Can product sales be explained by internet search traffic?  

The case of video games sales

Oliver Schaer  
Nikolaos Kourentzes  

Lancaster Centre for Forecasting
Forecasting challenges:

- Fast technological lifecycles
- High and fast-paced (often global) competition on technological, provider and application level
- Abrupt change in consumer and user behaviour
- Often very limited data availability in particular for new product launches

Can search traffic popularity help us to understand sales and improve forecasts?
Google Trends as an indicator

Example on a technology search query

www.google.com/trends
Incorporating search traffic data into forecasting models

Existing research:

- Adoption of hybrid cars (Jun 2012)

- Automobile sales, unemployment claims, travel destination planning and consumer confidence (Choi and Varian 2012, 2009a&b)

- Box-office revenue, video game sales and billboard rank (Goel et al. 2010)

- Box-office revenue (Kulkarni et al. 2011)

- Hotel room demand (Pan et al. 2012)

- Popularity of online social networks (Franses 2014; Bauckhage and Kersting 2014; Cannarella and Spechler 2014)
Incorporating search traffic data into forecasting models

Models used contained ARX(1,1), bivariate and multivariate regression with non-univariate inputs.

Little has been done in relation to product sales.

No attention paid to lead order selection of inputs.

Models are often not well assessed on their forecasting performance.

- i.e. lack of adequate benchmark models
Research questions

I. Dependency of sales on online traffic? Is there causal relationship?

II. Is search traffic information leading, lagging or contemporaneous?

III. Is the causality consistent over the product lifecycle?

IV. Does it offer any improvement in forecasting accuracy?
How will we attempt to answer the questions

Use sales data from the video games industry (across lifecycle of products)

Experimentally identify best lead or lag order of search data for describing product sales (across lifecycle)

Test whether leading search information is useful in improving the forecasting accuracy over the Christmas period
Datasets used

Video game sales from VGchartz
- Global physical sales information of 100 popular game titles launched between 2005 and 2014 at a weekly frequency
- Aggregate across various gaming platforms such as PC, XBox, PS3 or Wii

Search Traffic Popularity from Google Trends
- Weekly global search traffic popularity information
- Game title used as search traffic keyword
Google Trends data vs. Actual Sales

**Call of Duty 3**

- **Actual Sales**
- **Google Trends**

**Wii Sports**

- **Actual Sales**
- **Google Trends**
Experiment I – Motivation

Target
- Find the lead/lag with the highest impact for a given point on the product lifecycle

In order to:
- Observe the frequency of leads and lags over the lifecycle
- Identify if the selection of lead/lag changes over time
- Investigate if there is an effect for titles that have predecessors

Note that our aim is to identify the lag structure, not build the best model yet.
Experiment I – Design

Linear Regression model

\[ y_t = \beta_0 + \beta_1 x_{t-i} + \epsilon \]

\( y_t = \) Video Game Sales \hspace{1cm} \( x_{t-i} = \) Google Trends \hspace{1cm} \( i = \) \{\
  \begin{align*}
  -5, \ldots, -1, & \quad \text{Lead} \\
  0, & \quad \text{Contemporaneous} \\
  1, \ldots, 5 & \quad \text{Lag}
  \end{align*}\}

\( AIC \) as model evaluation criteria

(Montgomery et al. 2012; Hyndman and Khandakar 2008)

Rolling life-cycle window with fixed size of 12 observations (3 Months – quarterly performance)
Experiment I – How is it calculated?
Experiment I – Results

Change of the lead/lag structure over the lifecycle

Leading and lagging phases during the lifecycle: GTA San Andreas
Experiment I – Results

Shift from short lead to long lead over time

Percentage of Leads and Lags over time
Experiment I – Results

New released games series titles tend have longer lead during the launch phase

Assissin's Creed Serie

Battlefield Serie

Game launches within serie

Game launches within serie

Week since launch

Week since launch

Lag 5

Lead 5
Experiment II - Motivation

Target
- Forecast sales for the Christmas period (4 weeks) with games launched between January and June
- Focus on Christmas because of peak sales

In order to:
- Evaluate whether search traffic popularity data can increase the short-term forecasting accuracy after the game has already been launched.
Experiment II – Design

Total of 26 series

In-sample consists of 12 observations

Forecasting horizon set to 4 weeks

Validation of 4 short-term forecasting models (e.g. Choi and Varian 2012; Pan et al. 2012; Goel et al. 2012)

- AR(p) model
- ARX(p,1) with search traffic data as explanatory variable
- LM with search traffic data as explanatory variable
- Naïve
Experiment II – Results

<table>
<thead>
<tr>
<th>Error Metric</th>
<th>Naïve</th>
<th>AR(p)</th>
<th>LM</th>
<th>ARX(p,1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MASE</td>
<td>3.671</td>
<td>4.652</td>
<td>5.89</td>
<td>3.607</td>
</tr>
<tr>
<td>GMRAE</td>
<td>1.000</td>
<td>1.216</td>
<td>1.6</td>
<td>1.005</td>
</tr>
</tbody>
</table>

ARX(p,1) outperforms AR(p)
Naïve not outperformed substantially

Is it worth to consider search engine data?
Conclusion

I. Search traffic popularity is highly correlated with video game sales

II. Leading information contained in many cases. Order changes across lifecycle

III. Short leading information in launch phase; long leads towards end of lifecycle

IV. Limited forecasting accuracy gains on initial experiments

V. Selection of search term can be difficult
Further Research Questions

What is the value of search data for the launch phase?

Alternative or further online sourced explanatory variables i.e. views from video trailers on YouTube

An important question in the industry is when to launch a title. Can online search information help?
Thank you!

o.schaer@lancaster.ac.uk

@oliverschaer
References


