CSGO Skin Case Market Analysis

Gavin Sears
sgsears@haverford.edu

James Simbolon
jsimb@haverford.edu

Abstract

Confidence intervals are a way of measuring uncertainty with projected data. This interval presents itself as a region surrounding the projected data, with a percent value attributed to it. The percent value represents the percentage of confidence intervals that will contain the true projected values.

1 Introduction

The freemium video game Counter Strike: Global Offensive has a vast and untamed market for cosmetic items. These purely aesthetic in-game items are skins for weapons with unique and/or unusual colors and patterns, and can fetch a high price depending on various factors, including popularity and a predetermined wear value called a float value. One such way of getting these items is through cases. Cases are a virtual item that will randomly give the player one weapon skin. There are a few dozen types of cases in the game. Each kind of case has some attributes of interest: The quantity of cases available to buy, the price to buy the case, the price to open the case, and the list of cosmetic items available from that case.

2 Subject of analysis

Our goal is to create our own analysis of the market for these cases using data found by various sources. Our analysis will be on the relationship between the highest item price one might see in a million cases vs the price of the case.

2.1 Hypothesis

Our hypothesis is that the value of the case will rise with the value of the most expensive skin and be strongly correlated, but not be as strongly correlated with the average value, and perhaps not even rise with the average value.

2.2 Motivation

In CSGO, we suspect that people make purchases based off of what case has the most valuable or famous skins in it, as opposed to what case has the best return on your money. Simply finding a relationship between price and the highest value item does not prove this statement, but it does at least reveal a potential pattern, which could be studied further. One of us (James) is also is a longtime CSGO fan, and thinks this data would be interesting to see.

2.3 Data Used

In order to do this analysis, we will first need to collect our data. To begin, we get data from the game (Valve, 2012) so that we have some of the possible wear values, as well as the name and the rarity of each item. From there, we must also reference an article which does an analysis on the odds for wear values of items (STEP7750, 2020). This article will help us simulate the case openings later. Another

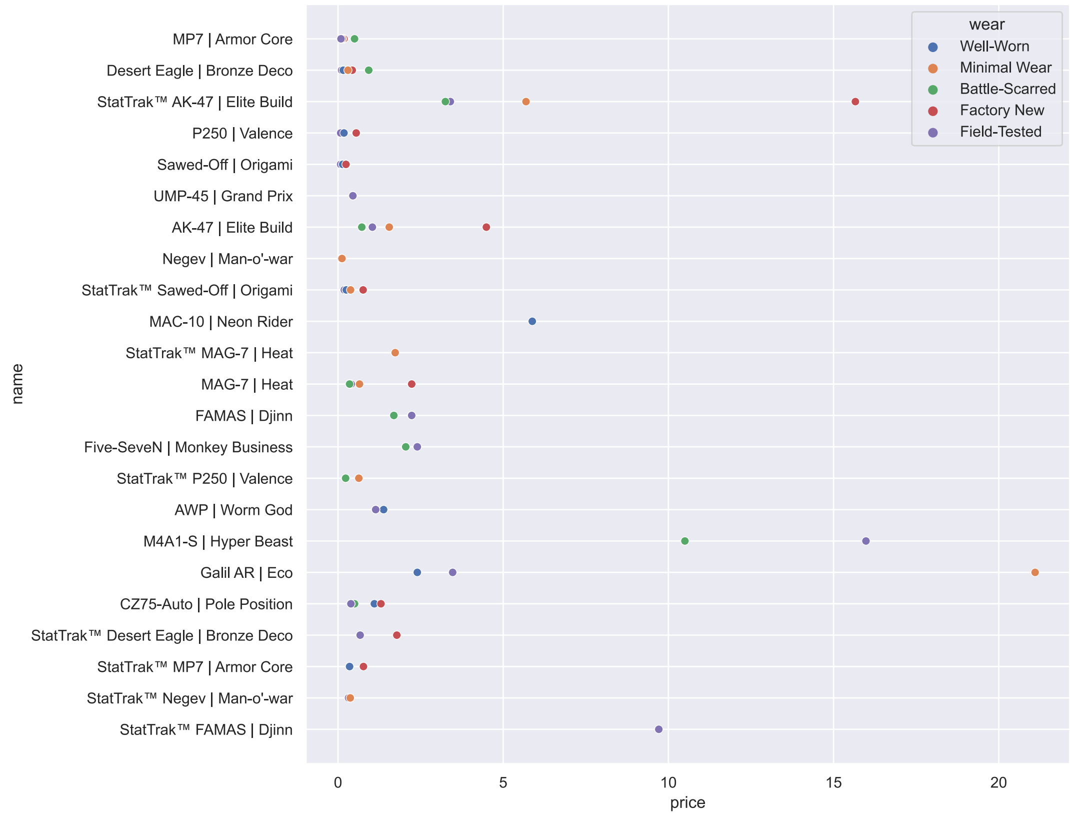

Figure 1: 300 Case openings in chroma2 simulated

value for simulating the cases is the odds for each rarity of item. This can be found on a post by the CSGO server (guó fú, 2017). Lastly, there is an API on Bitskins as well as Steam available for us to get the prices of cosmetic items, as well as the volume and the costs of purchases for cases. Using all of this information, a simulation of the games case opening process was created, and one million cases were open from nine different types of cases. From here the data was combined and compiled until a database with the following header was obtained: case_name, case_price, case_volume, hash_name, name, wear, price, volume. The only unintuitive names here are volumes, which is the number of that item available for purchase, and the hash name, which is the market listing name, and includes the wear, alongside other features. The case names were decided beforehand, and are used to organize the lines. The case price and volume is determined by the place with the most complete case listings, which was steam. The hash names are universal, and were generated using the basic formula that the game goes by. The game formula begins with whether or not there is a quality called StatTrak, and lists that. Then, there is the name of the type of weapon. this name is separated by the name of the design by a vertical bar. Then in parentheses we have the wear grade. This wear value is determined by the aforementioned floats, but instead of looking at the exact float values on the market, we take the lowest price within the wear grade. The lowest price is what people will want to buy anyways, so this seemed like a fair way to ensure a more complete dataset. These prices and volumes were also obtained from Bitskins, which has an incomplete listing of cases, but an expansive listing of weapon skins. There is a master database containing every case, and there are also smaller ones which have only the simulated case openings from one case. The first analysis, as well as a visual to understand the database is a simple scatterplot. The following is a snippet of 300 case openings from the chroma2 case openings, which has variable wear values (to note: the wear values have unfortunately been unordered. Going least to most worn, we have the following: Factory New, Minimal Wear, Field-Tested, Well-Worn, and Battle-Scarred) Each dot represents a case opening, and the colors of the dots denote wear values. The x value denotes price in US dollars, and the y axis denotes the name of the weapon skin. For reference, 300 chroma2 cases would cost 249 dollars. If one were to be counting the price of the items received, the total would be 257.10 dollars, with an average of a skin valuing 0.86 per case. Of course, while this is enlightening, the true goal of this project was not to take the averages of the cases.

3 Analysis

After creating the database, we tried various methods to analyze the data. The very first of these methods was a correlation: As one can see, we have (understandably) strong relationships between the prices of

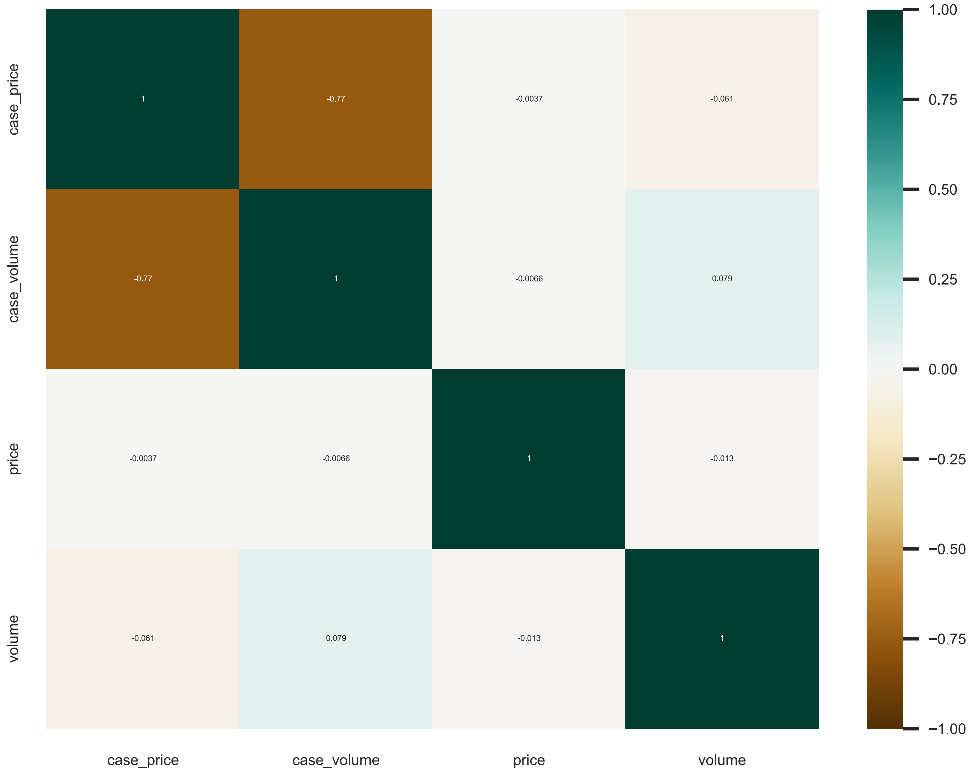

Figure 2: Heatmap of the database

the cases and the volume sold. The less cases being sold, the more valuable they are. This only makes sense. Interestingly enough, however, the same is not true for the weapon skins. There does not seem to be a very strong correlation between the price of a skin and the quantity listed. This must mean that even if rare, valuable skins are being sold in low quantities and high costs, there must also be skins being sold at low quantities and low prices, as well as high quantities and high prices, to cause there to be no correlation. Lastly, if we try to see how the skin price affects the price of the case, we see a value of -0.0037. This does not seem very promising for our original hypothesis, but it does tell us that as skin prices go up, case prices go down somewhat. It seems as though perhaps many of the cases have similarly-priced items, and this may cause the higher case prices to be somewhat meaningless looking solely at the prices. The next, more revealing set of visuals are some linear regressions, beginning with case price versus the average price of the items obtained in the case after one million simulations:

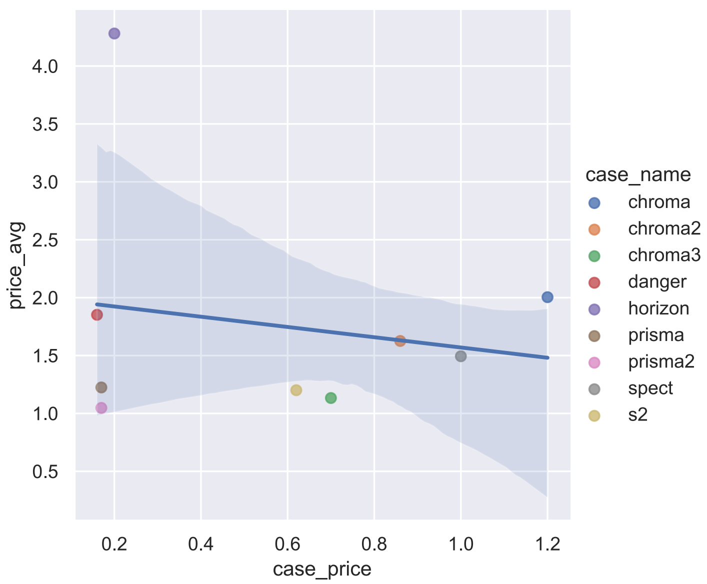

Figure 3: Case price versus the average price of skins obtained from said case

Each dot represents a type of case (which is denoted via color), and the x axis is the price of the case, while the y axis is the average price of the items obtained from the case. There is also a linear regression applied with a confidence interval. It seems as though the regression doesn't catch onto any real trend, as this data is quite irregular, with only nine effective data points. If we are to interpret it directly, however, it seems that the higher the average price of the items, the lower the case price is. The horizon case in particular is an outlier, as has the highest average item price, and yet has the lowest cost. This seems to support the hypothesis made earlier. The next graphic depicts another linear regression, this time between the price of a type of case and the highest price of an item obtained in the simulations from that case: Each dot in this visual also represents a case, and all cases are also denoted by color. The

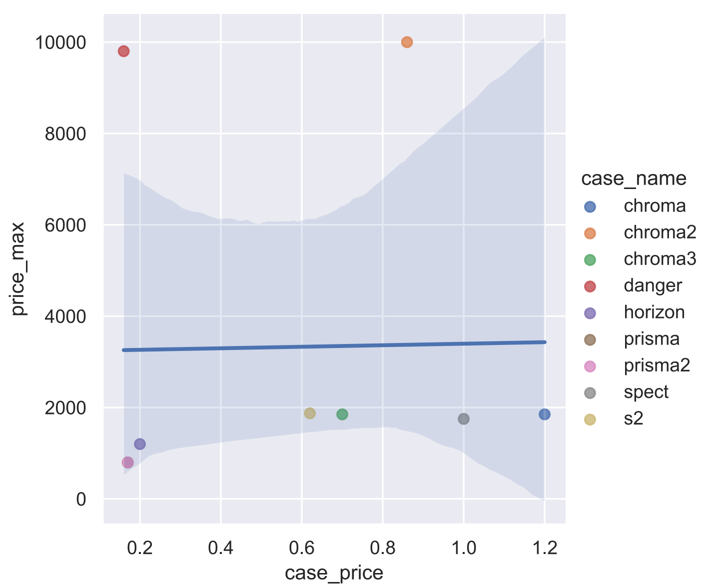

Figure 4: Case price versus the maximum price of skins obtained from said case

price of the case is on the x axis, while the price of the most expensive item obtained from the case is on the y axis. This linear regression appears more helpful, but according to the confidence interval, there is more uncertainty than before. It appears that, according to said linear regression, there is a positive relationship between the highest price and the case price. However, this relationship does not seem incredibly significant. However, this also supports the original hypothesis to a degree

4 Conclusions

It seems as though our findings were somewhat in support of our hypothesis. The average skin price was in fact lower with cases that have high prices. However, it seemed as though high prices didn't make much of a difference. One might conclude that the price of the case depends solely on quantity available, which would be reasonable. It is an odd thought, however, that the prices of the skins are not affected by the scarcity. One would think perhaps that originally there was a larger number of said cases, and so now there is a normal or even large amount of those skins available. Regardless, there is more potential to be found in this database, but for now, our conclusion is that our findings have somewhat confirmed our biases.

References

guó fú. 2017. CSGO gài shuài gōng shì. Technical report, CS:GO guó fú yùn, sep.

STEP7750. 2020. Analysis of float value and paint seed distribution in cs:go. *blog.csgofloat.com*.

Hidden Path Entertainment Valve. 2012. Counter-strike: Global offensive.