Forecasting diffusion with pre-launch online search traffic data

Oliver Schaer
Nikolaos Kourentzes
Robert Fildes

IIF Workshop on Forecasting New Products and Services

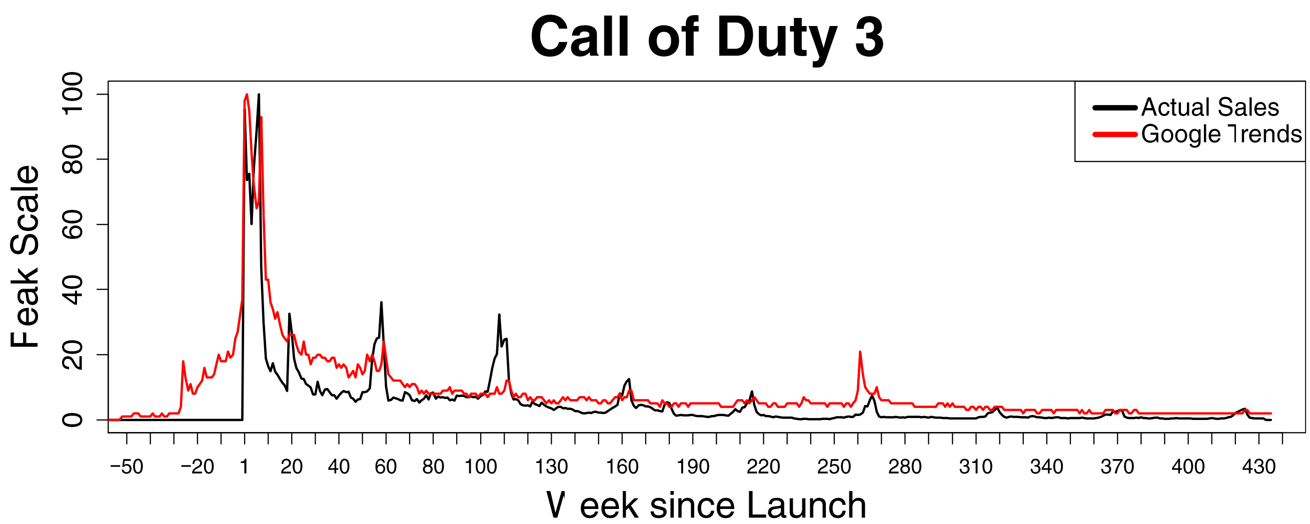
12th May 2016

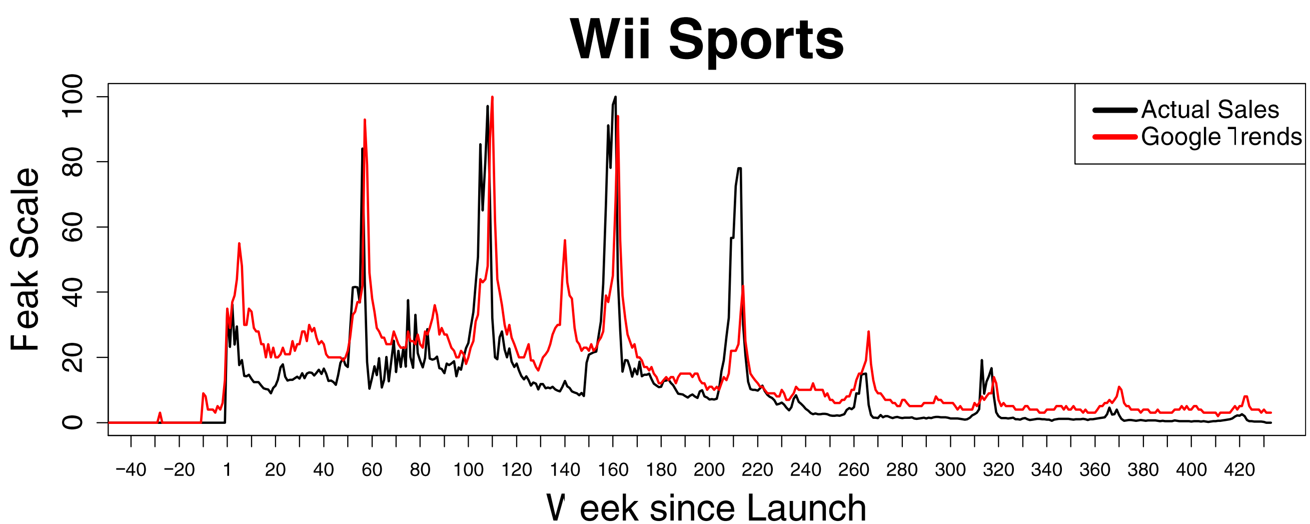
Lancaster Centre for Forecasting
Lancaster University Management School

LUMS

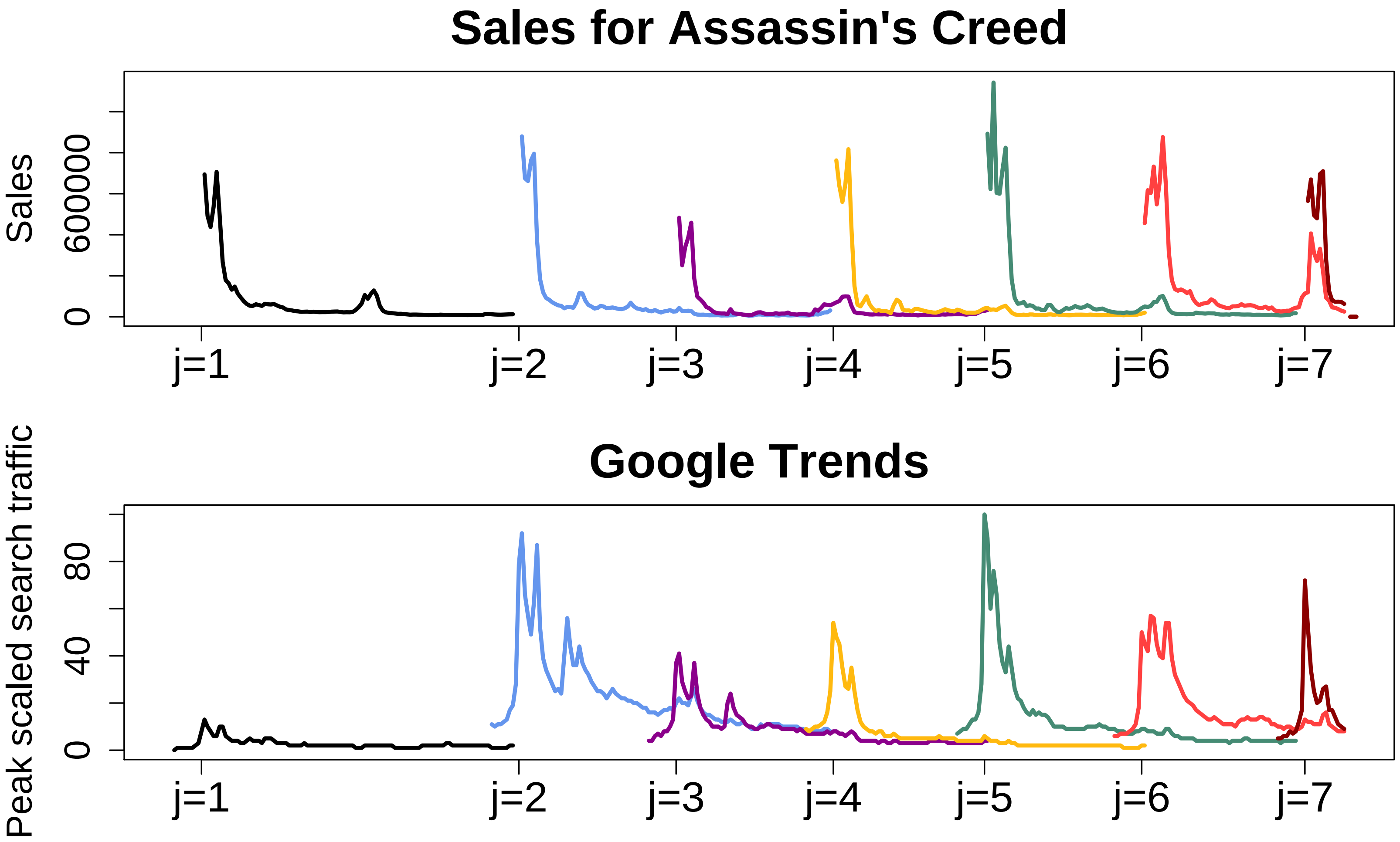
Lancaster Centre for Forecasting

The game life-cycle





Generations of games



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

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?

Experiment – Motivation

Model

- Bass model (Bass 1969)

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

Note that the aim is not to build the best possible fit but examine whether search traffic information improves the forecasts.

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

Model set up

Bass model (Bass 1969) for generation j = 2, 3…J

$$Y_{j,t} = Y_{j,t-1} + p_{j-1}(\hat{m}_j - Y_{j,t-1}) + q_{j-1}\frac{Y_{j,t-1}}{\hat{m}_j}(\hat{m}_j - Y_{j,t-1})$$

$$\begin{aligned}
Y_{j,t} &= \text{cumulative number of adaptors at time } t \text{ for generation } j \\
p_{j-1} &= \text{coefficient of innovation from the previous generation } j-1 \\
q_{j-1} &= \text{coefficient of imiation from the previous generation } j-1 \\
\hat{m}_j &= \text{estimated market size for generation } j
\end{aligned}$$

OLS parameter estimation (Bass 1969) with further minimisation of MSE as suggested by Lilien (2000) using BOBYQA optimisation algorithm (Powell 2009).

Google Trends handling

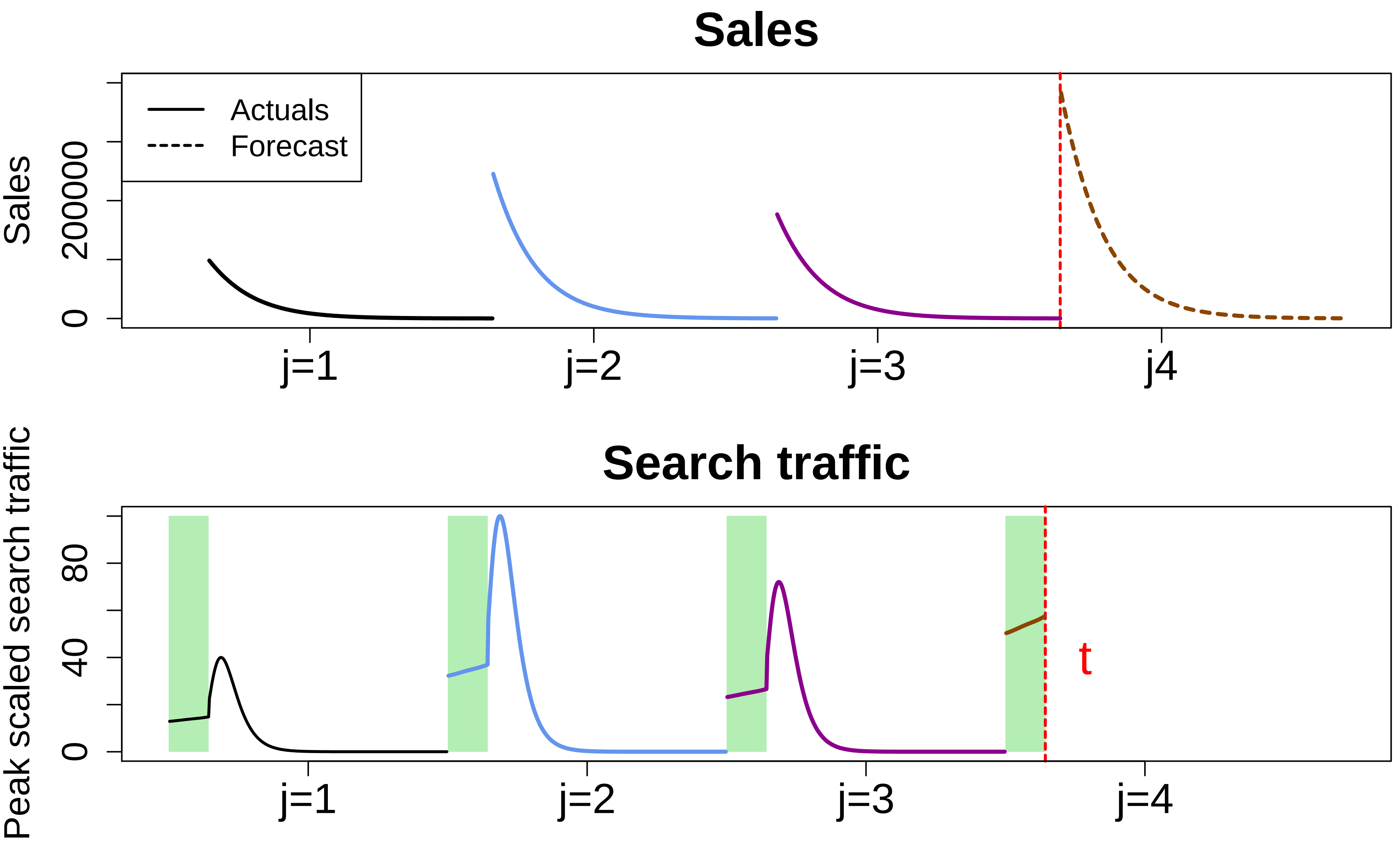
GT = topic search data

Peak scaled at time *t* to the highest observed search traffic in the series

$$\mathrm{GT}_j = \sum_{i=1}^{w} \mathrm{GT}_{j,(t-l)+w} \qquad w = \text{Window Size} \qquad l = \text{Lead Time}$$

For the experiment we use Window Size of 6 and lead time with 1 and 6 weeks

Estimation process



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

Numbers of generations needed for model estimation vary

Relative Mean Absolute Error and median across series

Google Trend model selection

Percentage Increase: $$m_j = \frac{\mathrm{GT}_j}{\mathrm{GT}_{j-1}} * m_{j-1}$$

Linear trend: $$m_j = \alpha_0 + \mathrm{GT}_j + \epsilon_j$$

AR(1) + Percentage Increase: $$m_j = \alpha_0 + \alpha_1 m_{j-1} + \beta_1 \frac{\mathrm{GT}_j}{\mathrm{GT}_{j-1}} + \epsilon_j$$

AR(1) + Google Trend: $$m_j = \alpha_0 + \alpha_1 m_{j-1} + \beta_1 \mathrm{GT}_j + \epsilon_j$$

PI	Linear Trend	AR(1) PI	AR(1) GT
1.000	1.079	1.031	1.070

<1 = better

No. Series = 6, Window Size = 6, Lead Time = 1

Performance across generations

Assassin's Creed

Generation	Naïve	Naïve Diff.	Linear Trend	AR(1)	Optimal
Game 1	NA	NA	NA	NA	NA
Game 2	**0.066**	NA	NA	NA	0.062
Game 3	1.746	1.925	1.925	NA	0.986
Game 4	1.223	1.993	1.382	4.347	1.012
Game 5	1.020	**0.925**	1.083	1.071	0.937
Game 6	1.165	1.334	1.103	**0.998**	0.76
Game 7	1.035	1.087	1.025	**0.898**	0.627

<1 = better

Window Size = 6, Lead Time = 1

Performance across series

Lead time 1 week

Naïve	Naïve Diff.	Linear Trend	AR(1)	Optimal
1.020	1.087	1.086	1.070	0.866

<1 = better

No. Series = 6, Window Size = 6

Lead time 6 weeks

Naïve	Naïve Diff.	Linear Trend	AR(1)	Optimal
1.007	1.143	1.088	1.033	0.861

<1 = 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.

Next steps

- Increase sample size
- Further improve Google Trends estimation, i.e. take the shape of the search traffic (Xiong & Bhradwaj 2014)
- Use machine learning for identifying market sizes and remaining model parameters for new game series (Lee et al. 2014).
- Look at diffusion models with higher flexibility to allow for better model fit such as the Gompertz or Weibull (Meade & Islam 2006, Moe & Fader 2002) or two stage models (Van den Bulte & Joshi 2007).
- Include the multi-generation and leapfrogging effect (Jiang & Jain 2012)

Great Aletsch Glacier, Switzerland

Thank you!

o.schaer@lancaster.ac.uk
@oliverschaer

References

Bass, F.M. (1969). A new product growth for model consumer durables. Management Science, 15, 215-227.

Fildes, R. Goodwin, P., Lawrence, M. and Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. International Journal of Forecasting, 25, 35-47.

Goodwin, P., Dyussekeneva, K. and Meeran, S. (2012). The use of analogies in forecasting the annual sales of new electronics products. IMA Journal of Management Mathematics, 1-16.

Goodwin, P., Meeran, S. and Dyussekeneva, K. (2014). The challenges of pre-launch forecasting of adoption time series for new durable products. International Journal of Forecasting, 30, 1082-1097.

Islam, T. and Meade, N. (1997). The Diffusion of Successive Generations of a Technology: A more General Model. Technological Forecasting and Social Change, 56, 48-60.

References

Jiang, Z. and Jain, D.C. (2012). A Generalized Norton-Bass model for Multigeneration Diffusion. Management Science, 58, 1887-1897.

Kim, T., Hong, J. and Kang, P. (2015). Box office forecasting using machine learning algorithms based on SNS data. International Journal of Forecasting, 31, 365-390

Kulkarni, G. Kannan, P.K. and Moe, W. (2012). Using online search data to forecast new product sales. Decision Support Systems, 52, 604-611.

Lee, H., Kim, S.G., Park, H-W. and Kang. P. (2014). Pre-launch new product demand forecasting using the Bass model: A statistical and machine learning-based approach. Technological Forecasting & Social Change, 86, 4—64.

Lilien, G. Y., Rangaswamy, A., and Van den Bulte, C. (2000). Diffusion models: Managerial applications and software. In: Mahajan, V., Muller, E., and Wind, Y. (Eds.), New-Product Diffusion Models, 295-311. Massachusetts: Kluwer.

Meade, N. and Islam, T. (2006). Modelling and forecasting the diffusion of innovation – A 25-year review. International Journal of Forecasting, 22, 519-545.

References

Moe, W. and Fader, P.S. (2002). Using Advance Purchase Orders to Forecast New Product Sales. Marketing Science, 21, 347-364.

Norton, J.A. and Bass, F.M. (1987). A Diffusion Theory Model of Adoption and Substitution for Successive Generations of High Technology Products. Management Science, 33, 1069-1086.

Powell, M.J.D. (2009). The BOBYQA algorithm for bound constrained optimization without derivatives. Department of Applied Mathematics and Theoretical Physics, Cambridge England, technical reportNA2009/06.

Van den Bulte, C. and Joshi, Y.V. (2007). New Product Diffusion with Influentials and Imitators. Marketing Science, 26, 400-421.

Xiong, G. and Bharadwaj, S. (2014). Prerelease Buzz Evolution Patterns and New Product Performance. Marketing Science, 33, 401-421.