are the outstanding buy and sell orders in the market. These are buy and sell orders that have gone unfulfilled so far. They make up about a third of the data for most of the items on the market. However this data is more useful as it is the only part of the data that is the actual bids. This data is split into two parts, the buy orders and the sell orders, each of which has the price and the cumulative number of buy orders that would be willing to purchase at this price. The fact that this gives us separate supply and demand only data points should help during identification.

## 2 The Model

For any given item, assume that there is a mass of consumers who have valuations based on some distribution  $F_V$ . With some exogenous probability, some consumers are endowed with



~~As a result there is~~ the same distribution of valuations in those endowed as well as those who are not endowed, even if there are different numbers of people who are endowed.

~~However, since this process of granting items is random, it is in no way efficient. There is no process that leads to the individuals who value the good most receive it under the current function. In order to achieve this efficiency, there is a market instituted, taking the form of a double auction.~~

It is known that double auctions converge rapidly to a competitive environment. The data pulled is relatively poor for extracting valuations from the bids, as for much of the data the bids are not observed; as a result attempting to identify valuations from the bids would not work well. While the result is certainly possible for double auctions under conditions such as sealed-bid, and one buyer and seller, there has been no identification, based on a dominant or equilibrium argument in the continuous double auction.

As a result of this, we will abstract from the dynamics and the mechanism of the double auction, and due to the large amount of traffic, focus on its convergence into a competitive market. If the market is efficient, then we will see a matching between buyers and sellers such that after the trades, only those with the highest valuations have the items.

## 2.1 The Matching Problem

Since it is known that the planner's problem of maximizing total welfare, and the decentralized market are equivalent for this problem, we may look at either interchangeably to provide context for the problem.

The surplus generated by any exchange between a buyer and a seller is given by the valuation of the buyer minus the valuation of the seller. A central planner, who wishes to maximize the total surplus then faces the question of finding a path between the Cartesian product of buyers valuations and sellers valuations that maximizes:

This process simplifies to a continuous linear program, and while it does have a solution and a duality gap of zero, it is quite unwieldy to work with. It is easier to consider a discretised alternative: breaking the distribution of valuations into discrete chunks. Instead of a distribution of types of buyers and sellers, there are finitely many, which represent the mass that is contained in their quantiles. One major advantage of this approach is that it is very simple to handle the question of there being different numbers of buyers and sellers.

While under the continuous linear program, one would have to ensure that the manifold was measure-preserving after controlling for the percent of the population who has the item, in the case of the discrete version, one must only increase or decrease the number of buyers and sellers.

Under the discrete model, we consider  $J$  quantiles of the valuations for the buyers, and  $J$  quantiles of the valuations for the sellers. The valuations are therefore given by the inverse distribution function applied to the index of the buyer divided by  $J$ , or the index of the seller divided by  $J$ . The linear program for discrete planners problem is as follows:



$$\max_{\alpha_{i,j}} \sum_{i=1}^I \sum_{j=1}^J \left[ F_V^{-1} \left( \frac{i}{I} \right) - F_V^{-1} \left( \frac{j}{J} \right) \right] \alpha_{i,j}$$

subject to:  $\forall j, 1 \leq j \leq J \quad \sum_{i=1}^I \alpha_{i,j} \leq 1$

$\forall i, 1 \leq i \leq I \quad \sum_{j=1}^J \alpha_{i,j} \leq 1$

The constraints serve to require that each individual make at most one exchange. Of interest as well is the dual of the problem, which is specified below.

$$\min_{x,j} \sum_{i=1}^I x_i + \sum_{j=1}^J y_j$$

subject to:  $\forall i, j; 1 \leq j \leq J, 1 \leq i \leq I$

$$x_i + y_j \geq F_V^{-1} \left( \frac{i}{I} \right) - F_V^{-1} \left( \frac{j}{J} \right)$$

The solutions to this problem form the shadow prices of the exchange, or the amount of surplus that a buyer or seller takes based on his or her type.

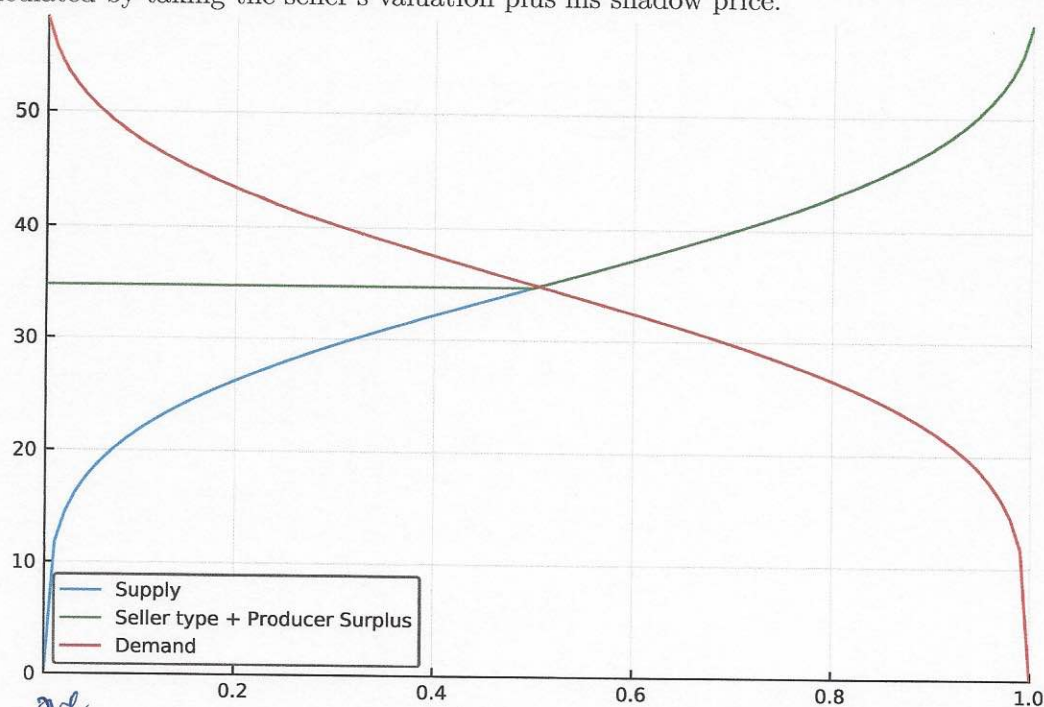
### 2.1.1 Results

One important thing to note about the Objective function is that despite the transformation of the inverse ~~CDF~~, it remains the buyer's valuation less the seller's valuation, and this function is both super-modular and sub-modular. This implies that for this matching problem, both PAM and NAM are supported. After some inspection, one can see that while the process will determine which portion of the sellers and buyers match, any permutation within that is just as optimal. *even though*

*That said,* However, the dual of the problem does have a unique solution, as it is the shadow price for the type of the seller and the buyer. These values are the producer and consumer surplus for each type. Since it is a competitive equilibrium, there is one price supported, as the good is homogeneous, and the matching is occurring between valuations for the good. The seller's valuation plus his shadow price will be equal to the competitive price for all sellers who do exchange.

For equal sized buyer and seller valuations, this gives the intuitive result that the lower half of the distribution of sellers will sell to the upper half of the distribution of buyers, and we will have the efficient result. As the size of the seller's mass shrinks, with the rarity of the item increasing, more of the sellers choose to sell, and the receiving end of the distribution of buyers shrinks, as the price increases. This is demonstrated below for valuations that are

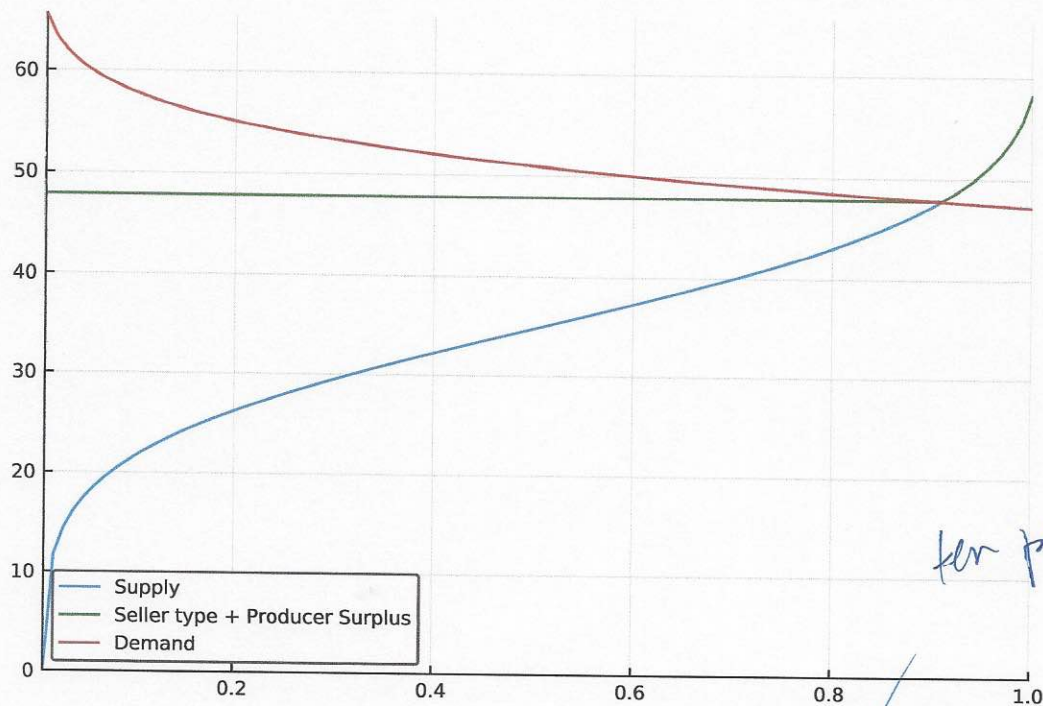
distributed normally, with mean 35, and standard deviation of 10. The equilibrium price is calculated by taking the seller's valuation plus his shadow price.



If we consider the decentralized market version of the problem, all buyers are indifferent between the sellers they choose, as they must give up the producer surplus to the seller, and as a result face a constant price to buy from any seller type.

When the proportion of the population are buyers is increased to ten times the proportion of the population that are sellers, we see the result change:





The distribution of buyers has become truncated by the difference in the number of buyers and sellers. To maintain the efficient outcome, only the top 10% of the buyers are able to purchase, and 90% of the sellers are now selling. The result is a much higher price.

While we would want to put down this change in the price to constant demand, but a decrease in supply, the distribution of sellers has remained constant, and in fact more of them are selling now. Within the context of this matching model, the change in the relative sizes of the population of suppliers acts to truncate the buyers rather than lower the supply. It is important to note that these are not exactly supply and demand in the normal sense, as instead of quantity, the x-axis is the proportion of the sellers that exchange.

### 2.1.2 Equilibrium

As a result of the lens in which this market is viewed, we have a slightly different sort of equilibrium. While all the desirable properties of an equilibrium hold, notably efficiency, and being in the core, we are only examining exchanges in one good, so it remains a partial equilibrium.

Let us assume that the valuations of the players are distributed normally, as in the examples above. Then the supply function can be written as  $q = \Phi\left(\frac{p-\mu}{\sigma}\right)$  and the demand function can be written as:  $q_{1-\xi} = 1 - \Phi\left(\frac{p-\mu}{\sigma}\right)$ . Where  $\xi$  is the percent of people endowed with the item. In equilibrium, the quantity of buyers and sellers are equal:

$$\Phi\left(\frac{p-\mu}{\sigma}\right) = \frac{1-\xi}{\xi} \left(1 - \Phi\left(\frac{p-\mu}{\sigma}\right)\right)$$

$$p^* = \mu + \sigma \Phi^{-1}(1-\xi)$$

*which* This tells us, that in equilibrium, the price that the market supports is the average valuation plus a component that depends on the rarity of the item. Essentially, this claims that the price is controlled by some universal notion of value, such as the design of the skin, as well as a rarity element that drives price up or down depending on how easy it is to obtain.

## 2.2 Identification

*are* For some fixed  $\xi$ , this model gives a deterministic price for some distribution of valuations. If we were to claim that the randomness in this model arises from some unobserved error, then it remains unidentified? *exist* We would find that  $p^* = \mu + \sigma \Phi^{-1}(1-\xi) + U$ . *There are three* There are three primitives, but any estimates of the price would only have a single dimension. *There are no* There are no published numbers of  $\xi$  and the mechanism for determining it is complicated at best, and impossible with currently available data.

### 2.2.1 Estimating $\xi$

*Although* If we were able to estimate  $\xi$ , the problem becomes one of regression, and the covariates suffer from measurement error. This would lead to biased and inconsistent estimates of the coefficients, *and* while clever rearrangement of the model might allow for estimation, it is quite difficult to estimate  $\xi$  outside of the model. Crude estimates of  $\xi$  may be able to be obtained using the number of crates sold and the probabilities of each item being unboxed by the item. However there are several complications that make this almost impossible to handle.

- *There is no data* There is no data concerning the actual inventories of active players, as players are able to set their inventories as private, preventing anyone from seeing their contents
- Items can be combined into other items of higher quality, and there is no data on the percentage of times this has been done.
- The actual drop rate of the items is unknown, and the amount of possible drops is limited to a only two per week per player. There are no reliable estimates of the drop rate, nor what factors affect it.

Since rare items are obtained almost exclusively through opening loot boxes, one could obtain an estimate of the percentage of people endowed with the item by taking the number of the lotteries sold and multiplying it by the probability of obtaining that particular item in the lottery. However the error cannot be quantified, and any regression coefficients remain biased and inconsistent.



### 2.2.2 Another Estimation Strategy

If all estimates of  $\xi$  are unsatisfactory, as I believe, then one method of estimating  $\mu$  and  $\sigma$  is ~~allowing~~ the deviations in the price not be caused by random additive shocks, but instead by randomness contained within  $\xi$ . By assuming a distribution on  $\xi$ , ~~we may allow~~ for the randomness in the price, even when all other covariates contained in the mean do not change. Effectively, we buy identification by taking a very strong stance on how the endowments are distributed among the population, and that all the noise in the price is caused by these deviations in the distributions among the population.

Computationally, this result is not very clean, ~~as~~ unless it happens that the distribution is uniform on the entire unit interval, the distribution of  $p$  will not take a very "nice" form. Effectively, this identification assumption has quite a lot of power in determining the form of the valuations, and calls for a very strong assumption on the nature of  $\xi$ . However, as stated above, there is very little information known about  $\xi$ , as it is unobserved, and taking a strong stance about the nature of it is at best guess-work.

### 2.2.3 Using the Quantity sold to approximate

All of the calculations so far have only used the price data, but we may be able to use the quantity sold for a useful calculation. Since the amount that is sold is determined by the percentage of the population that receives the item, and the distribution is important only for calculating the price that the item costs, we may use the quantity of an item sold divided by the number of active players to determine the percentage of the population that has exchanged. While this cannot account for exchanges that did not take place on the market, it is still the best estimate that can likely be gathered from the data.

Since we know that for each item sold, there are different qualities sold at market, and the probability of obtaining each quality is known, we may form the estimate for each of the different qualities. This allows us several values of  $\xi$  observed, for which we will have to assume that the mean is constant. However, this forces a zero restriction of quality on the mean of the valuations, which is a rather unreasonable assumption. By approach the model this way, it claims that the differences in prices between the different qualities of items is driven solely by the probability of them being dropped. This is unreasonable. Any attempt to put indicators for the quality inside the mean will cause there to be colinearity in the covariates, and linear squares will not produce a result. If however, we are willing to accept this misspecification error as small enough to not cause problems, or if we only examine the highest qualities among which there is almost no discernible difference, we can estimate this model using linear regression.

Since only the drop rate will be measured will be measured with error, we need only rearrange the regression so that the drop rate is the dependent variable, for which measurement error does not induce bias and inconsistency. The estimable model would then be:

$$\Phi^{-1}(1 - \xi) = \frac{p^*}{\sigma} - \frac{\mu}{\sigma}$$

$$\Phi^{-1}(1 - \xi) = \beta_0 + \beta_1 p^*$$



This method can be estimated using linear regression, and the values of  $\beta$  can be adjusted to determine the true values of the  $\mu$  and  $\sigma$  for the distribution. However, all these results are driven by forcing the quality to have no effect on the mean, and the magnitude of this error cannot be observed. What we would like to seek is another way to observe changes in  $\xi$  that does not require such a strong assumption.

## 2.3 A Dynamic Approach

One possible way to handle the identification is to use the only covariate that has a zero restriction on the mean: the time. Let us consider a series of time intervals, in which there is a matching device. In each interval, a percentage of the population is awarded the item, and the matching device functions as above. We may use the same strategy as above, estimating  $\xi$  using the quantity sold over the total number of players in the time interval, and this in fact may be more precise than the estimate used above. However this number can change over intervals, giving us the changes in  $\xi$  needed to identify the mean and the standard deviation in our model.

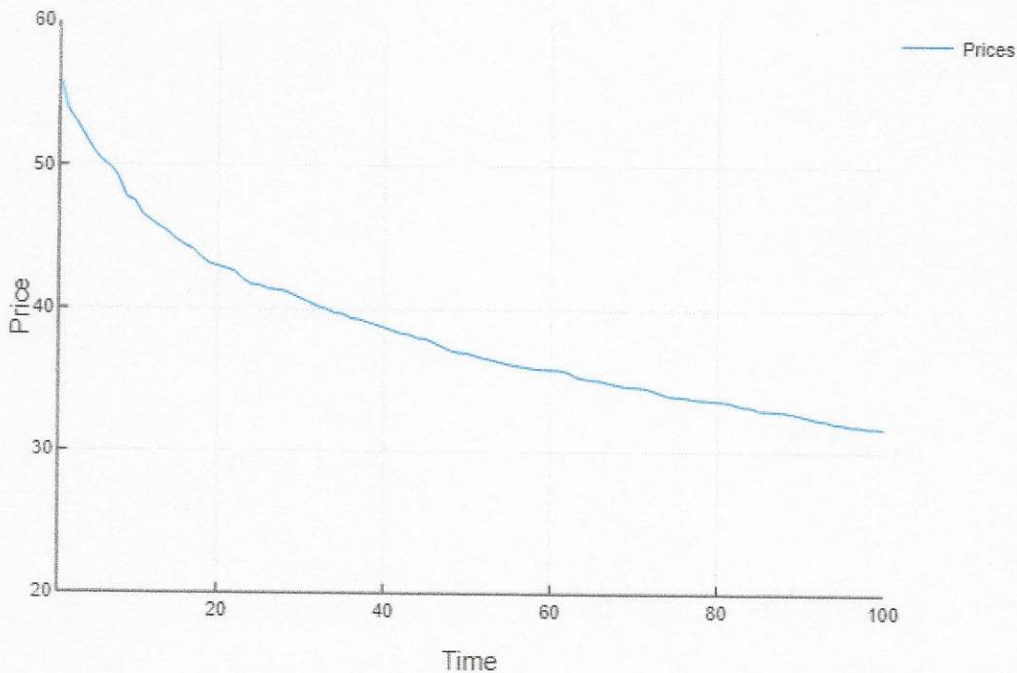
Let us first consider the model with no entrants. After the initial exchange, those that do not have the item are random attributed the item again, but their distribution is no longer the initial distribution, it has been conditioned on losing the top portion of its mass. Therefore, the distribution of those that are possible sellers is a mixture of this truncated distribution, and the top portion that left the potential buyers. However, in this model, the top portion of those that have it will never sell the item, as the valuations of those that do not are all strictly below them. So we may consider the seller distribution to be a percentage of the buyers. The process then repeats, albeit with a slightly truncated portion of the valuation function.

This model also more captures more elements of the market than the original, as it can explain the behavior observed of a high initial price, and it slowly dropping to some equilibrium level. With an explanation of the dynamics of the process in place, we can look at the entire lifetime of the item, and we only have to control for the truncation of the valuations for the demand.

As long as there is no entrance of individuals into the model, the price will necessarily decrease. For the same valuation function as previous examples: Normally distributed with  $\mu = 35$ ,  $\sigma = 10$ , with a drop rate of 0.01 per interval and  $N = 1000$ , a simulation of the price over these intervals is plotted below.

$\$N\$$





One useful result of doing this is that we may be able to get a more precise estimate of the drop rate, by looking at the number sold in the first interval that the item was on the market, as it is far less influenced by exchange and other unobserved factors. This number divided by the total number of active players will likely give a much better estimate of the proportion of players who receive the item per interval.

### 2.3.1 Specification

Let us be specific with the notation used in this model. For each time period  $t$ , the drop rate to individuals estimated is given by:  $\xi_t$ . The price observed in that period is  $p_t$ . In the first time period, everything proceeds according to the previous model. However in the second time period, ~~we allow for the top  $\xi_0$  percent to exit the model~~, so

The supply function has become

$$Pr[Z < z | Z < F_V^{-1}(1 - \xi_0)] = \frac{F_V(z)}{F_V(F_V^{-1}(1 - \xi_0))} = \frac{F_V(z)}{1 - \xi_0}$$

$$q = \frac{\Phi\left(\frac{p - \mu}{\sigma}\right)}{1 - \xi_0}$$

The demand function is adjusted by both  $\xi_0$  and  $\xi_1$ .

$$q\left(\frac{\xi_1}{1 - \xi_1}\right) = 1 - \frac{\Phi\left(\frac{p - \mu}{\sigma}\right)}{1 - \xi_0}$$



This pattern continues, yielding:

$$\begin{aligned}
 q_s &= \frac{\Phi\left(\frac{p-\mu}{\sigma}\right)}{\prod_{t=0}^{T-1}(1-\xi_t)} \\
 q_d &= \frac{1-\xi_T}{\xi_T} \left[ 1 - \frac{\Phi\left(\frac{p-\mu}{\sigma}\right)}{\prod_{t=0}^{T-1}(1-\xi_t)} \right] \\
 p^* &= \mu + \sigma \Phi^{-1} \left[ \prod_{t=0}^T (1-\xi_t) \right]
 \end{aligned}$$

Since  $\xi_t$  is estimated with error, we must still ensure that it appears in the dependent variable instead of as a covariate, so the model that we must estimate is as follows:

$$\Phi^{-1} \left[ \prod_{t=0}^T (1-\xi_t) \right] = \beta_0 + \beta_k 1_{\{Quality\}} + \beta_p p^* + U$$

From this, we may obtain our estimates of  $\mu$  and  $\sigma$ , controlling for the changes in the average valuation caused by the different qualities. Any other covariates can be added to the model as well, and controlling whether or not they affect the mean or the standard deviation by multiplying the indicator by  $p^*$ .

$$\Phi(p^*)$$