Steam Video Game Demand Curves

by

William-Elijah Clark

ECO 4412

Econometrics

12/01/2016

1. Introduction

Demand curves are a staple of economics, and are fundamental for understanding what output would place a firm at a market clearing equilibrium, how elastic or inelastic a specific good is, understanding how much utility a consumer derives from consumption, etc. However, demand curves are only theoretical without data and econometric analysis, and cannot be made without them. Thus, the intended goal of this paper is to derive a demand curve from a specific market.

In regards to the specific market that was selected for this paper, there are a few reasons behind why the Steam PC video games market was selected. Entertainment as a whole is subject to risk taking. While larger entertainment firms have the capital to procure their own economic models and the capital to take risks (which *Blockbusters: Hit-Making, Risk-Taking, and the Big Business of Entertainment* covers in more depth)[1], smaller firms typically do not have economic models available for use. Risk-taking is blind without economic models to rely on. Independent game firms are included in these smaller firms, and are no exception for blind risk-taking.

This said, I think that the construction of an open-source demand curve for video games would be beneficial to these specific firms, even if the resulting demand curve in this paper is simple. Once a single model is established, a firm could formulate their own supply function, and could match the price of their game accordingly. Furthermore, once the first rudimentary model exists, it could be further tailored to a firm's specific genre, and could be reconstructed with different data that is specific to that game genre.

While no model could account for subjective qualities of a video game (for instance, the artwork inside a game, the game's plot, innovation in context of game mechanics that have never been implemented etc.), it would still solve the problem of blind risk-taking that most independent

games firms have. It could also be argued that the risks these firms take are heightened by dealing with elastic goods, which would require deliberate pricing in order to maximize profit. That said, the most common market that independent game studios sell their products is the Steam PC games market. This is why this market is relevant to this problem, and to this specific model.

As a note in this paper, certain variables (and thereby, their data points) are proxy variables. This is due to the fact that ValveCorp (the company that owns the Steam marketplace) does not publish their sales data.[2] The closest approximation is an Application Programming Interface [API] called SteamSpy, which is an automated bot that crawls around the site, collects data on how many Steam Users own specific games, and then publishes them.[3] All other data is freely available on the Steam Marketplace, and should not be subject to any significant measurement errors.

2. Relevant Economic Concepts and the Model

To start, basic demand curves rely on either an exchange in price and quantity, or they rely on an exchange of good bundles. In theory, the simplest demand curve would look like this,

$$Q_D = \alpha + \beta P_D$$

Wherein Q_D is quantity demanded, P_D is the price at a specific quantity, α is the choke price, and β is the coefficient on price. This would theoretically work in context of one attempted simple linear regression model,

$$\ln(\textit{ownbe}) = \alpha + \beta \textit{pfull}$$

wherein *ownbe* is the number of sales for a game *before* any discounting occurs, *pfull* is the price of a game before discounting occurs, α is the choke price, and β is the coefficient on price.

However, with the collected data, r-square is 0.08046, and poorly explains the relationship between these two variables, as shown in *figure 1*. It also fails to go through diagnostics, as the regression line and the residuals are clearly non-normal, as shown in *figure 2*.
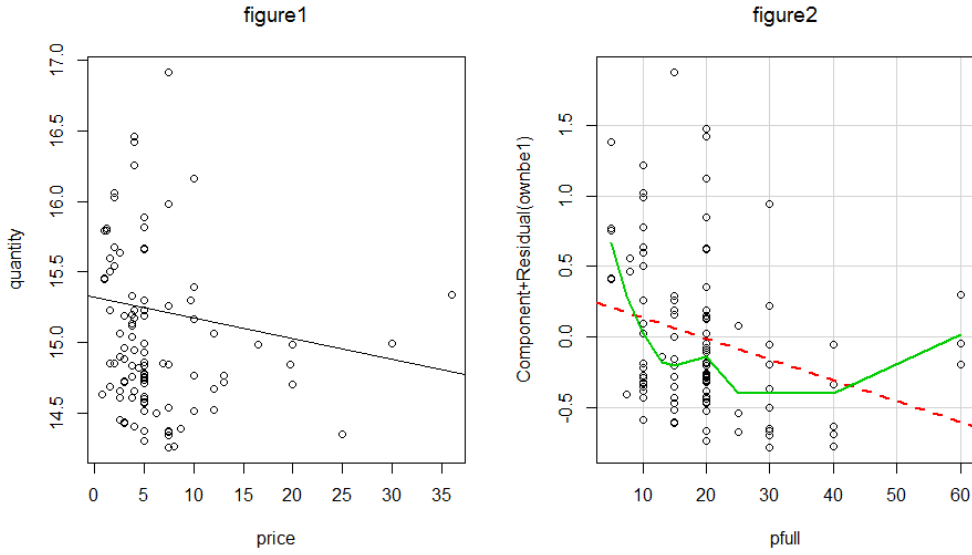
Due to this ill-suited result, the economic model required additional explanatory variables to properly derive a demand curve.

An additional obstacle was found with the appropriated model that was constructed based off of Pradeep and Harikesh's fish markets[4],

$$pfull = \alpha + \beta_o ownbe + \delta genre$$

wherein *pfull* is price before discounting, *ownbe* is the number of games owned before discounting, *genre* is the game genre(or previously, a type of fish), α is the choke price, β_0 = coefficient for quantity purchased, and δ coefficient for the specific genre/fish.

This did not work as well as anticipated, and resulted in the same problems as the simple linear regression model above.

While several multi-variable regression models were constructed after these two, the model with the best fit was the following (after going through the process that Marafi outlines)[5],

$$\ln(ownaf) = \alpha + \beta_1 pfull + \beta_2 pdisc + \beta_3 \ln(ownbe) + \delta genre + \beta_4 month$$

wherein *ownaf* is the number of sales for a game *after* any discounting occurs, *pfull* is the price of a game before discounting occurs, *pdisc* is the discounted price of a game at the time of a sale, *ownbe* is the number of sales for a game *before* any discounting occurs α is the choke price, *genre* is the game genre, *month* is the amount of time that has occurred since a game's release date (in months), α is the choke price, and β is the coefficient on each independent variable other than genre (with β_1 for *pfull*, β_2 for *pdisc*, β_3 for *ownbe*, and β_4 for *month*).

Additionally, δ is divided into δ_0 for action-adventure, δ_{misc} for miscellaneous genres, δ_{rpg} for role-playing games, δ_{shoot} for shooting games, and δ_{strat} for strategy games. As a note on these specific genre classifications, the four non-miscellaneous game genres were the game categories with the largest number of observations within the data set. All other game genres were folded into the category of miscellaneous (many of which had n<5 observations, and would not be able to be regressed.) Furthermore, specific sub-genres within a genre were folded into their main genres

(e.g; first-person shooters, third-person shooters, and top-down shooters were all collected as shooter) in order to create a model that works within the scope of this project.

In addition to the construction of the demand curve, we can also consider price elasticity in this model, as follows,

$$\text{elasticity} = (\Delta Q / \Delta P) * (P/Q),$$

which may be further defined in this model as the following;

$$elast_1 = \alpha * \overline{pfull} / \overline{ln(ownbe)}$$

and

$$elast_2 = \alpha * \overline{pdisc} / \overline{ln(ownaf)} \ ,$$

wherein $elast_1$ and elast2 are the elasticities of goods at full price and discounted price, respectively, α is the regression coefficient, *pfull* is the price before discounting, *ownbe* is the quantity before discounting (in logarithmic form), *pdisc* is the price after discounting, and *ownaf* is the quantity after discounting (in logarithmic form).
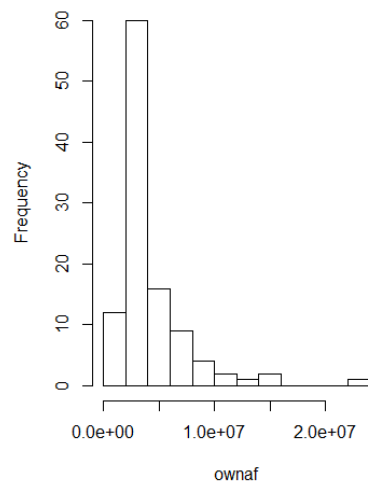
4. Estimating the Model

The following histograms represent each of the six initial variables collected, with most following either a t-student distribution or a normalized distribution, save for genre.
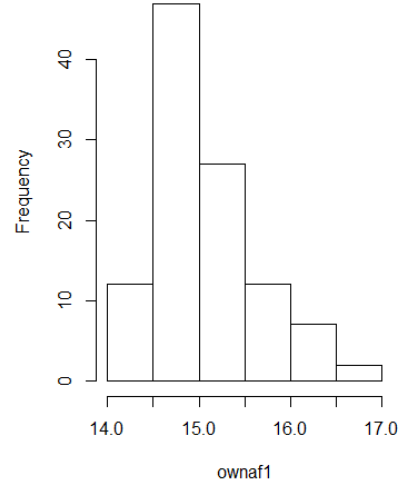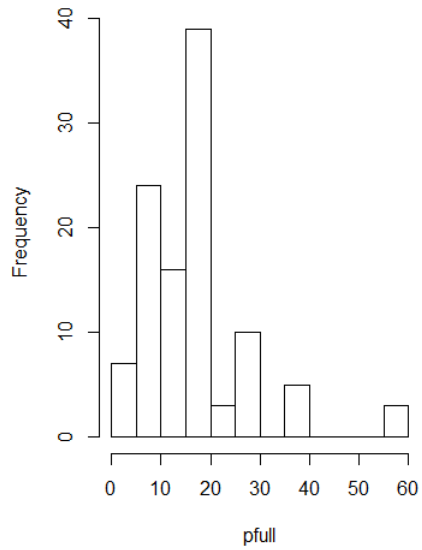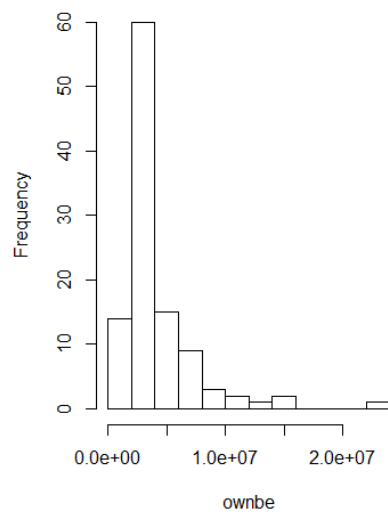
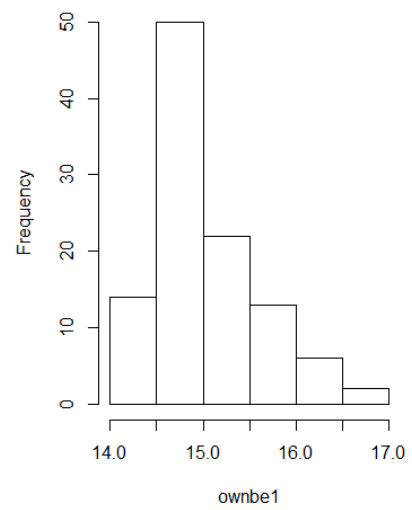**Histogram of genre**
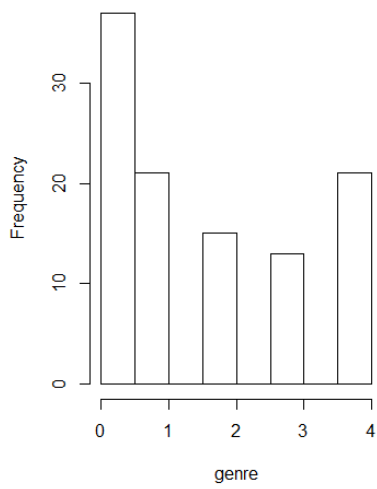
Histogram of ownaf

Histogram of ownaf1

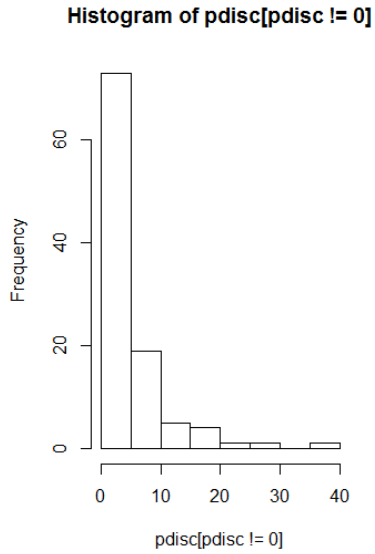**Histogram of pfull**

Histogram of ownbe

Histogram of ownbe1

Histogram of pdisc[pdisc != 0]

From the n=107 observations of 6 variables collected into this model's dataset, the resulting values for the coefficients were α = 0.2069e-01 , β_1 = -8.456e-04 , β_2 = 2.720e-03 , β_3 = 9.876e-01 , β_4 = -1.612e-05 , δ_{misc} = 2.613e-02 , δ_{rpg} = 1.611e-02 , δ_{shoot} = 7.402e-03 , and δ_{strat} = 1.531e-02 .

With the mean of *pfull, pdisc, ln(ownbe), and ln(ownaf)* [as 19.06953, 6.181923, 15.04096, and 15.07137, respectively], *elast$_1$* =0.02623161, and *elast$_2$* = 0.008486554. Furthermore, we can take

$$\exp(elast_1)$$

and

$$\exp(elast_2)$$

to arrive at an elasticity of 1.026579 for non-discounted games, and 1.008523 for discounted games, respectively.

In addition to an r-square value of 0.9948, *figure 3* shows the diagnostics for this regression curve, and figure 4 shows a (mostly) normal QQ-plot. While there is one outlier that has

8

a standardized residual of approximately 8, the rest of the data appears to fit a normalized

distribution, and stops at 2.
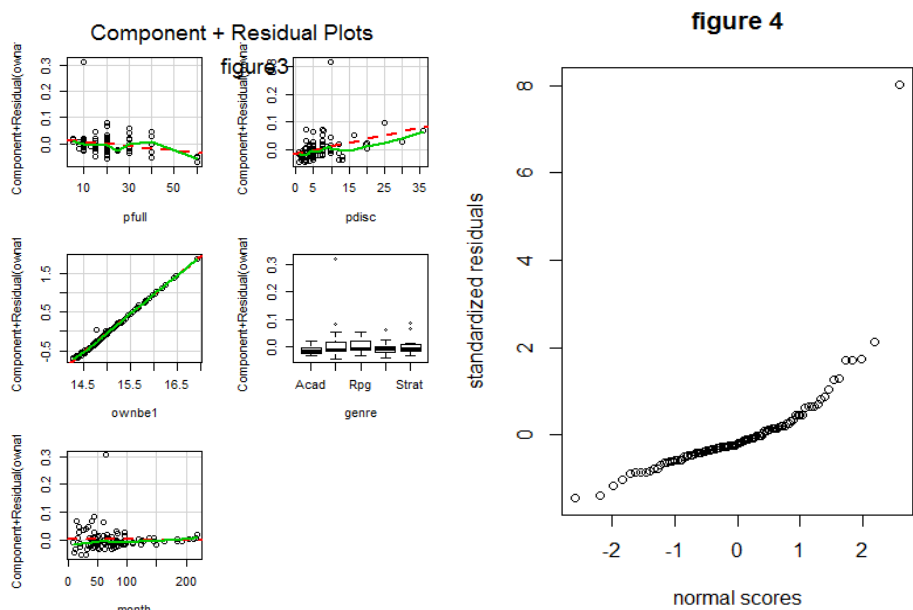
Testing for Variance Inflation Factors implies that (in spite of having five independent

variables) there is not multicollinearity in this model. The closest variable that has the potential to

create problems was *pfull*, although this is predominantly a problem with sqrt(vif). Additionally,

when it is to the exponent of 1/(2*df), this is a non-issue, as presented in the following tables.

VIF

	GVIF	Df	GVIF^(1/(2*Df))
pfull	3.654226	1	1.911603
pdisc	3.386242	1	1.840174
ownbe1	1.389795	1	1.178896
genre	1.635442	4	1.063419

month 1.849916 1   1.360116

SQRT(VIF)

        GVIF Df GVIF^(1/(2*Df))

pfull 1.911603 1   1.382607

pdisc 1.840174 1   1.356530

ownbe1 1.178896 1   1.085770

genre 1.278844 2   1.031222

month 1.360116 1   1.166240

      In terms of autocorrelation, the Durbin-Watson Test failed to reject the assumption that there is no autocorrelation, with a DW statistic of 1.91. Additionally, with a p-value of 0.498, the likelihood of rejecting the null hypothesis would require an α value of approximately 0.5. Considering that such an alpha value is unlikely, there is most likely not autocorrelation.

One potential problem with this model is that the residuals appear to be a left-skewed t-distribution, as shown in figures 5 and 6 below. This would be problematic without other

Distribution of Studentized Residuals   **QQ Plot**
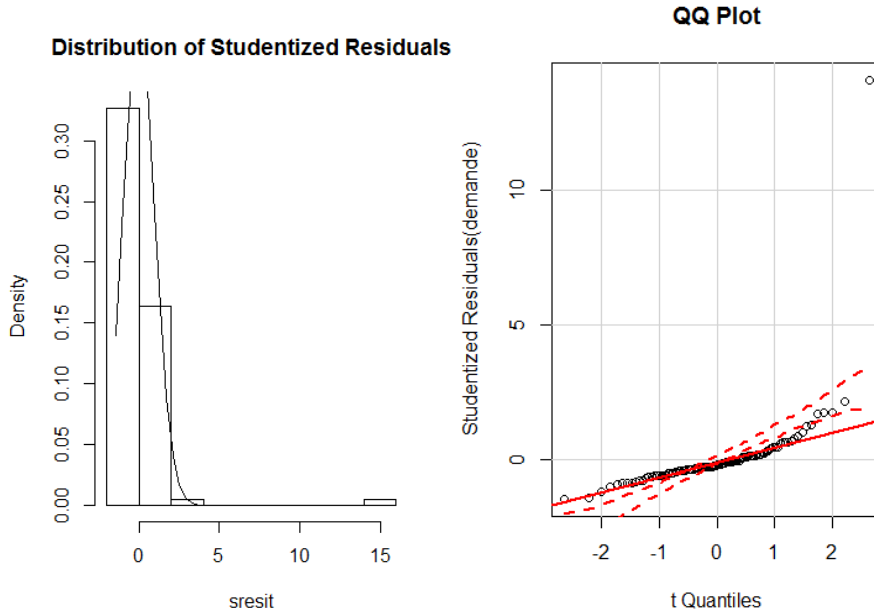
diagnostics.

4. Economic Interpretation

On the note of elasticity, it can easily be said that both discounted and non-discounted games have an elasticity greater than 1, and are thus, both elastic goods. That said, non-discounted games appear to be somewhat more elastic than discounted games. With the fact that elasticity is increasing when price increases, this implies that this is, indeed, a downward sloping demand curve, where a higher price correlates with lower quantity demanded. While this is no surprise in context of a normal good that it non-vital, it is unexpected that discounted games are close to unit elastic. With this said, the hypothesis

$H_0 = E_d \geq 1$ for $price_i \leq \mu$; $E_d \leq 1$ for $price_i \geq \mu$

$H_1 = Not\ H_0$

is a bit close to call in regards to the null hypothesis. However, as the elasticity is in a relatively close margin to 1 at the mean in both cases, it could be concluded that we fail to reject the null hypothesis. This may require more stringent hypothesis testing in the future, in order to better analyze this problem.

5. Conclusion

Unexpectedly, this project seemed to require more than two independent variables in order to have a regression model—with a reasonable R-Square value— that served as a useful demand curve. While different demand curves may require differing numbers of variables in order to be useful, it appears that multicollinearity is not as significant of a concern in this case, as it possibly is in other cases.

While this is not within the immediate scope of this specific project, further expansion of this model could occur with data collected within a specific genre or sub-genre that a games firm could plan to supply to the market, allowing the elimination of δ*genre* as a variable altogether. The variable *month* could also be eliminated, from the perspective of an games firm that is concerned about pricing at the launch of a game—due to the value of month being zero, in this case, and the lack of price decay that would occur over time. Finally, if a firm in this market has a supply curve available for use, the firm should be able to calculate its optimum price, based off of either this demand curve, or a modified version of this demand curve.

Bibliography

1) Elberse, Anita. *Blockbusters: Hit-making, Risk-taking, and the Big Business of Entertainment.* New York: Henry Holt, 2013. Print.

2) Galyonkin, Sergey. "Steam Sales." SteamSpy. Sergey Galyonkin, n.d. Web. 08 Nov.

2016. <http://steamspy.com/sale/>.

3)Marafi, Salem. "Price Elasticity with R." *Salem Marafi.* N.p., 10 June 2014. Web. 01 Dec. 2016. <http://www.salemmarafi.com/code/price-elasticity-with-r/>.

4) Orland, Kyle. "Introducing Steam Gauge: Ars Reveals Steam's Most Popular Games." Ars Technica. Ars Technica, 15 Apr. 2014. Web. 08 Nov. 2016.

5) Orland, Kyle. "Steam Gauge: Addressing Your Questions and Concerns." Ars Technica. Ars Technica, 16 Apr. 2014. Web. 08 Nov. 2016.

6) Parfitt, Ben. "Digital Charts Won't Pick up Steam." MCV UK. MCV, 21 Apr. 2011. Web. 08 Nov. 2016.

7) Pradeep K. Chintagunta, Harikesh S. Nair and R. Sukumar, "Measuring Marketing-Mix Effects in the 32/64 Bit Video-Game Console Market"Journal of Applied Econometrics, Vol. 24, No. 3, New Econometric Models in Marketing (Apr. - May, 2009), pp. 421-445

8) "Steam: Game and Player Statistics." Steam: Game and Player Statistics. Valve Corporation, n.d. Web. 08 Nov. 2016. <http://store.steampowered.com/stats/>.

9) "Steam Search." Steam Search. Valve Corporation, n.d. Web. 08 Nov. 2016. <http://store.steampowered.com/search/?filter=topsellers>.

10) Wolfgang Härdle, Alan Kirman, "Nonclassical demand: A model-free examination of price-quantity relations in the Marseille fish market", Journal of Econometrics Volume 67, Issue 1, May 1995, Pages 227–257

References

[1] Elberse, Anita. *Blockbusters: Hit-making, Risk-taking, and the Big Business of Entertainment*. New York: Henry Holt, 2013. Print.

[2] Parfitt, Ben. "Digital Charts Won't Pick up Steam." MCV UK. MCV, 21 Apr. 2011. Web. 08 Nov. 2016.

[3] Galyonkin, Sergey. "Steam Sales." SteamSpy. Sergey Galyonkin, n.d. Web. 08 Nov.

2016. <http://steamspy.com/sale/>.

[4] Pradeep K. Chintagunta, Harikesh S. Nair and R. Sukumar, "Measuring Marketing-Mix Effects in the 32/64 Bit Video-Game Console Market"Journal of Applied Econometrics, Vol. 24, No. 3, New Econometric Models in Marketing (Apr. - May, 2009), pp. 421-445

[5] Marafi, Salem. "Price Elasticity with R." *Salem Marafi*. N.p., 10 June 2014. Web. 01 Dec. 2016. <http://www.salemmarafi.com/code/price-elasticity-with-r/>.