The Relationship Between Internet Marketing, Search Volume, and Product Sales

Honors Research Thesis

Presented in partial fulfillment of the requirements for graduation *with honors research distinction in Economics* in the undergraduate colleges of The Ohio State University

By

Nicholas Lincoln

The Ohio State University
June 2011

Project Advisor: Professor Richard Steckel, Department of Economics

Abstract

This paper determines whether internet advertisement, and a product's online popularity, as measured in search queries, can predict sales revenue. To test for correlations, the sales data, ad spending, and Google keyword search volume for Apple's iPod and iPhone was collected, and developed into a finite distributed lag model.

The model for the iPod's sales revenue shows that there is a strong seasonal effect on sales, and neither the internet popularity, as measured by Google searches, or advertisement spending has a statistically significant effect. The iPhone's sales revenue is shown to be significantly influenced by the internet popularity, and its lag. The iPhone's revenue is not significantly affected by advertising or seasonality. The results of this study could be used to determine the effectiveness of advertisement on consumer interest in a product, on the internet. Similar models could be able to determine whether Google search volume can predict the sales revenues of other products.

I. Introduction

Internet advertisement spending has been growing in recent years, and is projected to overtake television and radio advertisements within a decade. Market research by Experian has shown that over 60% of consumers of every age group research a product online before making a purchase, and about the same percentage compares prices online before making a purchase (Experian Inc., 2010). Many companies are now researching the market potential and predictive power of social media. Hewlett-Packard, for example, has demonstrated that it is possible to predict movie ticket sales with data from the social network Twitter (Asur and Huberman, 2010). Several other studies have looked at internet metrics, like the number of page visits, or click through rates, for improving ad effectiveness (Canny, Chen, and Pavlov, 2009; Caruso, Giuffrida, and Zarba, 2011). There have also been studies on Google metrics specifically, to investigate the effectiveness of search ads (Omidvar, Mirabi, and Shokri, 2011). Surprisingly however, there have been no studies on the use of the actual keyword search volume for a product, to predict its revenue. By understanding the relationship between search volume and product revenue, ad effectiveness could be improved.

This paper contends that the sales revenue of some products can be predicted by keyword search volume, like the Apple iPhone, while the sales of others, like the iPod, may not be. The possible factors responsible for these results will be discussed. This study also suggests that, contrary to most products, the ad spending has no significant impact on the sales revenue for these two products. The lack of correlation between ad spending and sales revenue will be examined, and the role of Apple's unique marketing practices will be revealed as the most probable explanation. Finally, the seasonality of iPod sales will be discussed, and the models for all these effects will be specified and reviewed.

II. Related Work

In 2010, Hewlett-Packard built a model to forecast box-office revenues, from the rate and sentiment of tweets on the social network Twitter, and found that the model could outperform market-based predictors. The study concluded that content on social media web sites, like Twitter, could be used to predict real world outcomes (Asur and Huberman, 2010). A key finding was that the pre-release promotions for movies were actually not predictive of their performance in the box office. The study defined pre-release promotions as hyperlinks to trailers and other promotional material, and link forwards from one user in the network to others. Tweet rate, rather than pre-release promotions, was found to be a better predictor of box-office success. The tweet rate was defined as, "the number of tweets referring to a particular movie per hour" (Asur and Huberman, 2010). High tweet rates corresponded to more successful films.

In the HP study, the tweet rate was a measure of the popularity of a movie on the social network Twitter. A distinction should be between the measure of popularity used in the HP study, the tweet rate, and the measure of popularity used in this study, the Google search volume. The HP study measured the tweet rate during a critical period, which was defined as one week before the movie was released, to two weeks after. The reason that HP only measured this critical period was the short time span that movies were shown in theaters. In contrast to the HP research, this study measures Google search volume over a 6-year time period, in which the products being observed sold continuously. Nevertheless, the concept of popularity affecting sales revenue is the same in both studies.

III. Data

The data for this study was collected from Apple's annual Form 10-K reports, and Google Insights for Search, for the 6-year time period between January 2004 and December 2010. All of the data is measured quarterly. Google search data was used because, at the time of this study, Google was the highest trafficked search engine on the web, so it would likely provide the best representation of internet popularity by search volume (Wornack, 2011). Apple products were chosen to be studied because of their popularity and brand awareness with consumers. The Apple iPod was released in 2001, so its search and revenue data was measurable from the beginning of the 6-year period. The iPhone was released in 2007, so its search and revenue data was zero at the beginning of the 6-year period. However, its statistical model was adjusted for this. Since the iPhone was released within the measured time period, it was a good indicator of how pre-release search volume was related to the sales revenue that the product generates.

To acquire the most accurate representation of popularity on the internet, Google AdWords was used to find the top twenty keyword alternatives to "iPod" and "iPhone." These highest searched keyword alternatives were incorporated into the search volume data, along with the model names of each iPod and iPhone released during the 6-year time interval. Duplicate keywords that appeared in the top keyword alternatives list, such as "new iPhone" and "iPhone new," were only included once.

Weekly search volume data was collected on the sets of keywords for both products. The weekly data was averaged to create quarterly data. Products that were not released until the middle of the study had values of zero for the search volume until their releases. To prevent this from skewing the data, these products were only included in the data once the first non-zero value was measured. They were included in the data from that point after, even if their search volumes went back to zero.

Google Insights for Search provided scaled data, so the search volumes were not measured in absolute terms. The scale put a value of 100 on the week that the highest search volume was recorded, and every other value was in proportion to that number. For example, a value of 50 would indicate that during that week, people were half as likely to search for the keyword or its derivatives, than during the week in which the value was 100. The advantage to using the scaled data was that the popularity of products with very different search volumes, could be still be compared. Scaling the data put it into proportions, so a search value of 23 for both the iPod and iPhone, for example, indicates that during that week, both products had the same relative popularity. The use of scaled data allows the impact of popularity on sales revenue to be measured in relative terms, irrespective of absolute search volume. A product with a high absolute search volume might have high sales revenue, but in relative terms, that product might not be as popular as a product with a smaller absolute search volume.

Quarterly sales revenue of the Apple products, and advertisement spending figures were taken from annual Form 10-K reports. The data for both of these variables was measured in millions of US dollars. Advertisement spending was collected as a whole figure for all mediums, rather than a specific figure for online advertisement. This allowed for the possibility that consumers could have learned about the product from any form of advertisement, and then searched for it online.

IV. Setting up the Hypothesis Test

This experiment can be broken into two components: the affect of advertising on internet popularity, and the affect of internet popularity on sales revenue. If both of these relationships are found to have correlations, then by the transitive property, advertising affects sales revenue, and should be included with popularity, in a combined model. There has been extensive research into how advertising affects sales revenue, and it has been concluded that current advertising influences current sales, as well as sales into the future (Weiss, 1995). Due to the possibility of current advertisements influencing future sales, a lagged term should be included in the model to measure its effect. While there is no research on the influence of internet popularity on sales, it is reasonable to suspect that current popularity, like advertising, has both present and future effects on sales. The inclusion of a lagged term for popularity can measure this effect.

The hypothesis to be tested is whether internet popularity influences sales revenue, and, if advertising is correlated with internet popularity, whether both advertising *and* internet popularity influence sales revenue. The null is that there are no correlations between advertisement spending, internet popularity, and sales revenue.

V. Developing the Model

Due to the current and future effects of advertising and popularity, a linear finite distributed lag model was developed to test the hypothesis. Let S stand for sales revenue at time t. Let P represent internet popularity, and A represent advertisement spending. Finally, let α be a constant, and let u represent all the unmeasured determinants of sales revenue. Writing the equation with P and A both lagged one quarter gives,

$$S_t = \alpha_0 + \beta_1 P_t + \beta_2 P_{t-1} + \delta_1 A_t + \delta_2 A_{t-1} + u_t. \quad (1)$$

This equation can be modified with a time trend to account for the increasing tendency of sales revenue for the iPod and the iPhone, which gives,

$$S_t = \alpha_0 + \beta_1 P_t + \beta_2 P_{t-1} + \delta_1 A_t + \delta_2 A_{t-1} + \gamma t + u_t. \quad (2)$$

Equation (2) can be further modified for the iPod, which exhibited strong seasonality from year to year (see figure 1). The inclusion of dummy variables for each quarter, minus one, leads to the final model for the iPod,

$$S_t = \alpha_0 + \beta_1 P_t + \beta_2 P_{t-1} + \delta_1 A_t + \delta_2 A_{t-1} + \lambda Q_2 + \lambda Q_3 + \lambda Q_4 + \gamma t + u_t. \quad (3)$$

Equations (2) and (3) are the regression equations used for the iPhone and the iPod, respectively. The parameters to these models were estimated by ordinary least squares. As expected, the inclusion of dummy variables increased the goodness of fit for the iPod model, at the cost of three degrees of freedom (see tables 1 and 2). To prevent the results from being spurious, an augmented Dickey-Fuller test was performed to determine if the processes were non-stationary. The models for both the iPod and the iPhone were determined to be stationary processes.

VI. Results

The results of the iPod regression give the following model for sales revenue:

$$\begin{aligned}&iPod\_\mathrm{Re}venue = -154.18 + 21.38(popularity_t) + 17.74(popularity_{t-1})\\&-11.77(ads_t) + 17.73(ads_{t-1}) - 1776.16(Q1) - 1617.64(Q2) - 1617.67(Q3)\\&+8.31_t + u_t\end{aligned} \tag{4}$$

The time of year was the only statistically significant factor that influenced the iPod's sales revenue. As figure 3 shows, sales in the fourth quarter of each fiscal year escalated, and then dropped off in the first quarter of the next year. The regression output shows that the three dummy variables were all statistically significant, and their negative coefficients reinforce the fact that sales revenue was lower during those quarters than in the fourth (see table 2). The strong relationship between the fourth fiscal quarter and sales revenue is understandable, considering that it coincides with the Christmas season. Surprisingly, neither the internet popularity nor the ad spending had a statistically significant effect on sales. The adjusted R^2 and the probability of the F-statistic imply a strong goodness of fit for the model.

The results of the iPhone regression give a similar model to equation (4), but without seasonality:

$$iPhone\_Revenue = -1405.05 + 81.45(popularity_t) + 108.27(popularity_{t-1}) + 17.91(ads_t) + 13.73(ads_{t-1}) - 167.79t + u_t \quad (5)$$

Unlike the iPod, the iPhone was found to be significantly affected by both the current and lagged internet popularity, as well as the time trend (see table 3). Specifically, if the current popularity on the web were to increase by one unit, as measured by the Google Insights scale, the iPhone's sales revenue would likely increase by $81.45 million. If the lagged popularity on the web were to increase by one unit, the iPhone's revenue would likely increase by $108.27 million. As with the iPod, the ad spending and lagged ad spending were surprisingly insignificant. The adjusted R^2 and the probability of the F-statistic imply a strong goodness of fit for this model too.

VII. Discussion

The most surprising discovery was that advertising had no effect on sales revenue for either product. The most probable explanation for this is Apple's unique marketing strategy. According to research by Advertising Age, Apple spent $28 million to advertise the iPod and its features, when the product was first released in the fourth quarter of 2001. Apple then reduced the ad spending to merely $4.4 million for all of 2002 (Bulik, Cuneo, and Johnson, 2007). The strategy was to allow consumers to try the product, and let news spread via word of mouth. It was successful because of the intense brand loyalty of Apple's customers, and the simple and aesthetically appealing packaging that easily distinguished the iPod from its competitors. The iPod was marketed as a cool and hip product, and this idea was reinforced by the iPod's limited retail outside of Apple stores, which gave it exclusivity (Barrile, 2006). Apple planned to increase advertisement only if sales started to fall, and since features were repeatedly added over time, sales remained high, even as the product aged (Bulik, Cuneo, and Johnson, 2007). Since the iPhone was released in 2007, it has been marketed by the same method. Although advertisement spending has increased in absolute value over time, it has remained low in proportion to sales revenue. For this reason, it makes sense that the ad spending might not be a good predictor of sales revenue.

Another surprising discovery was that the internet popularity had no effect on the iPod's sales revenue. The most probable explanation for this is that by 2004, the iPod had already been in stores for nearly three years. Consumers did not need to research the product, or compare its price with competitors, because the price remained relatively constant during those years, even as new features were added (Bulik, Cuneo, and Johnson, 2007). Google searches would likely have been done by consumers wishing to research the product, buy the product, or seek news stories about it. If the novelty of the iPod had worn off by 2004, search queries may have been triggered only by the release of new models or news about the product, rather than by consumers' desire to buy it.

The iPhone was strongly influenced by both the current quarter's internet popularity and the previous quarter's. The popularity of the iPhone reached a peak about a year after its release, and reached its highest point in late 2010 (figure 4). The most likely explanation for the 2010 spike was the announcement that a new iPhone would be released for the Verizon Wireless network service provider. Until then, AT&T had been the only wireless service provider that the iPhone was compatible with. By hacking the iPhone, it was possible for consumers to access other wireless networks. In fact, some of the highest searched keywords that Google Adwords reported for the iPhone included the phrase "iPhone hacked," so a Verizon compatible iPhone would have encouraged consumers to research the new product. The iPhone's sales reached high points at the same time as the internet popularity, and the relationship was specified by equation (5).

One explanation for the influence that the lagged popularity had on the iPhone, is that consumers not only researched the phone, but the service contracts it had with wireless providers. Due to these contracts, which typically last two years, consumers would be more likely to research and compare the prices of the iPhone and its competitors. This would take more time than simply researching the phone by itself. Service contracts could explain the difference between the lagged popularity of the iPhone being significant, and the lagged popularity of the iPod being insignificant.

VIII. Conclusion

This paper has shown that the sales revenue of the Apple iPhone may be able to be predicted by its internet search popularity, as measured by Google search volume. The number of Google searches could indicate how interested consumers are in the product. Sales of the iPod were determined to be heavily influenced by the time of year, and were not predictable by internet popularity. While sales spiked during the Christmas season, they were not correlated with increased Google searches, possibly because of the length of time that the product has been on the market. Finally, Apple's advertisement spending was shown to be a poor predictor of sales revenue for both the iPod and the iPhone, possibly because of Apple's unique marketing strategy that promotes brand loyalty, and advertisement via word of mouth.

The purpose of this research was to determine whether the internet popularity of a product could predict its sales revenue, and if the amount of advertisement spending had any effect on popularity and sales. The results were inconclusive, as one product's popularity was shown to be correlated with sales, but the other product's popularity was not. Similar models could be developed for other products to test this hypothesis on a larger scale. If internet popularity, as measured in keyword search query volume, is determined to be a good predictor of sales revenue, then advertisers could improve the effectiveness of their ads.

References

Asur, Sitaram, and Bernardo A. Huberman. *Predicting the Future with Social Media*. Hewlett-Packard Development Company, L.P., 2010.

Barrile, Steve. *Ingredients for the Success of the Apple iPod: Marketing*. Melbourne: Warringal Publications, 2006.

Bulik, Beth Snyder, Alice Z. Cuneo, and Bradley Johnson. "Apple's Not Thinking Different." *Advertising Age* 78.26 (2007): *Academic Search Complete*. EBSCO. Web. 5 Mar. 2011.

Canny, John F., Ye Chen, and Dmitry Pavlov. *Large-Scale Behavioral Targeting*. Sunnyvale: Yahoo! Labs, 2009.

Caruso, Fabrizio, Giovanni Giuffrida, and Calogero Zarba. Behavioral Online Advertising. Ithaca: Cornell University Press, 2011.

Experian Inc. *The 2010 Digital Marketer: Trend and Benchmark Report*. Experian Marketing Services, February 2010.

Hanssens, Dominique M., and Amit Joshi. *Advertising Spending, Competition and Stock Return*. University of Central Florida, 2008.

Omidvar, Mohammad Amin, Vahid Reza Mirabi, and Narjes Shokry. "Analyzing the Impact of Visitors on Page Views with Google Analytics." *International Journal of Web & Semantic Technology* Vol.2, No.1, January 2011.

Weiss, Doyle L. *Does TV Advertising Really Affect Sales? The Role of Measures, Models, and Data Aggregation*. Journal of Advertising. 22 Sept. 1995.

Wornack, Brian. "Google, Microsoft Added Internet Search Market Share in December." Bloomberg. 13 Jan. 2011.

Table 1

Dependent Variable: iPod Revenue
Method: Least Squares

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Constant	-527.2871	520.9043	-1.012253	0.3224
iPod Popularity	101.8464	29.51803	3.450312	0.0023
Lagged iPod Popularity	3.155695	29.97599	0.105274	0.9171
Ad Spending	19.19611	11.07773	1.732855	0.0971
Lagged Ad Spending	-18.31012	13.46137	-1.360198	0.1875

R-squared	0.572819	Mean dependent var	1861.111
Adjusted R-squared	0.495150	S.D. dependent var	933.6804
S.E. of regression	663.4062	Akaike info criterion	15.99823
Sum squared resid	9682373.	Schwarz criterion	16.23820
Log likelihood	-210.9761	Hannan-Quinn criter.	16.06958
F-statistic	7.375104	Durbin-Watson stat	1.526380
Prob(F-statistic)	0.000631		

Table 2

Dependent Variable: iPod Revenue
Method: Least Squares
Included observations: 27 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Constant	1621.981	656.3727	2.471128	0.0237
iPod Popularity	21.37893	28.93144	0.738951	0.4695
iPod Lagged Popularity	17.74430	26.75992	0.663092	0.5157
Ad Spending	-11.77393	10.34664	-1.137948	0.2701
Lagged Ad Spending	17.73208	13.72590	1.291870	0.2127
Q1	-1776.159	450.3283	-3.944143	0.0010
Q2	-1617.635	369.9732	-4.372303	0.0004
Q3	-1617.674	290.4998	-5.568588	0.0000
Time Trend	8.313065	39.26956	0.211692	0.8347

R-squared	0.849706	Mean dependent var	1861.111
Adjusted R-squared	0.782909	S.D. dependent var	933.6804
S.E. of regression	435.0297	Akaike info criterion	15.24991
Sum squared resid	3406515.	Schwarz criterion	15.68185
Log likelihood	-196.8737	Hannan-Quinn criter.	15.37835
F-statistic	12.72070	Durbin-Watson stat	1.004298
Prob(F-statistic)	0.000006		

Table 3

Dependent Variable: iPhone Revenue
Method: Least Squares
Included observations: 27 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Constant	-1405.053	671.4158	-2.092673	0.0487
iPhone Popularity	81.44566	29.35410	2.774593	0.0114
iPhone Lagged Pop.	108.2701	32.91390	3.289495	0.0035
Ad Spending	17.90528	11.80102	1.517265	0.1441
Lagged Ad Spending	13.72669	14.13932	0.970817	0.3427
Time Trend	-167.7868	55.73355	-3.010516	0.0067

R-squared	0.961897	Mean dependent var	1918.074
Adjusted R-squared	0.952825	S.D. dependent var	2911.919
S.E. of regression	632.4664	Akaike info criterion	15.93026
Sum squared resid	8400288.	Schwarz criterion	16.21822
Log likelihood	-209.0585	Hannan-Quinn criter.	16.01589
F-statistic	106.0268	Durbin-Watson stat	2.689992
Prob(F-statistic)	0.000000		

Figure 1

iPod sales revenue regression output corresponding to table 1

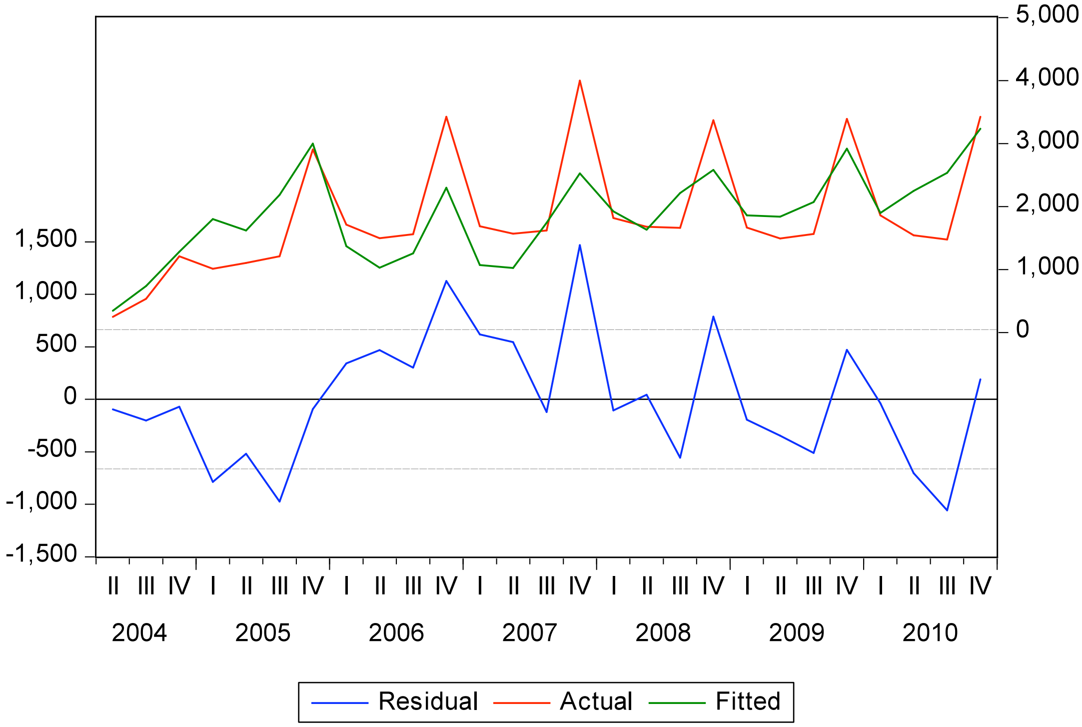


Figure 2

iPod sales revenue regression output corresponding to table 2

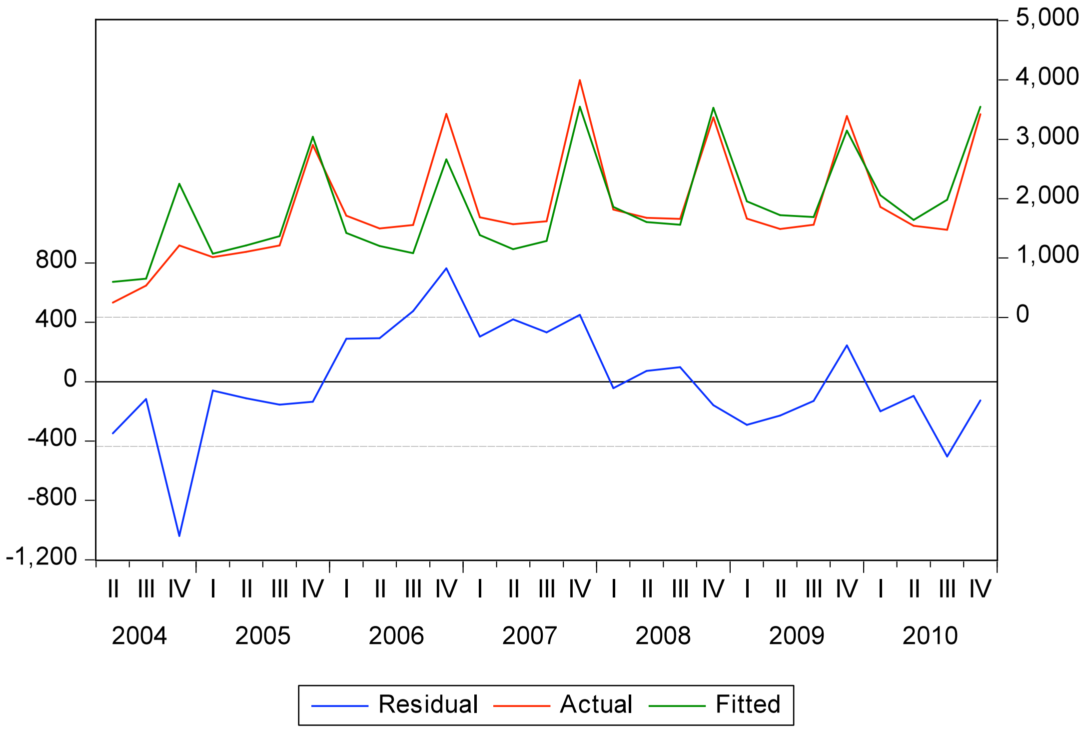


Figure 3

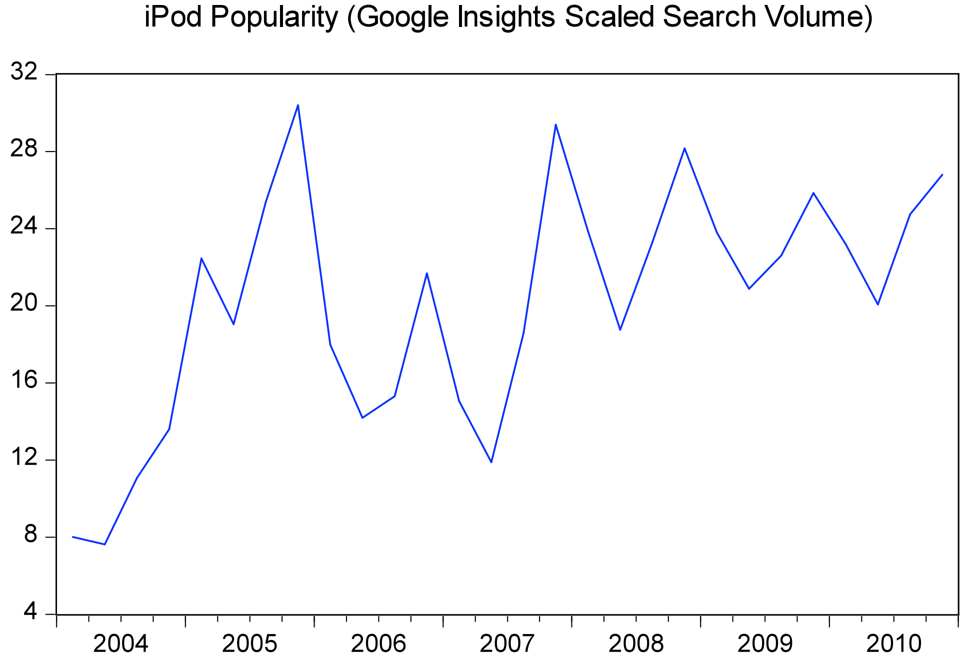

iPod Popularity (Google Insights Scaled Search Volume)
32
28
24
20
16
12
8
4
2004
2005
2006
2007
2008
2009
2010

Figure 4

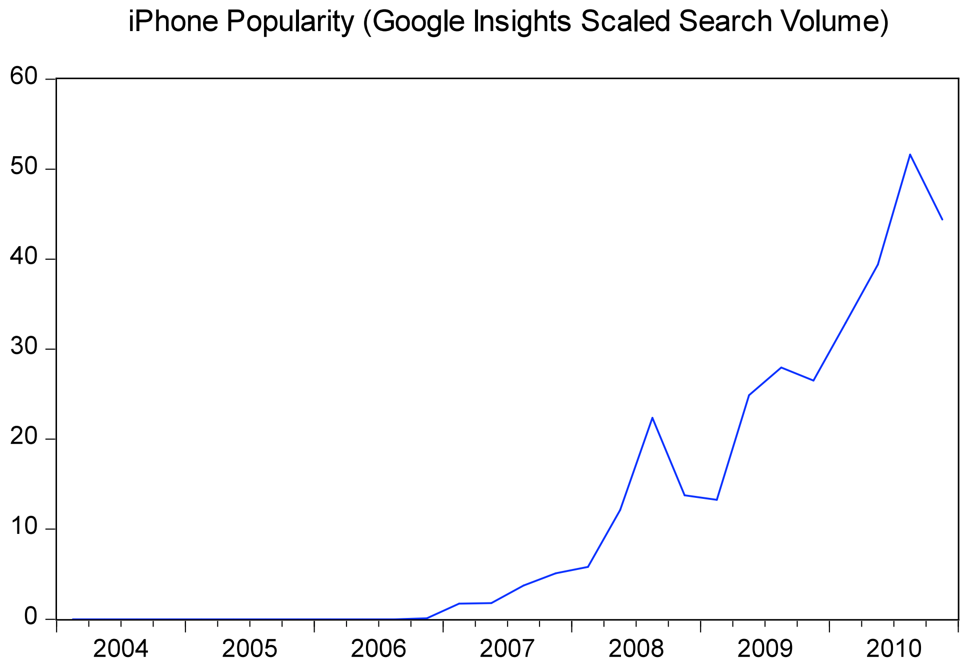
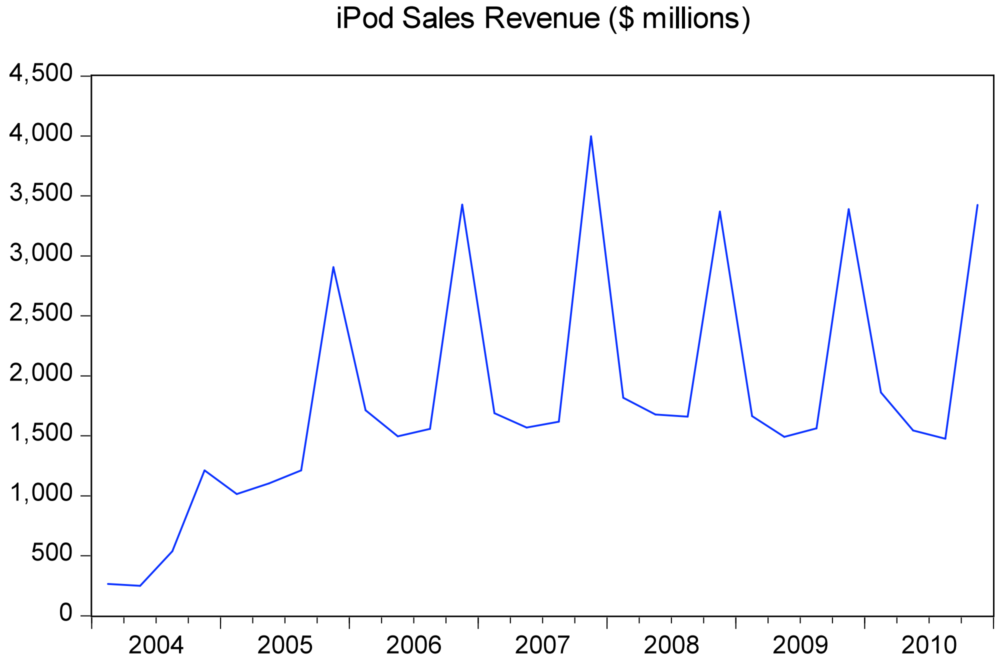

iPhone Popularity (Google Insights Scaled Search Volume)
60
50
40
30
20
10
0
2004
2005
2006
2007
2008
2009
2010

Figure 5



iPod Sales Revenue ($ millions)
4,500
4,000
3,500
3,000
2,500
2,000
1,500
1,000
500
0
2004
2005
2006
2007
2008
2009
2010

Figure 6

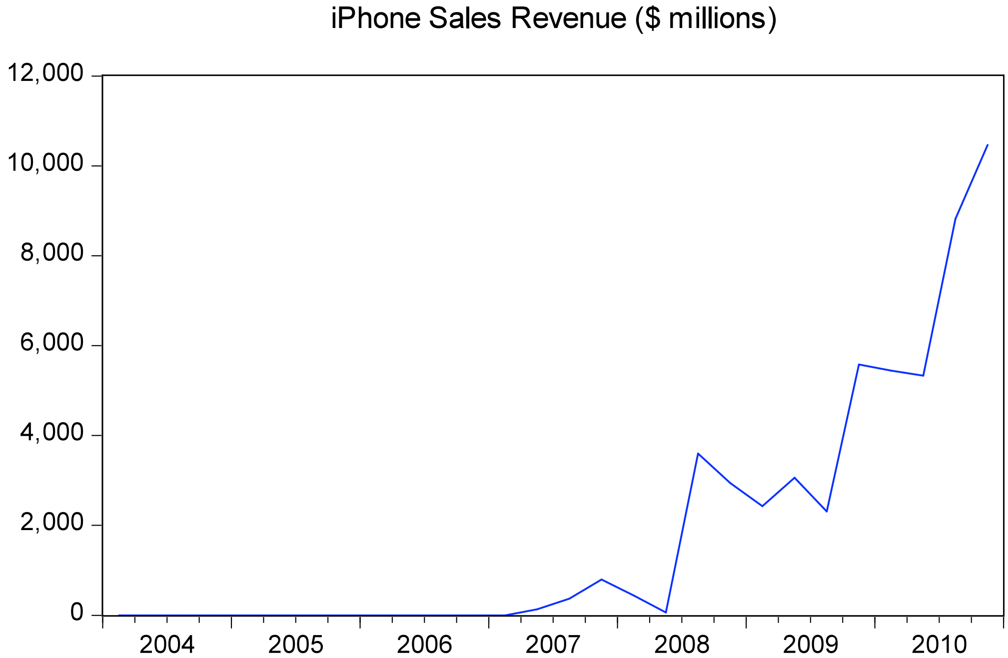
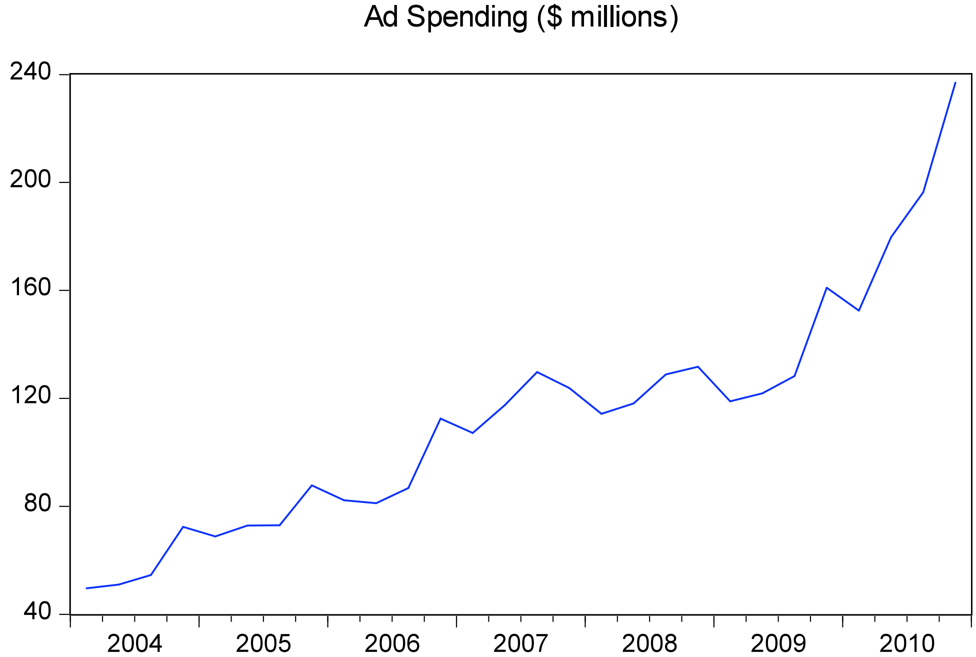

iPhone Sales Revenue ($ millions)
12,000
10,000
8,000
6,000
4,000
2,000
0
2004
2005
2006
2007
2008
2009
2010

Figure 7


Ad Spending ($ millions)
240
200
160
120
80
40
2004
2005
2006
2007
2008
2009
2010