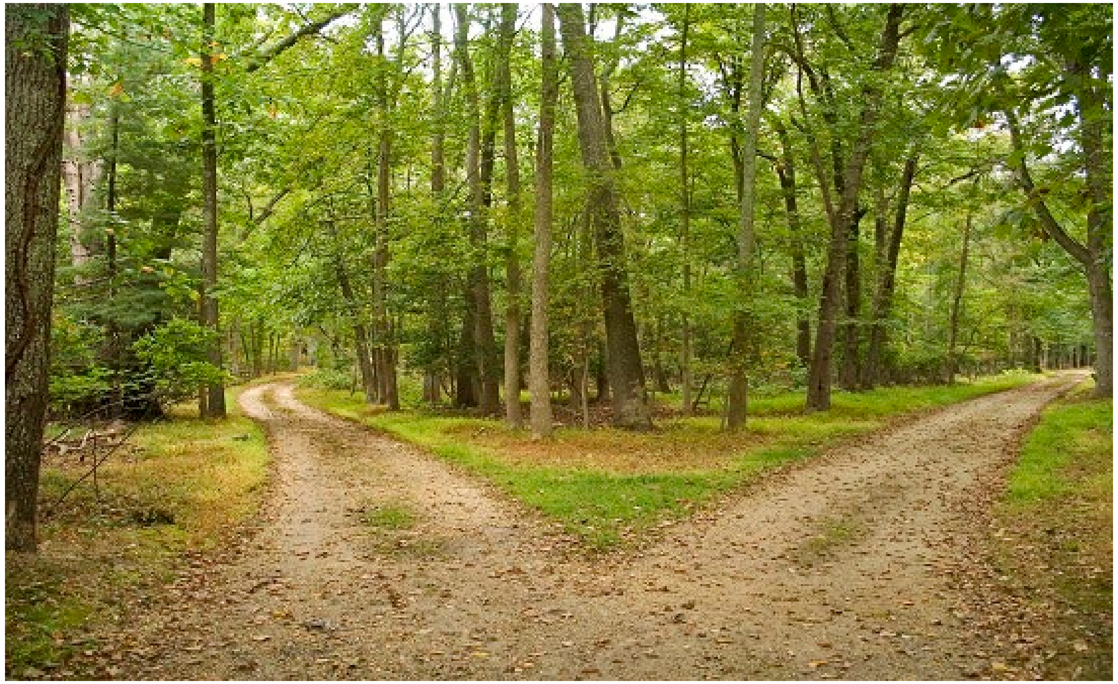


“Two roads diverged in a wood, and I—
I took the one less traveled by,
And that has made all the difference.”
- Robert Frost

Or did it…?

Carnegie Mellon University
Tepper School of Business

45-980: BIG DATA

Amr Farahat

Big Data, Models, and A/B Experiments

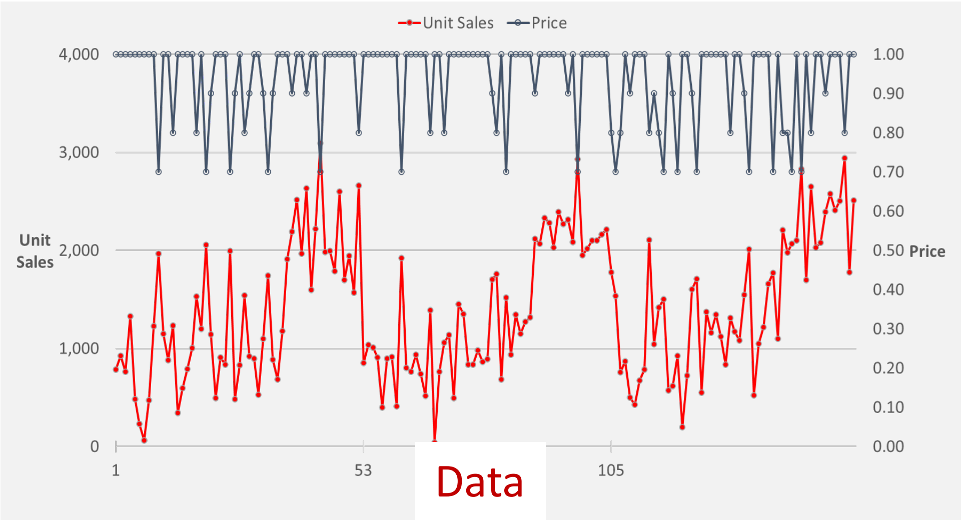
Recall: Extracting Value From (Big) Data

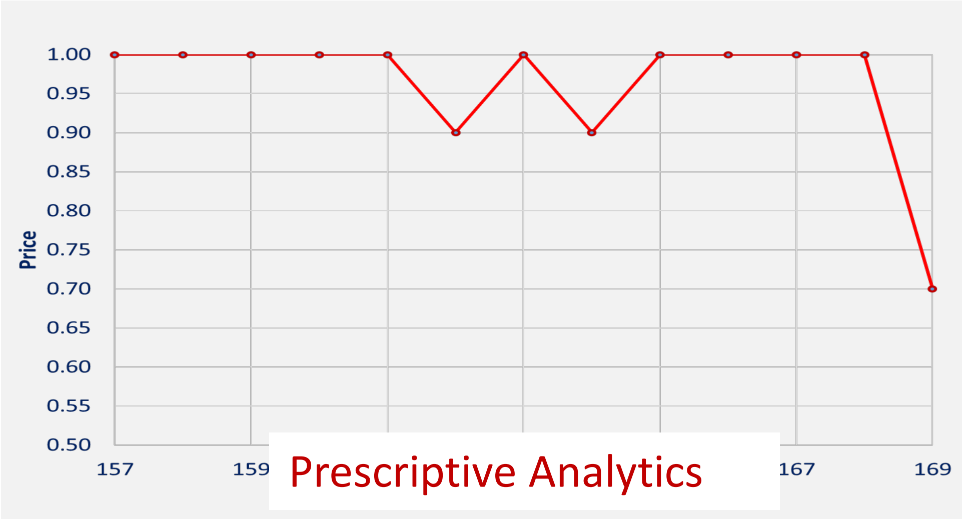
- The value of (big) data lies in our ability to utilize it to create business value.
- Information and data that doesn't impact a decision or doesn't improve the execution of a business process is of no value.
- Therefore, (big) data is only as valuable as the data-driven analytics we (or others) can utilize to create business value (monetary or not).

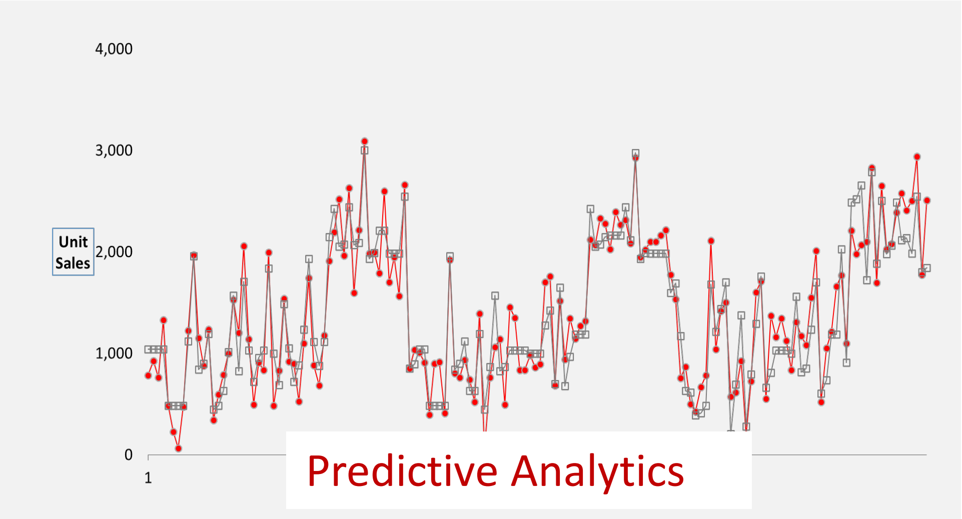
Recall: The Analytics Value Chain

	Descriptive / Diagnostic	Predictive	Prescriptive
Main questions:	*What is happening? What explains it?*	*What will happen if…?*	*What should we do?*
Main pre-occupation:	**Data**	**Data-Driven Models, Experiments, and Causal inference**	**Data and model-driven decision-making / Optimization**

A Typical Model Centric Approaches…







$$d_t = b_0 + b_1\ p_t + b_2\ p_{t-1} + b_3\ p_{t-2} + b_4\ t + s_2\ Season_2 + s_3\ Season_3 + \cdots + \ s_{13}\ Season_{13}$$

An Alternative Experiment-Based Approach…

	Descriptive / Diagnostic	Predictive	Prescriptive
Main questions:	*What is happening? What explains it?*	*What will happen if…?*	*What should we do?*
Main pre-occupation:	**Data**	**Data-Driven Models, Experiments, and Causal inference**	**Data and model-driven decision-making / Optimization**

A|B Testing Example

A|B Testing Example

A|B Testing Example

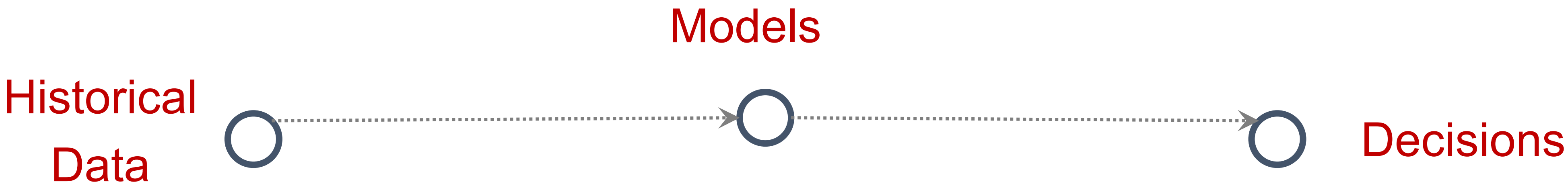
A/B Test Examples

Area	Subjects	Test	Control	Comparison Metrics
Medicine	Patients	Drug X	Placebo or Drug Y (e.g., the normal therapeutic for the disease)	• % surviving at a predefined timepoint • % reduction in biomarker levels
Public Health	Residents	Mosquito net	Nothing – "life as usual"	• % diagnosed with malaria in predefined timeframe
Retail	Stores	Childcare center within the store where shoppers can drop off their kids while shopping	No childcare center in store – "business as usual"	• Increase in Revenue from prior period • Increase in Shopper Visits from prior period • Increase in Average Basket Amount from prior period
E-Commerce	Website Visitors	Show product recommendations on home page (in addition to product pages)	Product recommendations only on product pages	• Increase in Revenue-per-Visitor • Increase in Average Basket Amount • Increase in Conversion Rate

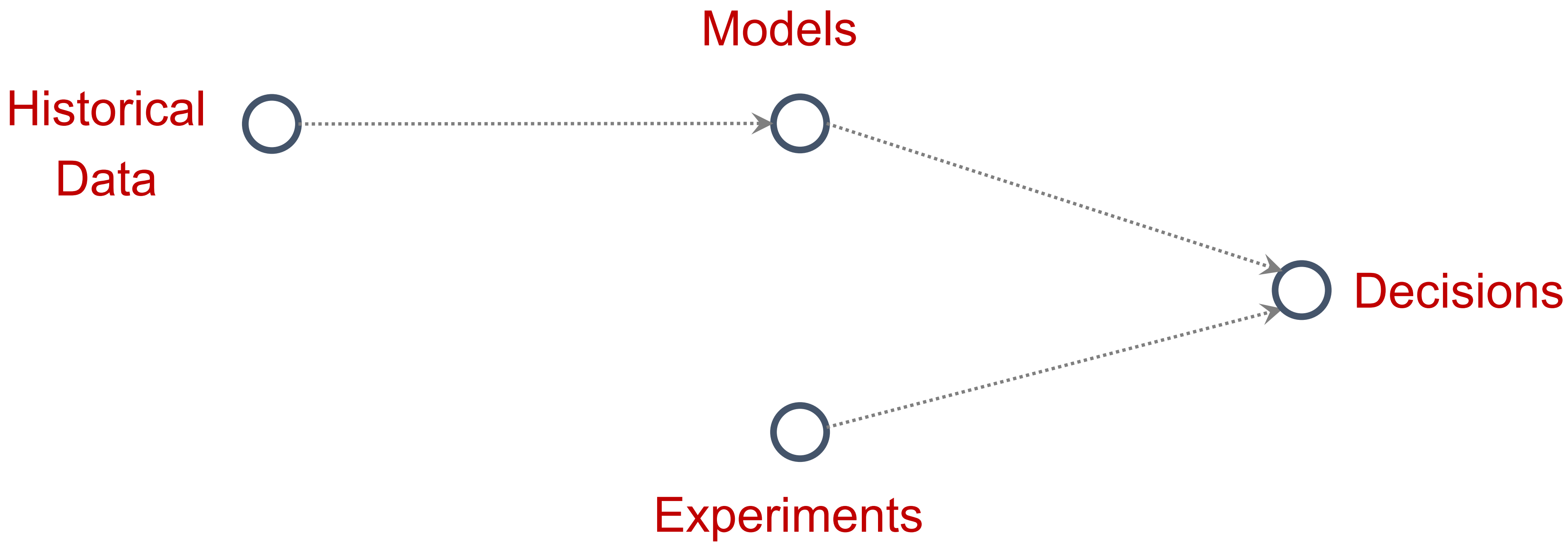
Table courtesy of MIT Sloan Professor Rama Ramakrishnan

AN EXPANDED VIEW OF BIG DATA, MODELS, AND EXPERIMENTS

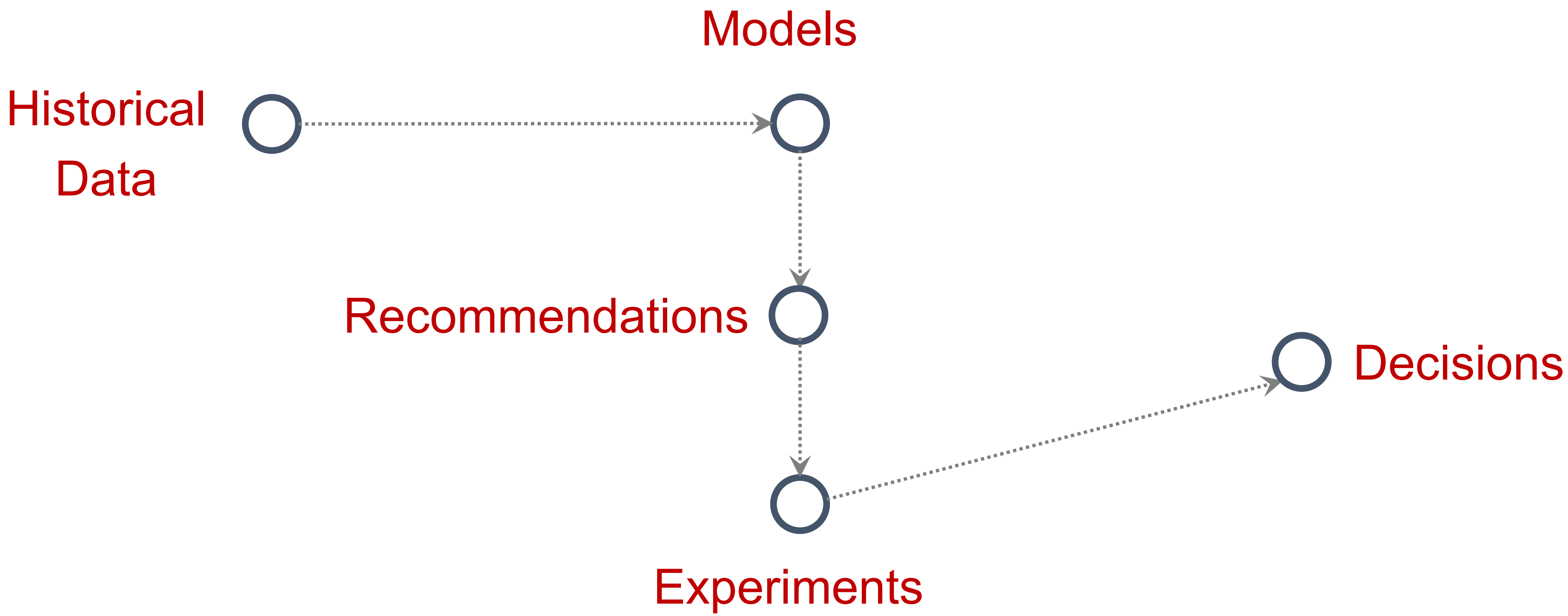
Traditional Model-Based View



Recognizing the Role of Experiments

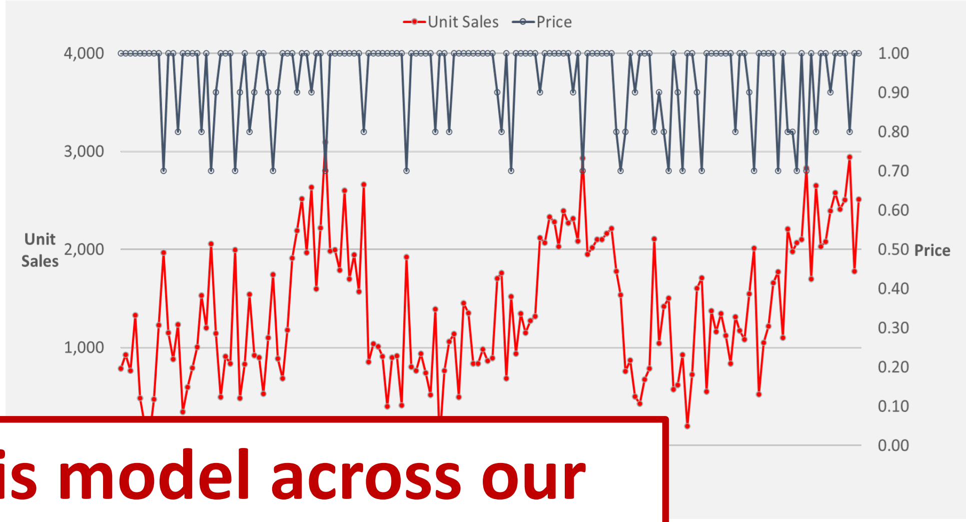


The Modern Perspective

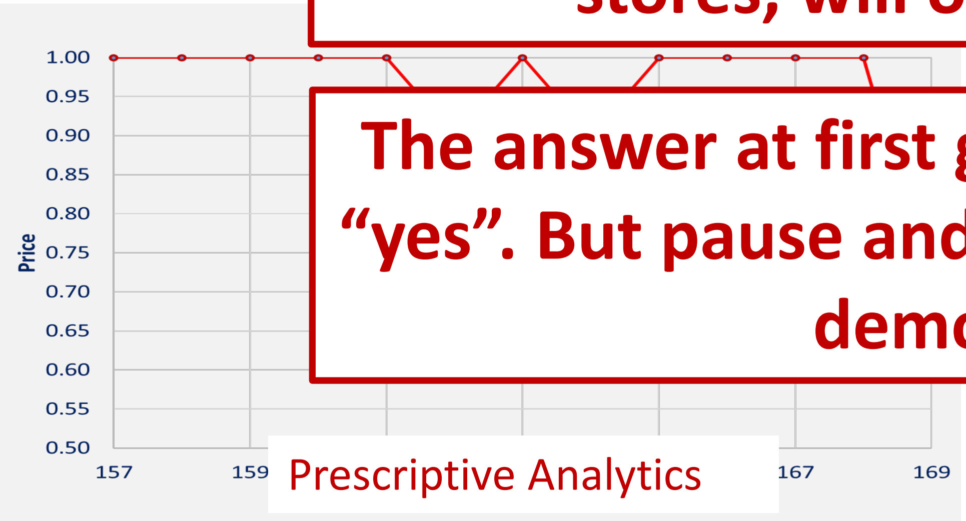


USING EXPERIMENTS TO ASSESS THE VALUE OF BIG DATA & ANALYTICS

Recall our Pricing Analytics in Retail Example



Suppose we implement this model across our stores, will our revenue increase?



Prescriptive Analytics

The answer at first glance should be an obvious "yes". But pause and think about how you would demonstrate this?

Predictive Analytics

A pretty good argument…

A. “We employed state-of-the-art, data-driven, predictive analytics”, and
B. “We optimized for the best possible decision.”

Now combine (A) and (B) and it should be obvious that our Analytics model delivers value…

A pretty good argument…

A. “We employed state-of-the-art, data-driven, predictive analytics”, and
B. “We optimized for the best possible decision.”

Now combine (A) and (B) and it should be obvious that our Analytics model delivers value…

- There’s definitely some truth in this argument…
- But let’s consider why the models might <u>not</u> deliver optimal decisions in practice:
  - All prediction models are approximations, even “state-of-the-art, data-driven” ones
  - All optimization models are also approximations, because:
    - they typically rely on prediction models (the demand equation in ours)
    - the model scope (variables, objective, constraints) is necessarily incomplete
  - We may have gotten unlucky because of the noise aspect of any model; but let’s focus here on “long-term” profit impact where noise presumably averages out

A More Objective Approach?

In practice, we have:

- multiple stores (different locations and characteristics), indexed by “i”
- selling multiple products (with their own demand equation), indexed by “k”
- over multiple weeks (exhibiting seasonality, evolving competition), indexed by “t”

Let $\mathbf{Y_{i,k,t}}$ denote the revenue from selling product k in store i in week t

To simplify things for the sake of pedagogy, let us

- fix the product k
- assume that if analytics delivers value next quarter then it will continue to deliver value into the reasonable future

So for now we will just consider $\mathbf{Y_i}$ = the profit in store i

"Treatment effect" and the key question

	Y_i *given* Analytics model deployment		Y_i *given* current system		Treatment effect = τ_i
Store i	$??	-	$??	=	**$??**
	profit of product k in store i at some given future period t when Analytics model is deployed		profit of product k in store i at some given future period t under current pricing system		value added from Analytics compared to current system

- How can we evaluate τ_i ?
- How can we be confident that $\tau_i > 0$?

“Retrospective” Approaches

Retrospective Approach 1 (Model-Based)

Consider a recent period, say Q4Y3 (part of our dataset)

- Use model to assess Analytics intervention
- Use model to assess current pricing system

	Y_i *given* Analytics intervention		Y_i *given* current system		Treatment effect = τ_i
Store i	$??	-	$??	=	**$??**

(Use the optimization model to determine prices and calculate profit using the regression demand model applied to Q4Y3)

(Use the actual prices set by the retailer based on the current pricing system. Use the regression demand model to calculate the corresponding demand and Q4Y3 profit)

Retrospective Approach 1 (Model-Based)

Consider a recent period, say Q4Y3 (part of our dataset)

- Use model to assess Analytics intervention
- Use model to assess current pricing system

	Y_i *given* Analytics intervention		Y_i *given* current system		Treatment effect = τ_i
Store i	\$26,940	-	\$25,759	=	**\$1,181 (+4.6%)**
	(Use the optimization model to determine prices and calculate profit using the regression demand model applied to Q4Y3)		(Use the actual prices set by the retailer based on the current pricing system. Use the regression demand model to calculate the corresponding demand and Q4Y3 profit)		

Retrospective Approach 1 (Model-Based)

Consider a recent period, say Q4Y3 (part of our dataset)

- Use model to assess Analytics intervention
- Use model to assess current pricing system

	Y_i *given* Analytics intervention		Y_i *given* current system		Treatment effect = τ_i
Store i	$26,940	-	$25,759	=	**$1,181 (+4.6%)**

(Use the optimization model to determine prices and calculate profit using the regression demand model applied to Q4Y3)

(Use the actual prices set by the retailer based on the current pricing system. Use the regression demand model to calculate the corresponding demand and Q4Y3 profit)

Problems with this approach:

- It assumes the demand model is perfect, but all models are approximations
- The model provides a "home-court advantage" to the Analytics approach

(P.S.: This approach is very prevalent, don't get me going on this …)

Retrospective Approach 2 (Model vs. Actual)

Again, consider a recent period, say Q4Y3 (part of our dataset)

- Use model to assess Analytics intervention (same as before)
- Use actual historical revenue to assess current pricing system

	Y_i *given* Analytics intervention		Y_i *given* current system		Treatment effect = τ_i
Store i	$26,940	-	$28,282	=	**$-1,342 (-4.7%)**

(Use the optimization model to determine prices and calculate profit using the regression demand model applied to Q4Y3)

(Simply use actual realized profit; i.e. multiply the prices set by the retailer by the observed demand and sum over all weeks in Q4Y3)

Retrospective Approach 2 (Model vs. Actual)

Again, consider a recent period, say Q4Y3 (part of our dataset)

- Use model to assess Analytics intervention (same as before)
- Use actual historical revenue to assess current pricing system

	Y_i *given* Analytics intervention		Y_i *given* current system		Treatment effect = τ_i
Store i	$26,940	-	$28,282	=	**$-1,342 (-3.9%)**
	(Use the optimization model to determine prices and calculate profit using the regression demand model applied to Q4Y3)		(Simply use actual realized profit; i.e. multiply the prices set by the retailer by the observed demand and sum over all weeks in Q4Y3)		

Problems with this approach:

- Inconsistent accounting - comparing apples (revenue estimated from Model) with oranges (actual realized revenue).
- Exacerbates the effect of the demand model noise ε. Perhaps the model was unlucky here?

Retrospective Approach 3a (Before vs. After)

- Pilot test model on a <u>future period</u>, say, Q1Y4 and record actual realized revenue
- Use <u>actual historical revenue</u> on a recent period, say Q4Y3 to assess current pricing system

	Y_i *given* Analytics intervention		Y_i *given* current system		Treatment effect = τ_i
Store i	\$?? (note: model estimate = \$11, 568)	-	\$28,282	=	**\$??**

(Apply Analytics intervention to Q1Y4 and record actual realized profits)

(Simply use actual realized profit; i.e. multiply the prices set by the retailer by the observed demand and sum over all weeks in Q4Y3)

Retrospective Approach 3a (Before vs. After)

- Pilot test model on a <u>future period</u>, say, Q1Y4 and record actual realized revenue
- Use <u>actual historical revenue</u> on a recent period, say Q4Y3 to assess current pricing system

	Y_i *given* Analytics intervention		Y_i *given* current system		Treatment effect = τ_i
Store i	\$?? (note: model estimate = \$11, 568)	-	\$28,282	=	**\$??**
	(Apply Analytics intervention to Q1Y4 and record actual realized profits)		(Simply use actual realized profit; i.e. multiply the prices set by the retailer by the observed demand and sum over all weeks in Q4Y3)		

Problems:

- Apples (Q1Y4) to Oranges (Q4Y3) in time periods. (Seasonality, hello!)
- This is probably the worst approach of all!

Retrospective Approach 3b (Before vs. After)

- Well, Q4Y3 is clearly NOT comparable to Q1Y4 because of seasonality. Perhaps, we should compare with Q1Y3 (especially since the demand model has no trend effect)

	Y_i *given* Analytics intervention		Y_i *given* current system		Treatment effect = τ_i
Store i	\$?? (note: model estimate = \$11, 568)	-	\$11,680	=	**\$??**
	(Apply Analytics intervention to Q1Y4 and record actual realized profits)		(Simply use actual realized profit; i.e. multiply the prices set by the retailer by the observed demand and sum over all weeks in Q1Y3)		

Problems:

- Noise in actual demand?
- There may be other demand factors that still differentiate historical performance from future performance?

Retrospective Approach 3c (Before vs. After)

- We could go a step further and average out the past 3 years' first quarter profit.

	Y_i *given* Analytics intervention		Y_i *given* current system		Treatment effect = τ_i
Store i	\$?? (note: model estimate = \$11, 568)	-	\$11,192	=	**\$??**

(Apply Analytics intervention to Q1Y4 and record actual realized profits)

(Simply use actual realized profit; i.e. multiply the prices set by the retailer by the observed demand and sum over all weeks in Q1Y1, Q1Y2, Q1Y3, then average the profits of these 3 past quarters)

Problems:

- If other demand factors still differentiate historical performance from future performance, we may have exacerbated the problem.

Summary So Far

- "Retrospective" approaches fail because:
  - The demand equation is considered as the source of truth rather than a useful approximation, and/or
  - The comparisons come down to apples-to-oranges

Summary So Far

- "Retrospective" approaches fail because:
    - The demand equation is considered as the source or truth rather than a useful approximation, and/or
    - The comparison is apples-to-oranges

- Noteworthy: if the underlying predictive model is very highly accurate (say, because the prediction problem is very simple), then some of the retrospective approaches might work sufficiently well.

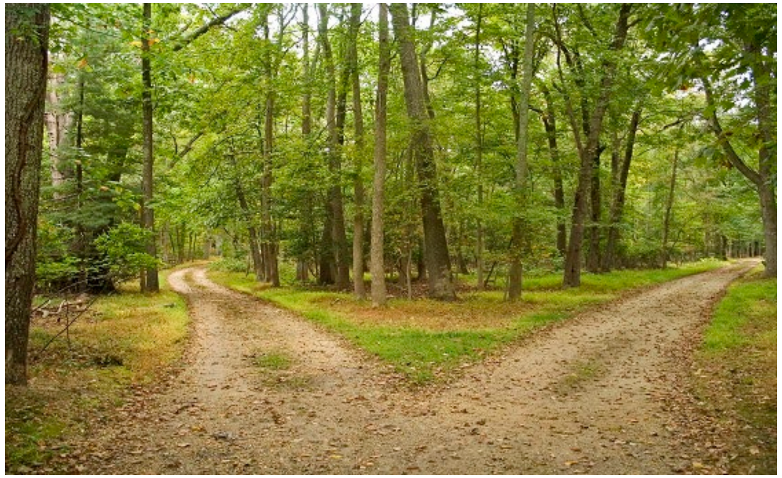
The Experimental Approach

Recall the Key Question

	Y_i *given* Analytics intervention		Y_i *given* current system		Treatment effect = τ_i
Store i	$??	-	$??	=	**$??**
	expected profitability of product k in store i at some given future period t when a particular Analytics tool is deployed		expected profitability of product k in store i at some given future period t under current pricing system		expected value added by Analytics compared to current system

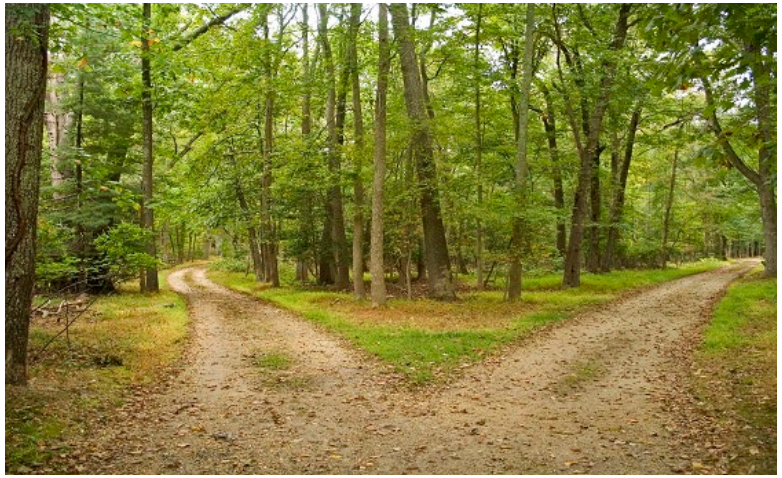
- The problem is that for any given store/product/time period we either deploy the Analytics intervention or keep using the current system – but we cannot do both.
- To fill out the above table we need two parallel universes, identical in every way except that in one universe we apply the Analytics intervention and in the other we stay with the the current system …

A Famous Line from Robert Frost

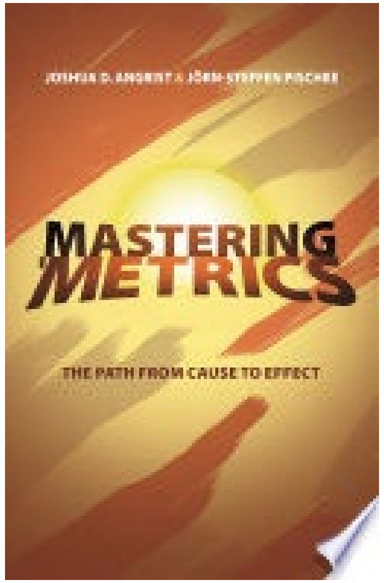


“Two roads diverged in a wood, and I—
I took the one less traveled by,
And that has made all the difference.”
- Robert Frost

Angrist on Frost



"Two roads diverged in a wood, and I—
I took the one less traveled by,
And that has made all the difference."
- Robert Frost



"The traveler claims his choice has mattered, but being only one person, he can't be sure. A later trip or a report by other travelers won't nail it down for him, either. Our narrator might be older and wiser the second time around, while other travelers might have different experiences on the same road. […] We can't know what lies at the end of the road not taken."
- Angrist and Pischke in Mastering 'Metrics: The Path from Cause to Effect

Experimental Approach: First Attempt

Experiment on two different stores with one as treatment and the other as control

- Pilot test model on Store i in Q1Y4 and record realized profit
- Keep current pricing system in Store i' in Q1Y4 and record realized profit

	Y_i *given* Analytics intervention	Y_i *given* current system	Treatment effect = τ_i
Store i	$\$y_i$	N\A	??
Store i'	N\A	$\$y_{i'}$	??
	(Apply Analytics intervention to Q1Y4 and record actual realized profits)	(Apply current pricing system to Q1Y4 and record actual realized profit)	Difference $\$y_i$ - $\$y_{i'}$ is taken as an estimate of Analytics intervention

Experimental Approach: First Attempt

Experiment on two different stores with one as treatment and the other as control

- Pilot test model on Store i in Q1Y4 and record realized profit
- Keep current pricing system in Store i' in Q1Y4 and record realized profit

	Y_i *given* Analytics intervention	Y_i *given* current system	Treatment effect = τ_i
Store i	$\$y_i$	N\A	??
Store i'	N\A	$\$y_{i'}$	??
	(Apply Analytics intervention to Q1Y4 and record actual realized profits)	(Apply current pricing system to Q1Y4 and record actual realized profit)	Difference $\$y_i - \$y_{i'}$ is taken as an estimate of Analytics intervention

Problems:

- Stores can have very different characteristics
- We need more samples!

The Experimental Approach: Second Attempt

Experiment on two sets of different stores with one as treatment group (say, Stores 1 and 4) and the other as control group (say, Stores 2, 4, and 5)

	Y_i *given* Analytics intervention	Y_i *given* current system	Treatment effect = τ_i
Store 1	$\$y_1$	N\A	??
Store 2	N\A	$\$y_2$	??
Store 3	N\A	$\$y_3$	??
Store 4	$\$y_4$	N\A	??
Store 5	N\A	$\$y_5$	??
Estimate of Average	$\mu_A = \$(y_1 + \$y_4) / 2$	$\mu_B = \$(y_2 + \$y_3 + \$y_5) / 3$	$\bar{\tau} = \mu_A - \mu_B$

The Experimental Approach: Second Attempt

Experiment on two sets of different stores with one as treatment group (say, Stores 1 and 4) and the other as control group (say, Stores 2, 4, and 5)

	Y_i *given* Analytics intervention	Y_i *given* current system	Treatment effect = τ_i
Store 1	$\$y_1$	N\A	??
Store 2	N\A	$\$y_2$	??
Store 3	N\A	$\$y_3$	??
Store 4	$\$y_4$	N\A	??
Store 5	N\A	$\$y_5$	??
Estimate of Average	$\mu_A = \$(y_1 + \$y_4) / 2$	$\mu_B = \$(y_2 + \$y_3 + \$y_5) / 3$	$\bar{\tau} = \mu_A - \mu_B$

- Can $\bar{\tau} = \mu_A - \mu_B$ be used as an estimate of the average value of the Analytics intervention compared to the current system?

The Experimental Approach: Second Attempt

Experiment on two sets of different stores with one as treatment group (say, Stores 1 and 4) and the other as control group (say, Stores 2, 4, and 5)

	Y_i *given* Analytics intervention	Y_i *given* current system	Treatment effect = τ_i
Store 1	$\$y_1$	N\A	??
Store 2	N\A	$\$y_2$	??
Store 3	N\A	$\$y_3$	??
Store 4	$\$y_4$	N\A	??
Store 5	N\A	$\$y_5$	??
Estimate of Average	$\mu_A = \$(y_1 + \$y_4) / 2$	$\mu_B = \$(y_2 + \$y_3 + \$y_5) / 3$	$\bar{\tau} = \mu_A - \mu_B$

- Potential problem: stores in the treatment and control groups might have very different characteristics and thus average profits

Treatment Effect vs. Difference in Group Means

	E[Y_i *given* Analytics intervention]	E[Y_i *given* current system]	Treatment effect, τ_i
Store 1	$\$y_1$	$\$x_1$	$\tau_1 = \$y_1 - \$x_1$
Store 2	N\A	$\$y_2$	???
Store 3	N\A	$\$y_3$	???
Store 4	$\$y_4$	$\$x_4$	$\tau_4 = \$y_4 - \$x_4$
Store 5	N\A	$\$y_5$	???
Estimate of Average	$\mu_A = \$(y_1 + \$y_4) / 2$	$\mu_B = \$(y_2 + \$y_3 + \$y_5) / 3$	

- Let x_i denote the (unobserved) counterfactuals in the treatment group.
- Consider the **average treatment effect on the treated**, $\tilde{\tau} = (\tau_1 + \tau_4) / 2$.
- How does $\mu_A - \mu_B$ relate to $\tilde{\tau}$?

Treatment Effect vs. Difference in Group Means

	E[Y_i *given* Analytics intervention]	E[Y_i *given* current system]	Treatment effect, τ_i
Store 1	$\$y_1$	$\$x_1$	$\tau_1 = \$y_1 - \$x_1$
Store 2	N\A	$\$y_2$	???
Store 3	N\A	$\$y_3$	???
Store 4	$\$y_4$	$\$x_4$	$\tau_4 = \$y_4 - \$x_4$
Store 5	N\A	$\$y_5$	???
Estimate of Average	$\mu_A = \$(y_1 + \$y_4) / 2$	$\mu_B = \$(y_2 + \$y_3 + \$y_5) / 3$	

$$\mu_A - \mu_B = \left(\frac{(y_1 + y_4)}{2} - \frac{(y_2 + y_3 + y_5)}{3}\right)$$

$$= \left(\frac{(y_1 - x_1) + (y_4 - x_4)}{2}\right)\left(\frac{(x_1 + x_4)}{2} - \frac{(y_2 + y_3 + y_5)}{3}\right)$$

$$= \left(\frac{\tau_1 + \tau_4}{2}\right)\left(\frac{(x_1 + x_4)}{2} - \frac{(y_2 + y_3 + y_5)}{3}\right)$$

$$= \tilde{\tau} + \left(\frac{(x_1 + x_4)}{2} - \frac{(y_2 + y_3 + y_5)}{3}\right)$$

Treatment Effect vs. Difference in Group Means

	E[Y_i *given* Analytics intervention]	E[Y_i *given* current system]	Treatment effect, τ_i
Store 1	$\$y_1$	$\$x_1$	$\tau_1 = \$y_1 - \$x_1$
Store 2	N\A	$\$y_2$	???
Store 3	N\A	$\$y_3$	???
Store 4	$\$y_4$	$\$x_4$	$\tau_4 = \$y_4 - \$x_4$
Store 5	N\A	$\$y_5$	???
Estimate of Average	$\mu_A = \$(y_1 + \$y_4) / 2$	$\mu_B = \$(y_2 + \$y_3 + \$y_5) / 3$	

- Consider the **average treatment effect on the treated**, $\tilde{\tau} = (\tau_1 + \tau_4) / 2$.
- How does $\bar{\tau}$ relate to $\mu_A - \mu_B$?

$$\mu_A - \mu_B = \tilde{\tau} + \left(\frac{(x_1 + x_4)}{2} - \frac{(y_2 + y_3 + y_5)}{3} \right)$$

Treatment Effect vs. Difference in Group Means

	E[Y_i *given* Analytics intervention]	E[Y_i *given* current system]	Treatment effect, τ_i
Store 1	$\$y_1$	$\$x_1$	$\tau_1 = \$y_1 - \$x_1$
Store 2	N\A	$\$y_2$	???
Store 3	N\A	$\$y_3$	???
Store 4	$\$y_4$	$\$x_4$	$\tau_4 = \$y_4 - \$x_4$
Store 5	N\A	$\$y_5$	???
Estimate of Average	$\mu_A = \$(y_1 + \$y_4) / 2$	$\mu_B = \$(y_2 + \$y_3 + \$y_5) / 3$	

- Consider the **average treatment effect on the treated**, $\tilde{\tau} = (\tau_1 + \tau_4) / 2$.
- How does $\bar{\tau}$ relate to $\mu_A - \mu_B$?

$$\mu_A - \mu_B = \tilde{\tau} + \underbrace{\left(\frac{(x_1 + x_4)}{2} - \frac{(y_2 + y_3 + y_5)}{3}\right)}_{\textbf{selection bias}}$$

The Experimental Approach Summary

Observed differences in average profit between model deployment and current approach

= Average treatment effect on the treated

+

Some bias due to the choice of stores

Eliminating Selection Bias with Randomness

	E[Y_i *given* Analytics intervention]	E[Y_i *given* current system]	Treatment effect, τ_i
Store 1			
Store 2			
Store 3			
Store 4			
Store 5			
Store 6			
Store 7			
Store 8			
Store 9			
Store 10			
Estimate of Average	$\mu_A = \$(y_1 + \ldots + \$y_5) / 5$	$\mu_B = \$(y_6 + \ldots + \$y_{10}) / 5$	

Non-random assignment

Eliminating Selection Bias with Randomness

	E[Y_i *given* Analytics intervention]	E[Y_i *given* current system]	Treatment effect, τ_i
Store 1	$\boxed{}$		
Store 2		$\boxed{}$	
Store 3	$\boxed{}$		
Store 4		$\boxed{}$	
Store 5	$\boxed{}$		Non-random assignment
Store 6		$\boxed{}$	
Store 7	$\boxed{}$		
Store 8		$\boxed{}$	
Store 9	$\boxed{}$		
Store 10		$\boxed{}$	
Estimate of Average	$\mu_A = \$(y_1 + \$y_3 + \$y_5 + \$y_7 + \$y_9) / 5$	$\mu_B = \$(y_2 + \$y_4 + \$y_6 + \$y_8 + \$y_{10}) / 5$	

Eliminating Selection Bias with Randomness

	E[Y_i *given* Analytics intervention]		E[Y_i *given* current system]	Treatment effect, τ_i
Store 1		**sample(c('T','C'),1)**		
Store 2		**sample(c('T','C'),1)**		
Store 3		**sample(c('T','C'),1)**		
Store 4		**sample(c('T','C'),1)**		
Store 5		**sample(c('T','C'),1)**		Random assignment
Store 6		**sample(c('T','C'),1)**		
Store 7		**sample(c('T','C'),1)**		
Store 8		**sample(c('T','C'),1)**		
Store 9		**sample(c('T','C'),1)**		
Store 10		**sample(c('T','C'),1)**		
Estimate of Average	$\mu_A = \$(y_1+\$y_2+\$y_4+\$y_7+\$y_9+\$y_{10}) / 6$		$\mu_B = \$(y_3+\$y_5+\$y_6+\$y_8) / 4$	

The Experimental Approach Summary

Observed differences in average profit between model deployment and current approach
= Average treatment effect on the treated
+
Some selection bias due to the choice of stores

- Selection bias can be eliminated using random assignment (and sufficient sample size)
- In other words, both treatment and control groups are on average identical in every aspect (known and unknown) except for the treatment (the Analytics intervention in this case).
- Random assignment also implies $\bar{\tau} = \tilde{\tau}$
- Notice that we have given up hope here of estimating the *individual* treatment effects. Instead, we're just looking to estimate the average treatment effect across stores.

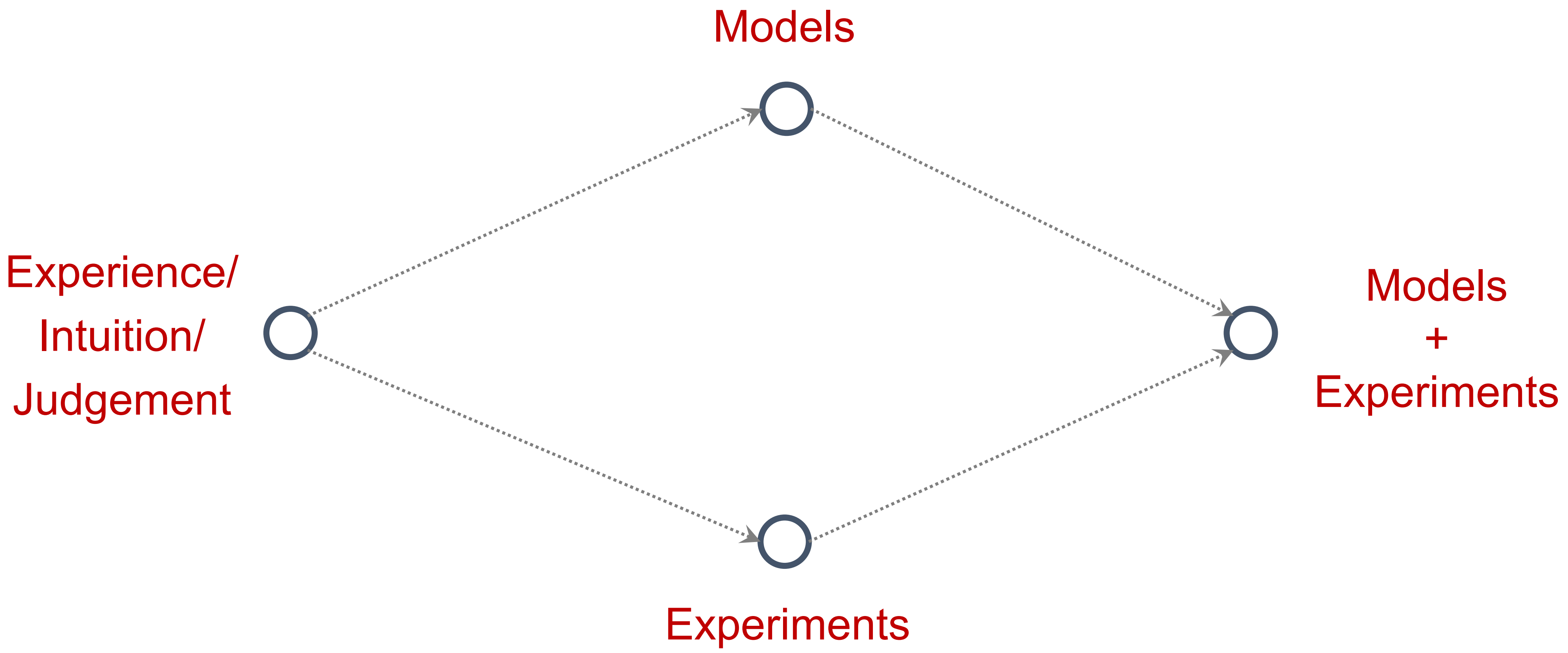
Business Experiments and The Analytics Edge (I)

- Randomized Control Experiments present a gold standard in *evaluating* the value of Analytics (i.e. the Analytics Edge).
  - When infeasible or too expensive, we may resort to alternative advanced approaches (including regression analysis and difference-in-differences).
  - Beware of the limitations of other simplistic retrospective approaches.

Business Experiments and The Analytics Edge (II)

- A/B testing is another important role of business experiments in *delivering* the Analytics Edge (a more direct alternative to the model-based approach).

Another Useful View of Analytics



Examples of Applied Research

Research Study 1

http://pubsonline.informs.org/journal/opre/

OPERATIONS RESEARCH
Vol. 65, No. 2, March–April 2017, pp. 446–468
ISSN 0030-364X (print), ISSN 1526-5463 (online)

The Impact of Linear Optimization on Promotion Planning

Maxime C. Cohen,[a] Ngai-Hang Zachary Leung,[b] Kiran Panchamgam,[c] Georgia Perakis,[d] Anthony Smith[c]

[a] Stern School of Business, New York University, New York, New York 10012; [b] College of Business, City University of Hong Kong, Kowloon, Hong Kong; [c] Oracle Retail Global Business Unit (RGBU), Burlington, Massachusetts 01803; [d] Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Contact: maxime.cohen@stern.nyu.edu (MCC); znhleung@cityu.edu.hk (N-HZL); kiran.panchamgam@oracle.com (KP); georgiap@mit.edu (GP); anthony.x.smith@oracle.com (AS)

Received: January 20, 2014
Revised: August 7, 2015; June 27, 2015
Accepted: September 23, 2016
Published Online in Articles in Advance: January 19, 2017

Subject Classifications: programming: integer, linear/applications, nonlinear/applications
Area of Review: Optimization

https://doi.org/10.1287/opre.2016.1573

Abstract. Sales promotions are important in the fast-moving consumer goods (FMCG) industry due to the significant spending on promotions and the fact that a large proportion of FMCG products are sold on promotion. This paper considers the problem of planning sales promotions for an FMCG product in a grocery retail setting. The category manager has to solve the promotion optimization problem (POP) for each product, i.e., how to select a posted price for each period in a finite horizon so as to maximize the retailer's profit. Through our collaboration with Oracle Retail, we developed an optimization formulation for the POP that can be used by category managers in a grocery environment. Our formulation incorporates business rules that are relevant, in practice. We propose general classes of demand functions (including multiplicative and additive), which incorporate the post-promotion dip effect, and can be estimated from sales data. In general, the POP formulation has a nonlinear objective and is NP-hard. We then propose a linear integer programming (IP) approximation of the POP. We show that the IP has an integral feasible region, and hence can be solved efficiently as a linear program (LP). We develop performance guarantees for the profit of the LP solution relative to the optimal profit. Using sales data from a grocery retailer, we first show that our demand models can be estimated with high accuracy, and then demonstrate that using the LP promotion schedule could potentially increase the profit by 3%, with a potential profit increase of 5% if some business constraints were to be relaxed.

Research Study 2

OPERATIONS RESEARCH
Vol. 60, No. 6, November–December 2012, pp. 1404–1422
ISSN 0030-364X (print) | ISSN 1526-5463 (online)

http://dx.doi.org/10.1287/opre.1120.1102

Clearance Pricing Optimization for a Fast-Fashion Retailer

Felipe Caro
UCLA Anderson School of Management, Los Angeles, California 90095, fcaro@anderson.ucla.edu

Jérémie Gallien
London Business School, London NW1 4SA, United Kingdom, jgallien@london.edu

Fast-fashion retailers such as Zara offer continuously changing assortments and use minimal in-season promotions. Their clearance pricing problem is thus challenging because it involves comparatively more different articles of unsold inventory with less historical price data points. Until 2007, Zara used a manual and informal decision-making process for determining price markdowns. In collaboration with their pricing team, we since designed and implemented an alternative process relying on a formal forecasting model feeding a price optimization model. As part of a controlled field experiment conducted in all Belgian and Irish stores during the 2008 fall-winter season, this new process increased clearance revenues by approximately 6%. Zara is currently using this process worldwide for its markdown decisions during clearance sales.

P.S.: Enjoy the Road not Taken!