Company X Revenue Analytics Case Study

Developed by: Adriano M. Yoshino – amyoshino@nyu.edu

CONTEXT

THE QUESTION

Company X is launching a change to its product. Because of the Android store release schedule, we have to launch the product change on the Company X Android app before we launch the change on iPhone.

1. What is the incremental revenue attributable to the product change? The change occurred for Android only on 2/1/2015.
2. Estimate what total incremental revenue it would drive if we had launched the same product change on iPhone, between 2/1/2015 - 2/10/2015.

DATASET

Data from 12/14/2014 to 02/10/2015 containing information about (for iPhone and Android):

- Number of users
- Average age of users
- Revenue
- Product change date (only for Android): 02/01/2015

IDEA / SOLUTION

INSPIRATION

The goal of this study is to infer the causal impact in revenues caused by the Promoted Trends product update for Android platform, then use this result to estimate the effect it would have caused if launched for iPhone at the same date. According to the literature (see references slide), a common approach to this problem is to determine the causal impact of a product change by using time-series information to forecast how would be the performance (revenues) if there were no intervention and then compare to the actual results.

To perform this estimation of causal effects in time-series, we are going to use the R package CausalImpact developed by Google in September 2015. The package is based on **Bayesian structural time-series and Markov chain Monte Carlo** algorithm for posterior inference. [More details about the algorithm on appendix A]

RESULTS

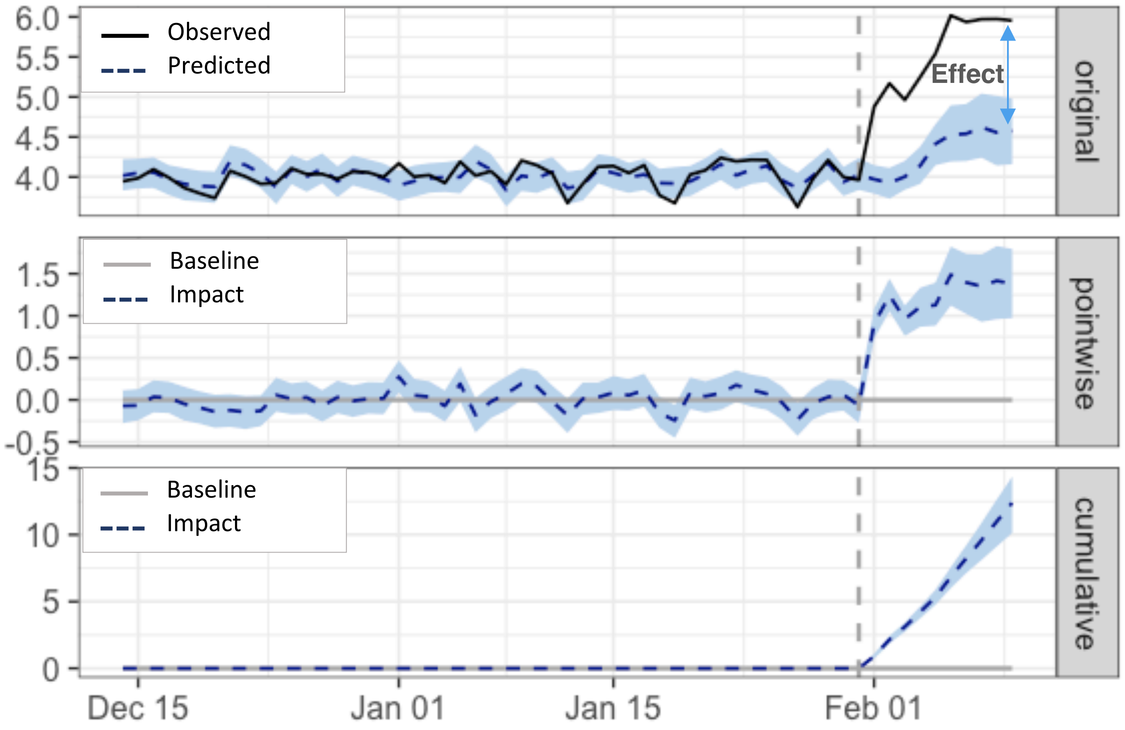
CALCULATING INCREMENTAL REVENUE FOR ANDROID

Implementation details on Appendix B

CausalImpact library handles this problem by:

- Predicting the revenue as if without impact (synthetic control)
- Comparing the results with actual revenue (observed values)

RESULTS:



Due to the product change, revenue increased 12.4 over 10 days

1.2 Avg. Effect

12.4 Cumul. Effect

29% Increase Effect

	Average	Cumulative
Actual	5.6	55.6
Prediction (s.d.)	4.3 (0.11)	43.3 (1.09)
95% CI	[4.1, 4.6]	[41.3, 45.5]
Absolute effect (s.d.)	1.2 (0.11)	12.4 (1.09)
95% CI	[1, 1.4]	[10, 14.3]
Relative effect (s.d.)	29% (2.5%)	29% (2.5%)
95% CI	[23%, 33%]	[23%, 33%]

Posterior tail-area probability p: 0.00102
Posterior prob. of a causal effect: 99.89837%

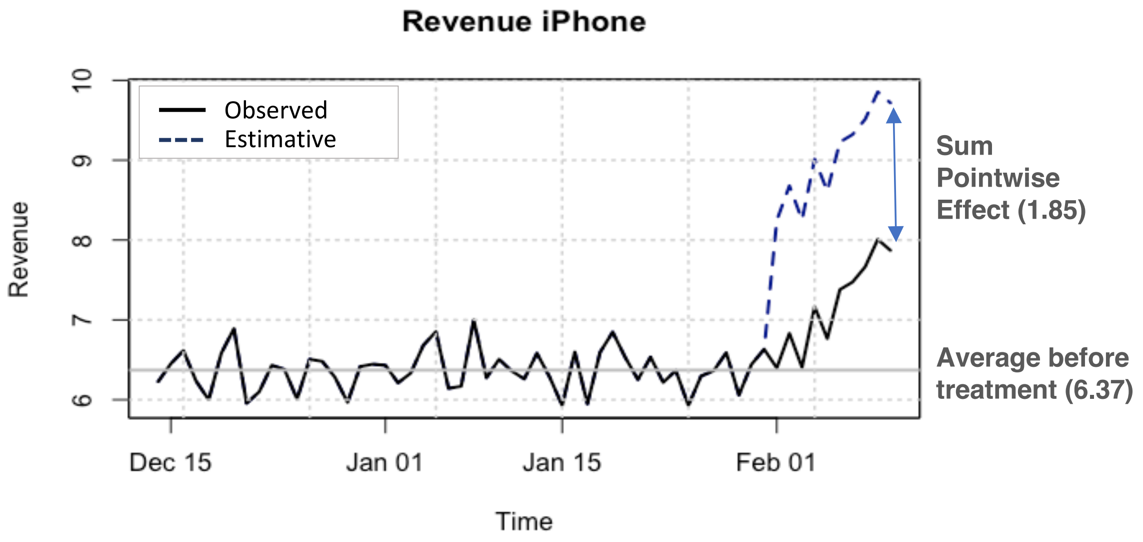
The probability of obtaining this effect by chance is very small (Bayesian one-sided tail-area probability p = 0.001). The causal effect can be considered statistically significant.

RESULTS

ESTIMATING REVENUE FOR IPHONE

Applying Effect

Assuming that the relative effect remains the same, we can estimate the impact of the product change if launched at the same time of Android:



Average before changing date = 6.37
Relative effect on Android = 29%

Pointwise effect = 6.37 x 0.29 = **1.85**

Results

29% Relative Effect

1.85 Avg. Effect

18.47 Cumul. Effect

Applying an increase of 29% over iPhone revenues we can estimate the incremental revenue attributable to the product change:

- 1.85 per day - 95% interval [1.53, 2.10]
- **18.5 total over period (10 days)** - 95% interval [15.3, 21]

CONCLUSION

SUMMARY

- The incremental Android revenue: 1.24 [1.02, 1.44] per day / 12.4 [10, 14.3] total
- In relative terms, Android revenues increased: +29% [+24%, +33%].
- Summing the points after the product change, the Android revenues had an overall value of 55.64 against expected a sum of 43.26 [41.23, 45.44] if there were no product changes.
- Applying an increase of 29% over iPhone revenues we can estimate the incremental revenue attributable to the product change as 1.85 [1.53, 2.10] per day / 18.5 [15.3, 21] total (10 days)

ADDITIONAL IDEAS / STEPS

- Use additional features to better predict the synthetic control
- Geo-located discretized data might help calculating more precisely the Causal Impact.
- Predict synthetic control using other algorithms (Random Forests, Gradient Boost Regression Trees, ARIMA...) and evaluate their performance against package standard model.

*95% interval represented by [x,y]

REFERENCES

CONSULTED MATERIALS:

1. “CausalImpact 1.2.1, Brodersen et al., Annals of Applied Statistics (2015). http://google.github.io/CausalImpact/”
2. Brodersen KH, Gallusser F, Koehler J, Remy N, Scott SL. Inferring causal impact using Bayesian structural time-series models. *Annals of Applied Statistics*, 2015, Vol. 9, No. 1, 247-274. http://research.google.com/pubs/pub41854.html
3. Scott, S. L. and Varian, H. R. (2014). Predicting the present with Bayesian structuraltime series. International Journal of Mathematical Modeling and Optimization 5 4–23. https://arxiv.org/pdf/1506.00356.pdf
4. Exploratory Blog. 2017. *An Introduction to Causal Impact Analysis*. [ONLINE] Available at: https://blog.exploratory.io/an-introduction-to-causal-impact-analysis-a57bce54078e. [Accessed 29 October 2017].

APPENDIX A

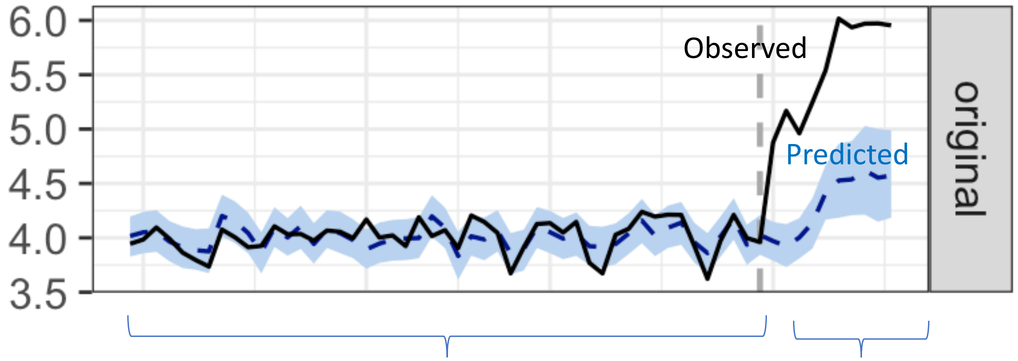
CAUSAL IMPACT PACKAGE IN R

From the author's website:

> "Given a response time series (e.g., clicks) and a set of control time series (e.g., clicks in non-affected markets or clicks on other sites), the package constructs a Bayesian structural time-series model. This model is then used to try and predict the counterfactual, i.e., how the response metric would have evolved after the intervention if the intervention had never occurred." [1]

HOW DOES IT WORK? (VERY SIMPLIFIED)

1. Combines Structural time series models + Bayesian spike-and-slab regression → average over a subset of the available predictors;
   - Predictors are other series/information that are not under effect of the treatment and can be used to model the behavior of the counterfactual.
2. Use Spike-and-slab to pick best parameters
3. Use Markov Chain Monte Carlo and Bayesian Averaging to fit the model and forecast.
4. Calculate differences between predicted and observed to measure impact.



Train model based on series not under effect of treatment. Spike and slab and MCMC helps picking the best parameters.

Estimate effect

"CausalImpact 1.2.1, Brodersen et al., Annals of Applied Statistics (2015). http://google.github.io/CausalImpact/"

APPENDIX B - IMPLEMENTATION

INCREMENTAL REVENUE FOR ANDROID

Training model

To train the model on CausalImpact package, the dataset should be organized as response variable in the first column and the covariates (predictors) on the following ones:

	revenue_android	num_users_android	average_age_android	revenue_iphone
2014-12-14	3.945254	22888.42	22.59757	6.227963
2014-12-15	3.984406	23339.89	26.73295	6.450311
2014-12-16	4.094981	23551.65	24.84183	6.615957
2014-12-17	3.991148	22521.53	24.52102	6.233813

Predictors

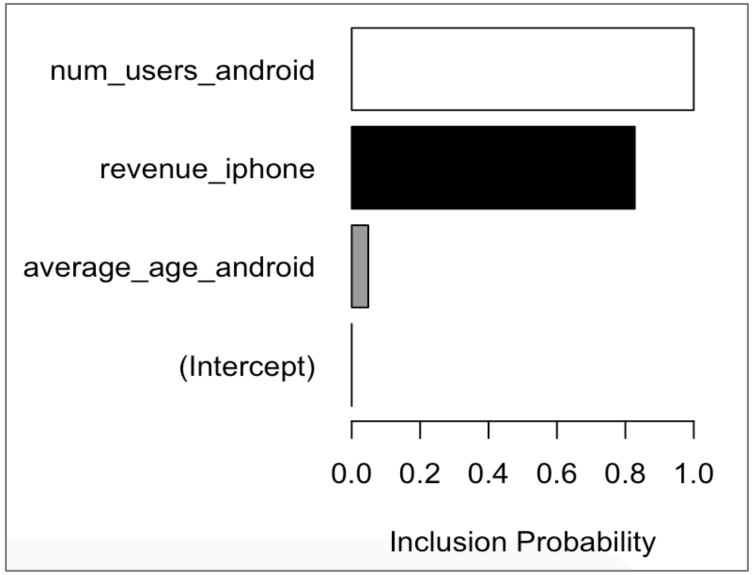
The predictors used to forecast the revenue as if without impact (synthetic control) are below and the impact of each feature in the model in the figure on the right:

- Number of users (Android)
- Average age (Android)
- iPhone revenue

Results

Results are obtained by the plot showing predicted curve x observed curve, pointwise increment and cumulated increment, and a table summarizing all information. (plots on Results slides).

Is it also possible to visualize the Bayesian marginal posterior inclusion probability for each predictor on the model:

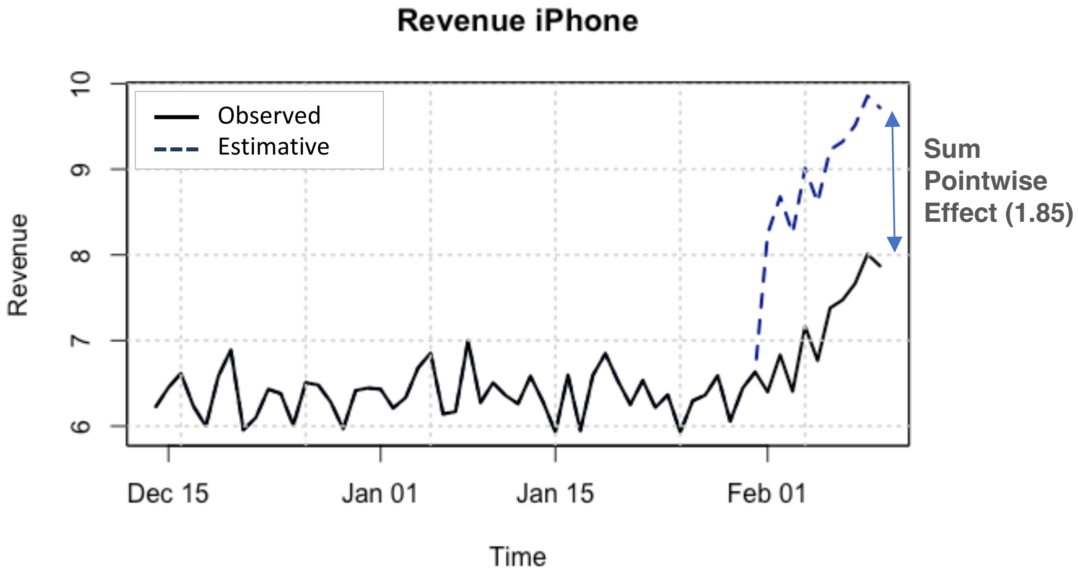


APPENDIX B - IMPLEMENTATION

ESTIMATING REVENUE FOR IPHONE

Applying Effect

Assuming that the relative effect remains the same, we can estimate the impact of the product change if launched at the same time of Android:



Average before changing date = 6.37
Relative effect on Android = 29%

Pointwise effect =
6.37 x 0.29 = **1.85**

Results

29% Relative Effect

1.85 Avg. Effect

18.47 Cumul. Effect

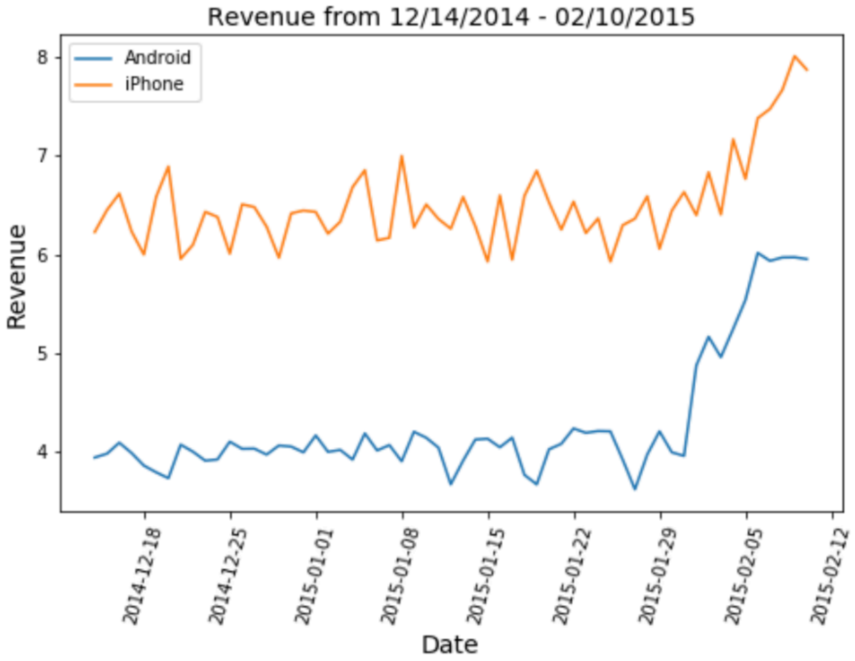
Applying an increase of 29% over iPhone revenues we can estimate the incremental revenue attributable to the product change:

- 1.85 per day - 95% interval [1.53, 2.10]
- 18.5 total over period (10 days) - 95% interval [15.3, 21]

APPENDIX B - IMPLEMENTATION

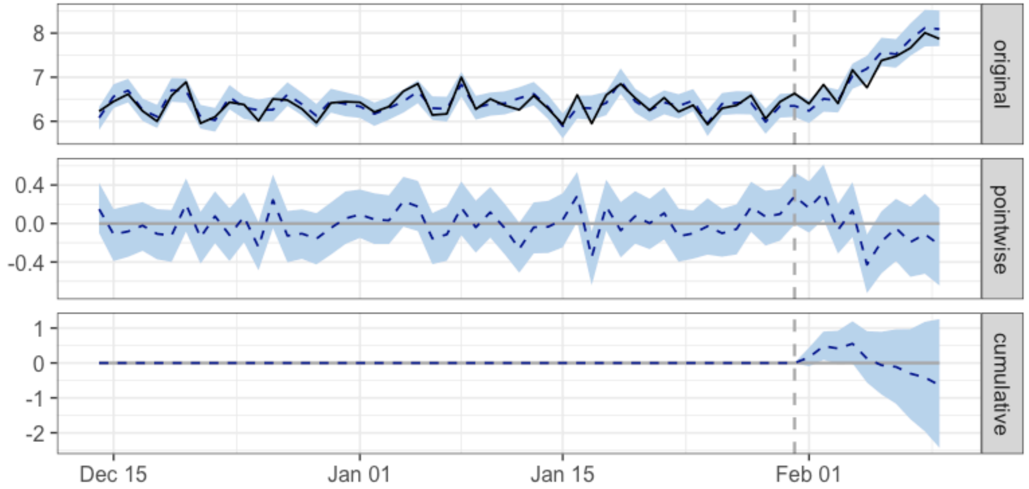
CURIOSITY: WHAT IF WE RUN THE ALGORITHM ON IPHONE NON TREATED DATE, WILL IT FIND SOME EFFECT?

By plotting revenue for Android and iPhone, it is possible to observe that both had some increase starting from 02/01/2015, however iPhone didn’t have released the update. Is this indicating an effect also?



Results

Looking at the plots below, there is no indication of effect (as expected). The summary of results show an effect of -0.064 in revenues



	Average	Cumulative
Actual	7.2	72.0
Prediction (s.d.)	7.3 (0.095)	72.6 (0.950)
95% CI	[7.1, 7.4]	[70.8, 74.5]
Absolute effect (s.d.)	-0.064 (0.095)	-0.637 (0.950)
95% CI	[-0.25, 0.12]	[-2.50, 1.19]
Relative effect (s.d.)	-0.88% (1.3%)	-0.88% (1.3%)
95% CI	[-3.4%, 1.6%]	[-3.4%, 1.6%]

Posterior tail-area probability p: 0.24349
Posterior prob. of a causal effect: 76%