

Using Internet search volumes for predicting product sales

Client

Nokia Markets, Strategy and Operational Excellence

Project Group

Joonas Ollila (Project manager)

Rasmus Hotakainen

Andreas Hübner

Susanna Siitonen

Pyry Ävist



Aalto University, School of Science and Technology
Mat-2.4177 Seminar on Case Studies in Operations Research
May 4, 2010

Contents

1	Introduction	3
2	Data sources	3
2.1	Comparison of data sources for search volumes data	3
2.2	Google search volume data sources	4
2.2.1	Google Insight for Search vs. Google Trends	4
2.2.2	Accuracy of trends data	5
2.3	Selecting the right search queries	5
2.4	Countries and products	6
2.5	Sales data	7
2.6	Data collection and conversion	7
3	Modeling search trends and product sales	8
3.1	Mathematical description of the model	8
3.2	Modeling Framework	8
3.3	Model validation	9
3.4	Analysis	9
4	Results	10
4.1	A product with good trends data	10
4.2	A product with bad trends data	11
4.3	General results	11
4.4	Summary of results	14
5	Discussion	15
5.1	Change in dynamics	15
5.2	Estimation of sales at product launch	15
5.3	Using other models than dynamic regression models	15
5.4	Other ways to improve the estimation	15
6	Conclusion	15

1 Introduction

The development of Internet search volumes can be examined with various services on the Internet. Some of them provide information about the search history in a given geographical region and enable the user to compare search volumes of different keywords. Thus it is possible to examine rising, falling and regular trends.

In this project we examine the usefulness of the history of internet search volumes (also referred as trends or search trends) for predicting product sales in the field of telecommunications. Our hypothesis is:

H 1. *There is a correlation between the search volumes linked to a cellular device and the actual sales volumes of this product.*

This hypothesis we test with 25 products from various manufacturers in 22 countries around the world. Some of the products have already been sold a longer time, some of them are just launched. We have in other words an extensive sample of sales data to investigate. Our key research objectives are:

1. To investigate which data source is the best for our purposes
2. To find a model that can predict at least the current monthly sales
3. To validate the chosen model and decide if it is usable in real life scenarios

Our task is to build a model for predicting product sales using Internet search volumes as the main tool. We test how well the products' sales correlate with Internet search volumes.

Our investigations deal with very recent developments in technology, hence there is only little existing work available. Shimshoni et al. (2009) examined the predictability of consumer behavior towards searches on Google. They argue that while there is a good portion of search queries and categories that show regular patterns or even seasonality, there are far more items that can not be predicted. Regarding our work the category "Telecommunications" showed only small predictability ratio and deviations were likely. Bear in mind that the authors investigated whole categories only. As we aim for single search queries we can not count at all on eventual repeated patterns in the search volumes although we can expect a product-life cycle-like development.

In Choi and Varian (2009) search volumes data from distinct search categories is used to estimate economic development. Regression models are used to forecast automotive sales in the US. Both sales and search volume data are included in the model. Again the authors focus on the big picture and do not investigate on specific search queries. Nevertheless we adopt and adapt the proposed models for initial analysis.

The structure of this report is as follows: In section 2 we select the most suitable data sources, examine search query structure and give insight into which countries and products we investigate. Section 3 contains the mathematical basis for our modeling work and deals with the used framework. In 4 we present our finding and deduct rules when models using search volumes data work and when not. Section 5 identifies further steps that could be taken to increase accuracy and soundness of search volumes based models and discusses those approaches. The conclusion is in section 6.

2 Data sources

2.1 Comparison of data sources for search volumes data

We rely on public and free of charge online data sources that provide data on user behavior towards Internet search over time. The sources should be used for searching for devices in our scope of research and provide us with formatted data we can use to train and evaluate our models. Our investigations brought up multiple potential sources. We name and evaluate those sources in Table 1.

After taking technical and substantial aspects into consideration, we come to the conclusion that Google Trends (respectively Google Insight for Search) will be the only source for the historical search volume data we will collect and analyze.

Table 1: Potential sources for trends data

Source	Description	Included in study
Google Trends (GTrends)	Google Trends provides detailed data on worldwide and local search volumes. Google is the World's leading Internet search engine and thus relevant for our research. The queries for search volume data can be customized by parameters like region and period. Results can be downloaded in a structured, machine readable format. http://google.com/trends	✓
Google Insight for Search (GI4S)	I4S is a more sophisticated interface to Google search volumes data and aims to advanced users and researchers. The main difference is that the search term categories Google defined can be explored. Furthermore the parameters can be set in a more detailed manner: Time periods can be set individually and regions can be adjusted more precisely. http://google.com/insights/search	✓most suitable for analysis
Bing Xrank	The Bing Xrank feature is the search volumes discovery tool for the Microsoft search engine Bing. It is similar to GTrends and GI4S but no results can be downloaded. http://bing.com/xrank	
Baidu Index	The same like Bing Xrank and GTrends/GI4S but as Baidu is a Internet search engine with a great market share in China it would be very interesting if data could be pulled from it. Unfortunately no formatted results can be downloaded and thus, this services is not appropriate for our project. http://index.baidu.com	
Yahoo Buzz	Yahoo Buzz shows trending search terms and topics for the Yahoo search engine. Its scope is very limited and it is aimed at casual users. http://buzz.yahoo.com	
Ebay Pulse	Ebay Pulse informs users about recent trends concerning search terms and auctions on Ebay. Search terms on Ebay may reflect customers demand for products in a good manner as it is actually a marketplace. However, there is no chance of viewing historic data or downloading formatted data sets. If Ebay introduces a services like this in the future, it may become a sound source. http://pulse.ebay.com	
Twitter Trends	Twitters search feature browses tweets in real time for given search queries. The data can be accessed via the API. Searching for tweets related to a certain region is possible but limited to the radius around a single spot. After checking for results on all given product names it became clear that available data is insufficient and therefore Twitter will not be chosen a source. http://search.twitter.com	

2.2 Google search volume data sources

2.2.1 Google Insight for Search vs. Google Trends

Google hosts two different services that allow exploring their user behavior towards search volumes¹. Those are GTrends and GI4S. The trends reports can be downloaded as CSV² files in regions worldwide and for each country and item. In the help section for GI4S it is said that both tools rely on the same data³. That is logical as Google provides only one front-end to its

¹The Terms of Use (see <http://google.com/intl/en/trends/terms.html> and <http://google.com/intl/en/insights/terms.html>) grant the usage of Google Trends Google Insight for Search for 'internal purposes' and 'research'. Section 5.3 of the Google Terms of Service (see <http://google.com/accounts/TOS>) prohibits the access to services provided by Google by any other interface than the one provided by Google itself.

²Comma separated values

³<http://google.com/support/insights/bin/answer.py?hl=en&answer=117683>

search services. Nevertheless we found terms for certain regions⁴ where GI4S gives results while GTrends states that those ‘terms do not have enough search volume to show graphs’⁵. These inconsistencies could be caused by different approximations and thresholds applied on the raw search data. In order to get a better view on the gaps between GTrends and GI4S we collected samples for three different terms and evaluated the correlation. After first retrieving data from GTrends we realized that compared to GI4S the collected data was very poor and contained many zero values. This observation lead us to the conclusion to use only data from GI4S for our research.

2.2.2 Accuracy of trends data

We discovered that some kind of noise is added to the data. The GI4S web page states ‘that several approximations are used when computing these results’⁶. Accordingly, those inconsistencies could also be caused by incomplete data. In order to get a better picture on the variability of those time series for a single search term, we retrieved a report on a single query on five consecutive days. Figure 1 gives the plot for this trends. Note that for most days the graphs are pretty similar except for the data collected on the 26th of February 2010 (the dashed line).

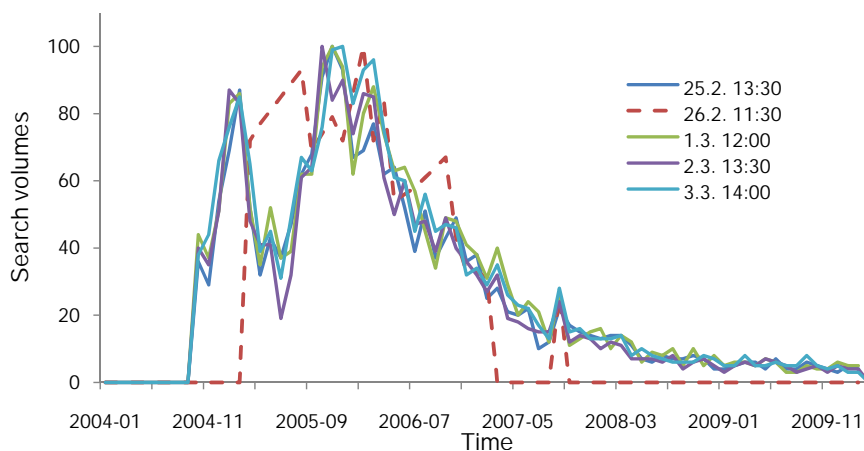


Figure 1: Trends data for *Motorola Razr V3* collected on different days

There is information on this issue provided⁷ in the GI4S help but in our opinion it does not explain the daily differences and even less the one abnormal result.

2.3 Selecting the right search queries

Names of the handheld devices we investigate on can be decomposed to the parts brand (manufacturer), model name or code and nickname.

Samsung
S5230
Star
 brand model name / code nickname

We examined the quality of search volumes results retrieved for different combinations of those parts and derived simple rules how they are used best to build up a search query. Table 2 presents our findings.

Google Insight for Search provides data on how Internet users have used certain search terms during a specific period of time. Different countries naturally have different languages and the products can be named differently. That is in different countries people use different queries for searching the same thing.

We want to know how people in different countries search for the products in our scope. Viewing the results in Google Search⁸ gives a good vision of what other names the product may

⁴e.g. *iphone 3g 8gb* in region France

⁵<http://google.com/trends?q=iphone+3g+8gb&ctab=0&geo=fr&date=all&sort=0> - Retrieving date: 2010-16-04

⁶<http://google.com/insights/search/>

⁷<http://google.com/support/insights/bin/answer.py?hl=en&answer=165883>

⁸<http://google.com/search>

Table 2: Rules for search queries

Pre-condition		Deducted rule
Product has a unique name worldwide (or in certain areas) / code	⇒	use the name / code
Product does not have a unique name or code and does not have nicknames	⇒	add brand
Product has multiple nicknames and a code	Searching worldwide (or areas with many names used)	⇒ combine code and nicknames with +-mark and add brand when needed ("nickname1" + "nickname2")
	Searching small area with only one nickname used	⇒ use that nickname straight and combine with brand if needed

have. To make sure that we have covered all the nicknames, a quick check in Wikipedia⁹ is helpful.

Samsung S5230 Star and *LG KP500 Cookie* are good examples of products that cause problems then retrieving search volumes on them. When searching with only the model name, the results are accurate but there are fewer hits and thereby the results do not always show. Using only the official nicknames is also problematic, since *Star* and *Cookie* usually describe different objects from phones. Adding the brand before the nicknames provides the best results in these cases. However in some countries the models have other nicknames, these can be found from Wikipedia. For example *Samsung S5230* is also called *Tocco lite*, *Samsung Star* and *Samsung Avila*.

All in all, when investigating the whole world, use the official nickname (*Samsung Star*). When a single country is in question, use the nickname in that country (e.g. *Samsung Avila* in Poland).

For less problematic products the procedure is to use the brand name together with the model name, or the model name alone. The difference between those two methods is illustrated in Figure 2. The pattern is very similar for both the search methods, but using only the product model name one gets a little bit more search volume since the product model name (for example the *S5230*) is not necessarily unique and can be, and often is used in other products or simply in various web documents. This is especially true with product model names consisting only of names.

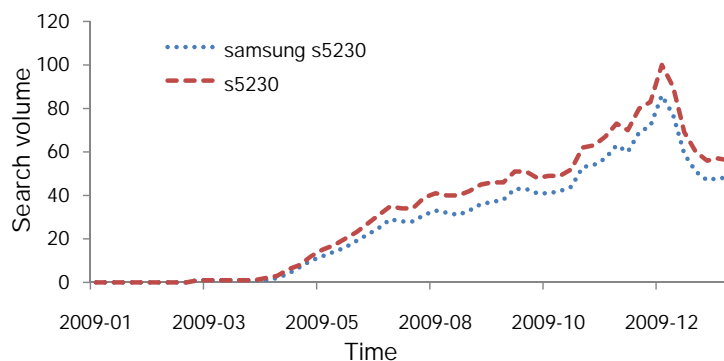


Figure 2: Example of Google Insight for Search result with product model and product model + brand

2.4 Countries and products

We received a list from our client of 22 countries to use. The countries are in alphabetical order: Argentina, China, Egypt, France, Germany, India, Indonesia, Iran, Italy, Nigeria, Pakistan, Philippines, Poland, Russia, Saudi Arabia, Spain, South Africa, Taiwan, Turkey, United Arab Emirates, United Kingdom and Vietnam.

⁹<http://wikipedia.org>

The selected countries are from all over the World, from several continents and each with very different financial and market development stages. The reason to select such a diverse list of countries is to preserve relevance in our work and find out if their trend-sale behavior is the same or if it differs somehow depending on the location or other measurement factors of the countries.

The actual investigation is made on 25 different handheld devices from 7 different manufacturers in the mobile and smart phone sector. The list of devices was provided by our client and it contains products that were launched during the last six years by major mobile phone manufacturers. The product list includes devices from various categories, including multimedia, business, smart phone, fashion and others. We are referring to the products as p0001-p0025 and p0026 is combined for multiple products from one brand as it will give more accurate trends data than a single product would provide on its own.

2.5 Sales data

We received the sales data from our client. The data covered the time period starting from 2006 and ending at the end of 2009. The data also contained monthly sales for the products in question; some of the products had sales data ranging up to four years and some products had data for only a few months. The number of available products varied greatly from country to country. For example, the data contained only 11 products that were sold in China, whereas practically all the products were available for purchase in Germany. Other countries that sold considerably less products, only 16 to 20, were Argentina, Egypt, India and Taiwan. The rest of the countries on the list all had over 20 of the products available. Also, as a minor curiosity, two of the products were sold in only about six countries and three other products were sold in only 14 of the 22 countries listed. Apart from the aforementioned anomalies, the rest of the products had good sales rates and only a few countries did not sell them at all. Curiously enough, only three products were sold in all of the listed countries. In total, there are 91 country-product combinations that have no sales data whatsoever.

One problem with the provided sales data was that there were some bad data points. Those bad data points in question had the same product name as the other, but a different product code. There were a total of six of these weird type data points; three in Italy, two in Argentina and one in South Africa. In some cases the same month already had other sales and they differed greatly from these and there was one case where there was not any data at all for the other product code. Our initial thought on the matter was to combine them to the other product code but quickly we reached the conclusion that it is better to leave them out completely because they might cause unforeseen problems at a later stage and make our sales data erroneous.

2.6 Data collection and conversion

While the sales data was provided in a Microsoft Excel file and had to be manually exported to format that could be read by the modeling framework the trends data was entirely downloaded manually from 11th of March to 29th of March 2010.

Out of 572 possible trends file 383 could be downloaded during the period from middle to the end of March 2010. The ones left out had no data at all so they were ignored right from the beginning. Out of those 383 items, files that were accidentally downloaded or corrupted files were filtered and we ended up with 293 trends respectively sales files for our investigations.

To accomplish the preparation of the trends data fast we prepared two CLI scripts that helped to run the tasks as a batch process.

In four steps sales and trends data files were filtered and adjusted:

1. Parsing trends data section out of the downloaded trends data file.
2. Reformatting trends data records
3. Converting weekly search volumes data to monthly.
4. Tailoring sales and trends data file so both had the identical number of data points.

Due to the fact that the sales data only contained monthly data points we converted the partly weekly trends data to monthly points. For this process we made sure that even accuracy was left the numbers for each month were properly calculated. Weeks which lied partly inside month a and month b (i.e. the first and last weeks of a month) were split and their value was weighted by the number of days belonging to the particular month lying in the week.

During step 4 overlapping data points were cut of so that both sales and trends data for a single product in a country started and ended with the same data.

3 Modeling search trends and product sales

Dynamic regression models and neural networks both use endogenous and exogenous data to estimate the dependent variable. Thus they are natural starting points for a study. We use a dynamic regression model to estimate the sales volumes, since its parameters are easier to interpret than those of a neural network model. The sales is the endogenous and historic variable while the Google search volumes data is the exogenous variable.

The time series we work with are not stationary, thus autoregressive (AR), moving average (MA) or combined ARMA-models are not the first choice. Product life cycles have some seasonality (peaks at Christmas almost always), however the variations are very big. Since a typical life cycle is less than four years this seasonality is difficult to model.

3.1 Mathematical description of the model

A dynamic regression model is a normal regression model where one or more of the explanatory variables are lags of the measured variable. We use also some normal regression models with trends values as explanatory variables. A model where only lags of the measured variable are used as explanatory variables becomes an AR model. A standard dynamic regression model looks like this:

$$y_t = \alpha_0 x_t + \alpha_1 x_{t-1} + \dots + \alpha_n x_{t-n} + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_m y_{t-m}, \quad (1)$$

where y_t is the examined (explained) variable at time t , $x_t \dots x_{t-n}$ and $y_{t-1} \dots y_{t-m}$ are the explanators. The coefficients $\alpha_0 \dots \alpha_n$ and $\beta_1 \dots \beta_m$ are estimated from the data.

The model in (1) measures in other words the value of variable y at time t . Our dynamic regression models look like this:

$$\text{Sales}_t = \alpha_0 \text{Trends}_t + \alpha_1 \text{Trends}_{t-1} + \alpha_2 \text{Trends}_{t-2} + \beta_1 \text{Sales}_{t-1} + \beta_2 \text{Sales}_{t-2}. \quad (2)$$

Models derived from (2) measure the sales volume at time t . We use different variations of this model, that is we predetermine the values of some coefficients α or β to be zero.

To measure the goodness of our models, we use a coefficient of determination for the prediction value. This measure of goodness (hereafter M) is calculated in the following way:

$$M = 1 - \frac{SSE}{SST} \quad (3)$$

where

$$SSE = \sum_{i=1}^T e_i^2 \quad (4)$$

and

$$SST = \sum_{i=1}^T (\text{Actual}_i - \overline{\text{Actual}})^2. \quad (5)$$

In (4) e_i is the error or the difference between actual sales and predicted sales. $\overline{\text{Actual}}$ is the mean value of the actual values. SSE and SST stand for square sum of errors and square sum total. The index i is the data point i and T is amount of data points which is the amount of months. Other measurements to measure goodness of our models such as the maximum error of the predictions could also be used but as we only want to compare models with each other the M value works fine.

3.2 Modeling Framework

Model Framework was implemented to provide a tool for data preprocessing and analysis for our project. It implements the mathematical models explained in 3.1. The Need for such tool rose from the sheer amount of datasets we had to analyze. With 22 countries, 25 products and multiple model variations we needed a tool in which we could quickly make changes to model parameters

and then run the entire data through it. Our final runs had approximately 5274 different model, country and product combinations. Using the framework we could produce results in less than an hour.

Main features of the analysis framework are the following:

- Input of data from our retrieval tools
- Data preprocessing
 - Creating new variables by delaying other variables (e.g. adding 2 months delayed version of sales)
 - Selecting target variable and variables which are used to explain the target variable
- Dynamic regression model parameter estimation
- Model validation
 - Sweeping validation where model is trained with values from $0, k$ and then the model is used to predict target sample $k + 1$. After that model is trained using values $1, k + 1$ and target sample $k + 2$ is predicted and so on.
 - Growing validation which is almost identical to the sweeping validation, but where we use first $0, k$ samples then $0, k + 1, 0, k + 2...$ thus it grows the dataset used for training in each iteration
- Output
 - Framework plots the results of validation runs and optionally outputs key performance indicators to CSV file

3.3 Model validation

The customary way to validate time series models is to divide the series in two parts. The first part containing 80% (or 90% or some other percentage) of the data is used to estimate the parameters of the model and the rest is used for testing and validating the model. This kind of validation does not work in our case, since the dynamics of the data change in time (due to different phases of the product life cycle). To get better results, we use n sequential data points to estimate model parameters, and then evaluate the next data point. Our n varies between 1 and 15. This idea gets pretty close to the customary 80-20 -procedure.

3.4 Analysis

The analysis of the data and the models was carried out in three steps:

- Step 1: Analyze some products in some countries to get a rough look on the data.
- Step 2: Analyze all products in all countries deeply to get detailed information about models.
- Step 3: Compile the results from Step 2 and represent them in tables and graphs.

The idea behind the first step was to find similarities between different brands and countries, to realize what challenges we had to face and to decide how the deeper analysis would be carried out. The analysis was done by training the models for the whole datasets in order to get results how the models behave for the whole product life cycle. We focused on the R^2 values of the models.

The second step of the analysis was carried out by using different sweep lengths to train the models and then try to predict the next months sales. By sweep length we mean how many values are used for estimating the model parameters. For example a model with sweep length of 2 uses 2 values to estimate the parameters. The reason for choosing the validation strategy based on sweeping instead of using the whole dataset to train the model was to get better results of how well the sales could be predicted using only the data that is available in the current state of the life cycle.

The third step of the analysis was carried out by choosing which models and sweeplengths we would take into account when compiling our results from Step 2. We would create a CSV file with the results in all the countries in order to have all the important results in a single file. Then we would calculate averages of M values and number of models with certain M values and represent the results in tables and graphs.

4 Results

In this section we present the results from the analysis. The results are presented from all three steps of the analysis. The focus will however be on the final results of the third step of the analysis.

According to the deep analysis in Step 2 with all the country-brand combinations we got some interesting results. One result was that models that only included current trends data or trends data with different lags were better than models that included both sales and trends. Especially in the beginning of the product life cycle where sales start it is wise to use a model that only is estimated from trends to avoid the model from predicting very high values. In the end where the sales and trends are stabilized a model that only uses sales lags could sometimes be better than using only trend and its lags. Overall the choosing of the model highly depends on the data quality of the trends. If the quality is bad all models are pretty bad, if the trends starts before the sales it is wise to use a model which only uses trends. In most situations a model with only trends is the best one, but there are problems with all models regarding the model dynamics at least at some point of the life cycle.

The short sweeplengths of 2 or 3 seemed better than long ones such as 15.

According to the results of the deep analysis we had to decide which models and country-product combinations we would take to the result CSV file. We decided that we would take the models presented in Table 3.

Table 3: Model name and type

Model name	Model type
Model 1	$Sales_t = \alpha_0 Trends_t$
Model 2	$Sales_t = \alpha_0 Trends_{t-1}$
Model 3	$Sales_t = \alpha_0 Trends_t + \alpha_1 Trends_{t-1}$
Model 4	$Sales_t = \alpha_0 Trends_t + \beta_1 Sales_{t-1}$
Model 5	$Sales_t = \alpha_1 Sales_{t-1}$
Reference Model	$Sales_t = Sales_{t-1}$

As can be seen from Table 3 we have three models with only trends and/or its lags, we have one model with trends and sales lagged one month and a model with only sales lagged one month. We will hereafter call the models as represented in Table 3. The sweeplengths we chose was the short ones 2 and 3 and the long one 15. The likely reason is that the product sales dynamics varies significantly throughout the product life-cycle phases. Thus, a short sweeplength captures best the dynamics of each phase. The reason for choosing the models and sweeplengths in such a way was to prove our results from the manual analysis. We used all the countries and brands to get results for different country-brand combinations. We also added a model called the Reference Model to the analysis. The Reference Model predicts that the sales next month is the same as the month before. The reason for adding such a simple model was to compare other models against it.

Regarding the country and brand combinations we found out that the quality of the data varies a lot and there are both models with good M values and models with bad M values.

4.1 A product with good trends data

In the Figure 3 is a product with good trends data. The product is p0018 in Italy. Model 1 was used for this particular example. The sweeplength used was 2. The model goodness value M for the model was 0.90 which is a good value.

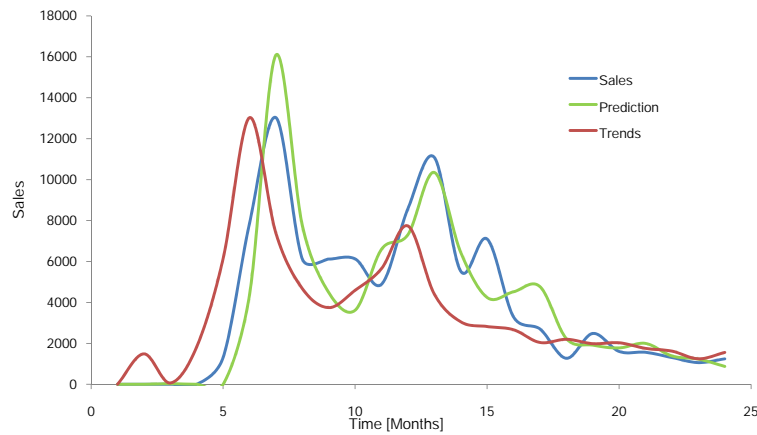


Figure 3: An example of good data: p0018 in Italy

From the Figure 3 we can see that the model predicts the sales very well. The sales have two peaks which the model can predict. The model also describes the dynamics in the sales well. The trends data for the model is very good and the sales lags by one month during the whole life cycle.

4.2 A product with bad trends data

In the Figure 4 is a product with bad trends data. The product is p0024 in United Arab Emirates. Model and sweplength are the same as in section 4.1. The model goodness value M for the model was 0 which means that the model does not describe the changes in the sales at all.

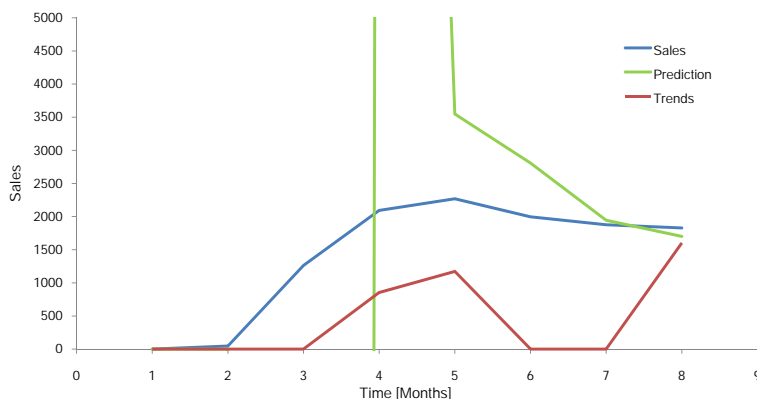


Figure 4: An example of bad data: p0024 in United Arab Emirates

From the Figure 4 we can see that the model has a high peak in the beginning which affects the whole model. The Sales value of the peak is much higher than the actual sales. The trends data is very bad and there is only data from a few months. The trends data does not follow the sales data at all and the model takes in garbage and gives out garbage.

4.3 General results

The first result we got was when we calculated the average M values for all the models with different country-brand combinations. The results are presented in the Figure 5

The numbers in Figure 5 represents the quality of the data in the country-brand. Number 1 means that the average Model goodness, M was over 0.3. The number zero reflects average M values between 0.15 and 0.30 while the value -1 means that the average M was under 0.15. The countries and brands are arranged so that the countries with better data are higher and the brands with better data are at the left. As you can see from the table most countries in Europe have good data while countries with emerging markets have worse. The brands with the best data are Apple

Country Brand	Apple	Sony Ericsson	Blackberry	LG	Samsung	Motorola	Vodafone
Germany	1	1	1	1	0	-1	-1
France	1	1	1	0	0	-1	-1
Italy	1	0	1	1	0	-1	-1
Indonesia	1	0	1	0	0	-1	-1
United Kingdom	0	1	1	0	-1	-1	-1
Spain	1	1	0	0	-1	-1	-1
Poland	0	1	1	0	-1	-1	-1
India	0	0	1	1	-1	-1	-1
South Africa	1	0	0	1	-1	-1	-1
Philippines	1	1	-1	-1	-1	1	-1
Turkey	1	0	0	-1	0	0	-1
Russia	0	1	-1	0	0	-1	-1
Taiwan	1	1	0	-1	-1	-1	-1
Vietnam	1	-1	0	0	-1	-1	-1
Saudi Arabia	1	0	-1	-1	-1	-1	-1
United Arab Emirates	1	-1	-1	-1	0	-1	-1
China	0	-1	-1	1	-1	-1	-1
Iran	1	-1	-1	-1	-1	-1	-1
Pakistan	0	-1	-1	-1	-1	-1	-1
Nigeria	0	-1	-1	-1	-1	-1	-1
Argentina	-1	-1	0	-1	-1	-1	-1
Egypt	-1	-1	-1	0	-1	-1	-1

1 M > 0.3
 0 0.15 M 0.3
 -1 M < 0.15

Figure 5: Model goodness in different country-brand combinations

and Sony Ericsson. It is worthwhile to notice that the Apple brand was made up of only one search word and Vodafone had only one product.

When we analyzed the country-brand combinations more deeply and calculated how many models there was with certain M values for different country-brand combinations we got results that corresponded with the results from the average M analysis. The first combination we investigated was all brands but Vodafone in the countries Germany, France and Italy. We used the Model 1 and Model 2 with sweeplengths 2 and 3. The result are in Figure 6

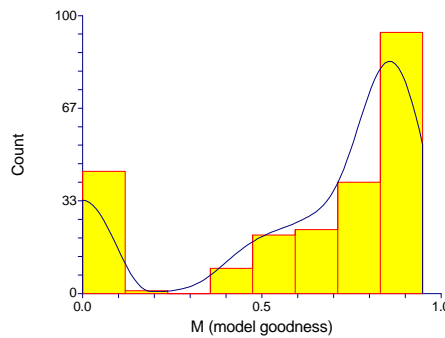


Figure 6: Number of models with certain M values in Germany, France and Italy

As you can see from the histogram there are mostly models with good M values. That means that Google search volumes data corresponds well to product sales in Germany, France and Italy.

The next combination we investigated was all brands but Vodafone in the countries Nigeria, Argentina and Egypt. The models we used was Model 1 and Model 2 with sweeplengths 2 and 3. The results are presented in Figure 7

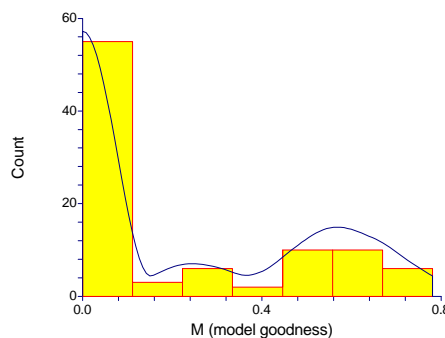


Figure 7: Number of models with certain M values in Nigeria, Argentina and Egypt

As you can see from the figure there are a lot of bad models. The models are much worse

than in the countries Germany, France and Italy. Google search volumes data is poor in Nigeria, Argentina and Egypt.

In both country analysis we had the same models, brands and sweeplengths to ensure that we only get results from how different countries differ from each other.

The next analysis was to have same countries, models and sweeplengths and vary the brands. The first combination we examined was the brands Apple and Sony Ericsson in the countries Germany, France and Italy. We use the Model 1 and Model 2 with sweeplengths 2 and 3. The results are in Figure 8

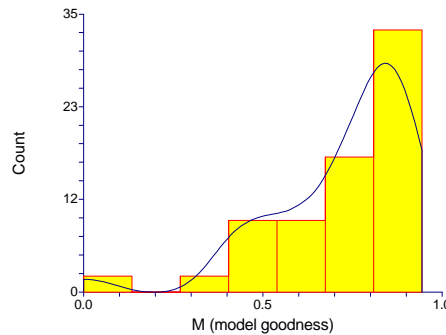


Figure 8: Number of models with certain M values with Apple and Sony Ericsson

Here, there are mostly good models which means that the brands Apple and Sony Ericsson have good models and therefore trends data related to these brands corresponds well to product sales.

The next combination we examined was the brands Motorola, Samsung and Vodafone in Germany, France and Italy. We used Model 1 and Model 2 with sweeplengths 2 and 3. The results are in Figure 9

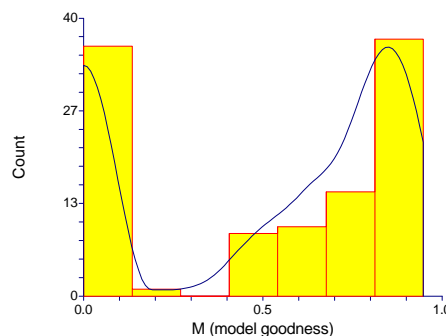


Figure 9: Number of models with certain M values with Motorola, Samsung and Vodafone

As you can see from the result there are plenty of models with bad M values. The brands Motorola, Samsung and Vodafone have many bad models and therefore data related to these brands corresponds badly to product sales.

After analyzing how different country-brand combinations differ from each other we analyzed the models and the sweeplengths. The first result was about the different models. For that purpose we calculated how many models there were with certain M values in the countries Germany, France, Italy, Indonesia, United Kingdom, Spain, Poland, India and South Africa with the brands Apple, Sony Ericsson, Blackberry and LG. We used only countries and brands with good quality trends data to see how the model should be chosen in cases with good data. We used all the sweeplengths. Each type of model was analyzed individually. The results of the analysis are represented in Figure 10

As you can see from the Figure 10 the Reference Model is the best one. The reason for that is that it is pretty good for most products but can not model the dynamics in the sales correctly. The Model 1 is on second place. It can probably model the dynamics at least for some parts of the life cycle for some products. The Model 3 has more models with M values between 0.5 and 0.7 than the best models but when we examine the interesting M values over 0.8 it has few models. The Models 4 and 5 are the worst ones and can not be used at all.

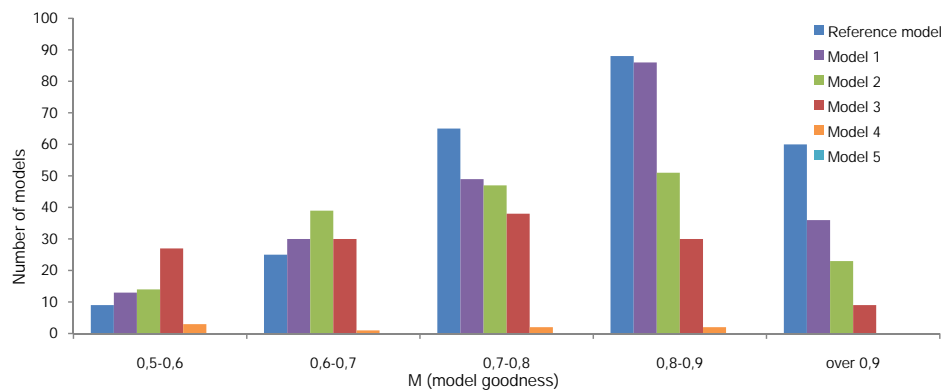


Figure 10: Number of models with certain M values for different model types

The next thing we studied was how the different sweeplengths differ from each other. For that purpose we calculated how many models there was with certain M values in the same countries and with the same brands as in the model analysis. We used the Reference Model, Model 1, Model 2 and Model 3 to examine how the sweeplength should be chosen when we have good data and use good models. Each sweeplength was analyzed individually. The results are represented the Figure 11

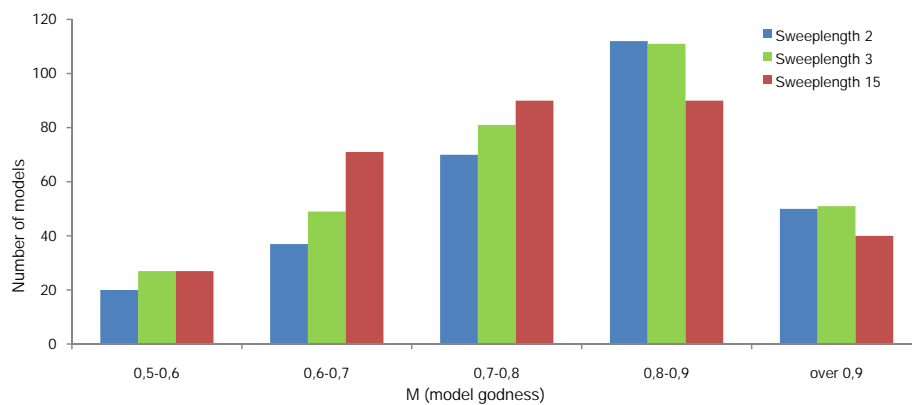


Figure 11: Number of models with certain M values for different sweeplengths

As you can see from Figure 11 the long sweeplength of 15 has more models with low M values than the short sweeplengths. When we examine the interesting and good M values over 0.8 we clearly see that there is more models with short sweeplengths. All in all the models with short sweeplengths are better.

4.4 Summary of results

Below a summary of the most important results is presented:

- The quality of Google search volumes data varies a lot between countries and products.
- Google search volumes data corresponds the sales better in European countries than in countries with emerging markets.
- The Reference Model works overall best but can not model the dynamics in the sales correctly. Model 1 comes in second place and can probably model the dynamics in the sales for some products in some countries. All the models have at least some problems modeling the dynamics in the sales.
- The models that uses only trends data and its lags are better than models with both trends and sales.
- Models with short sweeplengths as 2 or 3 are better than models with long sweeplengths as 15.

5 Discussion

During the analysis process we noticed several ways how the model could be improved. The purpose of this chapter is to give suggestions for further analysis.

5.1 Change in dynamics

There are different phases during a product life cycle, the dynamics change. We tackle this problem by using a short memory model, i.e. we use only the last two or three data points to estimate the model parameters and the next data point. There are other ways to deal with changing dynamics.

A simple solution would be to use different models in different phases. The sales during a product's life cycle vary a lot in the beginning whereas in the end the curve looks almost linear. In our analysis we also notice that AR-models, despite being useless in the beginning, capture the end of a product life cycle really well. It would thus be useful to use a dynamic regression model in the beginning and an AR-model in the end. Defining the point where one phase ends and another begins could however be problematic.

Another way to deal with this would be to increase the number of data points included in the model during time. That is after some predetermined amount of months we add one more point to the estimation. The changes in dynamics seem to be bigger in the beginning, this would motivate the use of more data points towards the end. To implement this small increase would be easier than changing the model totally at an undetermined point.

5.2 Estimation of sales at product launch

A challenge is too how to estimate sales volumes in the early face of the product life cycle. It is particularly difficult because we cannot know how many searches there are in absolute numbers, our values are relative. Even if we knew that, it would be hard to tell how many searches it takes to get a product sold, especially since we are talking about the first sales value and the product has not been sold before.

Investigating previous product's sales per trends ratios is one way to approach the problem. This could provide at least some kind of an estimate for the first sales value but the ratio however changes a lot between products.

Often the estimation jumps to volumes that are ten or even more times greater than the actual volume when estimating the second months sales value. The models we use could be improved by adding some feasible maximum value for the sales.

5.3 Using other models than dynamic regression models

In our analysis we focus on dynamic regression models. There are however other models that could be used for analyzing the future (or current) sales.

Neural network models could be used for this purpose. Our initial studies showed that they give about as good results as dynamic regression models. Going deeper into neural network may prove fruitful.

If we forget the trends aspect, a Bass diffusion model is often used for product sales forecasting. A Bass diffusion model estimates the sales based on assumptions of new innovators (persons being the first to purchase the product), imitators and ultimate market potential. The diffusion model could possibly be combined with the model we have used.

5.4 Other ways to improve the estimation

The search trends data we use has some unexplainable fluctuations. One interesting thing to investigate would be how these variations could be filtered. Another rather simple way to improve the estimation would be to get weekly sales data, now we were using monthly data. In many cases the trends data is weekly.

6 Conclusion

The best way to use Internet search volumes in predicting product sales is to use a dynamic regression model described with equation (1). The Internet search trends can in some rare cases

be used to assess future product sales. In several cases they do provide information about the sales at present. However there are a lot of cases when the search trends either do not exist or do not correlate significantly with the sales values. In these cases the trends cannot provide any help when assessing the product sales. Those cases can be identified using some simple rules which are listed in the results (see section 4.4).

Hypothesis 1 was thus partly accepted and partly abandoned. The results (e.g. Figure 5) clearly show that cases can be divided in really bad and really good cases. A good hypothesis for further work would be that the model in equation (2) provides better results when there are big fluctuations in the sales volumes than just assuming the sales to stay constant.

The most suitable data source for Internet search volumes is the Google Insights for Search. However, there are apparent differences between countries. Roughly it could be said that the bigger and more developed a country is, the better results our model provides. Possible reasons for this could be bigger amount of potential customers and also better Internet coverage. China is an exception to this rule, but then again Google is not the ruling search engine in China. When it comes to different product brands - the bigger the brand the better the results with our model.

All in all, search trends cannot yet replace other means of prediction. Estimating product sales will still be uncertain and there is room for unexplained fluctuations. Search trends are however a useful instrument for improving sales estimates when used properly.

References

- Choi, H. and Varian, H. (2009). Predicting the present with google trends. <http://googleresearch.blogspot.com/2009/04/predicting-present-with-google-trends.html>.
- Shimshoni, Y., Efron, N., and Matias, Y. (2009). On the predictability of search trends. <http://googleresearch.blogspot.com/2009/08/on-predictability-of-search-trends.html>.