On the Predictability of Search Trends

Yair Shimshoni Niv Efron Yossi Matias

Google, Israel Labs

Draft date: August 17, 2009

1. Introduction

Since *Google Trends* and *Google Insights for Search* were launched, they provide a daily insight into what the world is searching for on Google, by showing the relative volume of search traffic in Google for any search query. An understanding of web search trends can be useful for advertisers, marketers, economists, scholars, and anyone else interested in knowing more about their world and what's currently top-of-mind.

The trends of some search queries are quite seasonal and have repeated patterns. See, for instance, the search trends for ski in the US and in Australia peak during the winter season; or check out how search trends for basketball correlate with annual league events and how consistent it is year-over-year. When looking at trends of the aggregated volume of search queries related to particular categories, one can also observe regular patterns in at least some of hundreds of categories, like the Food & Drink or Automotive categories. Such trends sequences appear quite predictable, and one would naturally expect the patterns of previous years to repeat looking forward.

On the other hand, for many other search queries and categories, the trends are quite irregular and hard to predict. For example, the search trends for Obama, Twitter, Android, or global warming, and trends of aggregate searches in the News & Current Events category.

Having predictable trends for a search query or for a group of queries could have interesting ramifications. One could forecast the trends into the future, and use it as a "best guess" for various business decisions such as budget planning, marketing campaigns and resource allocations. One could identify deviation from such forecasting and identify new factors that are influencing the search volume like in the detection of influenza epidemics using search queries [Ginsberg etal. 2009] known as Flu Trends.

We were therefore interested in the following questions:

- How many search queries have trends that are predictable?
- Are some categories more predictable than others? How is the distribution of predictable trends between the various categories?
- How predictable are the trends of aggregated search queries for different categories? Which categories are more predictable and which are less so?

To learn about the predictability of search trends, and so as to overcome our basic limitation of not knowing what the future will entail, we characterize the predictability of a Trends series based on its historical performance. That is, based on the *a posteriori* predictability of a sequence determined by the discrepancy of forecast trends applied at some point in the past vs the actual performance.

Specifically, we have used a simple forecasting model that learns basic seasonality and general trend. For each trends sequence of interest, we take a point in time, *t*, which is about a year back, compute a one year forecasting for *t* based on historical data available at time *t*, and compare it to the actual trends sequence that occurs since time *t*. The discrepancy between the forecasting trends and the actual trends characterize the predictability level of a sequence, and when the discrepancy is smaller than a predefined threshold, we denote the trends query as *predictable*.

We investigate time series of search trends provided by Google Insights for Search (*I4S*), which represent query shares of given search terms (or for aggregations of terms). A *query share* is the total number of queries for a search term (or an entire search category) in a given geographic region divided by the total number of queries in that region at a given point in time. The query share represents the popularity of a query, or the aggregated search interest that users have in a query, and we will therefore use the term *search interest* interchangeably with *query share*.

The highlights of our observations can be summarized as follows:

- Over half of the most popular Google search queries were found predictable in 12 month ahead forecast, with a mean absolute prediction error of approximately 12% on average.
- Nearly half of the most popular queries are not predictable, with respect to the prediction model and evaluation framework that we have used.
- Some categories have particularly high fraction of predictable queries; for instance, *Health* (74%), *Food & Drink* (67%) and *Travel* (65%).
- Some categories have particularly low fraction of predictable queries; for instance, *Entertainment* (35%) and *Social Networks & Online Communities* (27%).
- The trends of aggregated queries per categories are much more predictable: 88% of the aggregated category search trends of over 600 categories in Insights for Search are predictable with a mean absolute prediction error of less than 6% on average.
- There is a clear association between the existence of seasonality patterns and higher predictability as well as an association between high levels of outliers and lower predictability.

Recently the research community has started to use Google search data provided publicly by Google Insights for Search (*I4S*) as auxiliary indicators for economic forecast. [Choi & Varian 2009] have shown that aggregated search trends of Google Categories can be used as extra indicators and effectively leverage several US econometrics prediction models. [Askitas & Zimmermann 2009] and [Suhoy 2009] have shown similar findings on German and Israeli economic data, respectively. Getting a better insight into the behavior of relevant search trends has therefore high potential applicability for these domains.

For queries or aggregated set of queries for which the search trends are predictable, one can use a forecasted trends based on the prediction model as a baseline for identifying deviations in actual trends. Such deviations are of particular interest as they are often indicative of material changes in the domain of the queries. We consider a few examples with observed deviation of actual trends relative to the forecasted trends, including:

- **Automotive Industry**
  We show that in the recent 12 months there is a positive deviation relative to the forecast baseline (i.e., an increased query share) in the searches of *Auto Parts* and *Vehicle Maintenance* while there is a negative deviation (i.e., a decrease in query share) in the searches of *Vehicle Shopping* and *Auto Financing.*

- **US Unemployment**
  In relation with the recent research that showed an improvement in prediction of unemployment rates using Google query shares [Choi and Varian 2009 b], we show that the search interest in the category of *Welfare & Unemployment* has substantially risen in the last year above the forecast based on the prediction model. We also show that the search interest in the category *Jobs* has significantly decreased according to the prediction model.

- **Mexico as Vacation Destination**
  We examine the large decrease in the query share of the category *Mexico* as a vacation destination, compared to the predictions for the last 12 months. We show that a similar deviation of (actual vs. forecast) query share is not observed for other related categories.

- **Recession Markers**
  We show several examples that demonstrate possible influences of the recent recession on search behavior, like an observed increase of query share for the category *Coupons & Rebate* compared to the forecast. We also show a negative deviation between the query share for category *Restaurants* compared to the forecast, where as the category *Cooking and Recipes* shows a similar positive deviation.

Outline
The rest of this paper is organized as follows. In Section 2 we formulate the notion of predictability and describe the method of estimating it along with the evaluation measures, prediction model and the time series data that we use. In Section 3 we describe the experiments we conducted and present their results. In Section 4 we examine the association between the predictability of search interest and the level of seasonality or internal deviation of the underlying search trends. Section 5 will present sensitivity analysis and error diagnostics and in section 6 we discuss the potential use of forecasting as a baseline for identifying deviations from regular search behavior; we demonstrate with some examples that the discrepancies from model predictions can act as signals for recent changes in the query share.

2. Time Series Predictability

In this section we define the notion of predictability as we use it in our experiments.

Predictability.
We characterize the *predictability* of a time series with respect to a *prediction model* and a *discrepancy measure*, as follows.

Assume we have:
- A time series $\mathbf{X}=\{x_{t-H}, \ldots, x_{t+F}\}$ with history of size H and a future horizon of size F
  Denote: $\mathbf{X}^H=\{x_{t-H}, \ldots, x_t\}$ and $\mathbf{X}^F=\{x_{t+1}, \ldots, x_{t+F}\}$
- A Prediction model: M, which computes a forecast $\mathbf{Y}=\boldsymbol{M}(\mathbf{X}^H)$,
  where $\mathbf{Y}=\{y_{t+1}, \ldots, y_{t+F}\}$
- A Discrepancy Measure: $\mathbf{D}=\mathbf{D}(\mathbf{X},\mathbf{Y})$
- A Threshold: $\mathbf{D'}$

Then, we say that **X** is *predictable* w.r.t $(\mathbf{M,D,D',t,H,F})$ iff $\mathbf{D}=\mathbf{D}(\mathbf{X},\mathbf{Y}) < \mathbf{D'}$.
The size of the discrepancy also characterizes the level of predictability of a series.

We will often refer to a trends sequence as predictable (or not-predictable) where the various parameters are implied by the context.

Data.
The time series that are used in the following experiments are based on Google Insights for Search[1] (*I4S*), which reports the query share for search terms in any time, location and category, as well as capable of reporting the most popular queries within a given time / location / category.

A *query share* is defined as the total number of queries for a search term or a set of terms (e.g., an entire search category) in a given geographic region, divided by the total number of queries in that region, at a given point in time. The *I4S* categories are organized in a tree-like hierarchical structure, with about 30 root level categories, that are further divided into sub-categories in a 3-level taxonomy, to a total of about 600 categories and sub-categories. Each search query is classified by *I4S* to a single category, nevertheless it will also be counted as a part of the query share to all its 'parent' categories. For each category, *I4S* calculates an *aggregated* times series which represents the overall query share of this category (i.e., the combined search interest of all the queries in the category).

In order to stay focused on the most influential patterns of the yearly seasonality and overall trend (direction), we are using time series of *monthly granularity* (i.e., one data point per calendar month) and refer to the entire available period (2004-2009). Obviously search trends with finer granularity (e.g., weekly or daily search data) do capture more patterns of search behavior within the intra-monthly and especially the day-of-week effect, however the fine resolution data is also noisier and thus calls for prediction models with higher complexity and a less homogeneous model space. We leave that for future research.

We have extracted time series of the entire available time range (2004-2009)[2] that consists of 67 data points, which were partitioned into two parts:

1. *The History Period* - 55 monthly data points (January 2004 - July 2008)
2. *The Forecast Period* - 12 monthly data points (August 2008 - July 2009)

Throughout the work, we will refer to 3 data sets of time series (with a similar format):

1. ***Country Data -*** Includes time series of the query shares for the 10,000 most popular queries in each of these countries: *USA, UK, Germany, France* and *Brazil.*
2. ***Category Data -*** Includes time series of the query shares for the 1,000 most popular queries in the US, for 10 major *I4S* categories: *Automotive, Entertainment, Finance & Insurance, Food & Drink, Health, Social Networks & Online Communities, Real Estate, Shopping, Telecommunications,* and *Travel.*
3. ***Aggregated Categories Data -*** Includes time series of aggregated query shares for about 600 *I4S* categories, which represent the normalized combined search volume in the US for each respective category.

Generic Prediction Model.
Our prediction process is based on the **STL** Procedure [Cleveland et.al. 1990], which is a filtering procedure based on locally weighted least squares for decomposing a given time series X into the *Trend*, *Seasonal* and *Residual* components.

STL is basically an EM-like algorithm that calculates the seasonal part assuming knowledge of the trend part (iteratively). To compute the forecast of the future values, we extrapolate the trend sub-series using regression, and use the last seasonal period of the seasonal component.

---

1. URL: http://www.google.com/insights/search/#
2. The time series data was pulled during July 2009, thus the value for this last month might change.

The STL procedure uses 6 configuration parameters, 3 of which are smoothing parameters for the three components, which in general should be chosen per time series. The prediction process in our experiments was using a fixed STL configuration for the forecast of all the time series. Given a sampled archive of search time series, we have used an exhaustive exploration and evaluation process that was searching for the best parameter set from a pre-defined set of optional parameter values. The optimality criterion was minimal mean absolute error and the output was a single parameter set w.r.t. the given sampled archive. By choosing to use a particular (fixed) configuration, rather than adjusting an individual parameters set for each given time series, we are adjusting the configuration to a large set of time series thus simplifying the prediction model and enabling much faster forecast.

Prediction Discrepancy function.
We define the discrepancy **D** as a combination of several error metrics between the forecasted trends $\mathbf{X}^F$ and the actual trends **Y**, as well as seasonal consistency metrics determined by difference in the auto-correlation between $\mathbf{X}^H$ and **X**.

Specifically, **D** is defined as a tuple:

$$\boldsymbol{D = < MAPE, MaxAPE, NMSSE, MeanAbsACFDiff, MaxAbsACFDiff >}$$

based on metrics defined below, we say that **D<D'** iff each of the entries in **D** is smaller than the corresponding entry in **D'**.

Error Metrics.
We use the following error metrics:

- The Mean Absolute Percent Error *(**MAPE**)*:
  $$\boldsymbol{MAPE} = (1/F) \sum_{i=(t+1,..,t+F)} (|x_i - y_i| / x_i)$$
- The Maximum Absolute Percent Error *(**MaxAPE**)*:
  $$\boldsymbol{MaxAPE = MAX}_{i=(t+1,..,t+F)} (|x_i - y_i| / x_i)$$
- The Normalized Mean Sum of Squared Errors *(**NMSSE**)*:
  $$\boldsymbol{NMSSE} = [(1/F) \sum_{i=(t+1,..,t+F)} (x_i - y_i)^2] / \mathrm{Var}(\mathbf{X})$$

Seasonal Consistency Metrics.
A necessary condition for a valid inference from the past of a time series to its future, is that the major characteristics of the underlying time series are maintained. In order to capture the dynamic consistency of a seasonal time series, we use a metric that is based on the time series lagged-dependencies that are represented by the Auto Correlation Function (ACF) coefficients.
We compare the sets of 12 ACF coefficients (corresponding to the first 12 lags) for two time series:

1. The complete time series $\mathbf{X} = \{x_{t-H},..,x_t,..,x_{t+F}\}$
2. The History period $\mathbf{X^H} = \{x_{t-H},..,x_t\}$.

A large difference between these sets of ACF coefficients suggests that the seasonal behavior of the recent period of F time steps is different than the seasonal behavior within the History period of the underlying time series. In such cases one could suspect that a prediction model which is based on the History period might yield erroneous prediction results.

We compute the following two metrics for each time series in order to represent the

consistency of its seasonal behavior and use them as additional information to assess the predictability potential of the underlying time series. Notice that these metrics are "stand alone" i.e., do require no prediction model.

- The Mean Absolute Difference between the Auto-Correlation Function (ACF) Coefficients of the complete time series (**X**) vs. the History period ($\mathbf{X^H}$).

  ***MeanAbsACFDiff* (X)** = Mean (Abs($ACF_{lag}(\mathbf{X})$ - $ACF_{lag}(\mathbf{X^H})$)), *lag=(1,..,12)*

- The Max Absolute Difference between the Auto-Correlation Function (ACF) Coefficients of the complete time series (**X**) vs. the History period ($\mathbf{X^H}$).

  ***MaxAbsACFDiff* (X)** = Max (Abs($ACF_{lag}(\mathbf{X})$ - $ACF_{lag}(\mathbf{X^H})$)), *lag=(1,..,12)*

Predictability Discrepancy Threshold.
We use the following threshold:

***D'* = < 25%, 100%, 10.0, 0.2, 0.4 >** .

Thus, we say that a given time series is predictable within the available time frame, w.r.t. the prediction model we use and the above error and consistency metrics, if all the following conditions are fulfilled:

1) The Mean Absolute Prediction Error (***MAPE***) < 25%
2) The Max Absolute Prediction Error (***MaxAPE***) < 100%
3) The Normalized Mean Sum of Squared Errors *(**NMSSE**)* < 10.0
4) The Mean Absolute Difference of the ACF Coef. Sets (***MeanAbsACFDiff***) < 0.2
5) The Max Absolute Difference of the ACF Coef. Sets (***MaxAbsACFDiff***) < 0.4

Predictability Ratio.
Given a set A of time series, denote its *predictability ratio* as the number of predictable time series in A, divided by the total number of time series in A.

3. Experiments and Results

Comparing the Predictability of Top Queries in Different Countries.
We have conducted an experiment to test the predictability of search trends regard to the 10,000 most popular search queries in five countries:

Country	***Predictability Ratio***	**Avg. *MAPE*** (for predictable queries)	**Avg. *MaxAPE*** (for predictable queries)
USA	54.1	11.8	27.1
UK	51.4	12.7	32.1
Germany	56.1	11.8	28.2
France	46.9	12.8	28.8
Brazil	46.3	13.7	30.5

Although the above results show some variability among the different countries, one can see that in general, about half of the time series that correspond to popular queries in Google Web Search are *predictable* with respect to the given prediction model and discrepancy function / threshold. One can see that among the *predictable* queries, the mean absolute prediction error (*MAPE*) is about 12% on average, while the maximum absolute prediction error (*MaxAPE*) is about 30% on average.

The Seasonally of Time Series.
Time series in general often include various forms of regularity, like a consistent trend (straight, upward or downward) or seasonal patterns (daily, weekly, monthly, etc). In seasonal time series, the amplitude changes along the time in a regular recurring fashion according to the relevant season. In many practical cases, it is common to use a seasonality *adjustment* where the seasonal component is subtracted from the time series before the analysis, where there are procedures that decompose time series into their seasonal and trend components [Cleveland and Tiao 1976], [Lytras etal. 2007]. We use such a decomposition to compute a metric that represents the relative portion of seasonality within a time series as follows:

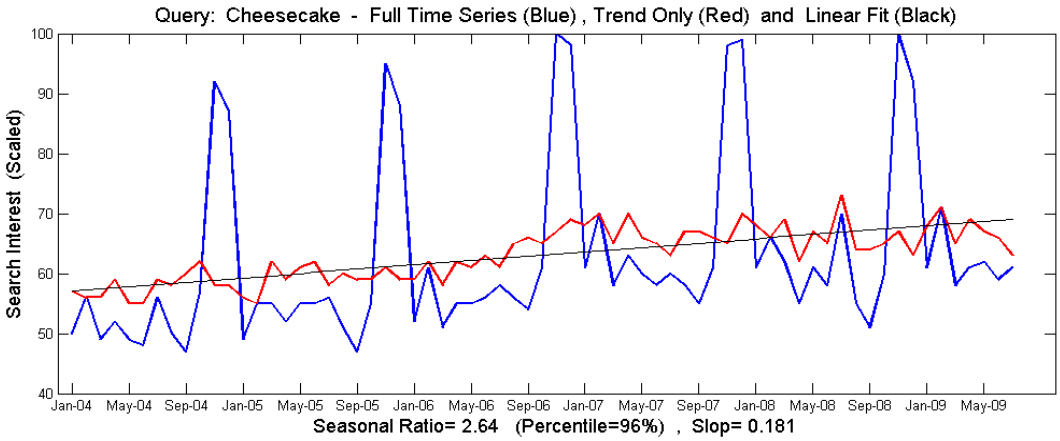
Given a time series $\mathbf{X} = \{x_1, \ldots, x_T\}$ and a decomposition of **X** into a seasonal component ***S*** and a *Trend* (i.e., directional) component ***Tr*** then:

$$\textbf{\textit{Seasonality Ratio}}(\mathbf{X}) = \left(\sum |\boldsymbol{S_i}| \right) \; / \; \left(\sum |\boldsymbol{Tr_i}| \right)$$

For each time series we forecast, we compute the respective seasonality ratio.

For example, let us examine the time series which represents the search interest for the query Cheesecake (in the US, 2004-2009). The blue curve in the following plot shows the original time series which has a significant seasonal component. The red curve is the *seasonality adjusted* time series; i.e., the trend component that is left after subtracting the seasonality component. It has an upward trend (with Slope:0.18) plus some variability. The seasonality ratio is 2.64 (which is on the 96 percentile of the 10,000 tested queries) and approximates the

ratio of the area between the red and blue curves, and the area underneath the red curve.



The Deviation of a Time Series.

In order to assess the extent to which a time series contains extreme values or outliers with large deviation from the overall pattern, we calculate for each time series the *deviation ratio.* In general, we compute the sum of the top values in the series divided by the total sum of the series, assuming that a large ratio would indicate the existence of considerable extreme values in the series. We normalize by the relative number of top values under consideration.

Given a time series $\mathbf{X} = \{x_1, \ldots, x_T\}$ and an integer w, s.t. $1<w<100$, let ***Prc***(**X**,*w*) be the w'th percentile of values of **X** (e.g., ***Prc***(**X**,*50*) = Median(**X**)). Then,

$$\textbf{\textit{Deviation Ratio}}\ (\mathbf{X},w) = \left(\sum_t (\mathbf{X}^w) / \sum(\mathbf{X})\right) / (1-(w/100)),$$

where $\mathbf{X}^w$ consists of all $\boldsymbol{x}_t$, s.t. $\boldsymbol{x}_t >= \textbf{\textit{Prc}}(\mathbf{X},w)$.

We use *w=90*. Notice that the normalization term, *(1-(w/100))* in the denominator, is setting the minimal ratio to be 1. Due to the relatively short time series (of 67 points) and since many cases show seasonal patterns with high and narrow peaks (e.g., like in the plot above), it is possible that these sharp peaks will be considered as outliers, although they are a regular part of the time series' recurring dynamics. To mitigate this, we computed the *deviation ratio* on the seasonal adjusted time series (i.e., on the *Trend* component that is left after the seasonal component is subtracted by the decomposition we have described above).

The Predictability of Search Categories.

In order to assess the predictability of categories, we have extracted the 1000 most popular queries in the US for a selection of 10 root level categories and tested their predictability. In the following table, we present the summary results, where the Predictability column on the left refers to the entire 1,000 queries (per category), and the two error metrics (*MAPE* and *MaxAPE* in columns 3 & 4) refer only to the sub-set of Predictable queries within each category. The seasonality and deviation ratios are also referring to the entire category sets.

The Predictability per category spans from 74% for the *Health* category, to 27% for the *Social Networks & Online Communities* category. In the third column we can see that the Mean Absolute Prediction Error (*MAPE*), varies from of 9% (in the *Health* category), to 14.1% (in the *Social Networks & Online Communities* category). The average *MAPE* for the 10 categories is 12.35%. Notice that the order of predictability ratio is not equal to the order of the *MAPE* error

since the Predictability is based on several other metrics as described in Section 2, however the correlation between them is high (*r= -0.85*).

The variability within the columns of seasonality ratio and deviation ratio represents the differences between the search profiles of the various categories, which correspond to the variability of the categories' predictability ratio. For example notice the relatively high seasonality ratio and low deviation ratio of the *Food & Drink* category which has 66.7% predictability ratio vs. the opposite situation of the *Entertainment* category that has 35.4% predictability ratio with a relatively low seasonality ratio and high deviation ratio.

Category Name	Predictability Ratio	MAPE *predictable queries*	MaxAPE *predictable queries*	Seasonality Ratio	Deviation Ratio
Health	**74.00**	9.00	20.00	0.73	1.58
Food & Drink	**66.70**	11.90	26.00	1.20	1.74
Travel	**64.70**	11.80	27.00	1.09	1.61
Shopping	**63.30**	12.40	28.00	1.21	1.78
Automotive	**57.60**	11.20	24.90	0.71	1.84
Finance & Insurance	**52.90**	13.30	30.60	0.65	2.00
Real Estate	**49.50**	12.90	29.90	0.72	1.82
Telecommunications	**45.60**	12.90	29.40	0.32	2.34
Entertainment	**35.40**	14.00	32.30	0.46	2.49
Social Networks	**27.50**	14.10	30.10	0.19	2.95

For the above summary results of the 10 categories, the correlation between the Predictability and the Seasonality Ratio is *r= 0.80* while the Deviation Ratio has a (negative) correlation of *r= -0.94* with the Predictability. In the next section we will further examine the association between these regularity characteristics and the predictability.

The Predictability of Aggregated Time Series that represent Categories.

We now show the results of an experiment of forecasting aggregated times series that represent the overall query share of categories (i.e., the combined search interest of all the queries in the category).

We ran the experiment on the aggregated time series of over 600 *I4S* categories and computed the average absolute prediction error over a period of 12 months ahead. We found 88% of the aggregated category time series to be predictable. The average *MAPE* for the entire set of aggregated category time series is 8.15%. (6.7% for Predictable queries only), with STD=4.18%. The Average Maximum Prediction Error (*MaxAPE*) for the entire set was 19.2% (16.6% for Predictable queries only).

In the table below, we show the prediction errors for the aggregated time series for the same 10 root categories we examined above. Notice that the prediction errors are now smaller, which was expected. However, we can also see that the order of the categories is not the same as the respective order in the table of the previous experiment. In general, the aggregated time series should have a higher predictability due to the noise reduction effect of the aggregation. The rightmost column shows the *MAPE* Reduction Rate, which is the relative improvement of the prediction error (*MAPE)* of the 1,000 queries per category (in the previous

experiment) and the single *MAPE* for the aggregated category time series here. All categories (except *Social Networks & Online Communities*) had their *MAPE* reduced, starting from 47% improvement for the *Finance & Insurance* category up to 85% for the *Food & Drink* Category.

Category	***MAPE***	***MaxAPE***	***Seasonality Ratio***	***Deviation Ratio***	***MAPE* Reduction Rate**
Food & Drink	1.76	4.52	0.70	1.18	0.85
Shopping	2.72	6.02	2.77	1.11	0.78
Entertainment	2.74	5.95	0.30	1.16	0.80
Health	2.99	7.69	1.04	1.11	0.67
Automotive	3.27	7.36	1.69	1.12	0.71
Travel	3.94	7.61	1.92	1.12	0.67
Telecommunications	5.2	9.07	0.74	1.20	0.60
Real Estate	5.62	12.8	2.95	1.11	0.56
Finance & Insurance	7.08	17.8	0.61	1.26	0.47
Social Networks	38.6	50.4	0.06	2.46	-1.74

The I4S classification into search categories is based on a hierarchical tree-like taxonomy where each category at the root level of the tree has several sub-categories under it. Thus, a combination of all the categories' prediction error into an overall evaluation of the prediction error, can consist of the average *MAPE* values of the 27 root level categories. However, a 'regular' (uniform) average which gives the same weight to each category, might be inaccurate. Therefore, we have computed a weighted average of the root categories' *MAPE*, where the weights are the overall relative search interest of each root category. The *MAPE Weighted Average* is 4.25%.

The following table shows the predication errors for the *I4S* root categories (sorted by the *MAPE*):

Root Category	***MAPE***	***MaxAPE***
Food & Drink	1.76	4.52
Beauty & Personal Care	2.2	7.41
Home & Garden	2.21	4.9
Photo & Video	2.34	8.31
Lifestyles	2.38	5.27
Games	2.59	4.45
Shopping	2.72	6.02
Entertainment	2.74	5.95
Business	2.91	11.5
Health	2.99	7.69
Local	3.24	5.49
Automotive	3.27	7.36
Reference	3.7	8.2
Industries	3.77	7.14
Recreation	3.81	7.58
Computers & Electronics	3.93	7.83

Travel	3.94	7.61
Internet	4.87	15
Telecommunications	5.2	9.07
Society	5.57	12.6
Real Estate	5.62	12.8
Sports	5.81	29.3
Arts & Humanities	6.98	11.8
Finance & Insurance	7.08	17.8
Science	10.1	15.5
News & Current Events	16.6	47
Social Networks	38.6	50.4
Average	**5.81**	**12.5**

Comparing the Predictability of a Category and its Sub-Categories.
It is reasonable to expect that a time series of the aggregated search of a set of queries should in general be more predictable than single queries. The larger the aggregation set is, the smaller would be the variability of the aggregated time series. This has implications on the predictability of categories vs. sub-categories, but also has implications regarding aggregated time series of group of queries such as campaign related queries or brand/topic related queries in general. In order to demonstrate this, we have explored the *MAPE* and *MaxAPE* prediction errors of the *I4S* category *Vehicle Brands* (in the *Automotive* category), compared to all its 31 'children' sub-categories.

The variability of Prediction Errors (*MAPE*) within the 31 vehicle brands sub-categories is substantial and varies from 3% to 38%. The average *MAPE* of the 31 brands is 11.4% (with STD=7.7%) which is quite similar to the average *MAPE* for the 1,000 most popular queries in the *Automotive* category (11.2%) as we presented above.

As expected, the average *MAPE* of the 31 sub-categories is larger than the *MAPE* of the aggregated time series of the *Vehicle Brands* category which is only 3.39%. We have also calculated the median *MAPE* (9.3%), as well as the weighted average *MAPE* (with relative search interest per category as weights) (9.7%). Both the median and the weighted average are lower than the regular average but still much larger than the *MAPE* for the overall aggregated category of *Vehicle Brands*.

4. Predictability vs. Seasonality and Deviation Ratios

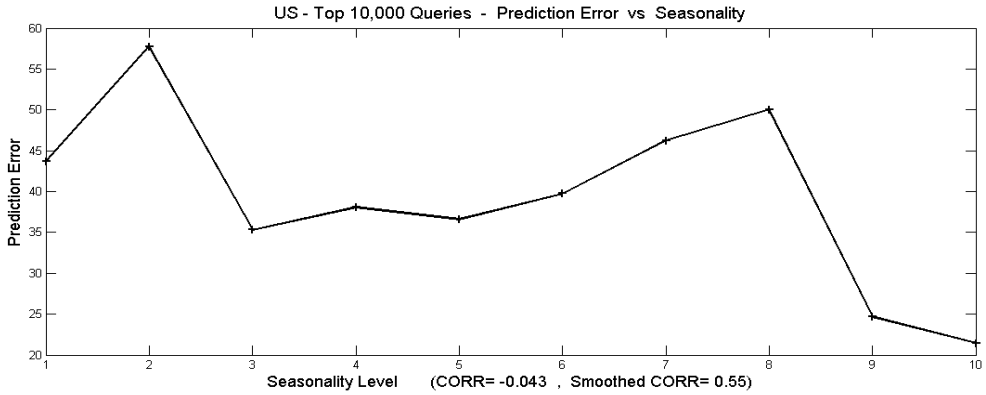
Among the 10 categories for which we have analyzed their 1,000 most popular queries, we calculated a correlation of *r= 0.80* between the Predictability and the Seasonality Ratio and *r= -0.94* between the Predictability and the Deviation Ratio (see table in Section 3).
Below, we examine the association between the these two time series' characteristics and the *MAPE* prediction error in the experiment we conducted on the 10,000 most popular queries in the US.

Seasonality and Prediction Errors.
Many patterns of search behavior have a strong seasonal component (e.g. holidays shopping, summer vacation, etc.) as implied from the specific market they are in. Occasionally, there is also a directional trend effect (up, down or changing) which might be less visually pronounced due to the confounding seasonal pattern. We have used the *Seasonality Ratio* (described above) as a representation for the 'level of seasonality' of the queries.

Among the 10,000 most popular queries in the US, the Seasonality Ratio varies in the rather large range [0.01,13], from time series with no seasonal component up to extremely seasonal time series. The median *Seasonality Ratio* is 0.4 and its mean value is 0.8.

We could see no significant correlation between the prediction error and the seasonality ratio. In order to visualize this possible association, we have sorted the values of seasonality ratio and created a ('smoothed') arrays of 10 average points[3]. Similarly, we have computed a 'smoothed' array of averages for the 10,000 corresponding *MAPE* prediction errors which were sorted according to the corresponding seasonality ratio.

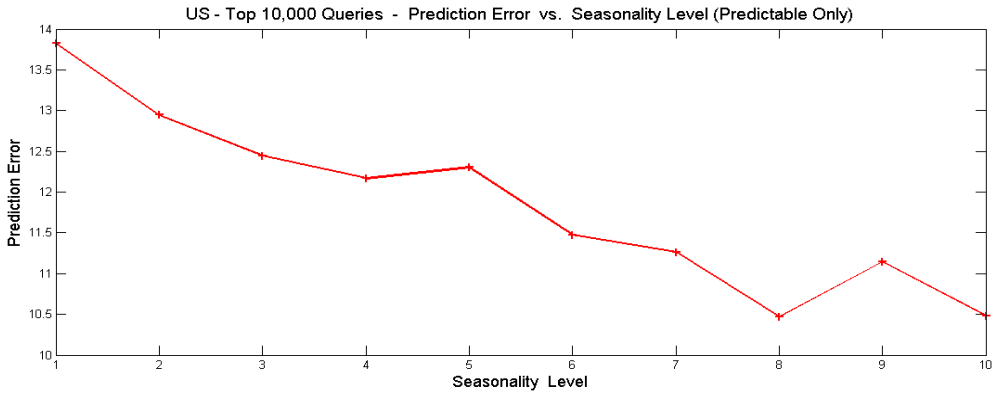
We show here a scatter plot of the 'smoothed' *MAPE* vs the 'smoothed' seasonality ratio. The plot shows a non-stable 'negative' association between prediction errors and the seasonality. The correlation coefficient between the 'smoothed' arrays is substantial (*r=0.55*), compared to the insignificant correlation we saw for the entire set.



---

3. Given a time series $\{Y_N\}$, $N=10,000$; $K=10$; $M=N/K=1,000$. We compute an array $A=\{A_1, A_2,.....,A_K\}$ of the averages of K consecutive non-overlapping windows of size M over the time series $\{Y_N\}$, such that $A_k = (1/M) \sum Y_i$, where $k=1,..,K$ and $i=\{1+(k-1)M,..,kM\}$.

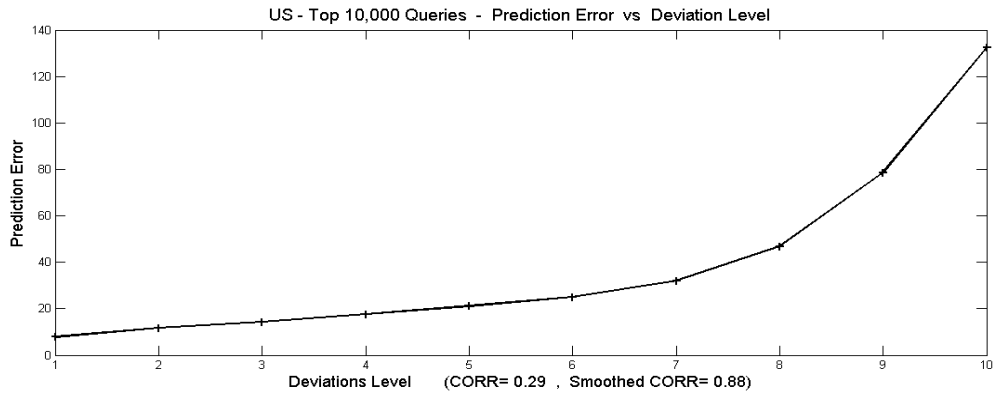
For the next plot we have repeated the same process - but for predictable time series only. The result shows a stronger 'negative' association between the *MAPE* prediction error and the seasonality ratio for Predictable queries.

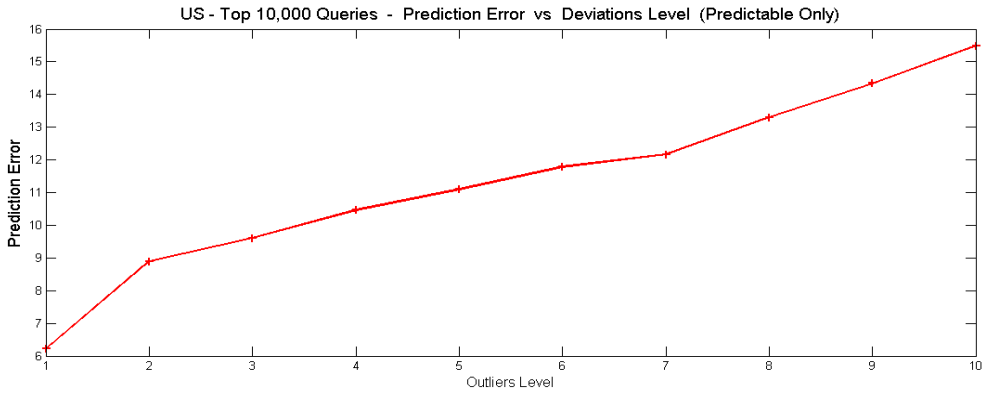


Deviation Ratio and Prediction Errors.

The *Deviation Ratio*, which represents the level of outliers and irregular extreme values in a time series was found to be associated with the Predictability of the search interest time series. For the 10,000 queries we tested, the average deviation ratio was 2.08 (STD=1.9). Only 5% of the Predictable time series had a deviation ratio in the upper quartile and 73% of the predictable time series had a deviation ratio under the median. The correlation coefficient between the deviation ratio and the the *MAPE* error was *r=0.29*. The average deviation ratio for the Predictable time series was: 1.50 where as for the non-Predictable queries the average was: 2.77.

We have applied the same process as above in order to visually demonstrate the association between *MAPE* and the deviation ratio. The following plot shows a clear positive association between the (sorted) 'smoothed' array of the deviation Ratio and the corresponding 'smoothed' array of prediction errors (*MAPE*). The correlation coefficient calculated for the 'smoothed' arrays was *r=0.88* (compared to *r=0.29* which was computed with the original values). Hence, we can say that the larger the deviation level in the time series, the larger is the prediction error. This can also be seen in the next plot for the the Predictable queries only.



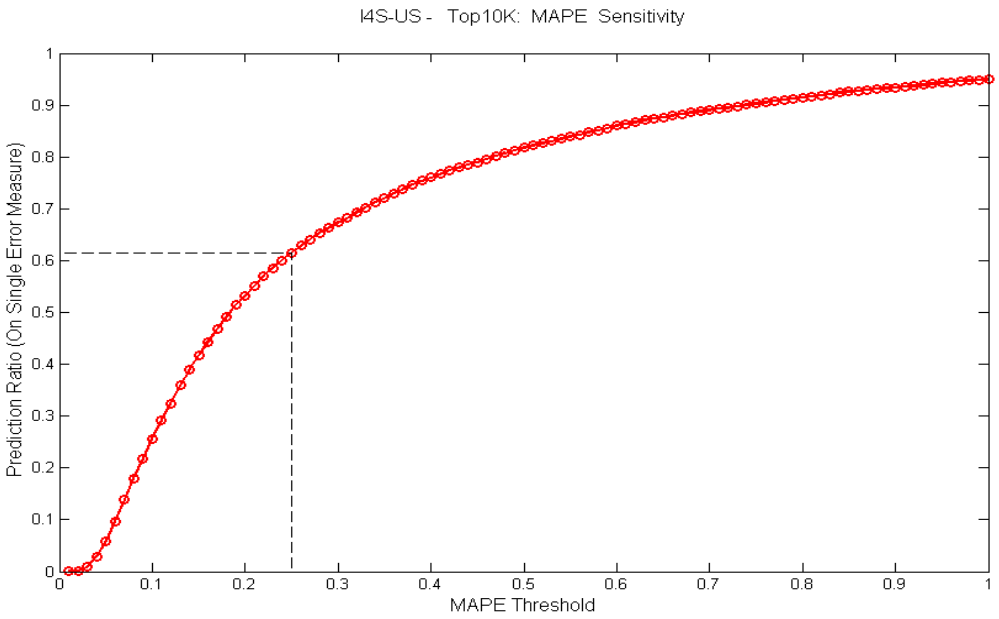


5. Sensitivity Analysis and Errors Diagnostics

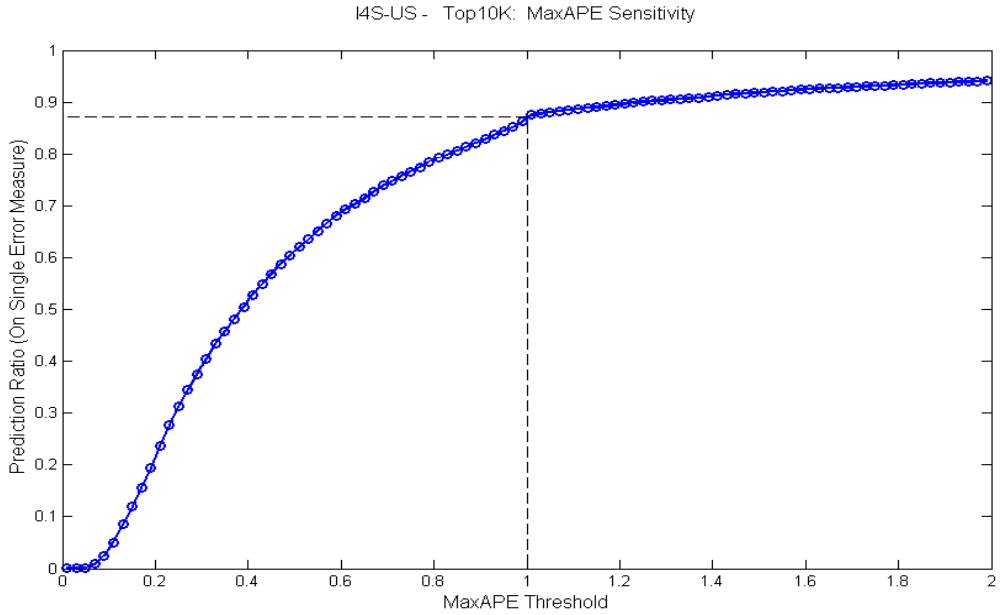
Sensitivity of the Predictability Thresholds.

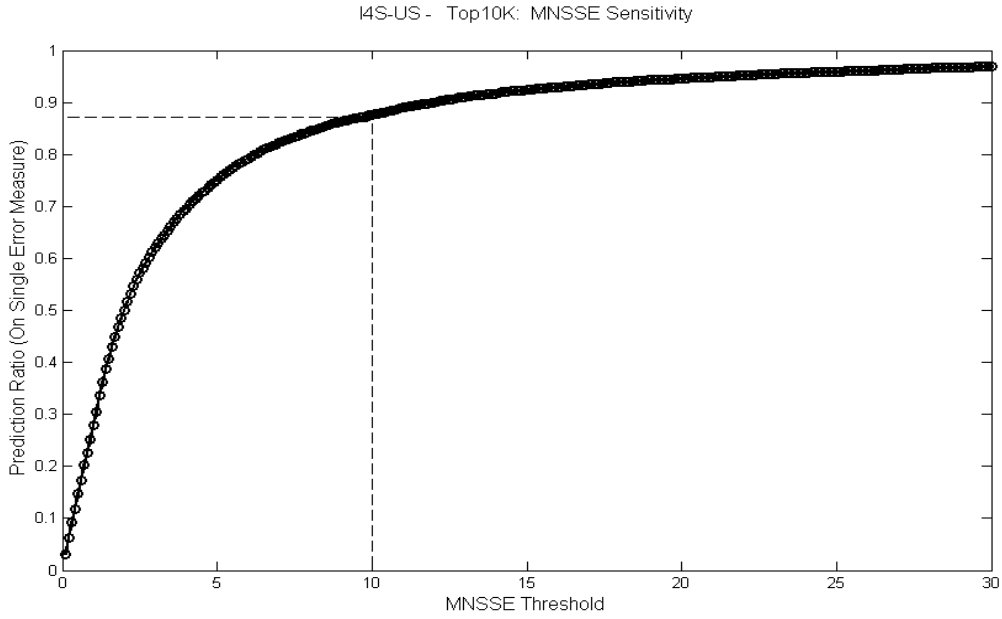
As described earlier, we have chosen a predefined set of thresholds which correspond to the three prediction error metrics (*MAPE*, *MaxAPE*, *NMSSE*) and two consistency metrics. These thresholds are responsible for the trade-off between the Predictability Ratio and the distribution of errors within the Predictable time series. In the following figure we see a sensitivity plot for the Mean Absolute Prediction Error (*MAPE*), that shows how the Predictability Ratio behaves as a function of the Predictability Threshold. We present a separate analysis for each error measure and not as a conjunction of all the conditions as appears in our Predictability definition.

The following plot shows that choosing a Predictability Threshold [*MAPE*<0.25] 'qualifies' more than 60% of the queries (for a single metric condition). Raising the *MAPE* threshold by 100% into 0.5, would imply that the Predication Ratio would rise by ~30% (using only the *MAPE* error metric). Raising the *MAPE* threshold even more, by 200% into 0.75, would imply that the Predication Ratio would rise by ~50% and will qualify approximately 90% of the queries.

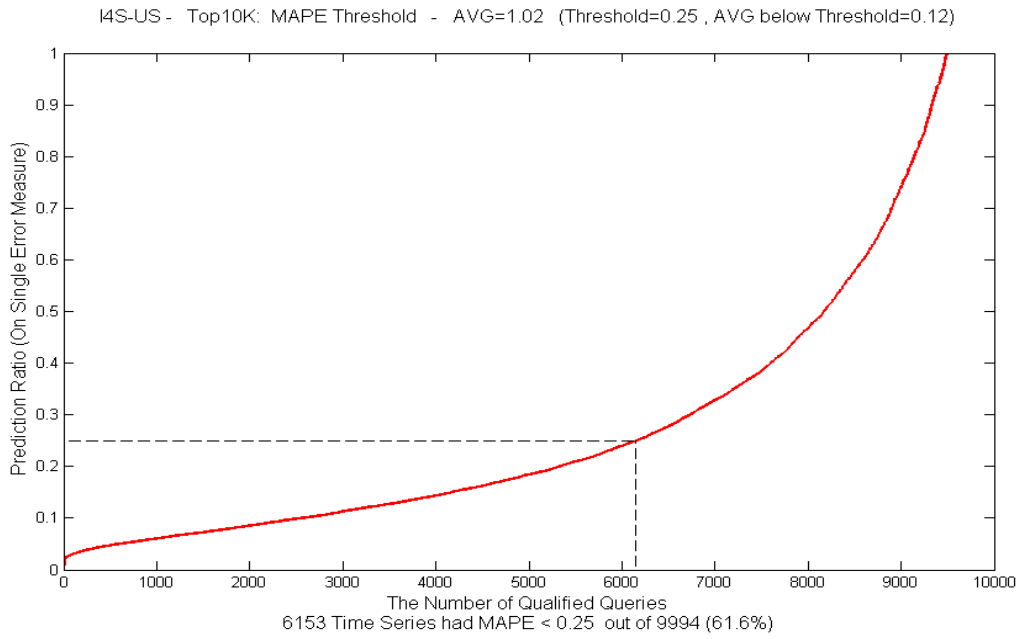


The next plots are the sensitivity plots for the *MaxAPE* and *NMSSE* error metrics. We can see that both chosen Predictability thresholds (1.0, 10.0) are located much farther into the "Predictable Region" and qualify almost 90% of the queries. Thus, in our experiments we use the *MAPE* as our primary 'filter' where the *MaxAPE* and the *NMSSE* play a secondary role.



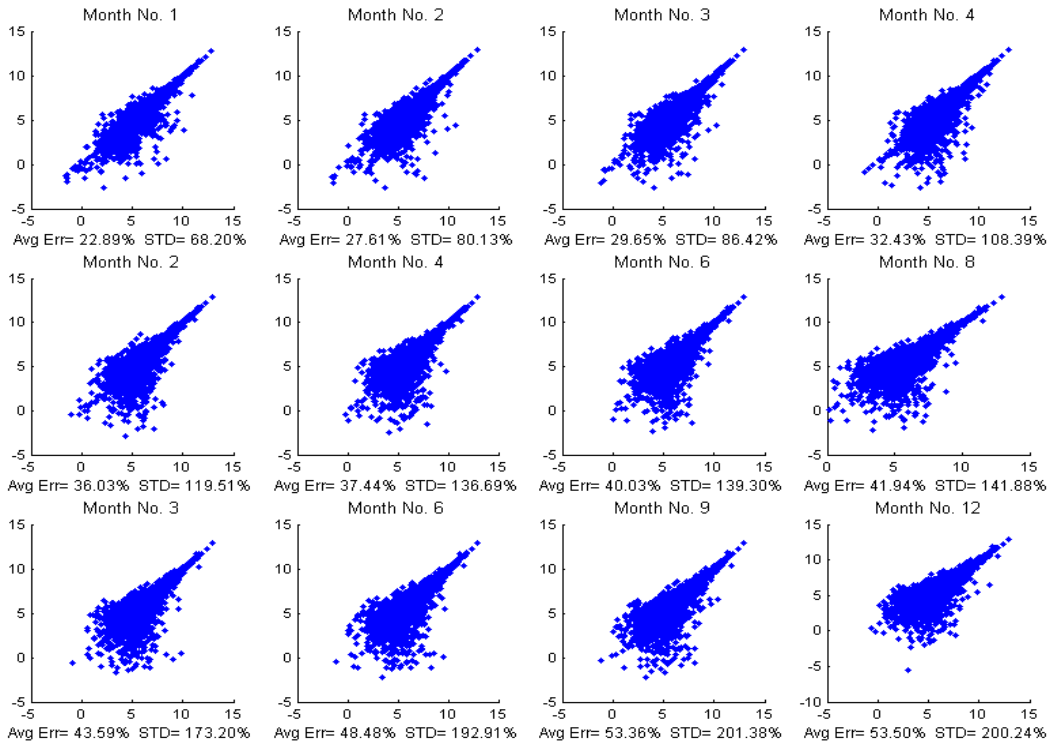


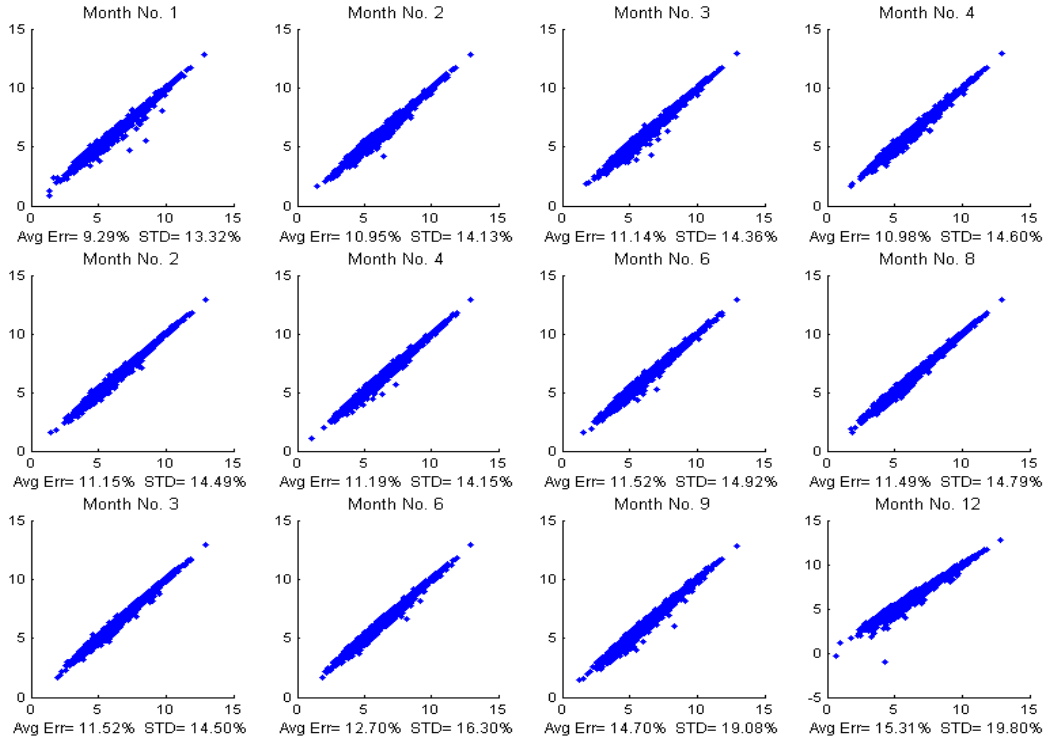
The following plot displays a similar presentation by showing the number of Predicable time series as a function of the Predictability Threshold (using only the *MAPE* error measure).



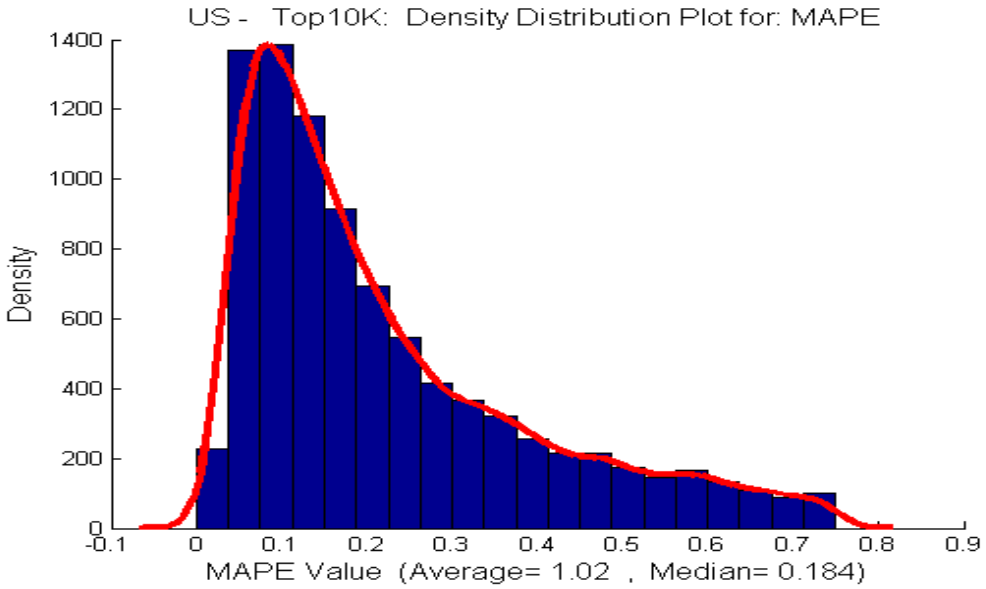
Prediction Errors Diagnostics.

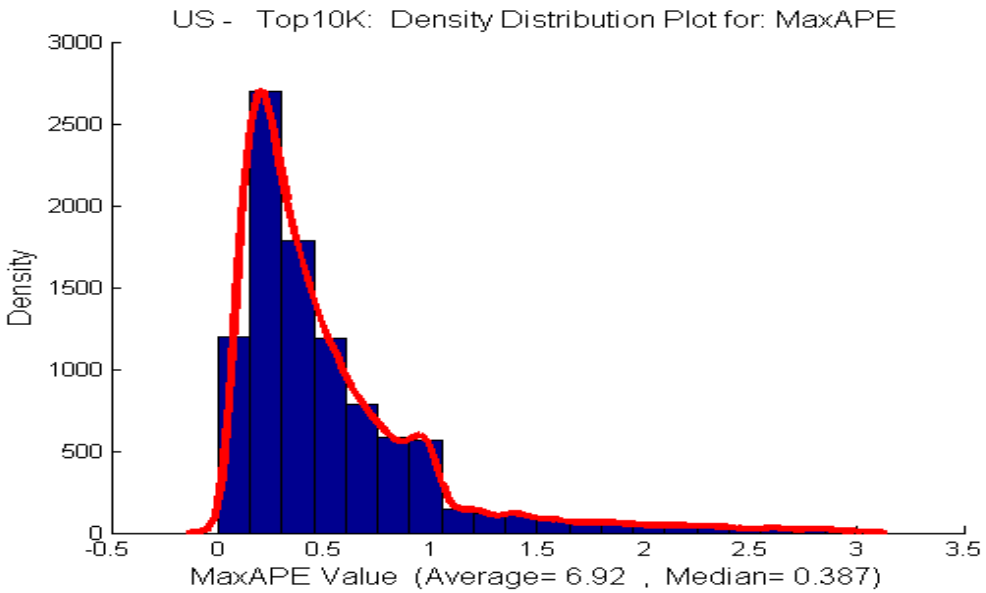
In this section we show diagnostics plots for the US data (top 10,000 queries). The following figure shows the actual values vs. the predicted values (in log scale), for each of the 12 months in the *Forecast Period*. The top 12 diagrams refer to the entire set of queries, followed by 12 diagrams for the Predictable queries only. One can clearly see the better prediction performance for the Predictable queries (at the bottom part) as expected. Notice that the performance for the different months deteriorates with time (higher average and STD of the prediction errors) especially towards the later months.


Month No. 1
Avg Err= 22.89% STD= 68.20%
Month No. 2
Avg Err= 27.61% STD= 80.13%
Month No. 3
Avg Err= 29.65% STD= 86.42%
Month No. 4
Avg Err= 32.43% STD= 108.39%
Month No. 2
Avg Err= 36.03% STD= 119.51%
Month No. 4
Avg Err= 37.44% STD= 136.69%
Month No. 6
Avg Err= 40.03% STD= 139.30%
Month No. 8
Avg Err= 41.94% STD= 141.88%
Month No. 3
Avg Err= 43.59% STD= 173.20%
Month No. 6
Avg Err= 48.48% STD= 192.91%
Month No. 9
Avg Err= 53.36% STD= 201.38%
Month No. 12
Avg Err= 53.50% STD= 200.24%


Month No. 1
Avg Err= 9.29% STD= 13.32%
Month No. 2
Avg Err= 10.95% STD= 14.13%
Month No. 3
Avg Err= 11.14% STD= 14.36%
Month No. 4
Avg Err= 10.98% STD= 14.60%
Month No. 2
Avg Err= 11.15% STD= 14.49%
Month No. 4
Avg Err= 11.19% STD= 14.15%
Month No. 6
Avg Err= 11.52% STD= 14.92%
Month No. 8
Avg Err= 11.49% STD= 14.79%
Month No. 3
Avg Err= 11.52% STD= 14.50%
Month No. 6
Avg Err= 12.70% STD= 16.30%
Month No. 9
Avg Err= 14.70% STD= 19.08%
Month No. 12
Avg Err= 15.31% STD= 19.80%

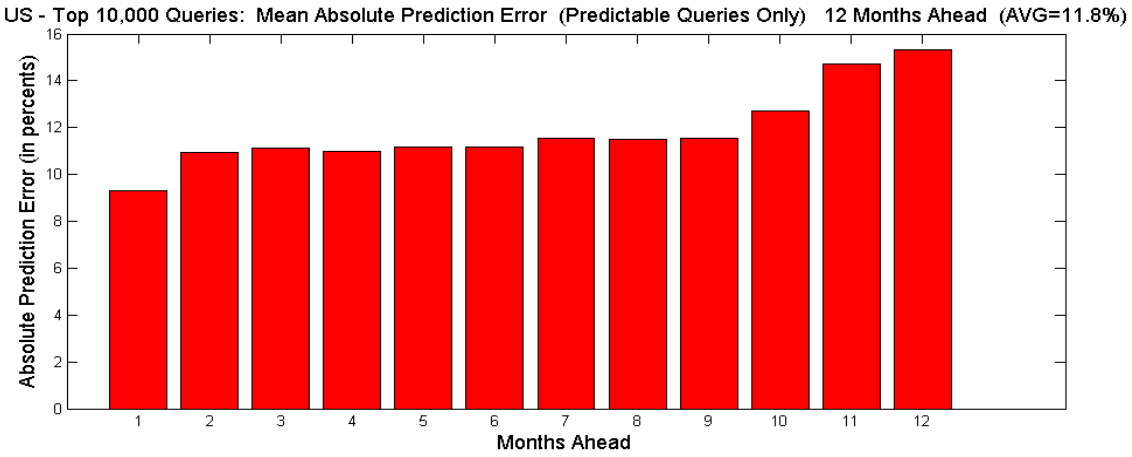
In order to learn more on the distribution of the average and maximum prediction errors within the top 10,000 most popular queries in the US, we present the histogram of the *MAPE* and *MaxAPE* error measures, with the density estimation superimposed (in red). We can see that both distributions are positively skewed and that the value of the average error is largely affected by the extreme error values. Notice that we have trimmed the data at 0.75 and 3.0 for MAPE and *MaxAPE* respectively (i.e., 3 x the chosen thresholds), to stay focused on the major part of the distribution.



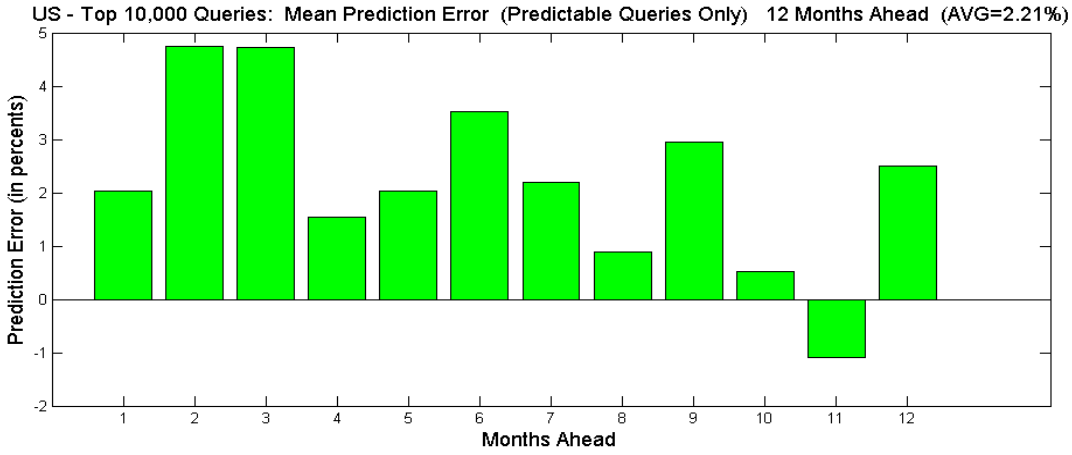


Comparison of the Forecast Performance along the Future Horizon.

Since in our experiments we are simultaneously predicting 12 month ahead, it is expected that the forecasts for the later months may have larger prediction errors. We have compared the prediction performance for the 12 consecutive month in the forecast period. The following plot shows the distribution of *MAPE* prediction errors for each future month. We are showing the average monthly *MAPE* for the Predictable queries only (among the 10,000 most popular in the US). Notice that the first month is predicted in greater accuracy than the rest, then there is an approximately constant error level for months 2-9, with some increase of the error rate in the last 3 months in the Forecast period.



The following plot shows the same type of diagram, but for the Mean Prediction Error (i.e., the 'directional' error measure with the sign). We can learn from this plot that there was a positive bias (upward) in the predictions along all months except the 11'th month. Such systematic tendency of the errors can be explained by a reduction of *query share* for many queries in the Forecast period (Aug 2008 - July 2009) due to the global economic crisis. Hence the actual search interest values were lower than expected by the prediction model that was based on the previous years.



In the following section we present examples of categories (and queries) regarding various markets and brands, for which the actual monthly *query shares* for the recent 12 months are different than model prediction.

6. Search Interest Forecasting as baseline for identifying deviations

The aggregated query share of the Google *Insights for Search (I4S)* categories were used in a recent work of Choi and Varian (2009), that showed how data taken from Google *I4S* could help to predict economic time series. For example, in the analysis on the US Retail Trade they have used the weekly aggregated time series of categories like: *Automotive, Computers & Electronics, Apparel, Sporting Goods, Mass Shopping, Merchants & Department Stores,* etc. In a later work [Choi and Varian 2009 b] have applied the same methodology on the U.S. unemployment time series using two sub-categories, *Jobs* and *Welfare & Unemployment*. They did not attempt to forecast the Google query share; rather, they have successfully used it as predictors for external economic time series. Other works have shown similar results, regarding the capability of aggregated categories' query share to predict econometrics and unemployment data from Germany [Askitas and Zimmerman 2009] as well as from Israel [Suhoy 2009].

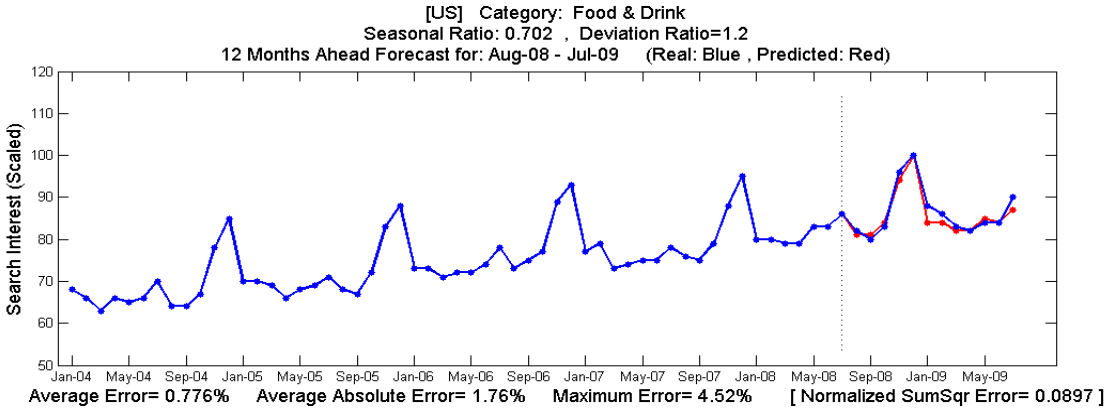
In the following, we will show time series of monthly query share of categories, where the forecast values (in red) were superimposed on the actual values (in blue). The errors made by the prediction model are expressing the deviation between the expected and the actual search behavior, which conveys a valuable information regarding the current state of search interest in the respective categories. Choi and Varian have shown that the users' search interest in several categories as represented by the aggregated query shares indeed have a short term predictive power regarding the actual underlying. The following plots show the aggregated time series of various categories that relate to some major US markets. These category plots, which are ordered by their average MAPE, vary in their Predictability level.

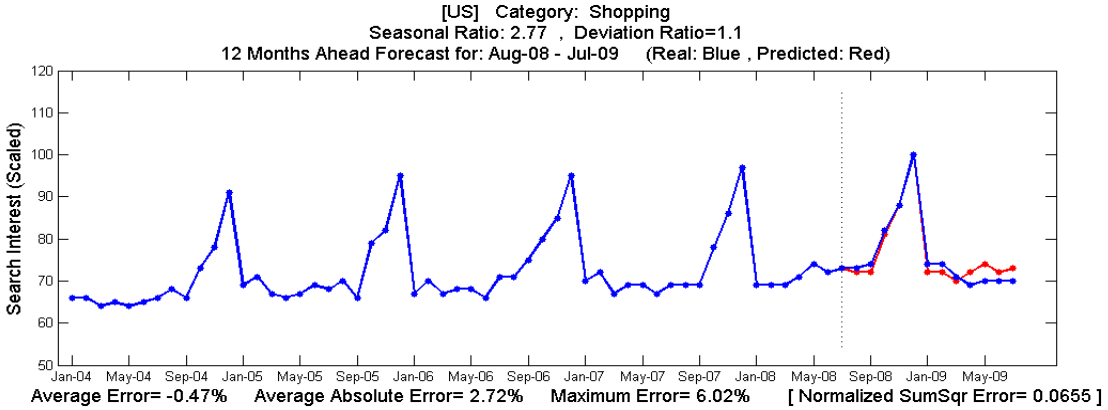
From the 10 category plots, we can see that many present a clear seasonal pattern. The first 7 time series showed a relatively low error rate (*MAPE*<6%), which is in accordance with the substantial regularity of search behavior of the respective categories that was maintained throughout the Forecast period. However, notice that the category of *Finance & Insurance* which shows a seasonal patterns with some medium irregularities (the seasonality ratio is well above its median), underwent a considerable change in the recent 12 months, highlighting an observed discrepancies between the predicted and the actual monthly search interest. The months of September-October 2008 which were low months in each year during the entire history period are observed as peak month in the Forecast period. This is an example where the prediction model could not anticipate the unexpected exogenous events.

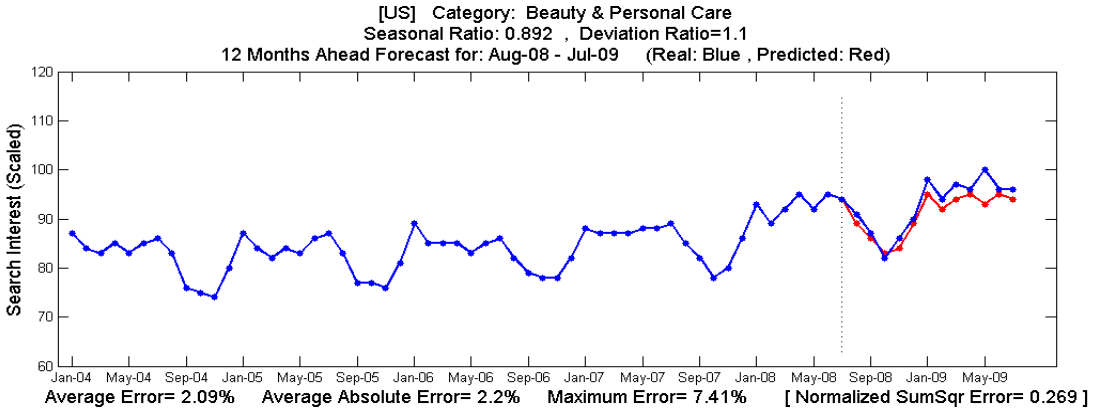
The category of *Energy and Utility* showed the most irregular search behavior (with the lowest Seasonality Ratio and the highest Deviation Ratio among the first 9 categories). In addition to the low regularity of its history, it seems that this category has also underwent a change in the dynamics of search interest, probably since mid year 2008. These contributed to the low prediction results for this category.

Another good example for lack of Predictability w.r.t. the prediction model, is the last plot of the *Social Networks & Online Communities* category that has shown a considerable exponential growth in the forecast period (due to the growing popularity of social networks like Facebook and Twitter), which could not be captured by the prediction model (notice the high deviation

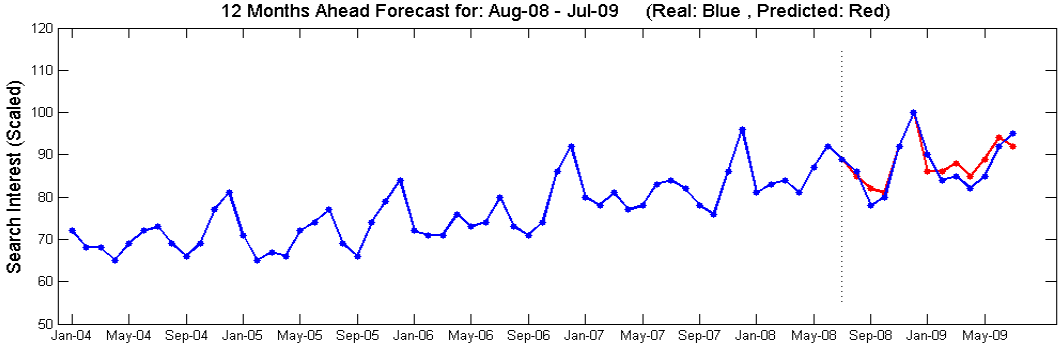
ratio). We will show below several other examples of the relation between the prediction performance and the external market events.





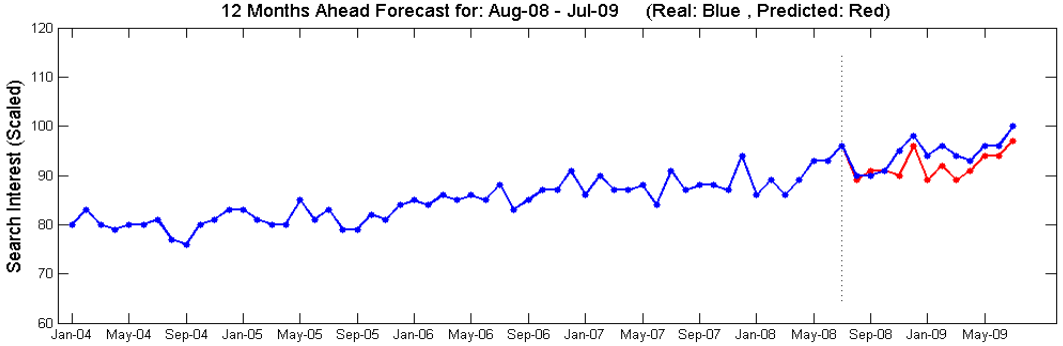


[US] Category: Games
Seasonal Ratio: 0.579 , Deviation Ratio=1.2
12 Months Ahead Forecast for: Aug-08 - Jul-09 (Real: Blue , Predicted: Red)



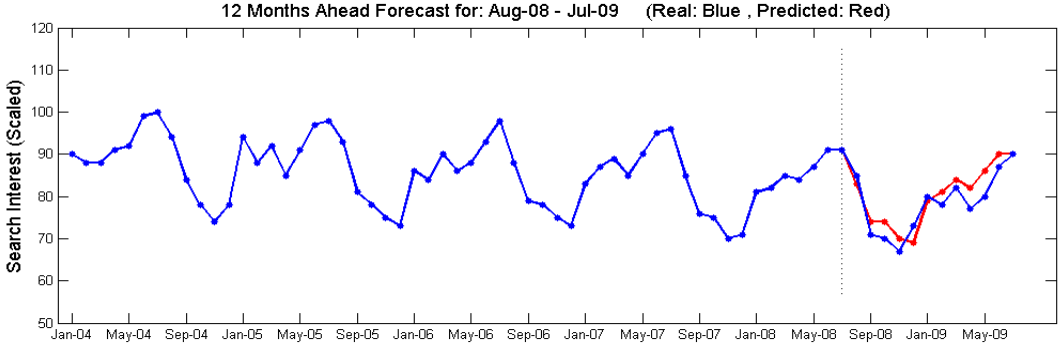
Average Error= -1.12% Average Absolute Error= 2.59% Maximum Error= 4.45% [Normalized SumSqr Error= 0.171]

[US] Category: Entertainment
Seasonal Ratio: 0.304 , Deviation Ratio=1.2
12 Months Ahead Forecast for: Aug-08 - Jul-09 (Real: Blue , Predicted: Red)



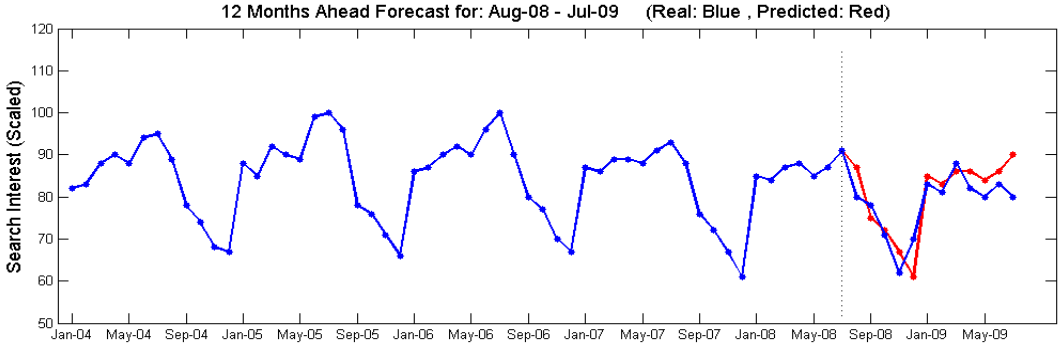
Average Error= 2.54% Average Absolute Error= 2.74% Maximum Error= 5.95% [Normalized SumSqr Error= 1.1]

[US] Category: Travel
Seasonal Ratio: 1.92 , Deviation Ratio=1.1
12 Months Ahead Forecast for: Aug-08 - Jul-09 (Real: Blue , Predicted: Red)



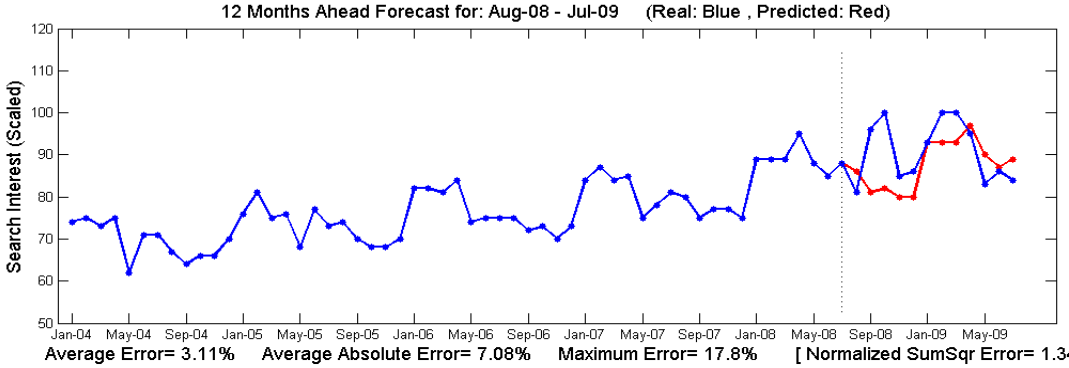
Average Error= -2.32% Average Absolute Error= 3.94% Maximum Error= 7.61% [Normalized SumSqr Error= 0.252]

[US] Category: Real Estate
Seasonal Ratio: 2.95 , Deviation Ratio=1.1
12 Months Ahead Forecast for: Aug-08 - Jul-09 (Real: Blue , Predicted: Red)



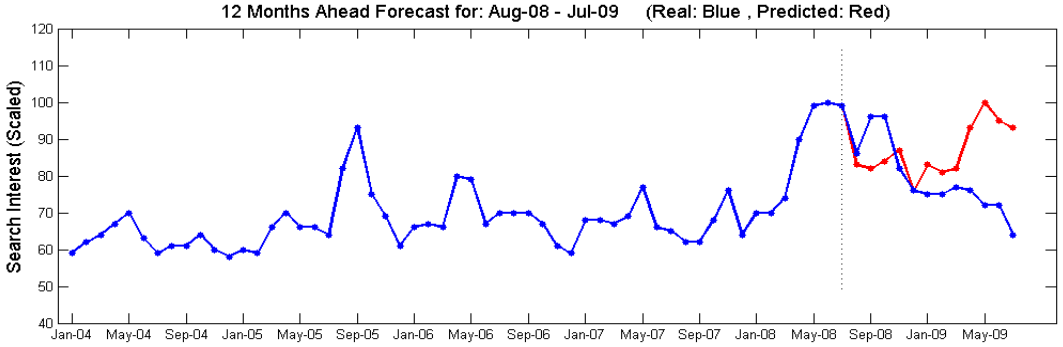
Average Error= -2.69% Average Absolute Error= 5.62% Maximum Error= 12.8% [Normalized SumSqr Error= 0.574]

[US] Category: Finance & Insurance
Seasonal Ratio: 0.611 , Deviation Ratio=1.3
12 Months Ahead Forecast for: Aug-08 - Jul-09 (Real: Blue , Predicted: Red)

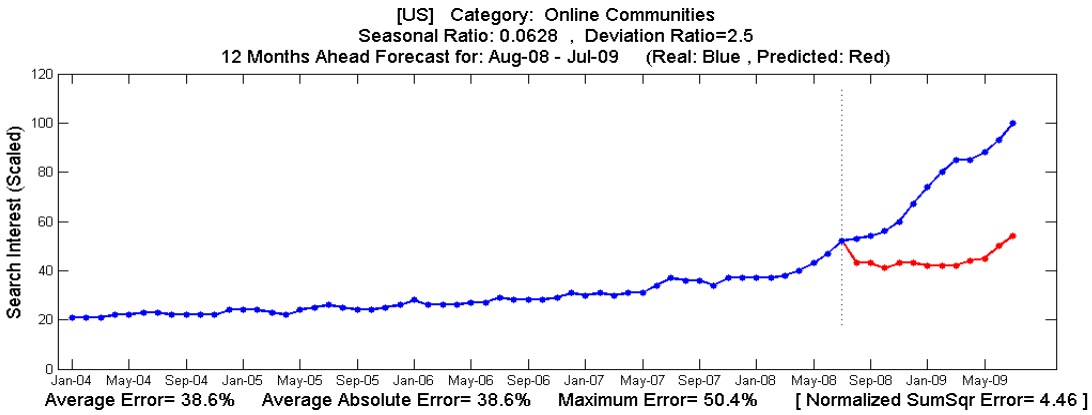


Average Error= 3.11% Average Absolute Error= 7.08% Maximum Error= 17.8% [Normalized SumSqr Error= 1.34]

[US] Category: Energy & Utilities
Seasonal Ratio: 0.49 , Deviation Ratio=1.4
12 Months Ahead Forecast for: Aug-08 - Jul-09 (Real: Blue , Predicted: Red)



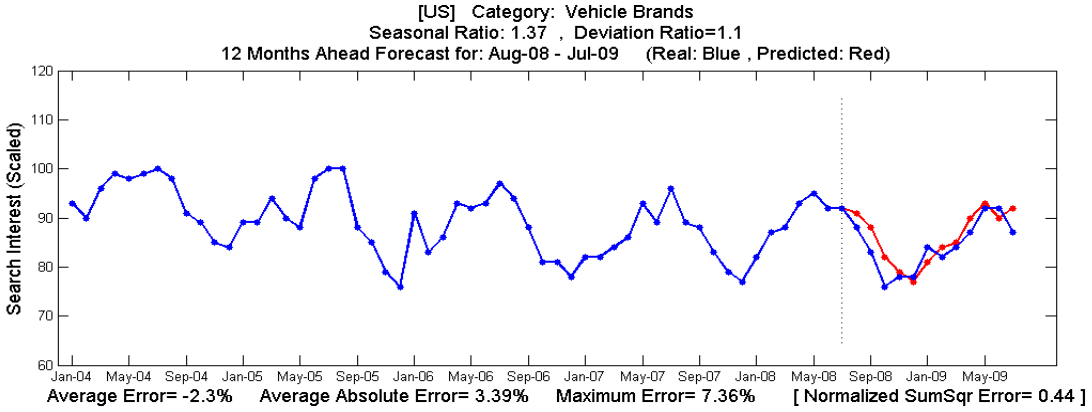
Average Error= -11.8% Average Absolute Error= 16.8% Maximum Error= 45.9% [Normalized SumSqr Error= 3.02]



Next, we show several examples where one can use the (posterior) prediction results in order to explore the changing dynamics of users' search behavior and possibly get insights on the relevant markets. Whenever we observe substantial prediction errors, i.e., discrepancies between the actual values vs the predicted values, we can conclude that the regularities in the time series (e.g., seasonality and trend) which were captured by the prediction model, were disturbed in the Forecast period. In cases where the actual values show a regularity that is not in accordance with the history's regularity, one could investigate the reasons for such deviation with relation to known external factors. It is important to emphasize that users' search interest is not necessarily always related to consumer preferences, buying intentions, etc. and can be related sometimes to news or or other associated events. A full discussion on the background and reasons for the following market observations is beyond the scope of this paper.

Example: The Automotive Industry.

We can see that the forecast for the entire Vehicle brands category for the 12 month period between Aug-08 and Jul-09[4] shows a relatively low prediction error rate of -2.3% on average. However, as we show below there are some noticeable deviations in different sub-categories.

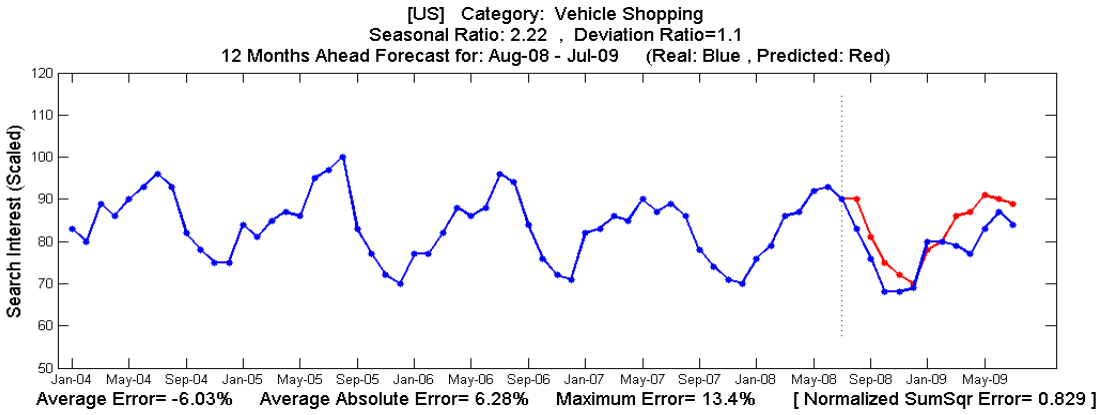


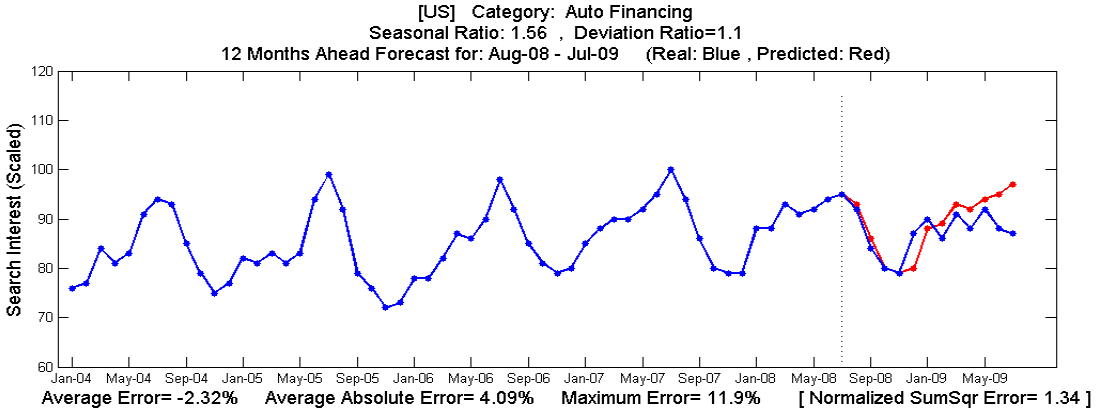
We can see in the next 4 plots that the category *Vehicle Shopping* shows an average negative deviation of 6% from the prediction model in the last 12 months and that the category *Auto*

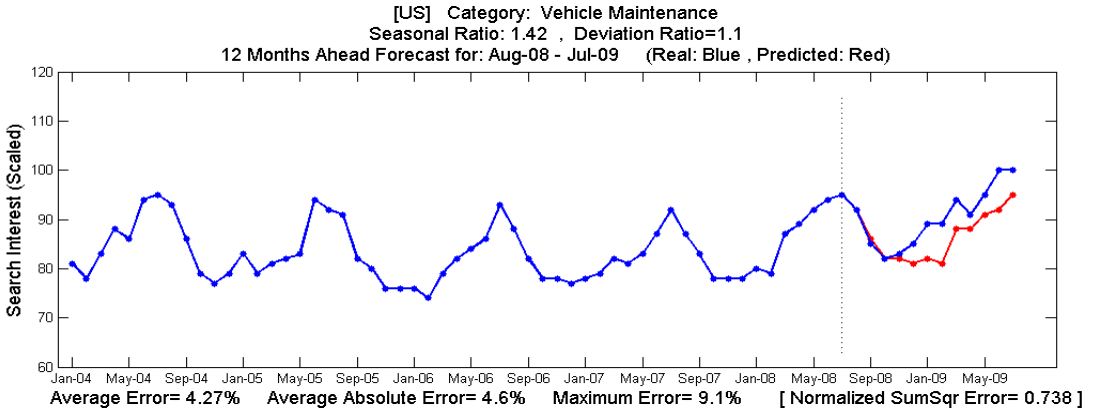
---

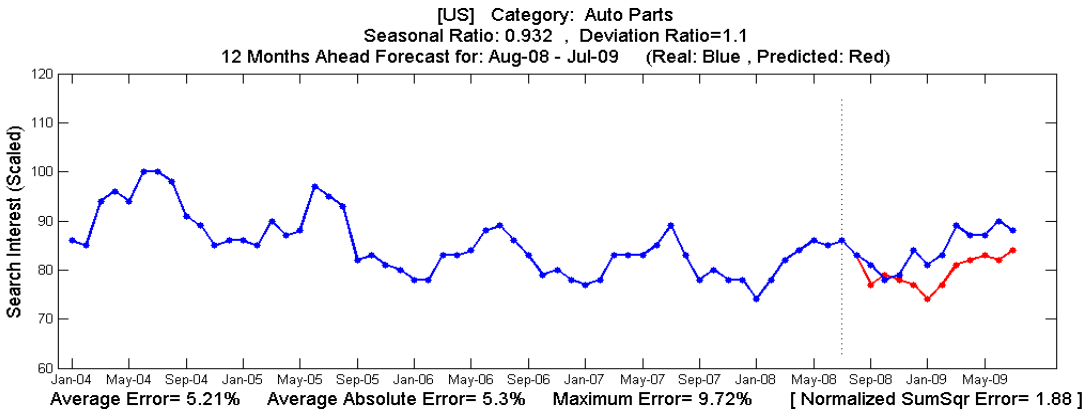
4. The time series data was pulled during July 2009, thus the value for this last month is partial and might be biased.

Financing is showing a small negative deviation with average of 2.3% respectively. Notice that both categories of *Vehicle Maintenance* and *Auto Parts* are showing a positive average deviation of 4.3% and 5.2% respectively, compared to the predictions.



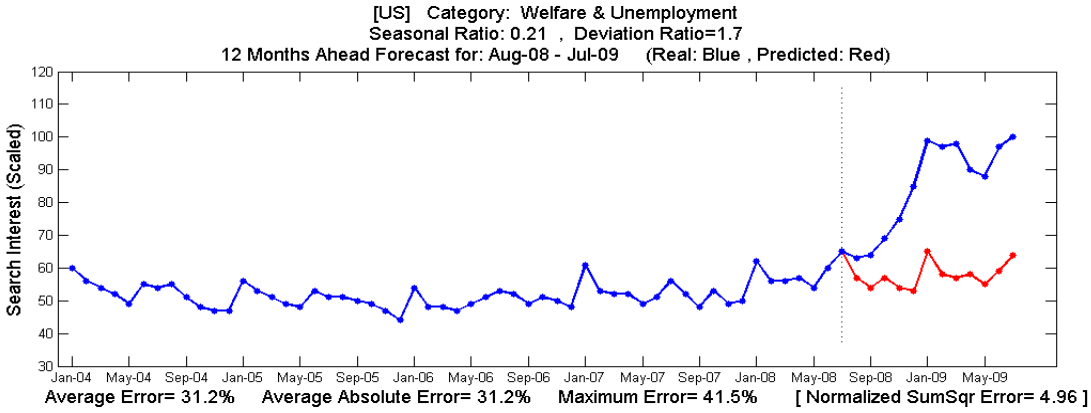




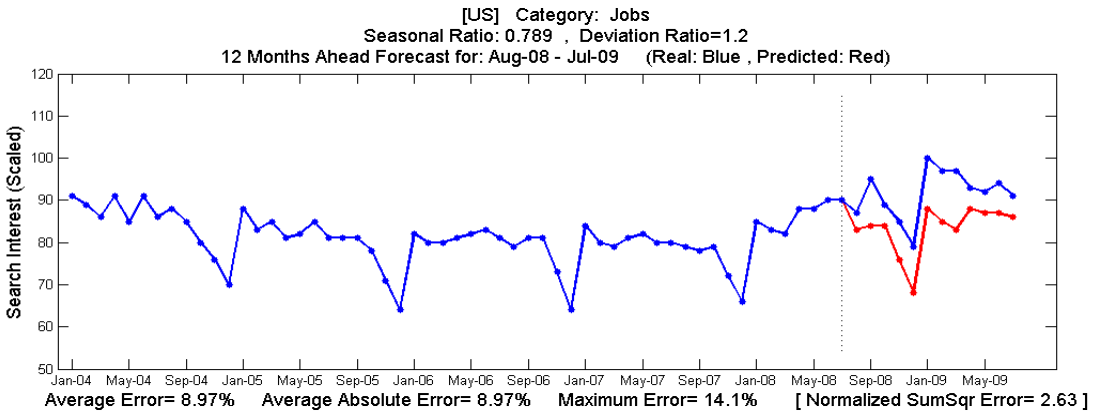


Example: US Unemployment.

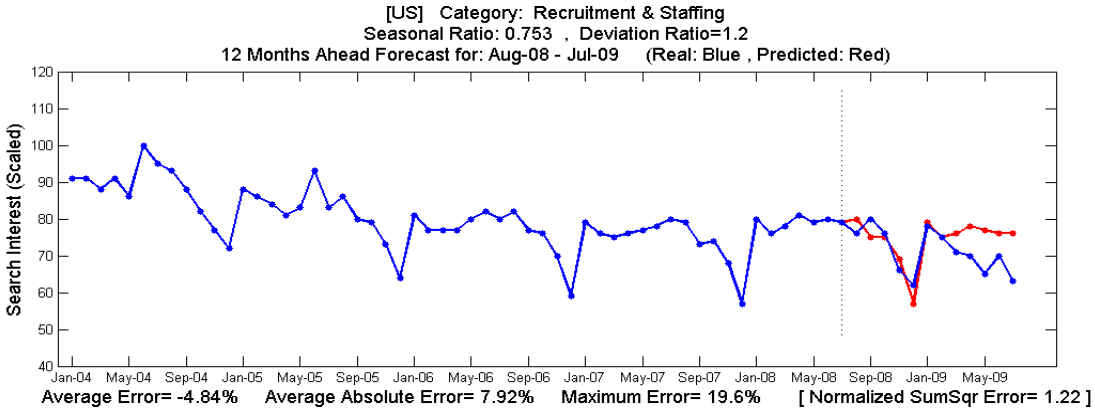
Choi and Varian (2009 b) have used weekly time series of the *I4S* aggregated categories *Welfare & unemployment* and *Jobs*, to help in short term prediction of "Initial Jobless Claims" reports which are issued by US Department of Labor. In the following plots, we show that the search interest the category *Welfare & Unemployment* has risen substantially above the forecast by the prediction model. The deviation of *Welfare and Unemployment* is systematic and relatively quite large. While the average *MAPE* for the entire set of (aggregated) categories' query shares is 8.1%, with STD 8.2%, the *MAPE* for *Welfare & Unemployment* is 31.2% which is 2.8 standard deviations above the overall average *MAPE*.



The actual monthly values for the aggregated query share of the category *Jobs* are also all higher than forecasted by the model. The time series shows a seasonal pattern with a distinguishable low value in December each year and a relatively constant level in between. At the end of the History period and throughout the Forecast, this regularity is shifted upwards by a confounding volatile factor, which causes large positive prediction errors. The Average Error is almost 9% per month.

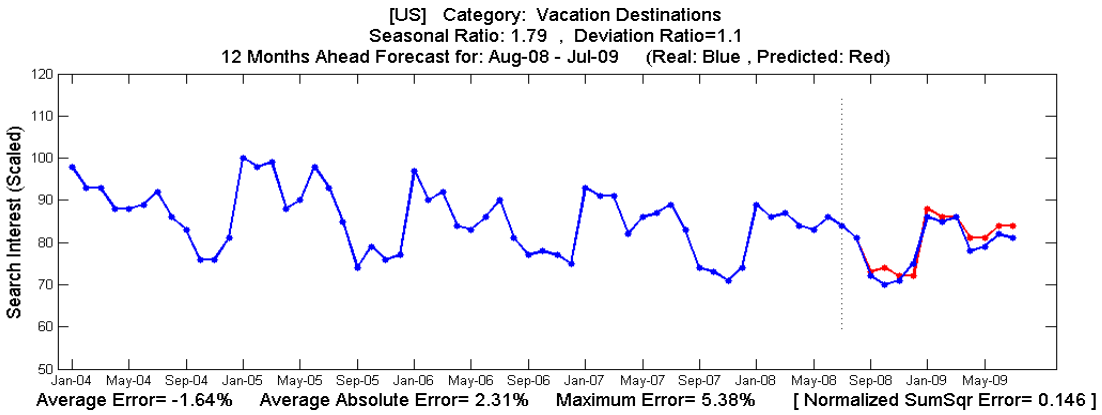


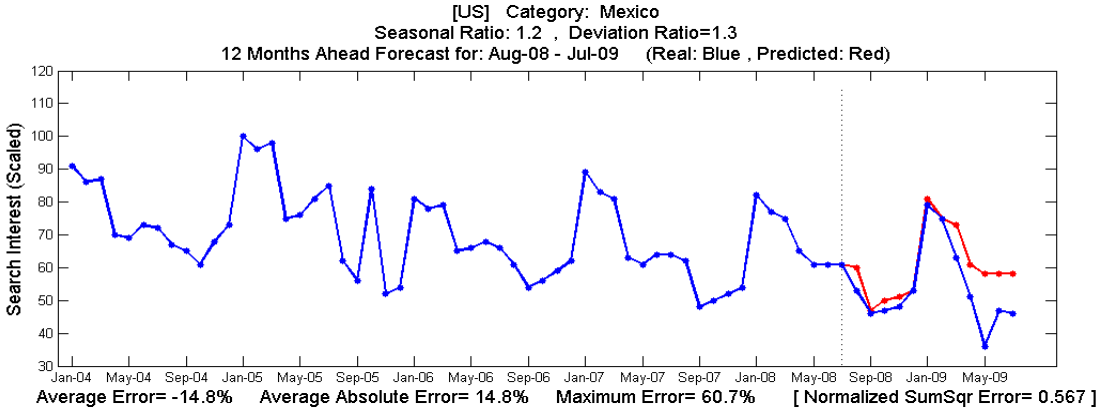
We present here also the aggregated query share of the category *Recruitment & Staffing*, for which we can observe a corresponding negative deviation where the model expectations are larger than the actual search interest values. Interestingly, despite a similar seasonal pattern as in the *Jobs* category, it seems that the change in the users' search behavior in this category has not started until March 2009. Beforehand the predictions were rather accurate and the average monthly deviation is therefore only about (-4.8%).

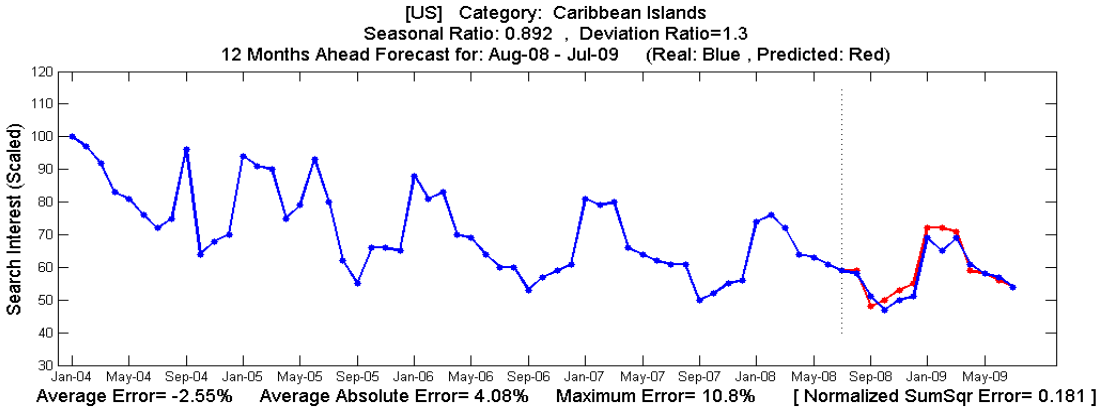


Example: Mexico as Vacation Destination.

In this example we show that the search interest for Mexico as a vacation destination has decreased substantially in the recent months. The I4S category *Mexico* is a sub-category of the *Vacation Destinations* category (in the *Travel* root category) which aggregates only the vacation related searches on Mexico. In the next plots we can see that the search interest in the category *Mexico* is down by almost 15% compared to the predicted. In comparison, we show the respective deviation in the entire category of *Vacation Destinations*, which is only -1.6% on average in the same forecast period. Notice for a reference that the search interest of another related vacation destination, the *Caribbean Islands* (with a similar seasonal pattern), also has not shown a deviation of similar magnitude (only -2.5%).

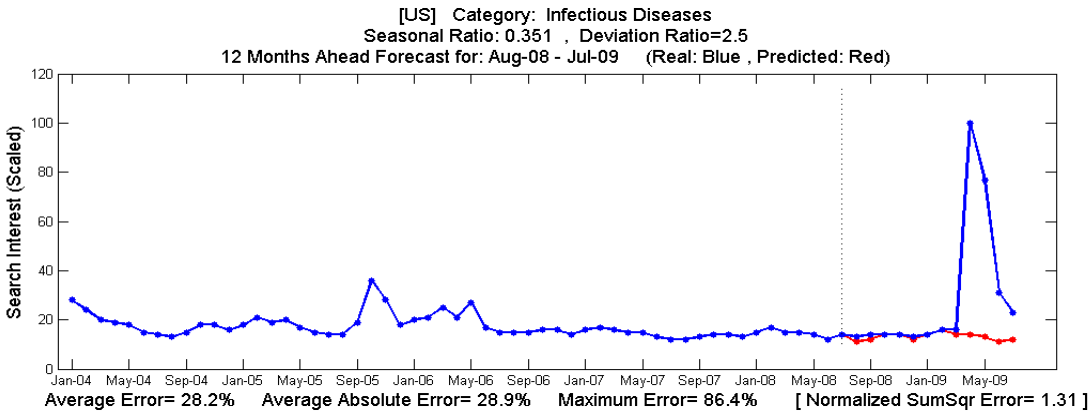






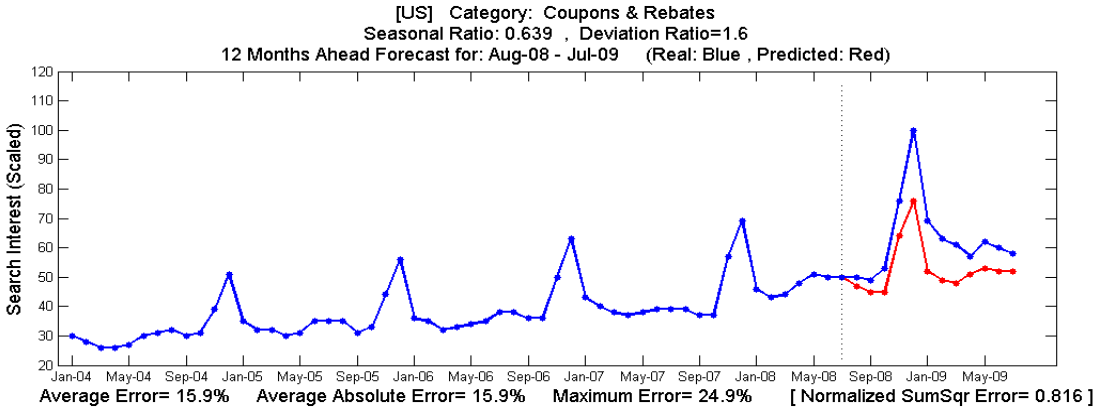
We considered the recent outbreak of the *Swine Flu* pandemic that started to spread in April 2009 as a possible contributor for such a negative deviation of *actual-vs-forecast* query share for *Mexico.* We examined the time series of the query share for *H1N1* and found it to be highly (*anti*) correlated (*r = -0.93*) with the observed deviations for *Mexico.*

As a reference, we show the aggregated query share for the category *Infectious Diseases*, demonstrating the magnitude of the search interest in this subject (in blue) that was spiking following the *Swine Flu* outbreak:

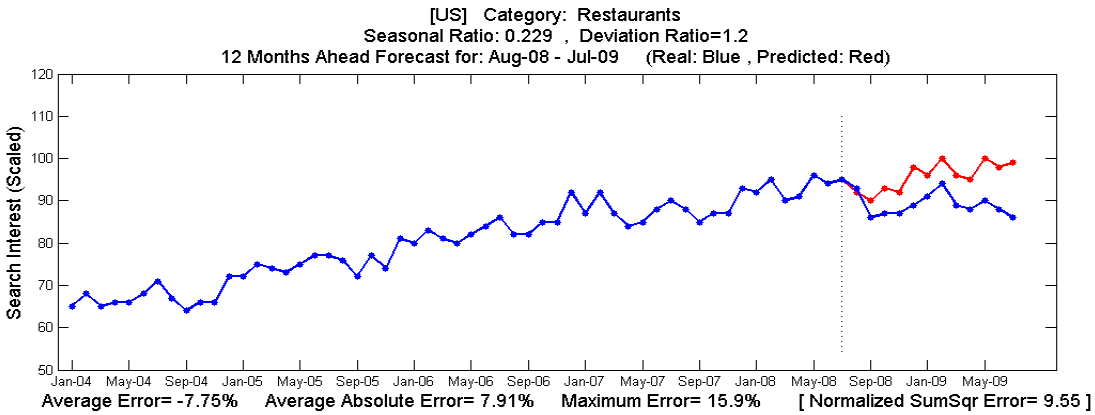


Example: Recession Markers.

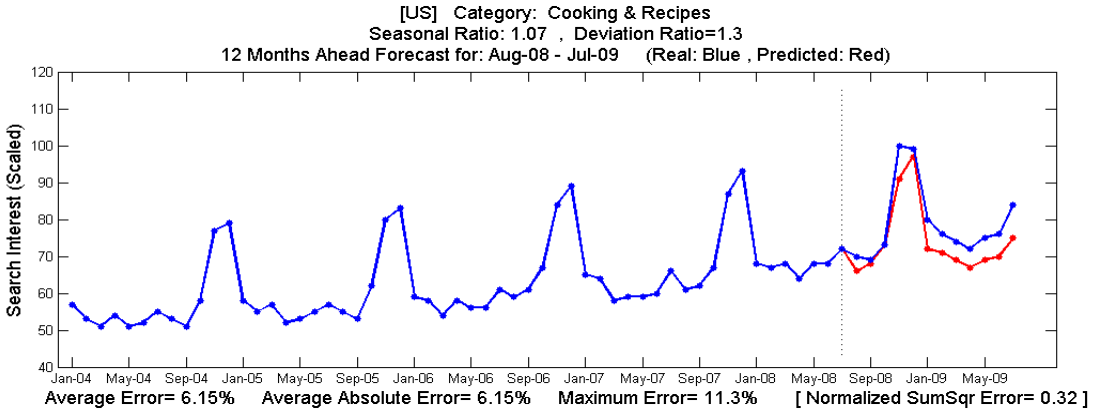
The following plots present the aggregated query share for some *I4S* sub-categories in subjects that might demonstrate the influence of the recent recession on search behavior of consumers, and often appear in articles and blog posts. The change in search interest for the category *Coupons & Rebates* is visible in the following plot, where one can see an average monthly deviation of 15.9% between the observed query share in the recent 12 month compared to the values predicted by the model. The model has captured the general seasonal pattern, however only accounted for a lower holidays peak and a much more moderate upward trend.



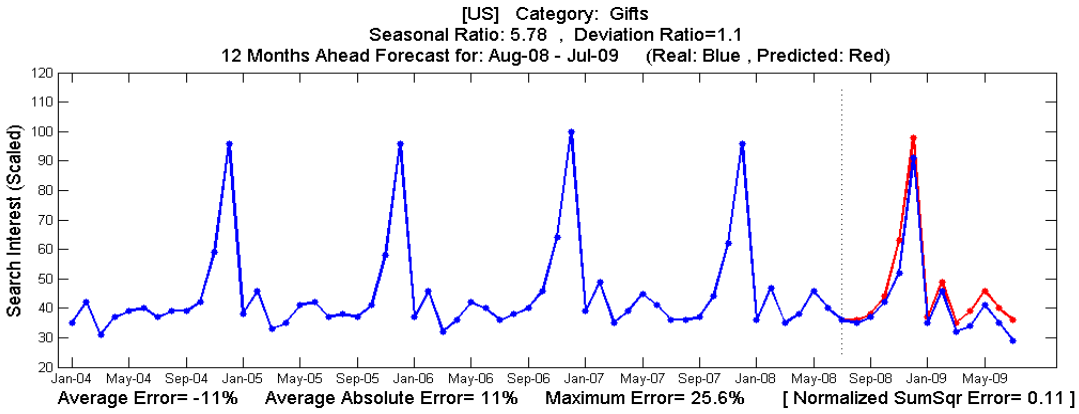
Next we see the observed query share of the *I4S* category *Restaurants,* that is systematically lower than the model predictions. The time series for the aggregated search interests in this category does not show a seasonal pattern, however there exist an upward trend since 2004, which was apparently broken at September 2008 hence causing negative *actual-vs-forecast* deviation with a an average of -7.8% per month.

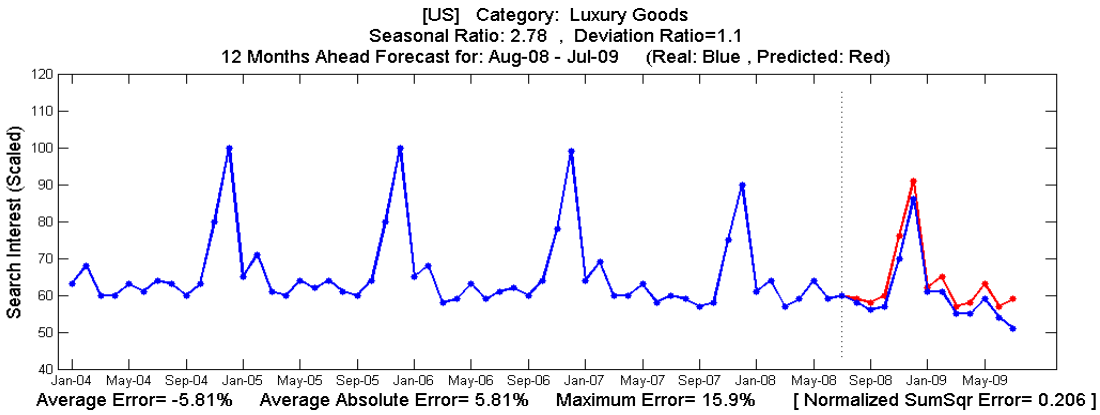


Below we can see for reference that the *Cooking & Recipes* category has a systematic positive deviation of *actual-vs-forecast* query share. The average monthly deviation of 6.15% represents a higher observed search interest in this category for the entire Forecast period compared to model prediction, with almost a constant deviation since January 2009.



Another example is the category *Gifts,* for which the query share has decreased in the recent 12 months compared to the model predictions, by 11% per month on average. Below we can also see that the category *Luxury Goods* is showing a negative deviation in the *actual-vs-forecast* query share, of 5.8% per month on average.





7. Conclusions

We studied the predictability of search trends. We found that over half of the most popular Google search queries are predictable w.r.t. the method we have selected, and that several search categories were considerably more predictable than others; that the aggregated queries of the different categories are more predictable than the individual queries and that almost 90% of I4S categories have predictable query shares. In particular we showed that queries with seasonal time series and lower levels of outliers are more predictable.

We considered forecasting as a baseline for identification of deviation of actual-vs-forecast, and considered some concrete examples for situations from the automotive, travel and labor verticals.

Further research can include an improved implementation of the prediction model as well as incorporating other forecasting models. We would also like to examine short-term forecasting in finer time granularity. Further analysis on actual-vs-forecast (including confidence estimation) could be conducted in various domains, like market analysis, economy, health, etc.

In conjunction with this study, a basic forecasting capability was introduced into Google Insights For Search, which provides forecasting for trends that are identified as predictable. Researchers, marketers, journalists, and others, can use *I4S* to get a wide picture on search trends which now also includes predictability of single queries and aggregated categories in any area of interest.

Acknowledgments
We would like to thank *Yannai Gonczarowsky* for designing and implementing the forecasting capabilities in I4S as well as *Nir Andelman*, *Yuval Netzer* and *Amit Weinstein* for creating the forecasting model library. We thank *Hal Varian* for his helpful comments. Special thanks to the entire team of *Google Insights for Search* that made this research possible.

References

[Askitas and Zimmerman 2009] Nikos Askitas and Kalus F. Zimmerman.
Google econometrics and unemployment forecasting. Applied Economics Quarterly, 55:107;120, 2009. URL http://ftp.iza.org/dp4201.pdf

[Choi and Varian 2009] Hyunyoung Choi and Hal Varian.
Predicting the present with google trends. Technical report, Google, 2009.
URL http://google.com/googleblogs/pdfs/google_predicting_the_present.pdf.

[Choi and Varian 2009b] Hyunyoung Choi and Hal Varian.
Predicting Initial Claims for Unemployment Insurance Using Google Trends. Tech. Report, Google, 2009. URL http://research.google.com/archive/papers/initialclaimsUS.pdf

[Cleveland and Tiao 1976] W.P. Cleveland and G.C. Tiao.
Decomposition of Seasonal Time Series: A Model for the Census X-11 Program, Journal of the American Statistical Association, Vol. 71, No. 355, 1976 pp. 581-587.

[Cleveland etal. 1990] R.B Cleveland, W.S. Cleveland, J.E. McRae and Irma Terpenning.
STL: A Seasonal-Trend Decomposition Procedure Based on Loess. Jou. of Official Stat., VOL. 6, No. 1, 1990 pp. 3-73.

[Ginsberg etal. 2009] Jeremy Ginsberg, Matthew H. Mohebbi, Rajan S. Patel,
Lynnette Brammer, Mark S. Smolinski & Larry Brilliant.
Detecting influenza epidemics using search engine query data. Nature 457, 1012-1014 (2009).
URL http://www.nature.com/nature/journal/v457/n7232/full/nature07634.html

[Lytras etal. 2007] Demerta P. Lytras, Roxanne M. Felpausch, and William R. Bell.
Determining Seasonality: A Comparison of Diagnostics From X-12-ARIMA (Presented at ICES III, June, 2007).

[Suhoy 2009] Tanya Suhoy.
Query indices and a 2008 downturn: Israeli data. Tech. Report, Bank of Israel, 2009.
URL http://www.bankisrael.gov.il/deptdata/mehkar/papers/dp0906e.pdf