Forecasting diffusion with pre-launch online search traffic data

Oliver Schaer
Nikolaos Kourentzes
Robert Fildes

Higher School of Economics
Saint Petersburg
25th May 2016

Lancaster Centre for Forecasting
Lancaster University Management School
Current challenges for forecasting product sales

Fast technological lifecycles

High and fast-paced (often global) competition on technological, provider and application level

 Abrupt change in consumer and user behaviour

Very limited data availability new product launches

Can search traffic popularity help us to improve sales forecasts?
The video game industry

Worldwide games revenue around £100 billions
Creates roughly 5% of the global entertainment revenue
U.S. companies alone employ more than 40,000 people
Challenges in the industry

Increasing developing costs

Video games are getting as expensive as movies

<table>
<thead>
<tr>
<th>Game</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Destiny 2014</td>
<td>£310m</td>
</tr>
<tr>
<td>Grand Theft Auto V 2013</td>
<td>£164m</td>
</tr>
<tr>
<td>Star Wars: The Old Republic 2011</td>
<td>£124m</td>
</tr>
<tr>
<td>Call of Duty Modern Warfare 2 2009</td>
<td>£124m</td>
</tr>
<tr>
<td>Final Fantasy VII 1998</td>
<td>£90m</td>
</tr>
<tr>
<td>Max Payne 3 2012</td>
<td>£65m</td>
</tr>
<tr>
<td>Tomb Raider 2013</td>
<td>£62m</td>
</tr>
</tbody>
</table>

Challenges in the industry

Shift towards the digital format

Recent Digital and Physical Sales Information

ESA, 2015. Facts, Sales, Demographic and Usage Data
The game life-cycle

**Call of Duty 3**

- Actual Sales
- Google Trends

**Wii Sports**

- Actual Sales
- Google Trends
Generations of games

Sales for Assassin's Creed

Google Trends

Sales

Peak scaled search traffic

j=1 j=2 j=3 j=4 j=5 j=6 j=7

0 40 80

j=1 j=2 j=3 j=4 j=5 j=6 j=7

0 600000
Using information from online sources

Using online explanatory variables for pre-launch forecasts

- Forecasting computer game sales using search traffic and social buzz for the opening sales (Xiong & Bharadwaj 2014)
- Box office sales using social network data (Kim et al. 2015) and search traffic (Kulkarni et al. 2012)

No application to life-cycle forecasting and parameter estimation for diffusion models.

I. Can search traffic data help in estimating diffusion model parameters?
What is a diffusion model

Measuring the adoption of a new product

- A common model in marketing is the Bass (1969) model
- Parameters for Innovators, Imitators and Market size
Ways to obtain model parameters in a pre-launch setting*

By judgement

- However, time consuming and problem with adjustment bias (Fildes et al. 2009)

Forecasting by analogy

- Using parameters from previous or similar products (Kim et al. 2014, Lillien et al. 2000, Norton & Bass 1987)

Market research

- Survey (Bass et al. 2001), Product attributes (Goodwin et al. 2012) or Pre-orders on CD albums (Moe & Fader 2002)

*See Goodwin et al. (2014) for a discussion on challenges with pre-launch forecasting
Experiment – Motivation

Model
- Bass model

Target
- Incorporate search traffic information into the analogy based forecast approach.

Aim
- Estimate market size parameter
- We are also interested in seeing whether there is lead time
The data

Video game sales from VGchartz
- Global physical sales at weekly frequency
- Using 6 games series with a total of 43 games
- Sales are aggregated across gaming platforms such as PC, Xbox, PS3 or Wii

Search Traffic popularity from Google Trends
- Weekly global search traffic popularity information
- Topic search with game title as keyword
Estimation process

Sales

- Actuals
- Forecast

Search traffic

Peak scaled search traffic

- j=1
- j=2
- j=3
- j=4

Sales

j=1 j=2 j=3 j=4
Sales
0 40 80

Search traffic

j=1 j=2 j=3 j=4
Search traffic

t
Benchmark models and accuracy measure

- Naïve: \( m_j = m_{j-1} \)
- Naïve + Difference: \( m_j = m_{j-1} + \Delta m_{j-1} \)
- Linear Trend: \( m_j = \alpha_0 + \alpha_1 j + \epsilon_j \)
- AR(1): \( m_j = \alpha_0 + \alpha_1 m_{j-1} + \epsilon_j \)

+ “optimal” fitted Bass model with actuals

Actuals contain two years of data

Numbers of generations needed for model estimation vary

Relative Mean Absolute Error and median across series
Google Trend model selection

Percentage Increase:

Linear trend:

AR(1) + Percentage Increase:

AR(1) + Google Trend:

\[
m_j = \frac{GT_j}{GT_{j-1}} \ast m_{j-1}
\]

\[
m_j = \alpha_0 + GT_j + \epsilon_j
\]

\[
m_j = \alpha_0 + \alpha_1 m_{j-1} + \beta_1 \frac{GT_j}{GT_{j-1}} + \epsilon_j
\]

\[
m_j = \alpha_0 + \alpha_1 m_{j-1} + \beta_1 GT_j + \epsilon_j
\]

<table>
<thead>
<tr>
<th>PL</th>
<th>Linear Trend</th>
<th>AR(1) PL</th>
<th>AR(1) GT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.000</td>
<td>1.079</td>
<td>1.031</td>
<td>1.070</td>
</tr>
</tbody>
</table>

<1 = better  No. Series = 6, Window Size = 6, Lead Time = 1
## Performance across series

### Lead time 1 week

<table>
<thead>
<tr>
<th>Method</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve</td>
<td>1.020</td>
</tr>
<tr>
<td>Naïve Diff.</td>
<td>1.087</td>
</tr>
<tr>
<td>Linear Trend</td>
<td>1.086</td>
</tr>
<tr>
<td>AR(1)</td>
<td>1.070</td>
</tr>
<tr>
<td>Optimal</td>
<td>0.866</td>
</tr>
</tbody>
</table>

\(<1 = \text{better}\)

No. Series = 6, Window Size = 6

### Lead time 6 weeks

<table>
<thead>
<tr>
<th>Method</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve</td>
<td>1.007</td>
</tr>
<tr>
<td>Naïve Diff.</td>
<td>1.143</td>
</tr>
<tr>
<td>Linear Trend</td>
<td>1.088</td>
</tr>
<tr>
<td>AR(1)</td>
<td>1.033</td>
</tr>
<tr>
<td>Optimal</td>
<td>0.861</td>
</tr>
</tbody>
</table>

\(<1 = \text{better}\)

No. Series = 6, Window Size = 6
Conclusion

Fully automated Bass model market size parameter estimation method that includes information from search traffic.

Google trend percentage increase market size estimation method outperformed most benchmark models.
Thank you!

St. Isaac’s Cathedral, St. Petersburg

o.schaer@lancaster.ac.uk
@oliverschaer


References


