Air Canada

AI Labs

AGIFORS - May 4th, 2021

Crystal.AI - Evaluating the impact

Methodology for performance evaluation

Greta Laage, Ivado Labs
Peter Wilson, Air Canada

Copyright © by Air Canada
Published by AGIFORS with permission

CRYSTAL AI

BIOGRAPHY

Greta Laage is a PhD candidate at École Polytechnique Montréal. She has been working with Ivado Labs and Air Canada on the Crystal.AI project. Her work focuses on demand forecasting for planning transportation systems and the impact on revenue.

Peter Wilson is managing the Revenue Management Advancement and Practice team at Air Canada. He has over 10 years of experience in RM working in the airline industry.

CONTENTS

Crystal.AI: Context

Problem Definition

Overview of methods used in practice and state of the art

Our Application and Specific Challenges

Counterfactual Prediction Models

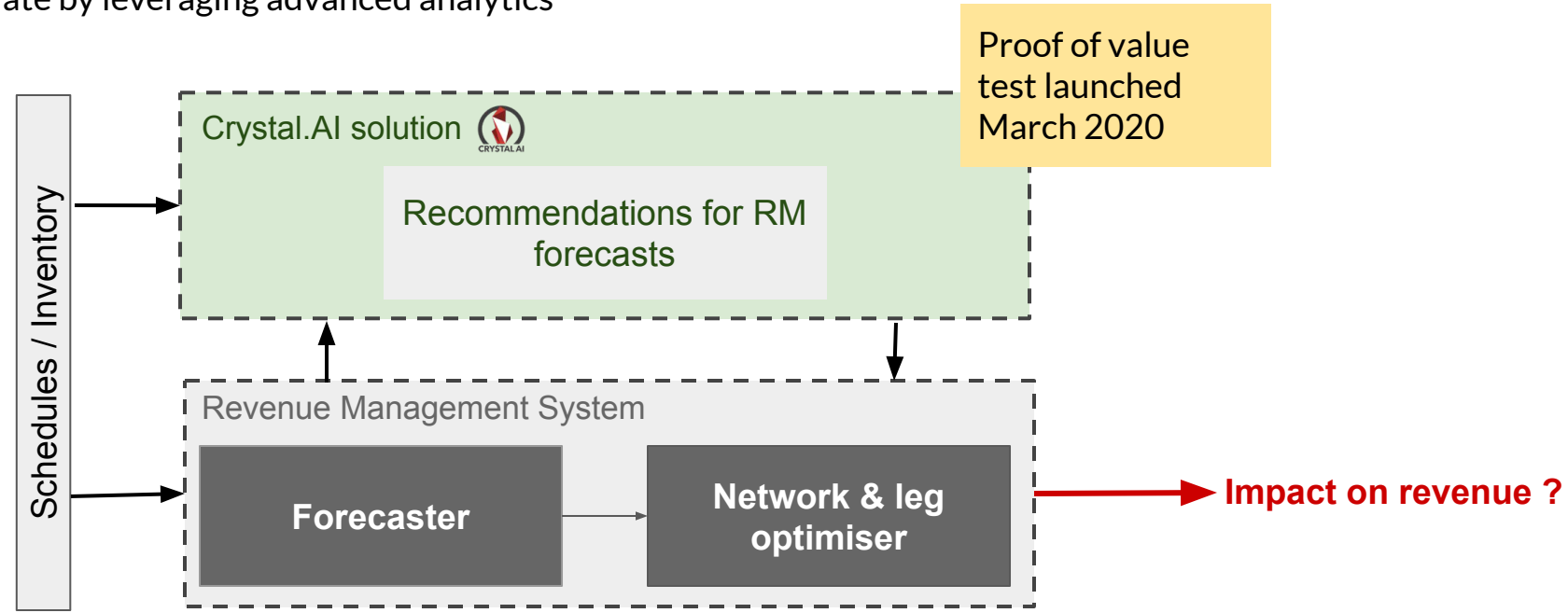
Experimental Setup and Results

Conclusion

Introduction

Crystal.AI

Improving the demand management workflow; making calibration quicker for demand managers and more accurate by leveraging advanced analytics



CONTENTS

Crystal.AI: Context

Problem Definition

Overview of methods used in practice and state of the art

Our Application and Specific Challenges

Counterfactual Prediction Models

Experimental Setup and Results

Conclusion

This is a joint work with Air Canada and Ivado Labs. The scientific paper associated with this presentation is available online: G. Laage, E. Frejinger, W. L. Hamilton, A. Lodi, G. Rabusseau *Estimating the Impact of an Improvement to a Revenue Management System: An Airline Application*, https://arxiv.org/abs/2101.10249, 2021.

Problem Definition

CONTEXT

Crystal.AI: AI-driven analytics solution for demand managers

- Adjustments to the RM forecasts
- Consequences on all subsequent decisions, e.g., pricing

OUR OBJECTIVE

Apply a change in a complex RM system

↓

Estimate its impact on a Key Performance Indicator, here revenue

CHALLENGES

- **Choice of reference point to compute the impact**
- **Small expected impact in percentage**
- **Extended period of time to have an effect on revenue**

CONTENTS

Crystal.AI: Context

Problem Definition

Overview of methods used in practice and state of the art

Our Application and Specific Challenges

Counterfactual Prediction Models

Experimental Setup and Results

Conclusion

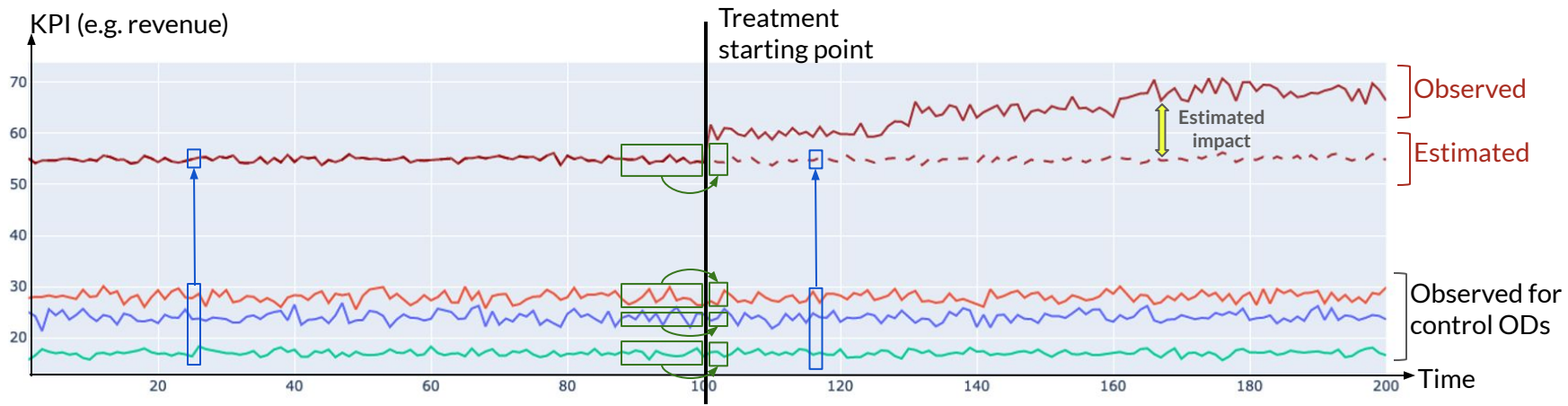
Methods used in practice

Method	Description	Pros/Cons
Change YoY	Analyse the Year over Year $ change	✓Easily computable for high-level purposes ✖Unreliable for measuring impact
Checkerboard	Apply change on alternating departure weeks	✖Subsequent dates can be affected ✖Difficult to manage
A-B Testing	Reference is the revenue of a set of ODs with similar behaviour	✖Important noise from inexistence of adequate ODs ✖Difficult to measure a small impact
Counterfactual predictions	Estimate what the revenue would have been keeping business as usual	✓Most promising for measuring small impacts ✓Adaptable for various treatments lengths

State of the art - Counterfactual Predictions

2 main methods,

- Horizontal: unconfoundedness literature Imbens and Rubin (2015).
- Vertical: synthetic controls, Doudchenko and Imbens (2016). Selected method, suited for long treatment periods.



Used for macroeconomic applications.

- Impact of the German reunification on per capita GDP of West Germany, Abadie et al. (2015)
- Impact of terrorist conflict on per capita GDP of Basque Country, Abadie and Gardeazabal (2003)

CONTENTS

Crystal.AI: Context

Problem Definition

Overview of methods used in practice and state of the art

Our Application and Specific Challenges

Counterfactual Prediction Models

Experimental Setup and Results

Conclusion

Our application and its challenges

- **Treatment**: new adjustments to RM forecasts.
- **Treated units**: OD pairs subject to the treatment.
- **Control units**: OD pairs unaffected by the treatment

Requirement: Both treated and control units should not suffer from large external shocks

Specificities of our application

- Microeconomic level
- Multiple treated units: a representative subset of the airline's network
- Large number of potential control units but potential spillover effects
- Small expected impact in percentage

CONTENTS

Crystal.AI: Context

Problem Definition

Overview of methods used in practice and state of the art

Our Application and Specific Challenges

Counterfactual Prediction Models

Experimental Setup and Results

Conclusion

Counterfactual Prediction models

$$Y_t = F(\mathbf{X}_t)$$

- Y_t Revenue of treated unit at day t
- $\mathbf{X}_t$ Revenue of multiple control ODs at day t
- F Counterfactual prediction model, ***trained on the pre-treatment period.***

Comparison of different models from the literature

- Synthetic controls and derivatives. F is a linear function: $F(\mathbf{X}_t) = \alpha_1 X_{1t} + \alpha_2 X_{2t} + \ldots + \alpha_n X_{nt}$
  - Synthetic Controls **SC (** Abadie and Gardeazabal, 2003) **.** The weights α form a convex combination.
  - Robust Synthetic Controls **RSC** (Amjad et al., 2018)**.** The revenue data are first denoised (via singular value thresholding), and the weights are estimated on denoised data.
  - Regression **CR** (Doudchenko and Imbens, 2016).
  - Regression with regularization **CR-EN** (Doudchenko and Imbens, 2016). Elastic-net regularization on the weights α.

- Feed-Forward Neural Network **FFNN**. F is a neural network, which can consider additional features in $\mathbf{X}_t$ such as temporal features.

CONTENTS

Crystal.AI: Context

Problem Definition

Overview of methods used in practice and state of the art

Our Application and Specific Challenges

Counterfactual Prediction Models

Experimental Setup and Results

Conclusion

Experimental Setup

Treated units

15 non-directional treated ODs, i.e., 30 directional treated ODs

Control units

317 carefully chosen control ODs to be unaffected by the treatment

Treatment Periods

- 15 pseudo-treatment period of 6 months to evaluate model performances
- Observed untreated *daily* revenue data from January 2013 to February 2020

Key Performance Indicator

Total revenue of the treated units over the *complete* treatment period

Setting for the model

- Predict the revenue of each treated unit at each day, and sum the predictions after. It allows to have predictions at OD-level, usually at the expense of accuracy.

Results - Daily predicted revenue

Absolute percentage error and squared error* at each day for each pseudo-treatment period:

$$\text{RMSE}_p = \sqrt{\frac{1}{T_1}\sum_{t=1}^{T_1}(\hat{Y}_t - Y_t)^2} \qquad \text{MAPE}_p = \frac{1}{T_1}\sum_{t=1}^{T_1}\frac{|\hat{Y}_t - Y_t|}{Y_t} \qquad T_1 = 180 \text{ days}$$

Average performance

	MAPE	RMSE
SC	4.6%	8.3
RSC	3.6%	6.5
CR	3.4%	6.0
CR-EN	8.6%	15.0
FFNN	4.6%	7.5

Takeaways

- Models have similar performance except CR-EN
- CR predicts the daily total revenue with 3.4% of error

*RMSE is scaled to the average daily revenue

Results - Daily predicted revenue at OD level

Absolute percentage error at each day for each treated OD*:

$$\text{MAPE}_i^{\text{od}} = \frac{1}{T_1} \sum_{t=1}^{T_1} \frac{|\hat{Y}_{it} - Y_{it}|}{Y_{it}} \qquad T_1 = 180 \text{ days}$$

Average performance

	MAPE^{od}
SC	13.1%
RSC	13.6%
CR	13.8%
CR-EN	16.0%
FFNN	13.3%

Takeaways

- Models have similar performance except CR-EN
- SC predicts the daily revenue of a treated OD with 13.1% of error

*Note on the FFNN: it relaxes independence assumption, and predicts simultaneously all treated ODs with multiple neurons in the output layer.

Results - Total predicted revenue

Error on the total revenue over *all treated ODs* and the *complete treatment period* at each pseudo-treatment period

$$\mathrm{tPE}_p = \frac{\sum_{t=1}^{T_1} \hat{Y}_t - \sum_{t=1}^{T_1} Y_t}{\sum_{t=1}^{T_1} Y_t}$$

Average performance

	$\lvert \mathrm{tPE}_p \rvert$
SC	3.3%
RSC	1.2%
CR	1.1%
CR-EN	2.5%
FFNN	1.0%

Detailed performance

Post-Treatment Periods in Setting S2

Takeaways

- High accuracy and stability for FFNN, RSC and CR
- FFNN can predict the total revenue with 1% of error

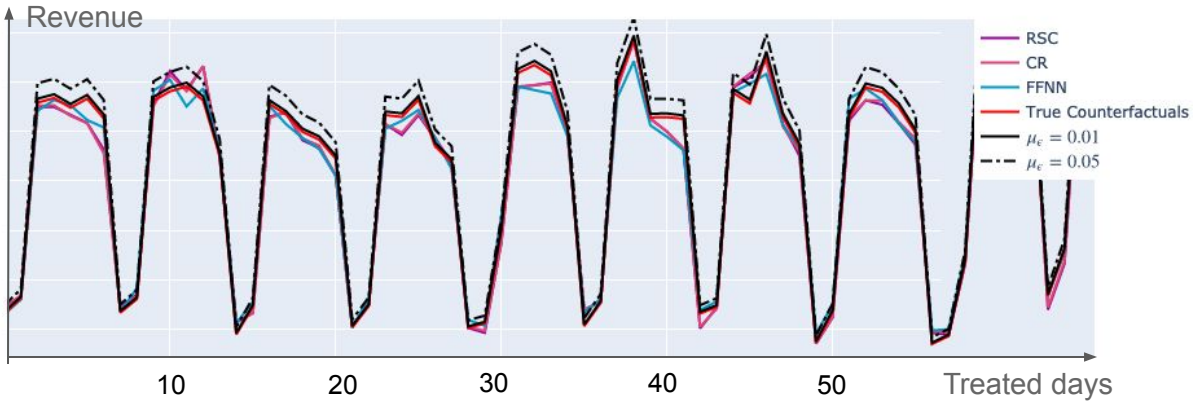
Results - Simulated Impact

Simulation study to estimate impact in period 2.

1. Simulate a treatment $\tilde{Y}_t = Y_t \times \epsilon, \quad \epsilon \sim lognorm(\mu_\epsilon, \sigma_\epsilon^2)$
2. Predict the counterfactual revenue $\hat{Y}_t$
3. Estimate the impact $\hat{\tau} = \sum_{t=1}^{T_1} \tilde{Y}_t - \hat{Y}_t$

Results

μ_ϵ	0.01	0.02	0.03	0.05
True	1.0%	2.0%	3.0%	5.1%
RSC	1.7%	2.6%	3.7%	5.7%
CR	1.5%	2.5%	3.5%	5.6%
FFNN	0.6%	1.6%	2.6%	4.7%



Takeaways

- Impact over/under estimated depending on the bias of predictions
- Accurate impact estimation, even with counterfactual prediction

CONTENTS

Crystal.AI: Context

Problem Definition

Overview of methods used in practice and state of the art

Our Application and Specific Challenges

Counterfactual Prediction Models

Experimental Setup and Results

Conclusion

Conclusion

- ★ New technique to estimate impact of a treatment: counterfactual prediction models
- ★ Applicable in various situations
- ★ Few requirements: only KPI data at OD level
- ★ Designed to estimate the revenue impact of Crystal.AI, but the industry was hit by COVID-19
- ★ To be leveraged in other change tests
  - Pricing structure/Fare brands/WTP
  - Overbooking calculation/no-show forecast method
  - Targeted advertising campaign

Appendix

Selection of control units: Unaffected by the treatment

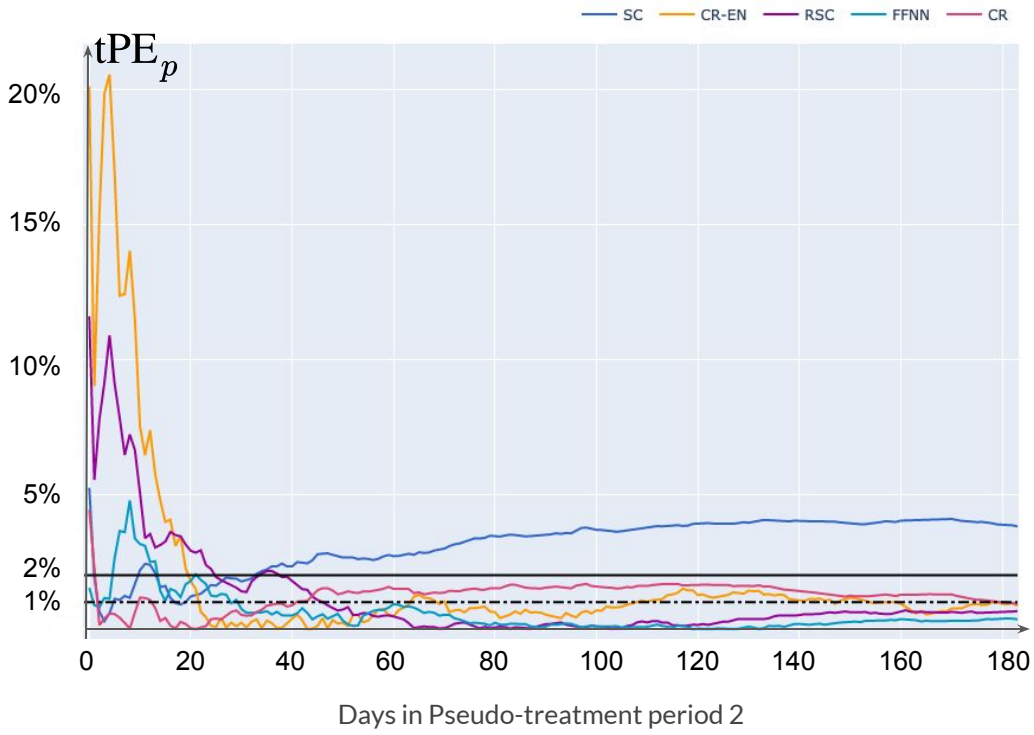
- Demand managers insights

- Discard potential ODs with 3 rules:
  - Geographic: discard OD pairs where either the origin or the destination is close to the treated ODs' airports.
  - Revenue ratio: discard the OD pairs where at least 5% of the itineraries have a leg identical to one of the treated ODs (the pricing of the treated OD pair can affect the pricing of the related itineraries)
  - Sparsity: discard seasonal ODs, i.e., those that operate only at certain times in the year, and sparse ODs.

- From the remaining ODs, we select the 40 most correlated ODs for each treated OD.

Appendix - Length of treatment period

We reached 1.0% of accuracy for the total revenue over a **6 months treatment period**. How is the accuracy for a **shorter** treatment period?



Takeaways

- 1% and 2% thresholds of absolute error are reached after 1 month for most models
- The treatment period can be substantially reduced

Appendix: Confidence Intervals

Settings

- ❑ Setting S1
- ❑ CR counterfactual predictions model

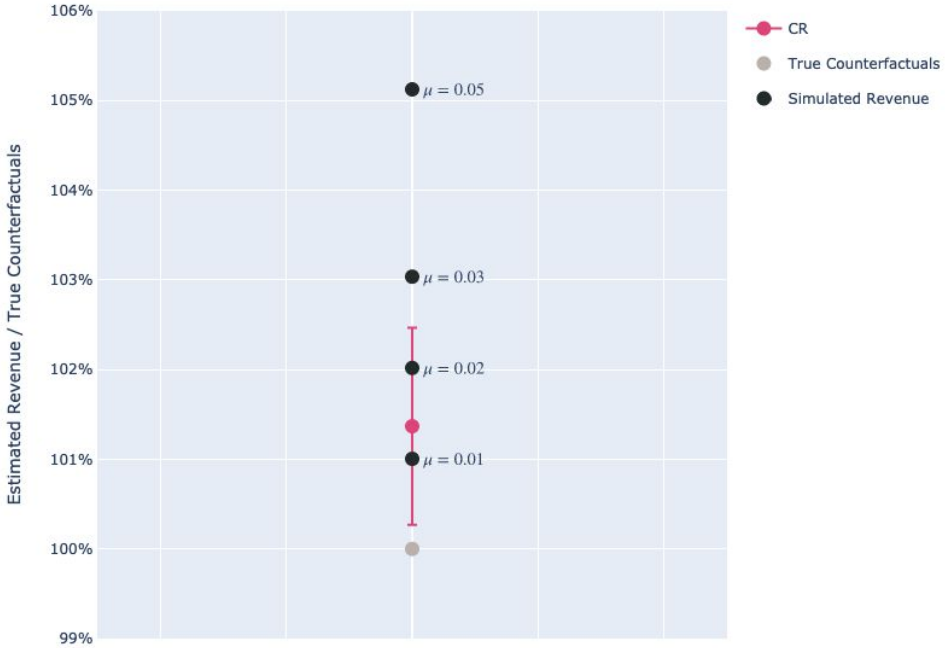
We can derive an exact formula for the confidence intervals, and compute them.

Assumptions

$$\hat{Y}_T = Y_t + \delta_t$$

Error term

Error terms
- ❑ independent
- ❑ identically distributed
- ❑ normally distributed



References

Abadie, A., Diamond, A., and Hainmueller, J. (2015) Comparative Politics and the Synthetic Control Method. *American Journal of Political Science*, 59(2):495–510.

Abadie, A. and Gardeazabal, J. (2003) The economic costs of conflict: A case study of the Basque Country. *American Economic Review*, 93(1):113–132.

Weatherford, L. and Belobaba, P. (2002) Revenue impacts of fare input and demand forecast accuracy in airline yield management. *Journal of the Operational Research Society*, 53(8):811–821.

Imbens, G. W. and Rubin, D. B. (2015) Causal inference in statistics, social, and biomedical sciences. Cambridge University Press.

Doudchenko, N. and Imbens, G. W. (2016) Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. Technical report, National Bureau of Economic Research.

Amjad, M., Shah, D., and Shen, D. (2018) Robust synthetic control. *The Journal of Machine Learning Research*, 19(1):802–852.