The Effect of Pricing on Consumers' Ratings: Insights from Beauty Products on Amazon

By KAYLIE NGUYEN, CYNTHIA CUI, AND LUCY ZHU

I. Introduction

A popular belief among consumers says "You get what you pay for" - that is, the more expensive something is, the better quality it would be. When consumers consider the value of a product, price is one of the most obvious features and an important index of quality (Scitovksy, 1945). Leavitt (1954) found that a higher price may sometimes increase buyers' willingness to pay. However, whether price is a truthful reflection of the quality of the product is up to debate. Gerstner (1985)'s empirical research on the relationship between price and quality found that the correlation is weak for many products and that higher prices appear to be poor signals of higher quality. However, previous studies, such as Geistfeld's (1982), also found that the price-quality correlation may vary across store types and may be dependent on the data sample. These studies, nonetheless, are mostly based on data from Consumer Reports or Consumers' Research Magazine (Geistfeld, 1982). Thus, it would be necessary to test the question in a more modern setting.

With the widespread use of e-commerce platforms such as Amazon, customers' perceptions of the quality of a product can be easily summarized by its ratings. As such, the aim of this paper is to investigate the relationship between price and customer reviews online. This would allow customers to make more informed decisions while purchasing products online. In addition, the research would provide companies with more understanding of how a consumer makes decisions - as Manag (2020) notes, both product price and customer reviews can influence a would-be consumer's willingness to pay. Tsai (2007) found that price is a major influence on customer satisfaction in the manufacturing industry. Furthermore, Li & Hitt (2010), who examined the impact of price on customer reviews while accounting for different review systems using data on digital camera sales, found that with unidimensional ratings, price has a substantial effect on product rating.

Using a regression model, our main explanatory variable of interest is price, along with product category, and the outcome variable is rating. The product ranking is included as a control and an interaction term between price and category is presented. We found that price is correlated with product ratings and an increase in price leads to a slight increase in rating. However, for luxury products, the increase in rating is smaller. We also tested for multicollinearity and presented evidence that indicates heteroskedasticity, though the results remain largely unaffected. To gain a deeper understanding of our results, we performed regressions separately for luxury and non-luxury beauty products. The results reveal that after controlling for ranking, increase in price leads to a smaller increase in ratings for luxury products than for non-luxury products, which corroborates our main finding.

II. Data Description

A. Description

The dataset used in the current analysis is from Ni & McAuley (2019), an updated version of the Amazon review dataset in 2014 to include reviews from 1996-2018. This dataset is observational and includes reviews and product metadata, which has descriptions, category information, price, brand, and image features. The most relevant variables are reviews, more specifically, product ratings, and price. The price of each product is recorded as numeric values while ratings range from one to five. A log transformation is performed on price to linearize the relationship and to make the coefficients more interpretable.

The bestseller rank of each product is also included as a control since it may affect customer ratings in addition to the price. The samples chosen are Amazon beauty products, as they provide a large yet manageable dataset while also having two categories for further investigation - in addition to All Beauty products, Luxury Beauty is included as well since the perception towards a product may also affect customer ratings, for which a dummy variable is constructed to represent whether a product is luxury or not.

B. Summary Statistics

A data.frame: 7 × 5

Variable	Min	Mean	Max	Sd
<chr>	<chr>	<chr>	<chr>	<chr>
product_avg_ratings	1	4.1	5	0.92
price	0.01	34	1000	47
main_cat	0	0.38	1	0.49
rank	35	617067	9549407	633782
logprice	-4.6	3.1	6.9	0.95
pct_rank	0	0.94	1	0.066
interaction	0	1.3	6.4	1.8

III. Model

A. Specifications

In this part, we will estimate the effect of price on the product ratings using our regression estimates with 5 different specifications. In the first specification, we want to study how much the product's price on its own will affect the rating of the product. We will perform a simple regression where we solely study the effect of price on product ratings:

(1) $$Y_i = \beta_1 \ln(X_i) + \beta_0 + \varepsilon_i,$$

where Y_i is the product rating for each product *i*, and X_i is the price of the product.

Our second explanatory variable of interest is the category in which the product belongs. Besides price, the category will also influence how a buyer views a product; especially when products are tagged with "Luxury," the buyer often automatically associate the product to a higher price level and generally expect better quality. However, in reality, not all "Luxury Beauty" products are more expensive than regular products. If the product is categorized as being luxury, raising the buyer's expectation but does not meet the standard, the rating might actually decrease. Therefore, in the second specification, we study the effect of both price and the category on product rating to see if being in different categories affect the buyer's expectation and thus opinion on the product. Our equation is as follows:

(2) $$Y_i = \beta_1 \ln(X_i) + \beta_2 C_i + \beta_0 + \varepsilon_i.$$

The only variable we add in this equation compared to the first one is C_i which is the dummy variable for the category. Level 0 corresponds to the product being in "All Beauty" category, and level 1 corresponds to "Luxury Beauty."

Since the categories may change a buyer's expectation of the product, it may also change how much the buyer evaluates the return on the price they paid for the products. In other words, the category not only impacts the rating directly, but the effect of price on rating may also differ for each category. We decided to add an interaction term between the price and category to study this effect.

(3) $$Y_i = \beta_1 \ln(X_i) + \beta_2 C_i + \beta_3 \ln(X_i) \times C_i + \beta_0 + \varepsilon_i,$$

Finally, we added two more specifications corresponding to the second and the third specifications, but here we will control for product's rank in the bestseller list on Amazon. We decided to add the rank as the control variable because seeing the ranking can have an impact on the buyer's perception of the product, such as valuing it higher due to the product being more popular and favored by others. The rank, therefore, also explains the buyer's rating and we want to control for it to eliminate any confounding effect. The equations for our last 2 specifications are:

(4) $$Y_i = \beta_1 \ln(X_i) + \beta_2 C_i + \beta_3 R_i + \beta_0 + \varepsilon_i,$$

where R_i is the percentage rank after transformation, and:

(5) $$Y_i = \beta_1 \ln(X_i) + \beta_2 C_i + \beta_3 \ln(X_i) \times C_i + \beta_4 R_i + \beta_0 + \varepsilon_i,$$

with the variables as given above.

In our model, we will focus primarily on the coefficient for the log(price), which is β_1. We will look at the magnitude, the sign, and the significance of the result to answer our research question. The assumptions for our model are no perfect collinearity, homoskedasticity, and the sample is large enough to avoid biasedness.

B. Table of Results

Comparison of Regression Results

	Dependent variable:				
	product_avg_ratings				
	(1)	(2)	(3)	(4)	(5)
logprice	0.094***	0.101***	0.138***	0.113***	0.151***
	(0.007)	(0.008)	(0.009)	(0.008)	(0.009)
main_cat		-0.043***	0.389***	-0.181***	0.260***
		(0.015)	(0.058)	(0.017)	(0.058)
logprice:main_cat			-0.131***		-0.134***
			(0.017)		(0.017)
pct_rank				1.993***	2.001***
				(0.117)	(0.117)
Constant	3.854***	3.847***	3.745***	1.999***	1.886***
	(0.023)	(0.023)	(0.027)	(0.111)	(0.111)
Observations	17,936	17,936	17,936	17,936	17,936
R^2	0.009	0.010	0.013	0.026	0.029
Adjusted R^2	0.009	0.010	0.013	0.025	0.029
Residual Std. Error	0.917 (df = 17934)	0.917 (df = 17933)	0.916 (df = 17932)	0.910 (df = 17932)	0.908 (df = 17931)
F Statistic	169.168*** (df = 1; 17934)	88.768*** (df = 2; 17933)	78.965*** (df = 3; 17932)	157.291*** (df = 3; 17932)	133.937*** (df = 4; 17931)
Note:					*p<0.1; **p<0.05; ***p<0.01

III. Discussion

A. Results

Regression 1.— The coefficient of “logprice” shows that, for a 1% increase in product price, the product average rating goes up by 0.094/100 stars.

Regression 2.— After considering the product category, the coefficient of “logprice” shows, under the same category, for a 1% increase in product price, the product average rating goes up by 0.101/100 stars. The coefficient of “main_cat” demonstrates that, for products with same price, moving from “All Beauty” category to “Luxury Beauty” category decreases the product average rating by 0.043 stars, which reflects the buyers’ expectation on the return for the price.

Regression 3.— The coefficient of “logprice” shows, for “All Beauty” products (“main_cat”= 0), a 1% increase in product price increases the product average rating by 0.138/100 stars. The coefficient on the interaction term is -0.131, meaning for the same change in price, being a luxury will lower the rating by 0.131 times the change in price compared to being a regular product. The coefficient of “main_cat” is 0.389, which means for products with same price, moving from “All Beauty” category to “Luxury Beauty” category changes the product average rating by 0.389-0.131*logprice stars, so the larger the price, the smaller the increase in product average rating and it can even be negative if the price is large enough.

Regression 4.— For products under same category with same ranking, a 1% increase in product price increases the product average rating by 0.113/100 stars. For products with the same price and ranking, moving from “All Beauty” category to “Luxury Beauty” category decreases the product average rating by 0.181 stars. Notice here, when we included the control variable of ranking, the increase in the R^2 can also justify our selection of this specification.

Regression 5.— For products with same ranking under “All Beauty” category, a

1% increase in product price increases the product average rating by 0.151/100 stars. The interaction term's coefficient is -0.134, meaning for the same change in price, "Luxury Beauty" products' average rating stars go down by 0.134 times the change in price compared to "All Beauty". The coefficient of "main_cat" is 0.26, which means for products with the same price and ranking, moving from the "All Beauty" category to "Luxury Beauty" category changes the product average rating by 0.26-0.134*logprice units of stars.

In summary, the results above show that there is a correlation between product price and ratings, and the increase in price induces different changes in average ratings for products from different categories. The coefficients for price are statistically significance, lending more evidence to the fact that price does have an effect on the overall rating. To further explore and verify the impact of product category on average ratings, we decided to conduct a robustness analysis by breaking down our categories of interest into two subgroups and run the regression analysis respectively.

Before presenting our robustness analysis, we first need to test if there's any multicollinearity or heteroskedasticity existing in our regression models. Though the above regression analysis assumed there is none, failure to prove this assumption will not change our core result because of the unbiasedness guaranteed by our large sample size.

B. Specification Check

Multicollinearity.—To test for multicollinearity, we conducted the VIF test on the four multiple regressions respectively, and the results are (in the order of Regression 2, Regression 3, Regression 4, Regression 5):

VIF Results

logprice	main_cat
1.138	1.138

VIF Results

logprice	main_cat	logprice:main_cat
1.578	17.192	19.343

VIF Results

logprice	main_cat	pct_rank
1.148	1.468	1.300

VIF Results

logprice	main_cat	pct_rank	logprice:main_cat
1.589	17.480	1.300	19.345

For all four multiple regressions, our key explanatory variable "logprice" has a low VIF score, demonstrating that "logprice" is not highly collinear with other control variables in our regression models. For all four multiple regressions, our key explanatory variable "logprice" has a low VIF score, demonstrating that "logprice" is not highly collinear with other control variables in our regression models.

However, in cases where "main_cat" and our interaction term enter the regression models, they both have high VIF scores. This is perceivable since the interaction term is computed by multiplying "logprice" with "main_cat".

After conducting the VIF test, we can see some of our control variables are highly collinear with one another, but this should not be an issue as the VIF scores for our key explanatory variable are at an acceptable level for all four multiple regressions.

Heteroskedasticity. —In these five models, the regressions were based on the homoskedasticity assumption. To see if this really holds, we plotted the trends of residuals relative to the change in “logprice” and found there’s heteroskedasticity existing in these models (see Appendix B). To further verify this, we conducted White’s test on the five regression models and found the coefficients on $\hat{y}$ and $\hat{y}^2$ to be significant in all five tests with F-statistics ranging from 32.2 to 577.5 (see Appendix C), thus we can be certain that we have heteroskedasticity in the five regression models.

To improve on the previous regression table, we calculated the robust standard errors and implemented them in the new regression table below.

Robust SE Regression Results

	Dependent variable:				
	product_avg_ratings				
	(1)	(2)	(3)	(4)	(5)
logprice	0.094***	0.101***	0.138***	0.113***	0.151***
	(0.008)	(0.009)	(0.011)	(0.009)	(0.011)
main_cat		-0.043***	0.389***	-0.181***	0.260***
		(0.014)	(0.054)	(0.016)	(0.054)
logprice:main_cat			-0.131***		-0.134***
			(0.017)		(0.016)
pct_rank				1.993***	2.001***
				(0.163)	(0.163)
Constant	3.854***	3.847***	3.745***	1.999***	1.886***
	(0.026)	(0.026)	(0.032)	(0.157)	(0.159)
Observations	17,936	17,936	17,936	17,936	17,936
R^2	0.009	0.010	0.013	0.026	0.029
Adjusted R^2	0.009	0.010	0.013	0.025	0.029
Residual Std. Error	0.917 (df = 17934)	0.917 (df = 17933)	0.916 (df = 17932)	0.910 (df = 17932)	0.908 (df = 17931)
F Statistic	169.168*** (df = 1; 17934)	88.768*** (df = 2; 17933)	78.965*** (df = 3; 17932)	157.291*** (df = 3; 17932)	133.937*** (df = 4; 17931)
Note:					*p<0.1; **p<0.05; ***p<0.01

Because of the large sample size, our coefficients are unbiased regardless of the existence of heteroskedasticity, and after correcting for heteroskedasticity, the major difference in the new table is the increase in standard errors, which makes the coefficient slightly less efficient.

C. Robustness

From our interpretation of the four multiple regressions’ coefficients, especially those of “main_cat” and “logprice:main_cat,” it suggests that an increase in product price induces different changes in product average ratings for products under different categories. To summarize, in regressions without the interaction term, moving from All beauty category to Luxury beauty category decreases the product average rating given the same product price and ranking; in models with the interaction term, they demonstrate that an increase in product price increases Luxury beauty products' average ratings less than it does for All beauty product average ratings.

Therefore, to test the robustness of our findings, we can alternatively perform regressions for products under different categories separately and compare the coefficients.

What we found is that, for all beauty products, a 1% change in price increases product average ratings by 0.138/100 stars while it's 0.007/100 stars for luxury beauty products. After adding the control variable of percentage ranking, a 1% change in price increases product average ratings by 0.149/100 stars while it's 0.029/100 stars for luxury beauty products.

Robust SE Regression Results for All Beauty

	Dependent variable:	
	product_avg_ratings	
	(1)	(2)
logprice	0.138^{***}	0.149^{***}
	(0.011)	(0.011)
main_cat		
interaction		
pct_rank		1.708^{***}
		(0.174)
Constant	3.745^{***}	2.158^{***}
	(0.032)	(0.169)
Observations	11,104	11,104
R^2	0.016	0.029
Adjusted R^2	0.016	0.029
Residual Std. Error	1.038 (df = 11102)	1.031 (df = 11101)
F Statistic	181.526^{***} (df = 1; 11102)	166.072^{***} (df = 2; 11101)
Note:		*p<0.1; **p<0.05; ***p<0.01

Robust SE Regression Results for Luxury Beauty

	Dependent variable:	
	product_avg_ratings	
	(1)	(2)
logprice	0.007	0.029^{**}
	(0.012)	(0.012)
main_cat		
interaction		
pct_rank		4.353^{***}
		(0.430)
Constant	4.134^{***}	-0.189
	(0.043)	(0.431)
Observations	6,832	6,832
R^2	0.0001	0.042
Adjusted R^2	-0.0001	0.041
Residual Std. Error	0.670 (df = 6830)	0.656 (df = 6829)
F Statistic	0.410 (df = 1; 6830)	148.219^{***} (df = 2; 6829)
Note:		*p<0.1; **p<0.05; ***p<0.01

The coefficients of "logprice" are considerably smaller for Luxury Beauty products than that of "All Beauty." This generally aligns with our finding that after controlling other variables in the model, an increase in product prices induces an increase for All Beauty products' rating but a smaller increase, or even a decrease for "Luxury beauty."

IV. Conclusion

To evaluate whether customers receive products of a quality befitting the amount paid, we investigated the relationship between price and customer reviews on a sample of Amazon beauty products. We found that a correlation is present between price and product ratings and that the increase in rating after an increase in price is dependent on whether the product is identified as luxury or not. After checking for multicollinearity and heteroskedasticity, we show that our finding is not majorly impacted. The robustness analysis results also support our main findings. As such, we would argue that customers do tend to be more satisfied with higher-priced items, though satisfaction lessens when the item is too expensive or is categorized as luxury, potentially because of higher expectations.

REFERENCES

Gerstner, E. (1985). Do Higher Prices Signal Higher Quality? Sage Journals. Retrieved April 16, 2023, from https://journals.sagepub.com/doi/pdf/10.1177/002224378502200210

Geistfeld, L. V. (1982). The Price-Quality Relationship—Revisited. The Journal of Consumer Affairs, 16(2), 334–346. Retrieved April 16, 2023, from http://www.jstor.org/stable/23859157

Leavitt, H. J. (1954). A Note on Some Experimental Findings About the Meanings of Price. The Journal of Business, 27(3), 205–210. Retrieved April 16, 2023, from http://www.jstor.org/stable/23507s72

Li, X., & Hitt, L. M. (2010). Price Effects in Online Product Reviews: An Analytical Model and Empirical Analysis. MIS Quarterly, 34(4), 809–831. Retrieved April 16, 2023, from https://doi.org/10.2307/25750706

Ni, J., Li, J., & McAuley, J. (2019). Justifying recommendations using distantly-labeled reviews and fined-grained aspects. Empirical Methods in Natural Language Processing (EMNLP) [Data set]. Retrieved April 16, 2023, from https://cseweb.ucsd.edu/~jmcauley/datasets/amazon_v2/

Scitovszky, T. (1944). Some Consequences of the Habit of Judging Quality by Price. The Review of Economic Studies, 12(2), 100–105. Retrieved April 16, 2023, from https://doi.org/10.2307/2296093

Sun, L., Zhao, Y., & Ling, B. (2020). The Joint Influence of Online Rating and Product Price on Purchase Decision: An EEG Study. Psychology research and behavior management, 13, 291–301. https://doi.org/10.2147/PRBM.S238063

Tsai, Jessica, et al. (2007). Customer Satisfaction's Durability in Question. Destination CRM. Retrieved April 16, 2023, from https://www.destinationcrm.com/articles/crm-news/daily-news/customer-satisfactions-durability-in-question-43512.aspx

VI. Appendix

A. Attribution

In the *Introduction* section, all the group members discussed the research question and motivation, while Cynthia Cui did the entire write up.

In the *Data Description* section, Kaylie Nguyen coded the data parsing and cleaning, and Cynthia Cui worked on the summary statistics. Cynthia Cui also wrote the subsection *Description*.

In the *Model* section, the whole group discussed the options for the specifications. Kaylie did the code for the model and regression table with edits from Lucy Zhu. Kaylie Nguyen wrote the model and specifications explanation.

In the *Discussion* section, the whole group discussed what to use for specification check and robustness. Lucy Zhu did the code for specification check and robustness part, as well as the writing for the whole section with edits by Kaylie Nguyen in *Results*.

Cynthia Cui wrote the *Conclusion* section.

Lucy Zhu did the *References*.

B. Heteroskedastic Residuals

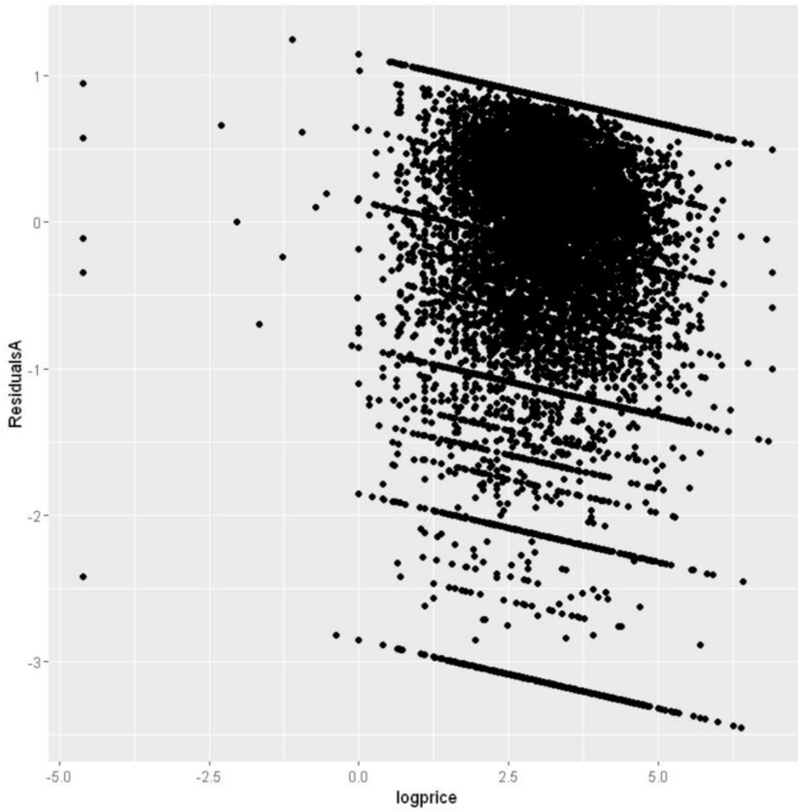

FIGURE 1 – RESIDUAL PLOTS OVER LOG PRICE FOR SPECIFICATION 1

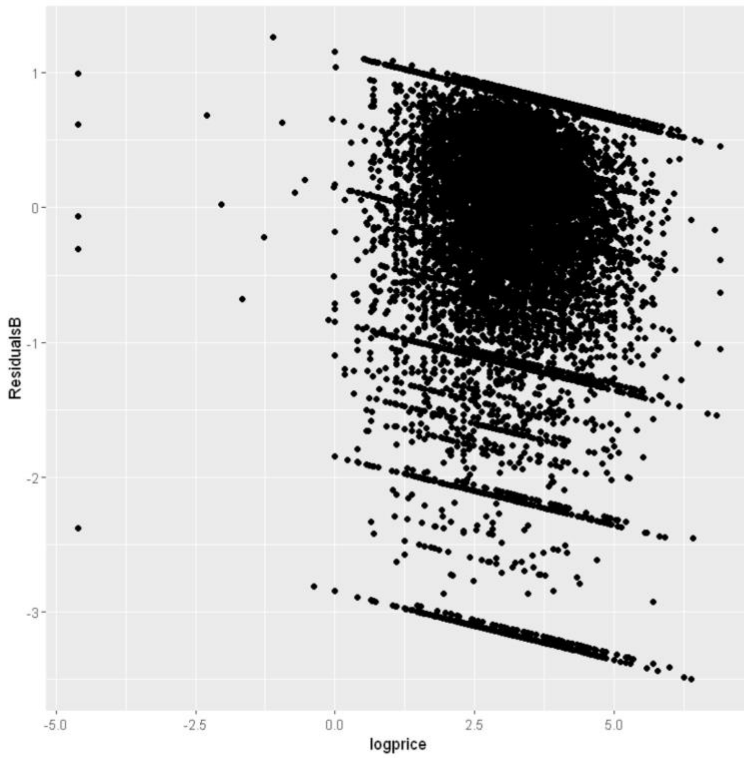

FIGURE 2 – RESIDUAL PLOTS OVER LOG PRICE FOR SPECIFICATION 2

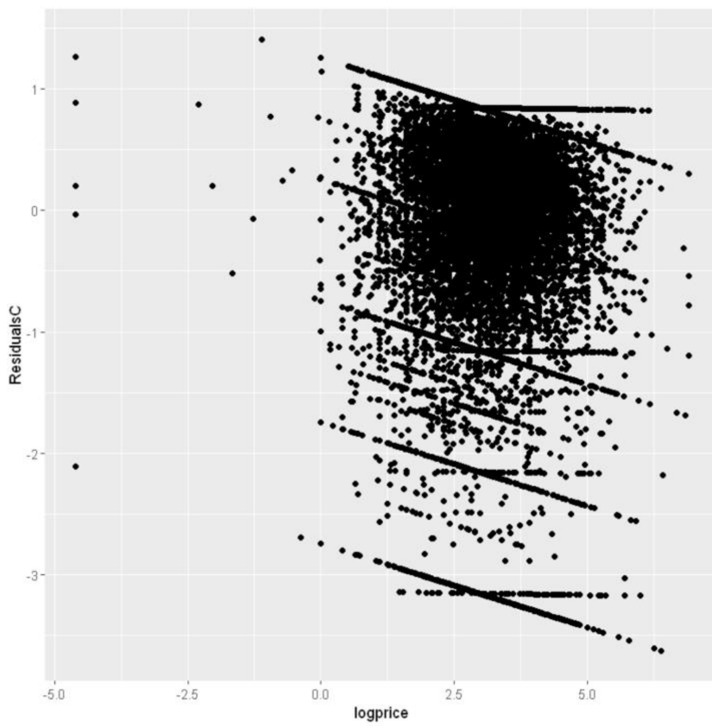

FIGURE 3- RESIDUAL PLOTS OVER LOG PRICE FOR SPECIFICATION 3

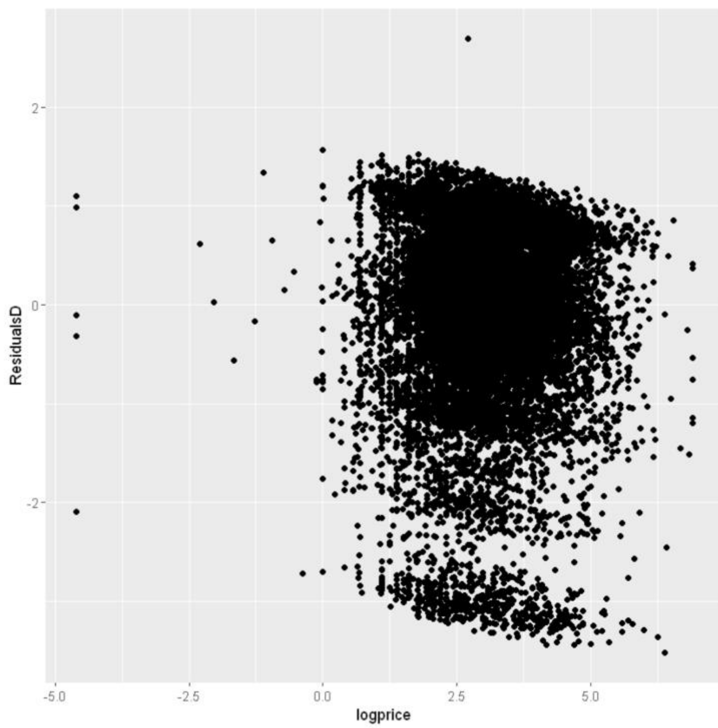

FIGURE 4- RESIDUAL PLOTS OVER LOG PRICE FOR SPECIFICATION 4

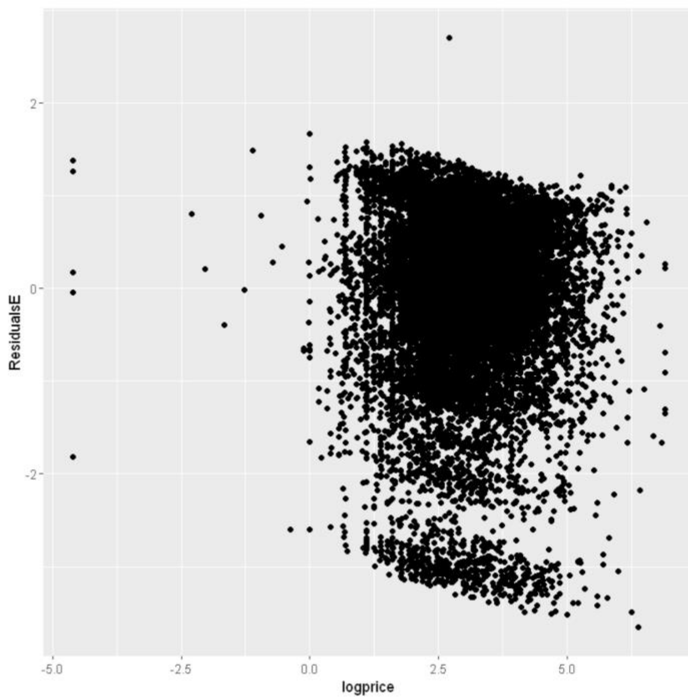

FIGURE 5- RESIDUAL PLOTS OVER LOG PRICE FOR SPECIFICATION 5

C. White's Test

```
Call:
lm(formula = res_sqA ~ y_hatA + I(y_hatA^2))

Residuals:
    Min      1Q  Median      3Q     Max 
-4.7111 -0.7402 -0.4957 -0.0512 10.8447 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 111.3903    15.4928   7.190 6.74e-13 ***
y_hatA      -52.3915     7.4875  -6.997 2.70e-12 ***
I(y_hatA^2)   6.2023     0.9046   6.856 7.30e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.801 on 17933 degrees of freedom
Multiple R-squared:  0.00537,   Adjusted R-squared:  0.005259 
F-statistic: 48.41 on 2 and 17933 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = res_sqB ~ y_hatB + I(y_hatB^2))

Residuals:
    Min      1Q  Median      3Q     Max 
-4.7219 -0.7567 -0.5041 -0.0416 10.7536 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 111.1735    14.1536   7.855 4.23e-15 ***
y_hatB      -52.9635     6.8322  -7.752 9.52e-15 ***
I(y_hatB^2)   6.3528     0.8245   7.705 1.38e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.814 on 17933 degrees of freedom
Multiple R-squared:  0.003579,  Adjusted R-squared:  0.003467 
F-statistic:  32.2 on 2 and 17933 DF,  p-value: 1.096e-14
```

FIGURE 6 – WHITE'S TEST RESULTS FOR SPECIFICATIONS 1 AND 2

```
Call:
lm(formula = res_sqC ~ y_hatC + I(y_hatC^2))

Residuals:
    Min      1Q  Median      3Q     Max 
-5.9354 -0.7522 -0.5134 -0.0645 11.3998 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  80.6760     7.9325  10.170   <2e-16 ***
y_hatC      -38.3319     3.8393  -9.984   <2e-16 ***
I(y_hatC^2)   4.5979     0.4648   9.893   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.813 on 17933 degrees of freedom
Multiple R-squared:  0.005888,  Adjusted R-squared:  0.005777 
F-statistic: 53.11 on 2 and 17933 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = res_sqD ~ y_hatD + I(y_hatD^2))

Residuals:
    Min      1Q  Median      3Q     Max
-6.2268 -0.6361 -0.3580 -0.0227 12.1638

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  51.0351     4.5651  11.179   <2e-16 ***
y_hatD      -21.8419     2.2389  -9.756   <2e-16 ***
I(y_hatD^2)   2.3438     0.2745   8.539   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.645 on 17933 degrees of freedom
Multiple R-squared:  0.06051,   Adjusted R-squared:  0.0604
F-statistic: 577.5 on 2 and 17933 DF,  p-value: < 2.2e-16


Call:
lm(formula = res_sqE ~ y_hatE + I(y_hatE^2))

Residuals:
    Min      1Q  Median      3Q     Max
-7.0664 -0.6333 -0.3697 -0.0279 13.1279

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  45.6000     3.6532  12.482   <2e-16 ***
y_hatE      -19.4407     1.7923 -10.847   <2e-16 ***
I(y_hatE^2)   2.0807     0.2199   9.462   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.648 on 17933 degrees of freedom
Multiple R-squared:  0.05806,   Adjusted R-squared:  0.05796
F-statistic: 552.7 on 2 and 17933 DF,  p-value: < 2.2e-16
```

FIGURE 7 – WHITE'S TEST RESULTS FOR SPECIFICATIONS 3,4 AND 5