Estimating the market potential pre-launch with search traffic

Oliver Schaer
Nikolaos Kourentzes
Robert Fildes

38th International Symposium on Forecasting
18th of June 2018
Motivation for generating pre-launch forecasts

Increased competition lead to shorter life-cycles and faster introduction of new products

Forecast needed to adjust marketing planning and operations

Difficult to estimate the market potential (Goodwin et al. 2014)

- Expert judgment bias when forecasting new products (Belvedere & Goodwin 2017; Tyebjee 1987)
- Consumer preferences change during pre-launch phase (Meeran et al. 2017)
Obtaining pre-launch market potential estimates

By expert judgement

• Survey (Lee et al. 2014, Kim et al. 2013, Bass et al. 2001)

By analogy

• Parameters from previous or similar products, i.e. product attributes (Kim et al. 2014, Goodwin et al. 2013, Lee et al. 2003, Lillien et al. 2000)

By market research

• Conjoint (Orbach & Fruchter 2011, Lee et al. 2006)
• Pre-orders (Moe & Fader 2002)
Obtaining pre-launch market potential estimates

By expert judgement
  • Survey (Lee et al. 2014, Kim et al. 2013, Bass et al. 2005)

By analogy
  • Parameters from previous or similar product attributes (Kim et al. 2014, Goodwin et al. 2013, Lee et al. 2003, Lillien et al. 2000)

By market research
  • Conjoint (Orbach & Fruchter 2011, Lee et al. 2006)
  • Pre-orders (Moe & Fader 2002)

Biased
No new information
Few data points and expensive
Not readily available
Demand forecasting with user-generated online information

Oliver Schaar *, Nikolaos Kourentzes, Robert Fildes

Centre for Marketing Analytics and Forecasting, Department of Management Science, Lancaster University Management School, UK

ARTICLE INFO

Keywords:
Google trends
Social media
Leading indicators
Product life-cycle
Search traffic
Electronic word-of-mouth

ARTICLE IN PRESS

International Journal of Forecasting [ ] [ ] [ ]

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

ABSTRACT

Recently, there has been substantial research on the augmentation of aggregate forecasts with individual consumer data from internet platforms, such as search traffic or social network shares. Although the majority of studies have reported increases in accuracy, many exhibit design weaknesses, including a lack of adequate benchmarks or rigorous evaluation. Furthermore, their usefulness over the product life-cycle has not been investigated, even though this may change, as consumers may search initially for pre-purchase information, but later for after-sales support. This study begins by reviewing the relevant literature, then attempts to support the key findings using two forecasting case studies. Our findings are in stark contrast to those in the previous literature, as we find that established univariate
Pre-release buzz (PRB)

PRB is the aggregate anticipation of consumers towards a new product (Houston et al. 2018)

- includes (i) communication, (ii) search, and (iii) participation in experiential activities

Why PRB supposed to work?

- Less information compared to post-launch forecasting available, i.e. univariate information
- PRB has smaller potential to have endogeneity than post-release phase
- Several studies which include pre-release buzz report increase in forecast accuracy.
Pre-release buzz forecasting literature

Forecasting with pre-launch buzz

<table>
<thead>
<tr>
<th>Study</th>
<th>Target variable</th>
<th>online source*</th>
<th>Measure</th>
<th>Use analogy</th>
<th>Early forecasts</th>
<th>Horizon†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liu (2006)</td>
<td>Box office sales</td>
<td>FOM</td>
<td>Vol.; Val.</td>
<td></td>
<td></td>
<td>1 w</td>
</tr>
<tr>
<td>Dhar and Chang (2009)</td>
<td>Music album sales</td>
<td>BLG</td>
<td>Vol.</td>
<td>x</td>
<td></td>
<td>3 w</td>
</tr>
<tr>
<td>Foutz and Jank (2010)</td>
<td>Box office sales</td>
<td>VSX</td>
<td>-</td>
<td>x</td>
<td></td>
<td>1 w</td>
</tr>
<tr>
<td>Asur and Huberman (2010)</td>
<td>Box office sales</td>
<td>TWR</td>
<td>Vol.; Val.</td>
<td></td>
<td></td>
<td>1 d</td>
</tr>
<tr>
<td>Wang et al. (2010)</td>
<td>Box office sales</td>
<td>FOM</td>
<td>Vol.</td>
<td></td>
<td></td>
<td>2 w</td>
</tr>
<tr>
<td>Hann et al. (2011)</td>
<td>Music album sales</td>
<td>P2P</td>
<td>-</td>
<td>x</td>
<td></td>
<td>1 w</td>
</tr>
<tr>
<td>Mülbacher et al. (2011)</td>
<td>Ski sales</td>
<td>FOM</td>
<td>Vol.; Val.</td>
<td></td>
<td></td>
<td>1 y</td>
</tr>
<tr>
<td>Kulkarni et al. (2012)</td>
<td>Box office sales</td>
<td>GTD</td>
<td>Vol.</td>
<td></td>
<td></td>
<td>1 w</td>
</tr>
<tr>
<td>Onishi and Manchanda (2012)</td>
<td>Box office sales</td>
<td>BLG</td>
<td>Vol.; Val.</td>
<td></td>
<td></td>
<td>1 d</td>
</tr>
<tr>
<td>Tian et al. (2014)</td>
<td>Cell phone service</td>
<td>BLG</td>
<td>Vol.; Val.</td>
<td></td>
<td></td>
<td>1 d</td>
</tr>
<tr>
<td>Xiong and Bharadwaj (2014)</td>
<td>Video game sales</td>
<td>BAU</td>
<td>Vol.</td>
<td></td>
<td></td>
<td>1 d</td>
</tr>
<tr>
<td>Craig et al. (2015)</td>
<td>Box office sales</td>
<td>FOM</td>
<td>Vol.</td>
<td>x</td>
<td></td>
<td>1 w</td>
</tr>
<tr>
<td>Gelper et al. (2015)</td>
<td>Box office sales</td>
<td>TWR</td>
<td>Vol.; Val.</td>
<td>x</td>
<td></td>
<td>1 d</td>
</tr>
<tr>
<td>Kim et al. (2015)</td>
<td>Box office sales</td>
<td>BLG</td>
<td>Vol.</td>
<td></td>
<td></td>
<td>1 w</td>
</tr>
<tr>
<td>Ding et al. (2017)</td>
<td>Box office sales</td>
<td>FBK</td>
<td>Vol.</td>
<td></td>
<td></td>
<td>1 w</td>
</tr>
<tr>
<td>Divakaran et al. (2017)</td>
<td>Box office sales</td>
<td>BLG</td>
<td>Vol.</td>
<td></td>
<td></td>
<td>1 w</td>
</tr>
<tr>
<td>Kim and Hanssens (2017)</td>
<td>Box office sales</td>
<td>GTD</td>
<td>Vol.; Vol.</td>
<td>x</td>
<td></td>
<td>1 w</td>
</tr>
<tr>
<td>This study</td>
<td>Video game sales</td>
<td>GTD</td>
<td>Vol.</td>
<td>x</td>
<td>x</td>
<td>52 w</td>
</tr>
</tbody>
</table>

* BAU = Baidu, BLG = Blog, FBK = Facebook, FOM = Forum, GTD = Google Trends, P2P = Peer-to-Peer Network, TWR = Twitter, VSX = Virtual Stock Exchange;
† d = days, w = weeks, m = months, y = year.
Investigating pre-launch buzz for sequential products

Most products nowadays are not entirely new but released as sequential.

• What is the relation of pre-launch buzz to the market parameter for sequential product releases?

• Pre-release buzz also able to improve long-term forecasts and what lead time can we achieve?

\[ \hat{Y}_t = \hat{m} \times F(\cdot) \]

\( \hat{Y}_t \) = cumulative number of sales,  
\( \hat{m} \) = estimated market potential,  
\( F(\cdot) \) = parametric model

Using direct analogy of previous releases requires very few training sample.
Generations of video games within a franchise

Sales for Assassin's Creed

Google Trends
Proposed methodology

Sales

- Actuals
- Forecast

Search traffic

Peak scaled search traffic

Graphs showing sales and search traffic trends over time.
Predicting the market potential with pre-release buzz (PRB)

<table>
<thead>
<tr>
<th>Model</th>
<th>Equation</th>
<th>Estimation point</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage PRB</td>
<td>$\hat{m}_j = (\Delta \text{PRB}<em>j \cdot m</em>{j-1}) \cdot \varepsilon_j$</td>
<td>$j \geq 2$</td>
</tr>
<tr>
<td>Factor PRB</td>
<td>$\hat{m}_j = \Delta \text{PRB}^{c_1}<em>j \cdot m</em>{j-1}^{1} \cdot \varepsilon_j$</td>
<td>$j \geq 4$</td>
</tr>
<tr>
<td>Factor PRB + int</td>
<td>$\hat{m}_j = c_0 \cdot \Delta \text{PRB}^{c_1}<em>j \cdot m</em>{j-1}^{1} \cdot \varepsilon_j$</td>
<td>$j \geq 5$</td>
</tr>
<tr>
<td>PRB + int</td>
<td>$\hat{m}_j = c_0 \cdot \text{PRB}^{c_1}_j \cdot \varepsilon_j$</td>
<td>$j \geq 3$</td>
</tr>
<tr>
<td>PRB + AR + int</td>
<td>$\hat{m}_j = c_0 \cdot \text{PRB}^{c_1}<em>j \cdot m</em>{j-1}^{c_2} \cdot \varepsilon_j$</td>
<td>$j \geq 5$</td>
</tr>
<tr>
<td>$\Delta$PRB + int</td>
<td>$\hat{m}_j = c_0 \cdot \Delta \text{PRB}^{c_1}_j \cdot \varepsilon_j$</td>
<td>$j \geq 4$</td>
</tr>
<tr>
<td>$\Delta$PRB + AR + int</td>
<td>$\hat{m}_j = c_0 \cdot \Delta \text{PRB}^{c_1}<em>j \cdot m</em>{j-1}^{c_2} \cdot \varepsilon_j$</td>
<td>$j \geq 5$</td>
</tr>
</tbody>
</table>

In addition we define the same models also in an additive way
Empirical evaluation on 56 franchises (257 games)

Google Trends with game title as keyword:

\[
PRB_j = \sum_{k=1}^{w} GT_{j,(t-l-w)}
\]

\(w = 4\) to 12 weeks Google Trends window \(l = 1\) to 12 weeks lead time

Set of parametric diffusion models

Smoothed adoption shape of \(j-1\) as non-parametric model

Average Relative Absolute Error on the cumulative adoption with \(m_{j-1}\) (Davydenko & Fildes 2013)

\[
\text{AvgRelAE}_h = \sqrt[n]{\prod_{r=1}^{n} \left( \frac{\text{AE}_{\text{Candidate model},r}}{\text{AE}_{\text{Benchmark},r}} \right)},
\]

\[\text{AE} = |y_{t+h} - \hat{y}_{t+h}|,\]
## Overall performance

### Overall results

<table>
<thead>
<tr>
<th>Model Description</th>
<th>Bass</th>
<th>Gompertz</th>
<th>Weibull</th>
<th>G/SGompertz</th>
<th>KNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage PRB (Mult.)</td>
<td>0.880</td>
<td>0.873</td>
<td>0.883</td>
<td>0.868</td>
<td>0.844</td>
</tr>
<tr>
<td>Percentage PRB (Add.)</td>
<td>0.999</td>
<td>0.988</td>
<td>0.963</td>
<td>0.963</td>
<td>0.955</td>
</tr>
<tr>
<td>Factor PRB (Mult.)</td>
<td>1.024</td>
<td>0.902</td>
<td>1.264</td>
<td>0.944</td>
<td>1.097</td>
</tr>
<tr>
<td>PRB (Mult.)</td>
<td>1.078</td>
<td>0.941</td>
<td>1.405</td>
<td>0.983</td>
<td>1.146</td>
</tr>
<tr>
<td>PRB + AR (Mult.)</td>
<td>0.841</td>
<td>0.813</td>
<td>2.608</td>
<td>1.129</td>
<td>0.835</td>
</tr>
<tr>
<td>Random Walk</td>
<td>1.000</td>
<td>0.993</td>
<td>0.971</td>
<td>0.970</td>
<td>0.960</td>
</tr>
<tr>
<td>With $m_j$</td>
<td>0.279</td>
<td>0.238</td>
<td>1.563</td>
<td>0.327</td>
<td>0.960</td>
</tr>
<tr>
<td>AR (Mult.)</td>
<td>1.366</td>
<td>1.611</td>
<td>4.557</td>
<td>2.079</td>
<td>1.958</td>
</tr>
</tbody>
</table>

Google Trends Window size 6
Lead time 1 to 12 weeks
Performance of GT percentage across lead times

Forecasting accuracy gains available 10 weeks before release
Managerial implications for pre-launch buzz:

• Automated pre-launch forecasting process improving accuracy by up to 15% compared to established analogy based methods
• Search traffic information providing up to 10 weeks lead time
• Good accuracy with simplest model that is available for the 2nd product generation

Further research

• Can we improve the accuracy by including trend direction
• Compare to other marketing variables such as ad spending
• Can this information also be used to allocate success of products by just looking into the pre-launch buzz?
Thank you!

Failing is fun
success is sweeter

o.schaer@lancaster.ac.uk
@oliverschaer
github.com/mamut86
References


References


References


