

ONLINE REVIEWS AND PRODUCT QUALITY IN THE WEB HOSTING  
INDUSTRY

---

A Thesis

Presented to

The Faculty of the Department

of Economics

University of Houston

---

In Partial Fulfillment

Of the Requirements for the Degree of

Master of Arts

---

By

Arthur R. Haney

August, 2011

ONLINE REVIEWS AND PRODUCT QUALITY IN THE WEB HOSTING  
INDUSTRY

---

An Abstract of a Thesis

Presented to

The Faculty of the Department

of Economics

University of Houston

---

In Partial Fulfillment

Of the Requirements for the Degree of

Master of Arts

---

By

Arthur R. Haney

August, 2011

## **ABSTRACT**

Consumers purchasing Internet-based services from companies that lack physical presences have a unique problem. They do not have access to the information that they require in order to discern the quality of the company's products. In lieu of alternatives, they may turn to publicly available reviews from other consumers to aid them in coming to a decision regarding product quality. Using hedonic regression analysis, this study examines the potential for online reviews to function as a proxy for product quality. Estimating our initial hedonic model using OLS reveals that the number of negative reviews posted for a company is associated with decreased prices. Estimating a new hedonic model that controls for economies of scale provides the same conclusion, though the effect is slightly decreased. The results of our estimates provide statistical support of our initial hypothesis.

## TABLE OF CONTENTS

ABSTRACT .....	iv
LIST OF TABLES .....	vi
1 INTRODUCTION .....	1
2 MARKET ANALYSIS .....	3
3 LITERATURE REVIEW .....	8
4 DATA .....	10
5 METHODOLOGY .....	16
An Issue of Identification .....	18
6 RESULTS AND DISCUSSION .....	20
7 CONCLUSION .....	26
APPENDIX I: CORRELATION TABLE .....	29
REFERENCES .....	31

## LIST OF TABLES

Summary Statistics .....	15
Results from OLS Estimation of Model 1 .....	23
Results from OLS Estimation of Model 2 .....	25
Correlation Table .....	30

## Section 1

### INTRODUCTION

Billions of people use the Internet, and the Internet contains over one trillion unique web pages (Alpert 2008). Behind the scenes, the sheer scale of the Internet has enabled an entirely different type of industry.<sup>1</sup> This industry consists of services companies that exist to serve the specific needs that are created by the Internet. The companies within this new industry create, deploy, and maintain the computer hardware and software that are required in order for the Internet to function. While old, familiar names such as I.B.M. and Microsoft have established themselves as the dominant providers of certain technological solutions, a large number of markets within this industry are not yet dominated by any single large company. Instead, thousands of individual services companies compete for millions of customers in a young, dynamic marketplace.

Consumers face a unique challenge within this new industry. How should a consumer differentiate between the thousands of companies offering nearly identical services? In a more typical industry, consumers would have easy access to information that provide clear indications of quality, and they would use this perceived quality to assist them in selecting a firm to purchase from. However, most of these conventional indicators of product quality are less reliable for Internet services companies, and it is reasonable to hypothesize that consumers must use an alternative method to discern quality. In this study, we will examine an alternative avenue that consumers may use to

---

<sup>1</sup> It is worth making a strong differentiation between Internet service providers (ISPs) and Internet services companies (ISCs). Internet service providers literally provide consumers with access to the Internet. Internet services companies provide consumers with services after they have gained access to the Internet.

discern product quality – an avenue that is similar to the Internet services industry in that it would not exist if it were not for the Internet. We will use a hedonic regression model of one of the Internet services markets to analyze the potential for the number of online reviews available for a company to act as a proxy for product quality in lieu of stronger available measures.

Using a hedonic regression model, it is possible for a researcher to determine the contributions of a product's attributes to the overall value of the product. Hedonics is used in many different places. The Bureau of Labor Statistics uses hedonics to control for quality within the C.P.I. Hedonics has long been used within urban economics for determining the value of product attributes that cannot be normally measured – things such as pollution and access to public transportation. Hedonics has also been used within the automotive industry, the wine industry, and many other industries. At this point in time, hedonics is a fairly established econometric technique, and it is a perfect tool for the purposes of this study.

## Section 2

### MARKET ANALYSIS

In order to build a hedonic regression model it is necessary for a researcher to have an in depth understanding of the product market that are analyzing. The reason for this is that, in order for a hedonic regression model to provide usable information, certain market conditions must be true. Products within a market must have such a diversity of attributes that consumers choosing among them face a continuous function. Firms within the market must be in a fully competitive state – there can be no monopoly or oligopoly power being exercised within a market. Products within the market can not be subdivided and sold to consumers as lesser versions of the same product (Rosen 1974, p.37). For the purposes of this study, the web hosting industry was selected for analysis.

When a computer used interacts with the Internet (e.g. by sending an email, visiting a web page, etc), their request is met by a server. A server is a computer system that indirectly or directly provides a service to users on the Internet. In 2008, over 44 million servers were required to service the demands of Internet users (Forest 2008). All servers reside in a facility that is called a data center. Data centers provide servers with a high quality connection to the Internet as well as cooling facilities and backup power sources – all of the things needed to ensure that servers are able to operate at a high level of reliability. While large companies and governments will usually maintain and operate their own data centers, smaller organizations and individuals are usually limited to acquiring servers through a leasing agreement with a web hosting company.

Web hosting companies are companies that either own and operate data centers or



lease space within another company's data center. According to statistics provided by the U.S. Census Bureau in their 2007 Economic Census, there are currently more than 17 thousand companies in the web hosting industry in the U.S. These companies receive over \$66 billion dollars in annual revenue and employ nearly 400 thousand employees. It is difficult to locate accurate statistics of the global web hosting industry. These statistics are only for the American web hosting industry. Only 13.1% of the world's Internet users reside in America as of 2009, so it is reasonable to conclude that the global web hosting industry is far larger than its American subset (Wolfram|Alpha 2009).

The web hosting market provides an excellent illustration of the difficulties that consumers encounter when trying to assess the relative quality of products offered by Internet services companies. Consumers are not able to visit the office of a web hosting company to make a purchase; instead, they must visit the company's web site. While the location, size, and décor of a physical office may provide consumers with valuable quality clues, the visual appearance of a web hosting company's web site is a very poor indicator of overall quality. The reason for this is that web hosting company templates are readily available that allow any web hosting company to present a similar professional appearance (Themeforest 2011).

Other quality indicators are equally unreliable or unavailable to a consumer within the web hosting market. While a consumer might be able to use interaction with a sales staff as an indication of quality, the sales processes of most web hosting companies are automated and most web hosting consumers will not be able to interact with the staff of a web hosting company until after they have purchased service. In a more conventional

market, the availability of trusted, well-known brand names would aid consumers in discerning product quality. However, due to the presence of thousands of firms offering similar services, brand names are not a particularly useful indication of product quality within the web hosting industry. In light of the difficulties that face consumers in attempting to discern product quality in this environment, it is reasonable to hypothesize that online reviews serve a particularly valuable function for consumers.

The web hosting market is ideal for the type of analysis that will be conducted within this study. There are several primary types of web hosting services – shared hosting, virtual private server hosting, cloud hosting, server colocation, and dedicated server hosting. Most web hosting companies provide at least two of these services. However, for the purpose of this study, it was necessary to choose a web hosting service that met the requirements of the hedonic regression model.

Shared hosting and virtual private server hosting are hosting services that allow for a single physical server to be divided among multiple customers. While shared hosting simply pools customers onto a single server, virtual private server hosting uses a technology known as virtualization to divide a single physical server into one virtual server for each customer. While shared hosting supports hundreds of customers on a single server, virtual private hosting is typically limited to approximately 50 customers per server. Unfortunately, neither shared hosting nor virtual private hosting were suitable for the creation of a hedonic regression model because many of the attributes for these products are unobservable without access to detailed transactional information.

Cloud hosting abstracts the allocation of server resources completely from the

limitations imposed by the physical server hardware. This abstraction allows for a consumer to pay for their server needs in an on demand basis. Cloud hosting is a relatively recent development; the first cloud hosting provider was Amazon.com with the launch of Amazon Web Services in 2006 (Hof 2006). For the purposes of this study, the cloud hosting product market was not suitable because the majority of the market is currently concentrated into the hands of just a few large companies. The reason for this is that cloud hosting currently depends on proprietary software, and this proprietary software requires a lengthy, expensive development cycle – putting outside of what most hosting companies are able to accomplish

Server colocation is when a customer leases space within a data center and builds their own data center infrastructure. Under server colocation, a customer is responsible for providing the server hardware as well as any networking equipment necessary to connect their servers to the data center's pre-existing infrastructure. Server colocation was not selected for analysis in this study because data centers conduct business primarily based on their physical locations, and product quality is readily apparent to customers that visit these locations. In addition to this, many of the product attributes for server colocation are unobservable or would require a large amount of detailed information to determine. For example, one of the key attributes of server colocation services is the security of the facility; however, obtaining information on the exact security measures in place at one location is not likely to be possible. This would preclude a hedonic regression model for server colocation as it would require a larger amount of data than that which would be easily obtainable.

Dedicated server hosting is when a customer is able to purchase a lease for server hardware that is owned by the hosting company. The web hosting company assumes all responsibility for the maintenance of the server hardware and networking infrastructure. Dedicated servers are not shared among customers. There is a tremendous amount of variety within the dedicated server market; it is quite conceivable that consumers face a virtually continuous function for product choice. Dedicated servers can not be subdivided and resold as dedicated servers. Due to the low barriers of entry, there is a large number of companies that provide a dedicated server hosting service, and no single company is large enough to exercise any amount of market power. For the purposes of this study, the dedicated server market was selected for analysis because it was the best fit for the assumptions required for a hedonic regression analysis.

### Section 3

#### LITERATURE REVIEW

There is a substantial amount of existing research that examines the correlation between product reviews posted online and sales. Basuroy, Chatterjee, and Ravid (2003) examined the effects of online reviews on the movie industry. They found that reviews were strongly correlated with sales, and negative reviews have a strong correlation than positive reviews. However, they also found that the strength of the correlation between negative reviews and sales decreases over time. Duan, Gu, and Whinston (2008) used panel data to analyze the impact of online reviews on movie box office performance. While they found that the actual ratings assigned by movie reviewers had no effect, box office sales are strongly correlated with the total number of online reviews.

Chevalier and Mayzlin (2006) examined review data from Amazon.com and BarnesandNoble.com. They concluded that improvements in online review scores lead to relative increases in a book's sales. They also found that negative (1-star) reviews had a stronger correlation with sales than positive (5-star) reviews. Hu, Liu, and Zhang (2008) used data from Amazon.com and a model that including qualitative measures of review quality. They discovered convincing statistical evidence that the impacts of reviews on book sales diminishes over time.

Previous research has focused primarily on the potential impact of online reviews on the sales of the product. However, in this study, we will use hedonic regression analysis to examine the possibility of online reviews functioning as an indicator of product quality. The U.S. Bureau of Labor Statistics usually includes some measure of

relative product quality in their hedonic regressions. They have used both broad classifications (Liegey 2003) as well as product brands to control for quality (Kokoski et al. 2001), and in both published research papers, the variables controlling for quality had significant coefficients. These results were important for the purpose of this study, because they indicate that product quality may have significant results within a hedonic regression model for a commodity market.

At this time, it is not possible to locate hedonic regression estimates of any web hosting industry products. Because servers have similar hardware characteristics to conventional desktop computers, hedonic estimates of personal computer products would have been useful in establishing the relative values of computer hardware characteristics. One of the more recent studies was published by the Bureau of Labor Statistics in 2001. While the age of the study decreases its usefulness in drawing conclusions regarding modern computer hardware, it is worth noting that the study found that computer characteristics such as cpu speed, memory, and hard drive space are statistically significant and have positive values for consumers (Holdway 2001). Because of these findings, it is reasonable to conclude that these same hardware characteristics will be important for dedicated server customers.

## Section 4

### DATA

The data for this study is cross-sectional and it consists of 710 unique observations. Each observation within the dataset is a single dedicated server and it includes variables for monthly price as well as hardware, software, service, and company characteristics. Observations were collected for 75 web hosting companies using the posted prices available from each company's web site. To avoid the complexity of using currency exchange rates, only companies that provided a price listing in U.S. Dollars were used for this study. In addition to this, results were limited to companies that provide a web site in English. The majority of the observations were collected from companies were selected from a list of web hosting companies available from WebHostingStuff.com. A smaller number of observations were collected by using Google searches to locate additional dedicated server companies. All observations were collected during the time period of May 2011 to July 2011. While dedicated servers can be leased on an annual or quarterly basis, all observations were restricted to monthly observations within the dataset. Summary statistics for the dataset are available in Table 1.

Collecting the hardware information for each dataset was, for the most part, relatively straightforward. The amount of memory and hard drive space for each server were included with price information for all of the companies used in this study. The type of disk controller used for each server was also provided with the pricing information for each dedicated server. A dummy variable was created for the case where this disk controller was a high speed controller. The reason that this was done is that high speed

drives are considerably more expensive than regular drives, and it was hoped that a dummy variable would help to address this pricing difference in the hedonic model.

Problems were occasionally encountered when collected the total gigahertz count for each server. While most companies within the dataset provided the model of each server's CPU, not all of the companies within the dataset included information on the total gigahertz provided by the CPU. Because of this omission, it was sometimes necessary to visit the CPU manufacturer's home page to collect the necessary additional information for each observation. When a company failed to provide the exact model for the CPU (instead providing a range of CPUs), a mid-grade CPU of the appropriate product family was selected. The distance between the gigahertz counts of CPUs within the same product family were relatively small, and only a few companies provided ranges instead of a single model. Therefore, the potential effects of this approximation on the CPU variable would be extremely limited.

The software characteristics collected for each observation were relatively limited. A dummy variable incorporates whether or not windows is provided “for free” with the server. An additional dummy variable was created to incorporate whether or not a control panel is included with a server. For this control panel dummy, it was necessary to assert a more specific definition than that which was typically provided by the companies within the dataset. Nearly all companies within the dataset advertise that they offer a free control panel. However, in most of these cases the control panel that they in fact refer to is the one that is freely available for the server's operating system. For the purposes of this study, the control panel dummy variable indicates the presence of a control panel that



requires a purchased license to operate.

The service characteristics for each server comprise a broad arrangement of variables that address various features and limitations of each server's service contract as well as some company characteristics. The variable for the amount of data transfer allotted to each server was collected with the price information for each server. The dummy variable for 24-hour phone support was collected by examining the support section for each company, and it literally indicates if the company provides a 24-hour support phone number.<sup>2</sup> The number of dedicated server base configurations provided by each company was also recorded for each observation. The number of base configurations was determined by counting the number of unaltered server configurations provided by each company. Providing more than a small number of base configurations requires a company to have a considerable infrastructure for tracking and managing server equipment, so this variable was collected as a proxy for the overall technological efficiency of a company's backend operation.

A dummy variable was created for whether a company provides a private network for its servers. A private network allows multiple servers owned by a single company to communicate with each other without impacting the server's data transfer allotment. An addition benefit of private networks is that they allow for decreased time for communication between servers. Very few companies provide private networks for servers by default (with no additional cost). Within our dataset, only 2 companies provide this feature.

---

<sup>2</sup> It is worth noting that nearly all companies in this dataset had a 24-hour phone number for sales, but 24-hour support was less common. Only 61% of the dataset included 24-hour phone support.

A dummy variable was created for whether or not a company includes server management at no additional cost with the server. The creation of this dummy variable was somewhat difficult, because all companies within the dataset advertise that they provide server management. However, the exact services that each company provides as part of its server management varies considerably. For the purposes of this study, a considerably more narrow definition was used. The management dummy variable only indicates the presence of total server management. To qualify as total server management, the management service must include software updates and installation service at no additional cost. Only 4.6% of the observations within the dataset meet this strict definition of server management.

The variable of interest for this study is the number of reviews posted for each company. The primary problem when determining how to collect the data for this variable was the lack of a credible review site for the web hosting industry. While several web sites exist that purport to collect reviews for web hosting companies, these companies make very little attempt to verify the veracity of the posted reviews. To ensure the quality of the reviews variable, the data for this variable was collected by using Google to conduct site-specific searches of the WebHostingTalk.com (WHT) discussion group. WHT is an online discussion group that is dedicated towards discussing the Internet services industry and related topics. WHT is home to over 386,000 members, and its members have made over 7.5 million posts. It was selected as the source of the reviews variable because the Web Hosting Talk usage policy requires verification for reviews, and the moderators of the forum investigate all posted reviews to verify their

authenticity.

The number of monthly unique visitors for each web site was collected by using the publicly accessible web site traffic data available from Compete.com to record 13-month averages for the traffic to each web hosting company's primary site. While Compete also provides statistics for total visits, the number of unique visitors to each web site provides a more reliable way to compare the traffic of multiple web sites. The reason for this is that individual web sites may redirect customers to a separate address for purchasing a server or for support inquiries. In this case, when a customer returns to the primary address this would be counted as another visit. The unique visitors statistic is better because, in the previously outlined scenario, the return to the original address does not increase the number of unique visitors to the site.

Compete's statistics are collected by tracking the activity of over 2 million users. They then use a proprietary statistical method to approximate a web site's total traffic. One shortcoming of their estimation technique is that web sites with a very small number of visitors are represented by Compete.com as having no traffic. However, only 40 of the observations within the dataset were subject to this shortcoming, so the downward bias that would result because of this is likely to be statistically insignificant.

**Table 1. Summary Statistics**

	<b>Units</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>Min</b>	<b>Max</b>
<b>Monthly Lease Price</b>	dollars	300.14	222.66	29.95	1699
<b>Hardware Characteristics</b>					
Memory	gigabytes	5.1	7.6	0.5	128
Disk	gigabytes	690	960	36	12,000
CPU (total gigahertz)	gigahertz	12.3	10.2	1.6	91.2
Server_CPU	-	0.611	0.488	0	1
High_Speed_Drive	-	0.041	0.198	0	1
<b>Software Characteristics</b>					
Windows_Included	-	0.052	0.222	0	1
Control_Panel_Included	-	0.114	0.318	0	1
<b>Service Characteristics</b>					
Base_Configurations	configurations	31	37	1	110
Data_Transfer_Limit	gigabytes	4,470	5,329	0	75,000
Includes_Private_Network	-	0.197	0.398	0	1
Includes_Management	-	0.046	0.211	0	1
24_7_Phone_Support	-	0.611	0.488	0	1
<b>Negative Reviews</b>					
Bad_Reviews	reviews	484	689	0	1,880
<b>Monthly Unique Visitors</b>					
Number_Monthly_Visitors	unique visitors	24,199	76,154	0	724,712

**Number of Observations = 710**

## Section 5

### METHODOLOGY

Our hedonic regression model for estimating the average effect on price of dedicated server attributes is specified as follows:

$$(1) \quad Price = B_0 + B_1 hardware + B_2 software + B_3 service + B_4 badreviews + \varepsilon$$

In model 1, the dependent variable is the monthly price of a dedicated server lease. The independent variables hardware and software are vectors of server-specific characteristics. The independent variable service is a vector of company characteristics that may affect the price of a dedicated server. The independent variable for bad reviews is the number of bad reviews posted about a particular company. We will estimate Equation 1 using ordinary least squares estimation with robust standard errors to correct for heteroskedasticity.

Included within the hardware vector are basic hardware attributes such as the amount of memory, the speed of the CPU, and the amount of hard drive space were included. To address the differences in pricing between consumer-grade hardware and server-grade hardware of similar performance, a dummy variable is included that is activated when the dedicated server is built using server-grade components.<sup>3</sup> To address the differences in price between standard and high performance drives, a dummy variable is included for high performance drives. In addition to this dummy variable, an interaction between this dummy variable and the amount of drive space within the server is also included. This interaction is included because the per-gigabyte prices for

---

<sup>3</sup> While the performance is not altered, manufacturers usually test server-grade components far more rigorously before being sending them out to retailers.

conventional hard drives is dramatically lower than the per-gigabyte prices for high performance drives.

Included within the software vector are two variables. Because windows licenses require more upfront expenditures than server operating systems, a dummy variable is included for whether or not a server includes windows. A dummy variable is also included for server control panel software that is included with a server at no additional cost. While there is considerable variation in the exact software that is included with dedicated servers, variables for these were not included because most of this software can be easily installed after the dedicated server lease has commenced.

The service vector contains variables that derive from the characteristics of the dedicated server hosting company. A dummy variable is included that is activated when a dedicated server includes private networking support at no additional cost. Another dummy variable activates when a dedicated server includes total management services at no additional cost. A dummy variable is also included for the availability of 24-hour phone support services. The number of base configurations provided by each web hosting company is included as a proxy for the efficiency of the company's backend operations.

While other research into the effect of online reviews often include positive reviews as well as numerous review score ratios, the limitations of the dataset available for this study require us to incorporate reviews in a more simplified fashion. Because we only have the total count of reviews rather than a definitive score for each review, there is a high degree of correlation between the number of negative and the number of positive reviews. Both negative and positive reviews increase as companies

This correlation would result in multicollinearity, which would make it fairly difficult for an ordinary least squares regression to provide valid coefficients for the variable of interest. Because we are examining online reviews as a proxy for product quality using a hedonic model, the omission of positive reviews is not necessarily a significant shortcoming of our model. For a customer that is seeking to lease a dedicated server, the cost of a negative event such as data loss or business disruption is tremendous, and the benefits of a positive event such as a productive interaction with customer service is substantially less. This means that it is reasonable to expect that the number of negative reviews available to a customer might constitute valuable information on the quality of a web hosting company's offerings.

#### An Issue of Identification

There is a potential problem with incorporating the number of negative reviews for a company as a variable within a hedonic regression model. This problem arises from the very nature of hedonic regression modeling. A hedonic regression model allows for the attributes of a product to be independently valued, but the values estimated for these attributes are the result of supply just as much as demand (Nerlove 1995, p.1699). This impacts our original model in several important ways. If there are economies of scale within the dedicated server market, then it may be possible that a negative coefficient does not reflect the consumer's value of negative reviews. Because a larger web hosting company is likely to have more negative reviews, a negative coefficient would instead reflect the fact that it becomes cheaper for a company to provide dedicated servers as the company increases in size. To address this potentially serious endogeneity issue, it was

necessary to incorporate an additional model into our study.

Our more advanced hedonic regression model, model 2, incorporates the variable for the number of unique visitors as a proxy for the relative size of the web hosting company:

$$(2) \quad \begin{aligned} Price = & B_0 + B_1 hardware + B_2 software + B_3 service + B_4 badreviews \\ & + B_5 uniquevisitors + \epsilon \end{aligned}$$

Because a web hosting company may offer more services than just dedicated servers, the number of unique visitors to a company's web site is only a rough indication of the number of dedicated server customers that a web hosting company may have. This means that the unique visitors variable may not be a significant indicator of the presence of dedicated server hosting economies of scale. However, it is reasonable to conclude that the number of a company's dedicated server customers is strongly correlated with the total size of the company. Thus, by controlling for company size, we will be able to solve the endogeneity problem produced by not addressing the potential effects of economies of scale on the coefficient estimated for the bad reviews variable, which is our variable of interest.



## Section 6

### RESULTS AND DISCUSSION

By estimating model 1 using OLS we are able to obtain the coefficient estimates. The resulting coefficients and standard errors can be found in table 2. The coefficient on the variable of interest, bad reviews, is statistically significant and negative. It indicates that each additional negative review for a dedicated server hosting company is associated with a price that is, on average, 5.8 cents lower. While 5.8 cents is not a large number in and of itself, we can obtain the effect size by multiplying this coefficient by the standard deviation (see table 1). This gives us an effect size of \$39.96 dollars, which is a significantly large effect given that the average monthly price of a dedicated server is \$300.14 dollars. This result would seem to provide statistical evidence in support our initial hypothesis, which was that negative reviews are a useful proxy for product quality.

With a few notable exceptions, the remaining coefficients estimated for model 1 have the expected signs and are statistically significant.<sup>4</sup> The notable exceptions to this are the coefficients for 24-hour phone support, disk space, and data transfer. We would expect to find that these coefficients are both statistically significant and positive. However, the actual estimates, in each case, are surprisingly different from expectations.

Rather than the positive coefficient that we would expect, our estimates for 24-hour phone support provide a negative coefficient that is statistically insignificant. We would expect the phone support coefficient to be positive because, intuitively, it seems that phone support would both cost web hosting companies more to provide and provide

---

<sup>4</sup> It is worth noting that value added features that are included with dedicated servers for “no additional cost” actually have highly significant, positive coefficients. Moreover, these coefficients correspond with the amount that a customer would expect to pay to add these features to a basic dedicated server plan.

greater value to the consumer. A possible explanation for this finding may be derived from the nature of the phone support variable. Because this variable is a dummy variable then the coefficient should be interpreted within the context of the alternatives. A common alternative to 24-hour phone support is a comprehensive online support system that is based on the usage of trouble tickets. It may be the case that customers value this alternative system more than they value having 24-hour access to phone support. Within the web hosting industry, most of the interactions between customers and companies within this industry occur strictly within the confines of the Internet. If this is indeed the case, then the convenience of being able to have their technical support queries met without needing to make a phone call may, indeed, provide greater value to consumers.<sup>5</sup>

While we would have expected the coefficient for data transfer to be significant and positive, our estimated coefficient for this variable is statistically insignificant. A possible explanation for this finding is that perhaps data transfer is a non-binding limit. The average limits on data transfer for the companies within our dataset is 4,470 gigabytes. Most web sites today consist of text, some images, and perhaps a few small videos. For these conventional web sites, it would take an extraordinarily large amount of traffic for them to reach the average data transfer limit. Larger sites with videos and other media content may find it easier to reach this limit, but these sites are actually fairly uncommon. If it is rare for consumers to reach this limit, then it is reasonable to conclude that consumers would have no real value for this excess data transfer.<sup>6</sup>

---

5 Left unaddressed by our explanation is the supply-side of this question. In other words, does it cost more for companies to provide a comprehensive online trouble ticket system? This is a question that we are not able to answer within this dataset, but would provide an interesting topic for future study.

6 This finding is actually an interesting demonstration of identification and the hedonic regression technique. The coefficients estimated in a hedonic regression are a product of both the demand and the supply curves within a market. While it is a cost to producing firms to provide it, the fact that

The coefficients estimated for the three disk space variables are also surprising. Both the variable for the amount of disk space and the dummy variable indicating a high performance drive are not significant. However, the coefficient calculated for the interaction between the amount of disk space and the high speed drive dummy variable is highly statistically significant. Indeed, the estimated effect of this interaction variable is \$36.18 dollars.<sup>7</sup> The results of a test for the joint significance of all three drive variables provides a p-value of 0.008. Despite two of these variables being statistically insignificant, this result indicates that the three variables have joint statistical significance.

A possible explanation for the lack of significance of the disk space variables is actually similar to the same explanation for the lack of significance of the data transfer limit variable. The average amount of hard drive space allocated to each server within the dataset is 690 gigabytes. This is enough space for approximately 35 days of high quality video or over 500,000 photographs. It is unlikely that the average dedicated server customer is able to take complete advantage of all of this space. Thus, the amount of disk space available with each dedicated server is a non-binding limit. If this is the case, then it would make sense for the coefficient estimated for hard drive space to lack significance. The lack of significance for the dummy variable indicating that a drive is high performance is most likely due to the fact that high performance drives have a per-gigabyte price premium over conventional drives. Under these interpretations, the coefficients for the drive space variables actually correspond fully with expectations.

---

consumers do not value transfer limits results in a very low significance for the estimated data transfer coefficient.

7 This estimate effect is calculated simply by multiplying the coefficient for the interaction variable by the standard deviations of

**Table 2. Results from OLS Estimation of Model 1***Monthly dedicated server lease price as dependent variable.*

	<b>Coefficient</b>	<b>Std. Err</b>	<b>t-value</b>	<b>p-value</b>
<b>Intercept</b>	45.79	9.49	4.82	0.000
<b>Hardware Characteristics</b>				
Memory	7.14	1.42	5.05	0.000
Disk	-0.006	0.01	-0.59	0.556
CPU (total gigahertz)	13.29	0.91	14.67	0.000
Server_CPU	50.99	10.95	4.66	0.000
High_Speed_Drive	-46.54	44.18	-1.05	0.293
Server_CPU*High_Speed_Drive	0.255	0.08	3.06	0.002
<b>Software Characteristics</b>				
Windows_Included	113.79	20.51	5.55	0.000
Control_Panel_Included	45.29	16.78	2.70	0.007
<b>Service Characteristics</b>				
Base_Configurations	1.48	0.25	5.95	0.000
Data_Transfer_Limit	0.0007	0.0008	0.86	0.389
Includes_Private_Network	55.63	17.06	3.26	0.001
Includes_Management	105.51	19.39	5.44	0.000
24_7_Phone_Support	-39.33	12.66	-3.11	0.002
<b>Negative Reviews</b>				
Bad_Reviews	-0.058	0.01	-5.77	0.000

**Number of Observations = 710, Adjusted R-squared = 0.7245**

While the results from estimating model 1 provide evidence in support of our hypothesis, the results from estimating model 2 allow us to examine the effects of including the relative sizes of the companies as an explanatory variable. The results from estimating model 2 using OLS are found in table 3. The addition of the new variable does not significantly change most of the coefficients within the regression. However, the inclusion of this additional variable does have an interesting effect on the coefficient estimated for the bad reviews. While the coefficient estimated for the monthly unique visitors variable is itself not significant at the 5% level, this is most likely due to monthly unique visitors being only a rough approximation of the number of total dedicated server customers that a company has.

The addition of the unique visitors variable to the second model decreases the estimated coefficient for the negative reviews variable by 6.01%. This change indicates that there was omitted variable bias in the original model due to not addressing the potential effects of economies of scale. However, our variable of interest remains both negative and statistically significant. While the estimated effect size decreases from \$39.96 to \$37.56 dollars, this is still a significant amount. This means that even after incorporating a measure of economies of scale within our original model, the enhanced model still provides statistical evidence in support of our hypothesis, which is that the number of negative reviews posted for each company provides a useful proxy for product quality.

**Table 3. Results from OLS Estimation of Model 2***Monthly dedicated server lease price as dependent variable.*

	<b>Coefficient</b>	<b>Std. Err</b>	<b>t-value</b>	<b>p-value</b>
<b>Intercept</b>	43.83	9.59	4.57	0.000
<b>Hardware Characteristics</b>				
Memory	7.14	1.44	4.95	0.000
Disk	-0.005	0.01	-0.50	0.618
CPU (total gigahertz)	13.33	0.91	14.69	0.000
Server_CPU	50.17	11.02	4.55	0.000
High_Speed_Drive	-43.90	44.12	-0.99	0.320
Server_CPU*High_Speed_Drive	0.249	0.08	3.01	0.003
<b>Software Characteristics</b>				
Windows_Included	109.91	20.71	5.31	0.000
Control_Panel_Included	53.51	18.57	2.88	0.004
<b>Service Characteristics</b>				
Base_Configurations	1.42	0.24	5.82	0.000
Data_Transfer_Limit	0.0007	0.0008	0.88	0.380
Includes_Private_Network	60.69	17.38	3.49	0.001
Includes_Management	112.05	17.32	6.47	0.000
24_7_Phone_Support	-35.58	13.00	-2.74	0.006
<b>Negative Reviews</b>				
Bad_Reviews	-0.055	0.01	-5.50	0.000
<b>Monthly Unique Visitors</b>				
Number_Monthly_Visitors	-0.0001	0.00005	-1.92	0.056

**Number of Observations = 710, Adjusted R-squared = 0.7251**

## Section 7

### CONCLUSION

It is important to note the limitations of the findings within this paper. Our hedonic regression analysis found a negative and significant coefficient for the variable indicating the number of negative reviews. This does not mean that the online reviews are responsible for the decrease in monthly dedicated server lease price in the same way that additional memory is responsible for the increase in dedicated server lease price. This difference in interpretation arises from the fact that we are using the number of negative reviews as a proxy of product quality. If we were able to incorporate product quality directly within our model, then we could reach the conclusion that product quality has a causal relationship with price. However, because we are limited to a count of the number of negative reviews, this precludes the possibility of a definitive causal interpretation.

As a topic for future research, there is a methodology that could have potentially provided an avenue for obtaining a more direct estimation of the causal effects of product quality. Rather than a cross-sectional dataset, this enhanced method would require panel data. Rather than using the number of negative reviews as a proxy for quality, the enhanced method would require scanning each review and determining a normalized score (perhaps on a 1 to 10 scale) for the review's assessment of the web hosting company – effectively normalizing the reviews data. After doing this, then a variable incorporating the average review score for each company would take the place of the original variable based on the number of negative reviews within our model. Then, by using panel data and a difference-in-difference technique, the effects of changing product

quality on the price of a dedicated server could be observed directly.

The primary problem with this enhanced methodology is that it would be challenging to construct a method for assigning a score to a review. A linguistic approach would have to be used to assign a score to each review. The linguistic algorithm would take each descriptive review and convert it to a quantified score by using word ratios, sentence structure analysis, and additional data. Because there are over 300,000 reviews that would need to be analyzed, it would be impossible to apply this algorithm by hand. A computer program would have to be written to process the reviews automatically and output a score. While this new reviews variable would probably provide a much stronger indication of product quality than a simple count of negative reviews, it would require a substantial investment of time to gather the necessary data and implement the new approach. Despite this, it would be interesting to explore this alternative methodology in future research.

An additional limitation of this study is also worth addressing, though it is not as exceptional as the first limitation discussed. The dataset that was collected for this study includes only web hosting companies with web sites written in English and prices denominated in dollars. It is possible that these restrictions may have introduced some selection bias. It is possible that the excluded companies have some fundamentally different characteristics from the companies that were included as part of this dataset. However, it is perhaps unlikely that this is the case because the dedicated server hosting product market is a competitive market, and it is also a market that is international. A customer located in Sweden can easily purchase from a web hosting company selling



dedicated servers in America. The prices that are set by a dedicated server hosting company in Sweden would have to remain competitive with this international competition. Because of this, it is reasonable to conclude that the possible selection bias (due to limitations on the dataset) are small.

In this study, we used a hedonic regression model to examine whether online reviews would function as a proxy for product quality within an Internet services company market, and our regression results provide statistical evidence in support of our initial hypothesis. Even after incorporating an additional explanatory variable to address potential economies of scale within the product market, the variable of interest, the number of bad reviews posted online for each company, remained statistically significant. While we are not able to determine a causal relationship, based on the results of our research, it seems that it is quite likely that online reviews are a powerful tool for discerning product quality within the dedicated server market. This finding is important because it opens the door to future research into online reviews as a proxy for product quality. It also opens the door for future research that can use online reviews to determine the nature of the relationship between product quality and the prices that consumers are willing to pay for a product.

## Appendix I

### CORRELATION TABLE

On the following page, a correlation table is provided for all of the variables that were used in this study. In addition to these variables, the correlations for the good reviews variable is also included, so that readers may examine the the reason that this variable was excluded to prevent multicollinearity. Due to the space constraints imposed by page size, the variable names are abbreviated in the correlation table. A legend for these abbreviations is provided below:

<b>Variable</b>	<b>Abbreviation</b>
Monthly Lease Price	prc
Memory	mem
Disk	dsk
CPU (total gigahertz)	cpu
Server_CPU	srv
High_Speed_Drive	spd
Windows_Included	win
Control_Panel_Included	cp
Base_Configurations	cfg
Data_Transfer_Limit	dtl
Includes_Private_Network	priv
Includes_Management	man
24_7_Phone_Support	ph
Good_Reviews	good
Bad_Reviews	bad
Number_Monthly_Visitors	nmv

**Table 4. Correlation Table***Please see previous page for abbreviations.*

	<b>prc</b>	<b>mem</b>	<b>dsk</b>	<b>cpu</b>	<b>srv</b>	<b>spd</b>	<b>win</b>	<b>cp</b>	<b>cfg</b>	<b>dtl</b>	<b>priv</b>	<b>man</b>	<b>ph</b>	<b>good</b>	<b>bad</b>	<b>nmv</b>
<b>prc</b>	1.00															
<b>mem</b>	0.53	1.00														
<b>dsk</b>	0.23	0.56	1.00													
<b>cpu</b>	0.78	0.51	0.26	1.00												
<b>srv</b>	0.43	0.13	0.03	0.41	1.00											
<b>spd</b>	0.08	-0.04	-0.03	0.02	0.11	1.00										
<b>win</b>	0.05	-0.02	-0.08	-0.08	0.02	0.14	1.00									
<b>cp</b>	0.04	0.03	0.07	-0.08	-0.11	-0.01	0.18	1.00								
<b>cfg</b>	0.26	-0.05	-0.18	0.23	0.32	-0.10	-0.16	-0.22	1.00							
<b>dtl</b>	0.01	0.17	0.14	0.05	0.00	-0.08	-0.06	-0.04	-0.13	1.00						
<b>priv</b>	0.33	-0.01	-0.17	0.27	0.39	-0.10	-0.12	-0.18	0.83	-0.17	1.00					
<b>man</b>	0.03	-0.08	-0.06	-0.12	-0.11	-0.01	0.04	0.36	-0.11	-0.01	-0.11	1.00				
<b>ph</b>	0.10	0.11	0.04	0.19	0.18	-0.04	0.07	-0.06	0.40	0.15	0.18	-0.15	1.00			
<b>good</b>	0.26	0.02	-0.11	0.24	0.35	-0.03	-0.13	-0.16	0.89	-0.09	0.82	-0.09	0.36	1.00		
<b>bad</b>	0.20	-0.01	-0.11	0.23	0.32	-0.03	-0.16	-0.19	0.84	-0.01	0.68	-0.12	0.37	0.90	1.00	
<b>nmv</b>	0.09	0.10	0.11	0.11	0.05	-0.05	-0.06	0.31	0.16	0.04	0.15	0.20	0.21	0.27	0.21	1.00

## REFERENCES

Alpert, Jesse, and Nissan Hajaj. "We Knew the Web Was Big..." *Official Google*

*Blog*. Google, July 2008. Web. 21 July 2011.

<<http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html>>.

Rosen, Sherwin. "Hedonic Prices and Implicit Markets: Product Differentiation in

Pure Competition." *Journal of Political Economy* 82.1 (1974): 34. *JSTOR*. Web. 13

May 2011. <[http://links.jstor.org/sici?sici=0022-](http://links.jstor.org/sici?sici=0022-3808%28197401%2F02%2982%3A1%3C34%3AHPAIMP%3E2.0.CO%3B2-U)

[3808%28197401%2F02%2982%3A1%3C34%3AHPAIMP%3E2.0.CO%3B2-U](http://links.jstor.org/sici?sici=0022-3808%28197401%2F02%2982%3A1%3C34%3AHPAIMP%3E2.0.CO%3B2-U)>.

Forest, William, James Kaplan, and Noah Kindler. Data Centers: How to Cut Carbon

Emissions and Costs. Issue brief. McKinsey & Company, Winter 2008. Web. 25

July 2011.

<[http://www.mckinsey.com/clientservice/bto/pointofview/pdf/BT\\_Data\\_Center.pdf](http://www.mckinsey.com/clientservice/bto/pointofview/pdf/BT_Data_Center.pdf)

>.

"Detailed Statistics - American FactFinder." 2007 Economic Census. U.S. Census

Bureau, 1 July 2011. Web. 29 June 2011.

<[http://factfinder.census.gov/servlet/IBQTable?\\_bm=y](http://factfinder.census.gov/servlet/IBQTable?_bm=y)>.

Wolfram|Alpha. "Number of Internet Users in America." Wolfram Alpha LLC. Web.

26 July 2011. <[http://www.wolframalpha.com/input/?](http://www.wolframalpha.com/input/?i=number+of+internet+users+in+america)

[i=number+of+internet+users+in+america](http://www.wolframalpha.com/input/?i=number+of+internet+users+in+america)>.

Wolfram|Alpha. "Number of Internet Users in the World." Wolfram Alpha LLC. Web. 26 July 2011. <<http://www.wolframalpha.com/input/?i=number+of+internet+users+in+the+world+in+2009>>.

Wolfram|Alpha. "Number of Internet Users in the World in 2007." Wolfram Alpha LLC. Web. 26 July 2011. <<http://www.wolframalpha.com/input/?i=number+of+internet+users+in+the+world+in+2007>>.

"Hosting - ThemeForest." Premium WordPress Themes and Web Templates - ThemeForest. ThemeForest. Web. 1 Aug. 2011. <<http://themeforest.net/category/site-templates/technology/hosting>>.

Hof, Robert. "Jeff Bezos' Risky Bet." Business News, Stock Market & Financial Advice. Businessweek, 13 Nov. 2006. Web. 21 July 2011. <[http://www.businessweek.com/magazine/content/06\\_46/b4009001.htm](http://www.businessweek.com/magazine/content/06_46/b4009001.htm)>.

Basuroy, Suman, Subimal Chatterjee, and S. Abraham Ravid. "How Critical Are Critical Reviews? The Box Office Effects of Film Critics, Star Power, and Budgets." Journal of Marketing 67.4 (2003): 103-17. Dr. Suman Basuroy, Publications. Oct. 2003. Web. July 2011. <<http://faculty-staff.ou.edu/B/Suman.Basuroy-1/Research-CV.html>>.

Duan, W., B. Gu, and A. Whinston. "Do Online Reviews Matter? — An Empirical Investigation of Panel Data." Decision Support Systems 45.4 (2008): 1007-016. SSRN. Web. July 2011. <[http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=616262](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=616262)>.

- Chevalier, Judith A., and Dina Mayzlin. "The Effect of Word of Mouth on Sales: Online Book Reviews." *Journal of Marketing Research* 43.3 (2006): 345-54. SSRN. Web. July 2011. <[http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=432481](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=432481)>.
- Hu, Nan, Ling Liu, and Jie Jennifer Zhang. "Do Online Reviews Affect Product Sales? The Role of Reviewer Characteristics and Temporal Effects." *Information Technology and Management* 9.3 (2008): 201-14. SSRN. Web. July 2011. <[http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1324190](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1324190)>.
- Liegey, Paul R. "Hedonic Model for DVD Players." U.S. Bureau of Labor Statistics, 22 Apr. 2003. Web. May 2011. <<http://www.bls.gov/cpi/cpidryer.htm>>.
- Kokoski, Mary, Keith Waehrer, and Patricia Rozaklis. "Using Hedonic Methods for Quality Adjustment in the CPI: The Consumer Audio Products Component." U.S. Bureau of Labor Statistics. Web. 16 Oct. 2001. <<http://www.bls.gov/cpi/cpiaudio.htm>>.
- Holdway, Michael. "Quality-Adjusting Computer Prices in the Producer Price Index: An Overview." U.S. Bureau of Labor Statistics, 16 Oct. 2001. Web. May 2011. <<http://www.bls.gov/ppi/ppicomqa.htm>>.
- "Compete: Help & FAQs | Compete." Compete | Compete. Compete, Inc. Web. 1 Aug. 2011. <<http://www.compete.com/help/q1>>.

Nerlove, M. "Hedonic Price Functions and the Measurement of Preferences: The Case of Swedish Wine Consumers." *European Economic Review* 39.9 (1995): 1697-716.

Web. 15 July 2011. <<http://agecon2.tamu.edu/people/faculty/capps-oral/agec%20635/Readings/Hedonic%20Price%20Functions%20and%20the%20Measurement%20of%20Preferences%20The%20Case%20of%20Swedish%20Wine%20Consumers.pdf>>.