Your MMM Is Broken:
Identification of Nonlinear and Time-varying Effects in Marketing Mix Models

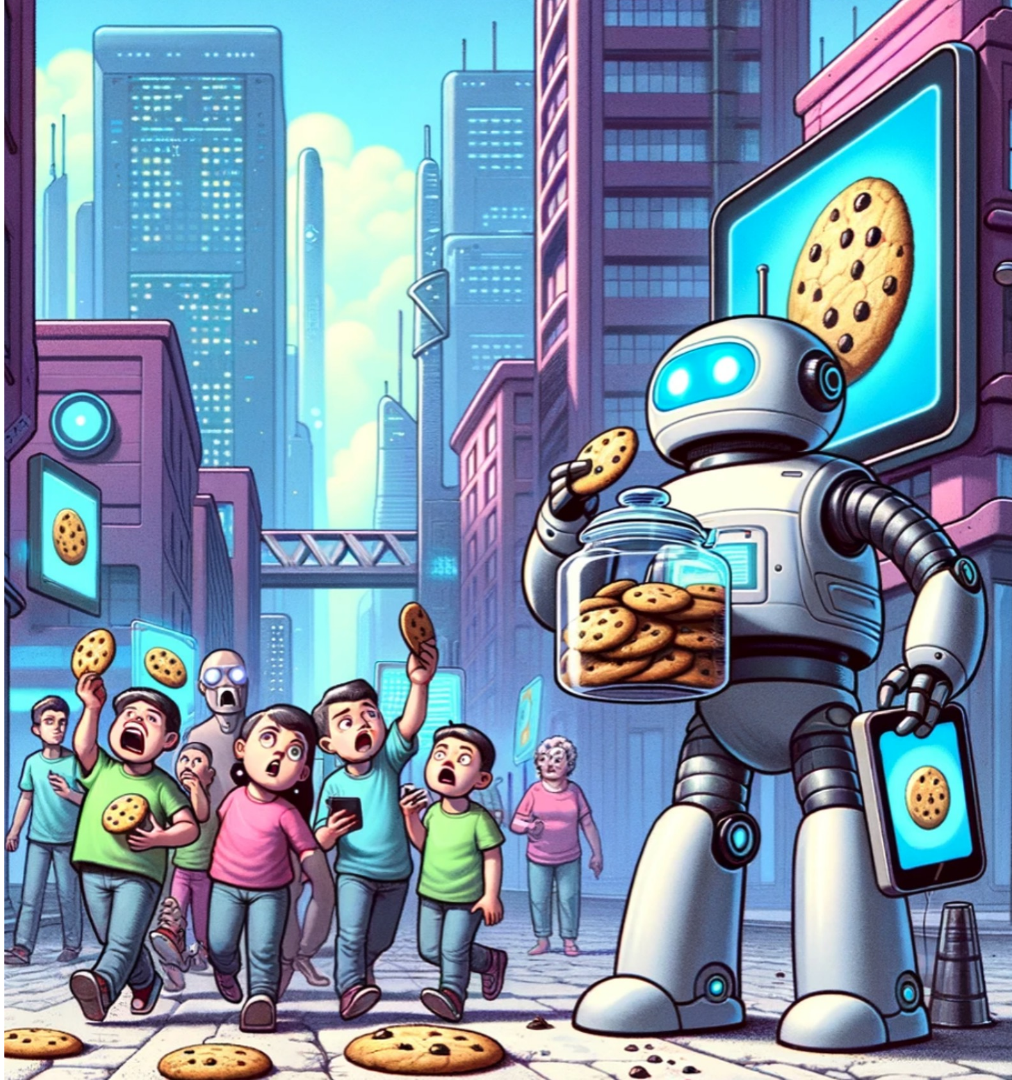
Ryan Dew

The Wharton School
University of Pennsylvania

Joint work with Nicolas Padilla (LBS) and Anya Shchetkina (Wharton)

The Cookie-free Future

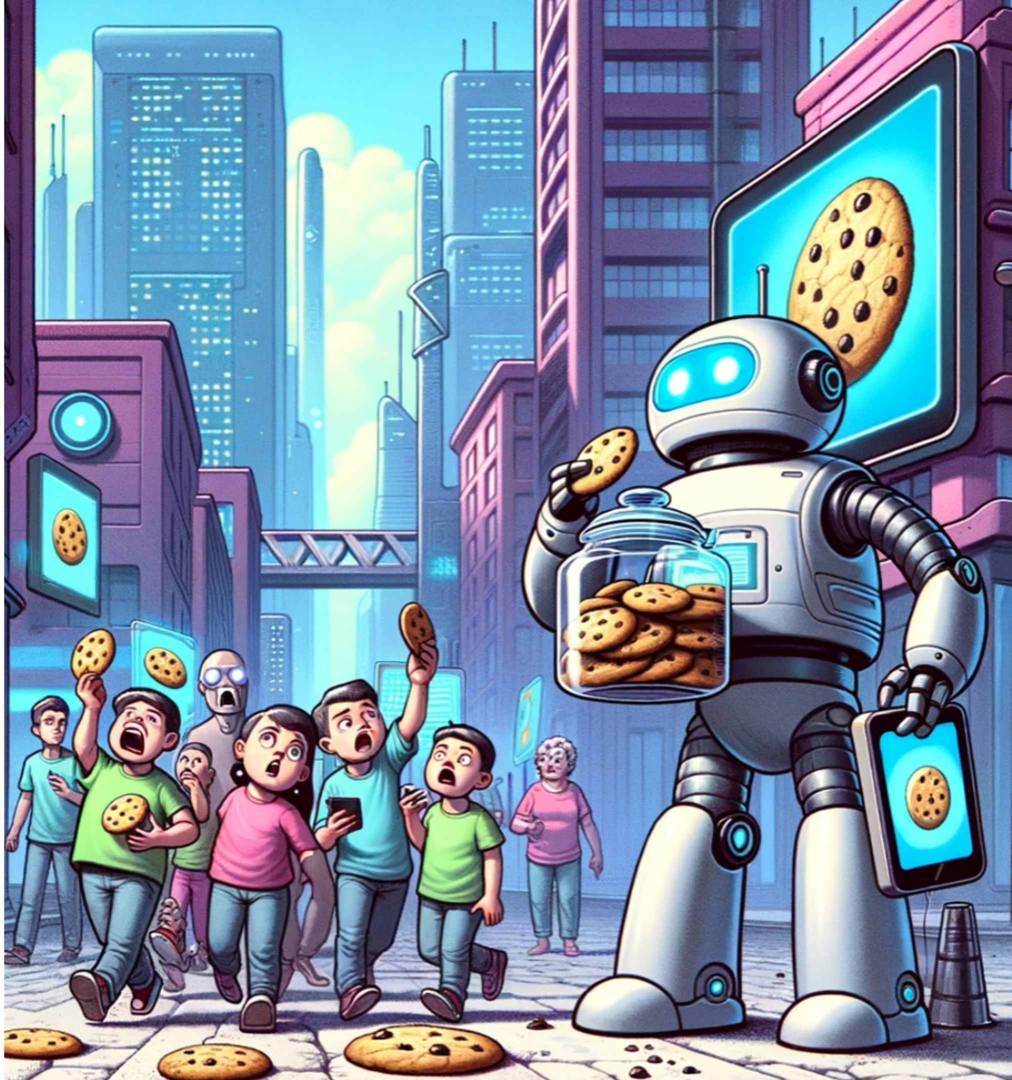
- For the past 10+ years, individual-level trackability has been central to measuring marketing efficacy



The Cookie-free Future

- For the past 10+ years, individual-level trackability has been central to measuring marketing efficacy

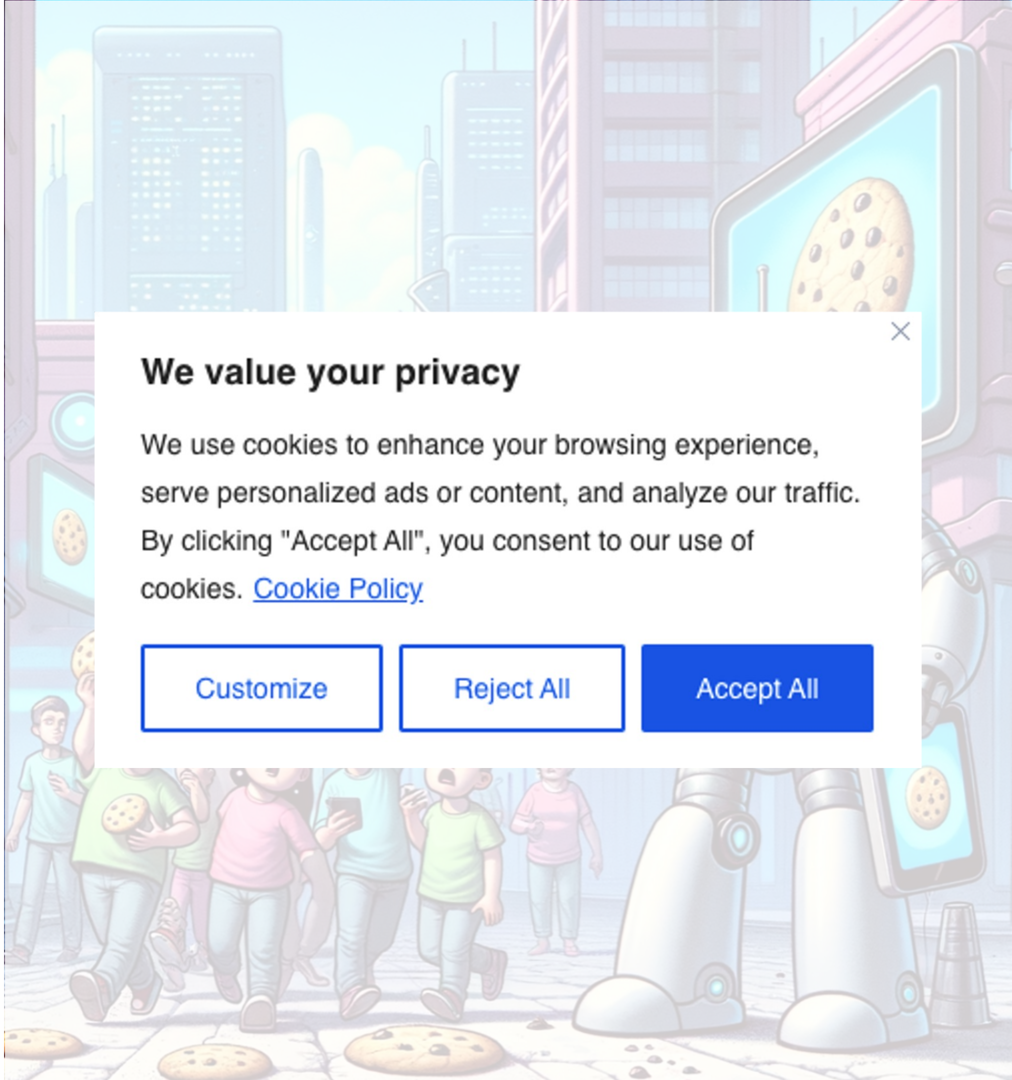
- Now? Gone!



The Cookie-free Future

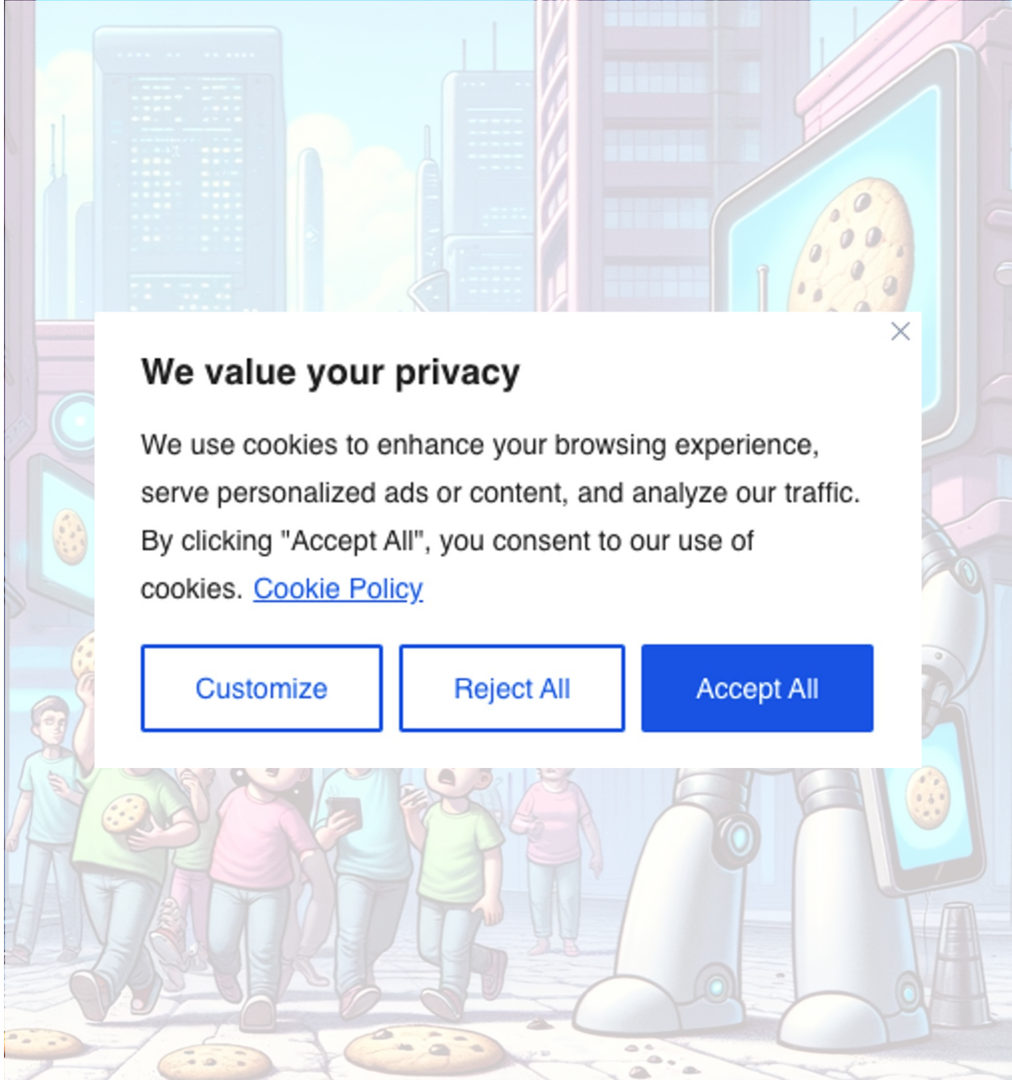
- For the past 10+ years, individual-level trackability has been central to measuring marketing efficacy

- Now? Gone!



The Cookie-free Future

- For the past 10+ years, individual-level trackability has been central to measuring marketing efficacy
- Now? Gone!
- Result: huge resurgence in interest in models based on **aggregate data**




THE
RETURN
OF THE
MARKETING MIX MODEL
PRODUCT
PLACE
PLACE
PROMOTION

[Submitted on 7 Jun 2021 (v1), last revised 5 Sep 2021 (this version, v3)]

Bayesian Time Varying Coefficient Model with Applications to Marketing Mix Modeling

Edwin Ng, Zhishi Wang, Athena Dai

Both Bayesian and varying coefficient models are very useful tools in practice as they can be used to model parameter heterogeneity in a generalizable way. Motivated by the need of enhancing Marketing Mix Modeling at Uber, we propose a Bayesian Time Varying Coefficient model, equipped with a hierarchical Bayesian structure. This model is different from other time varying coefficient models in the sense that the coefficients are weighted over a set of local latent variables following certain probabilistic distributions. Stochastic Variational Inference is used to approximate the posteriors of latent variables and dynamic coefficients. The proposed model also helps address many challenges faced by traditional MMM approaches. We used simulations as well as real world marketing datasets to demonstrate our model superior performance in terms of both accuracy and interpretability.

arXiv > stat > arXiv:2106.03322

Statistics > Applications

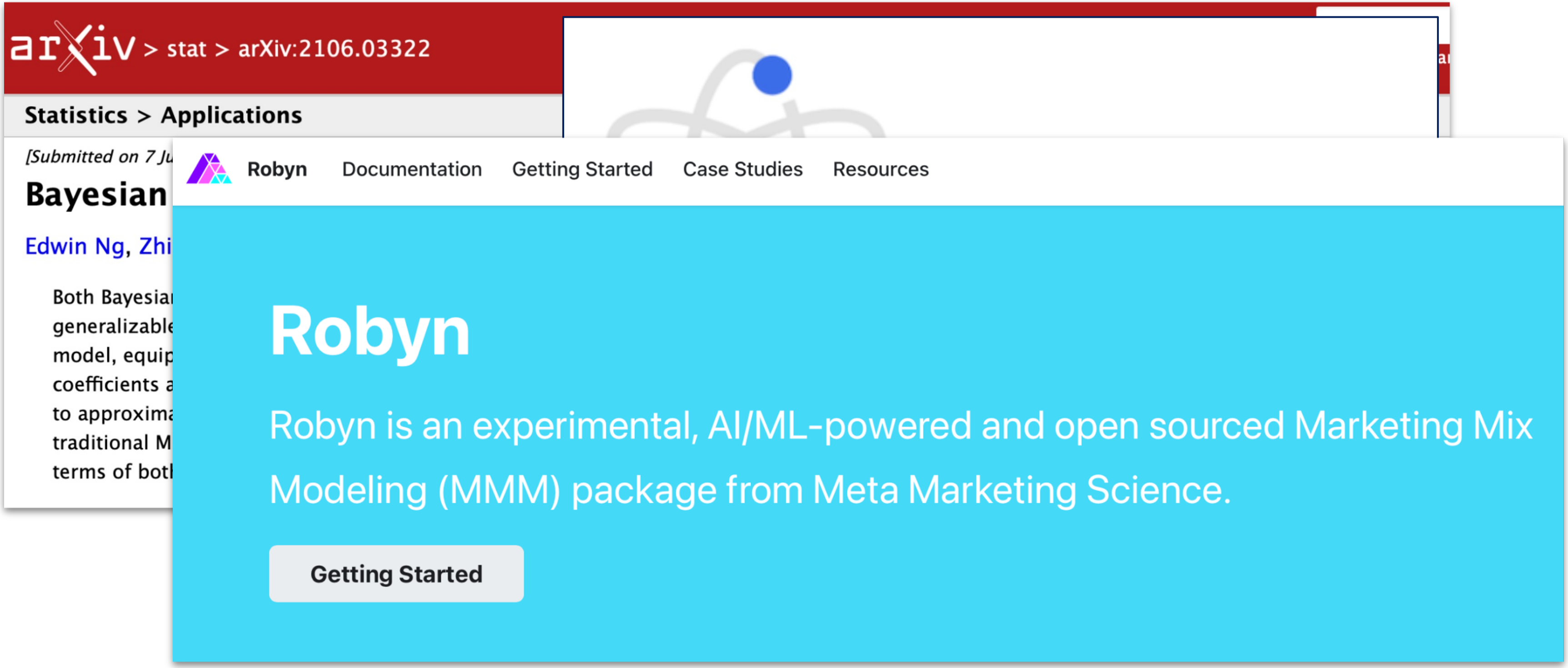
[Submitted on 7 Jun 2021 (v1), last revised 5 Sep 2021 (this versio

Bayesian Time Varying Coefficien

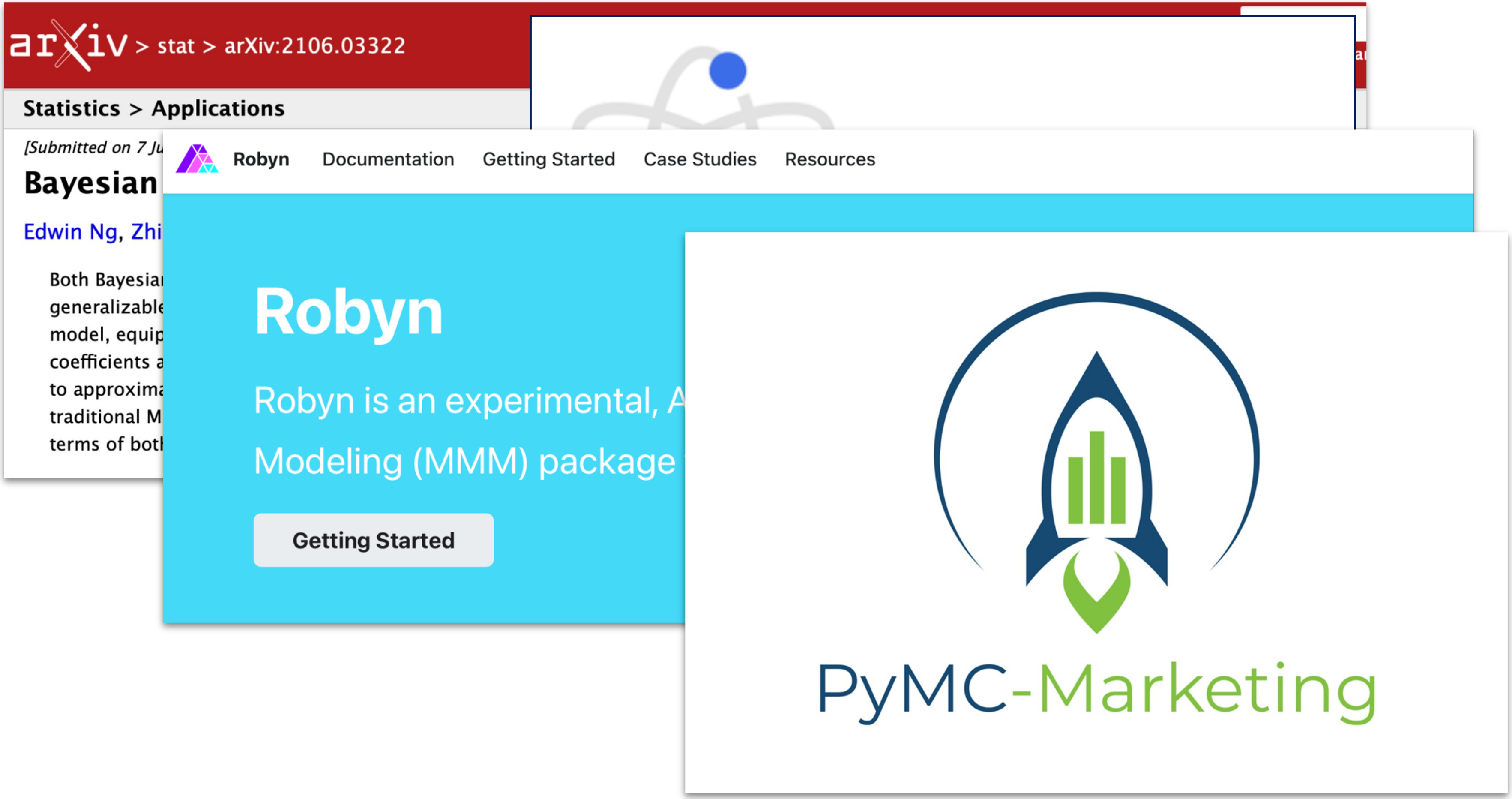
Edwin Ng, Zhishi Wang, Athena Dai

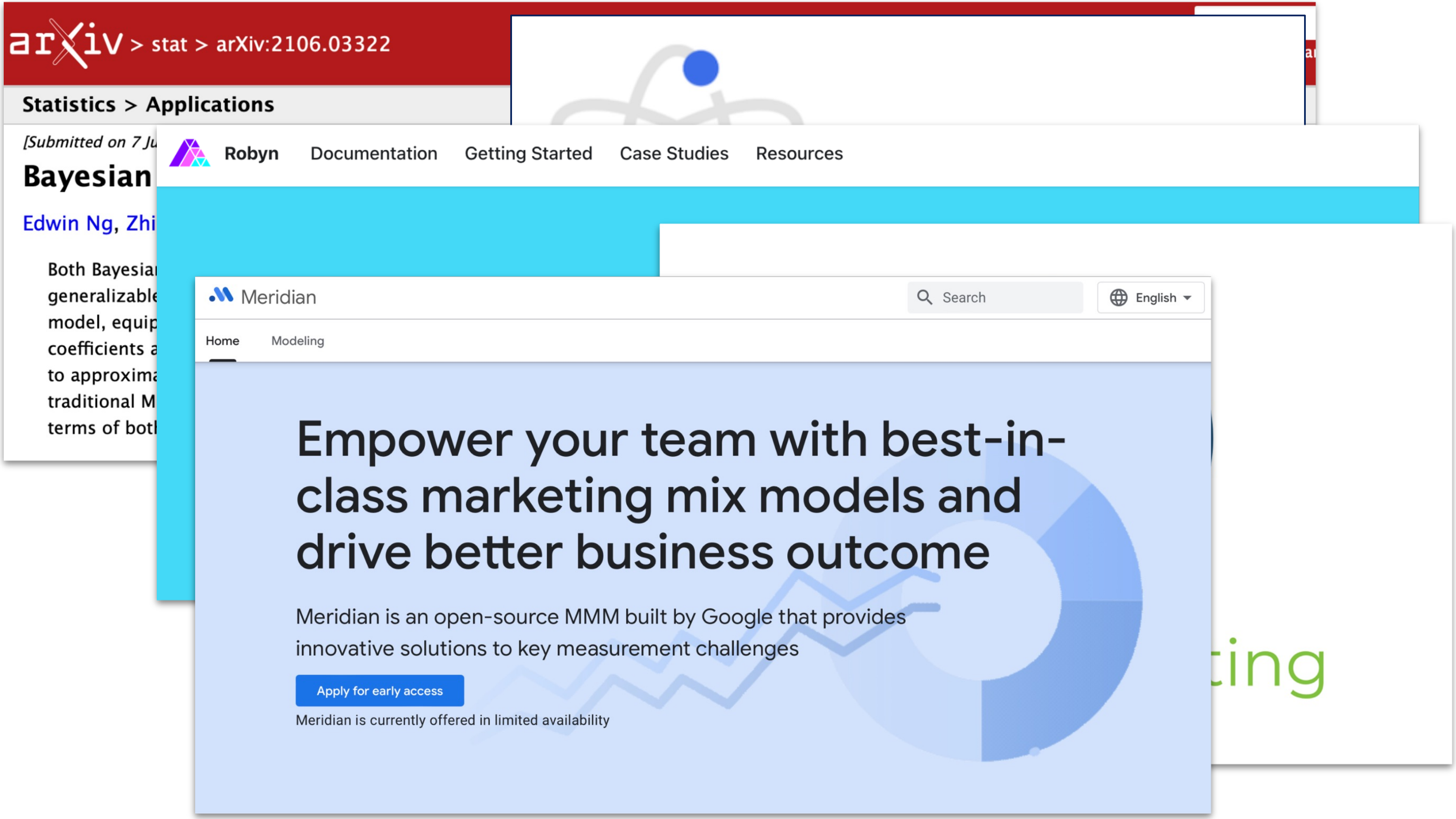
Orbit

Both Bayesian and varying coefficient models are very
generalizable way. Motivated by the need of enhancing
model, equipped with a hierarchical Bayesian structure
coefficients are weighted over a set of local latent vari
to approximate the posteriors of latent variables and dynamic coefficients. The proposed model also helps address many challenges faced by traditional MMM approaches. We used simulations as well as real world marketing datasets to demonstrate our model superior performance in terms of both accuracy and interpretability.



arXiv > stat > arXiv:2106.03322
Statistics > Applications
[Submitted on 7 Ju
Bayesian
Edwin Ng, Zhi
Both Bayesia
generalizable
model, equip
coefficients a
to approxima
traditional M
terms of both
Robyn
Documentation
Getting Started
Case Studies
Resources
Robyn
Robyn is an experimental, AI/ML-powered and open sourced Marketing Mix Modeling (MMM) package from Meta Marketing Science.
Getting Started


arXiv > stat > arXiv:2106.03322
Statistics > Applications
[Submitted on 7 Ju
Bayesian
Edwin Ng, Zhi
Both Bayesia
generalizable
model, equip
coefficients a
to approxima
traditional M
terms of bot
Robyn
Documentation
Getting Started
Case Studies
Resources
Robyn
Robyn is an experimental, A
Modeling (MMM) package
Getting Started
PyMC-Marketing



arXiv > stat > arXiv:2106.03322
Statistics > Applications
[Submitted on 7 Ju
Bayesian
Edwin Ng, Zhi
Both Bayesia
generalizable
model, equip
coefficients a
to approxima
traditional M
terms of both
Robyn
Documentation
Getting Started
Case Studies
Resources
Meridian
Search
English
Home
Modeling
Empower your team with best-in-class marketing mix models and drive better business outcome
Meridian is an open-source MMM built by Google that provides innovative solutions to key measurement challenges
Apply for early access
Meridian is currently offered in limited availability
ting

Marketing Mix Models

- Very long history in marketing
  Borden (1964), Palda (1965), Bultez and Naert (1979), Little (1979), Winer (1979), ..., Hanssens et al. (2003), ...

Marketing Mix Models

- Very long history in marketing
  Borden (1964), Palda (1965), Bultez and Naert (1979), Little (1979), Winer (1979), ..., Hanssens et al. (2003), ...
- At its heart: just regression

$$y_t = \sum_j \beta_j x_{jt} + \epsilon_t \quad \text{- or -} \quad \log y_t = \sum_j \beta_j \log x_{jt} + \epsilon_t$$

Marketing Mix Models

- Very long history in marketing
  Borden (1964), Palda (1965), Bultez and Naert (1979), Little (1979), Winer (1979), ..., Hanssens et al. (2003), ...
- At its heart: just regression

$$y_t = \sum_j \beta_j x_{jt} + \epsilon_t \quad \text{- or -} \quad \log y_t = \sum_j \beta_j \log x_{jt} + \epsilon_t$$

- ...with many well-documented **issues and extensions**

Marketing Mix Models

- Very long history in marketing
  Borden (1964), Palda (1965), Bultez and Naert (1979), Little (1979), Winer (1979), ..., Hanssens et al. (2003), ...
- At its heart: just regression

$$y_t = \sum_j \beta_j x_{jt} + \epsilon_t \quad \text{- or -} \quad \log y_t = \sum_j \beta_j \log x_{jt} + \epsilon_t$$

- ...with many well-documented **issues and extensions**
  - Long-term effects, often captured through **AdStock**: $S(x_t, L, \alpha) = \frac{\sum_{\ell=0}^{L-1} \alpha^\ell x_{t-\ell}}{\sum_{\ell=0}^{L-1} \alpha^\ell}$

Marketing Mix Models

- Very long history in marketing
  Borden (1964), Palda (1965), Bultez and Naert (1979), Little (1979), Winer (1979), ..., Hanssens et al. (2003), ...
- At its heart: just regression

$$y_t = \sum_j \beta_j x_{jt} + \epsilon_t \quad \text{- or -} \quad \log y_t = \sum_j \beta_j \log x_{jt} + \epsilon_t$$

- ...with many well-documented **issues and extensions**
  - Long-term effects, often captured through **AdStock**: $S(x_t, L, \alpha) = \frac{\sum_{\ell=0}^{L-1} \alpha^\ell x_{t-\ell}}{\sum_{\ell=0}^{L-1} \alpha^\ell}$
  - **Nonlinear** effects: $y_t = \sum_j f_j(x_{jt}) + \epsilon_t$

Marketing Mix Models

- Very long history in marketing
  Borden (1964), Palda (1965), Bultez and Naert (1979), Little (1979), Winer (1979), ..., Hanssens et al. (2003), ...
- At its heart: just regression

$$y_t = \sum_j \beta_j x_{jt} + \epsilon_t \quad \text{- or -} \quad \log y_t = \sum_j \beta_j \log x_{jt} + \epsilon_t$$

- ...with many well-documented **issues and extensions**
  - Long-term effects, often captured through **AdStock**: $S(x_t, L, \alpha) = \frac{\sum_{\ell=0}^{L-1} \alpha^\ell x_{t-\ell}}{\sum_{\ell=0}^{L-1} \alpha^\ell}$
  - **Nonlinear** effects: $y_t = \sum_j f_j(x_{jt}) + \epsilon_t$
  - **Time-varying** effects: $y_t = \sum_j \beta_{jt} x_{jt} + \epsilon_t$

Marketing Mix Models

- Very long history in marketing
  Borden (1964), Palda (1965), Bultez and Naert (1979), Little (1979), Winer (1979), ..., Hanssens et al. (2003), ...
- At its heart: just regression

$$y_t = \sum_j \beta_j x_{jt} + \epsilon_t \quad \text{- or -} \quad \log y_t = \sum_j \beta_j \log x_{jt} + \epsilon_t$$

- ...with many well-documented **issues and extensions**
  - Long-term effects, often captured through **AdStock**: $S(x_t, L, \alpha) = \frac{\sum_{\ell=0}^{L-1} \alpha^\ell x_{t-\ell}}{\sum_{\ell=0}^{L-1} \alpha^\ell}$
  - **Nonlinear** effects: $y_t = \sum_j f_j(x_{jt}) + \epsilon_t$
  - **Time-varying** effects: $y_t = \sum_j \beta_{jt} x_{jt} + \epsilon_t$
  - Endogeneity 😨

Marketing Mix Models

- Very long history in marketing
  Borden (1964), Palda (1965), Bultez and Naert (1979), Little (1979), Winer (1979), ..., Hanssens et al. (2003), ...
- At its heart: just regression

$$y_t = \sum_j \beta_j x_{jt} + \epsilon_t \quad \text{- or -} \quad \log y_t = \sum_j \beta_j \log x_{jt} + \epsilon_t$$

- ...with many well-documented **issues and extensions**
  - Long-term effects, often captured through **AdStock**: $S(x_t, L, \alpha) = \frac{\sum_{\ell=0}^{L-1} \alpha^\ell x_{t-\ell}}{\sum_{\ell=0}^{L-1} \alpha^\ell}$
  - **Nonlinear** effects: $y_t = \sum_j f_j(x_{jt}) + \epsilon_t$
  - **Time-varying** effects: $y_t = \sum_j \beta_{jt} x_{jt} + \epsilon_t$
  - Endogeneity 😨

Note: Terminology

Throughout the talk, dynamic = time-varying

What’s new? Firepower

What’s new? Firepower

Meridian

Wang et al. (2017), Jin et al. (2017), Sun et al. (2017), Zhang et al. (2023)

$$
\begin{aligned}
y_{g,t} = \mu_t + \tau_g + \sum_{c=1}^{C} \gamma_{g,c} z_{g,t,c} \\
+ \sum_{m=1}^{M} \beta_{g,m} HillAdstock\left(\left\{x_{g,t-s,m}\right\}_{s=0}^{L}\ ;\ \alpha_m, ec_m, slope_m\right) \\
+ \sum_{n=1}^{N} \beta_{g,n}^{(rf)} Adstock\left(\left\{r_{g,t-s,n} \cdot Hill\left(f_{g,t-s,n};\ ec_n^{(rf)},\ slope_n^{(rf)}\right)\right\}_{s=0}^{L}\ ;\ \alpha_n^{(rf)}\right) \\
+ \epsilon_{g,t}
\end{aligned}
$$

What's new? Firepower

Meridian

Wang et al. (2017), Jin et al. (2017), Sun et al. (2017), Zhang et al. (2023)

$$
\begin{aligned}
y_{g,t} = \mu_t + \tau_g + \sum_{c=1}^{C} \gamma_{g,c} z_{g,t,c} \\
&+ \sum_{m=1}^{M} \beta_{g,m} HillAdstock\left(\left\{ x_{g,t-s,m} \right\}_{s=0}^{L} \; ; \; \alpha_m, ec_m, slope_m \right) \\
&+ \sum_{n=1}^{N} \beta_{g,n}^{(rf)} Adstock\left(\left\{ r_{g,t-s,n} \cdot Hill\left(f_{g,t-s,n}; \; ec_n^{(rf)}, \; slope_n^{(rf)} \right) \right\}_{s=0}^{L} \; ; \; \alpha_n^{(rf)} \right) \\
&+ \epsilon_{g,t}
\end{aligned}
$$

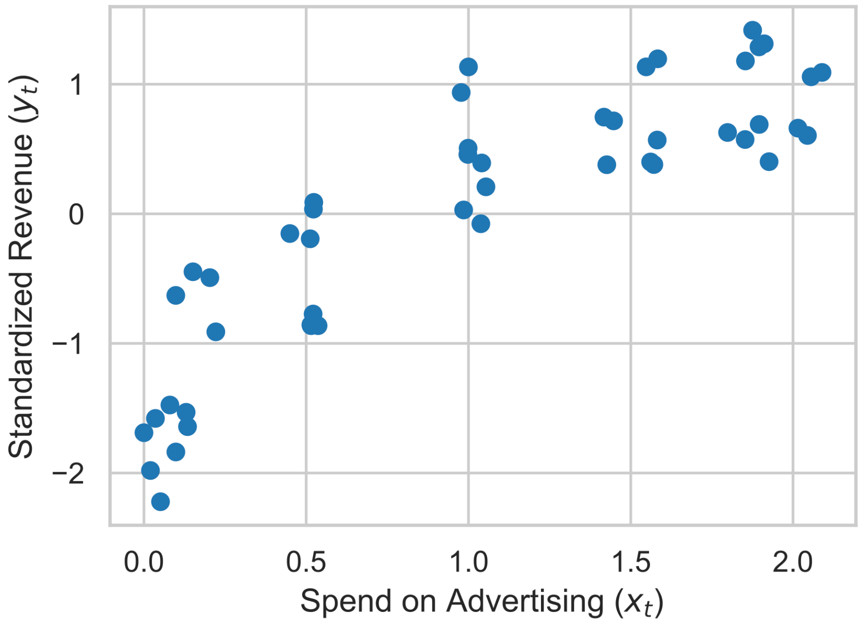
Uber's Orbit

Ng et al., (2021)

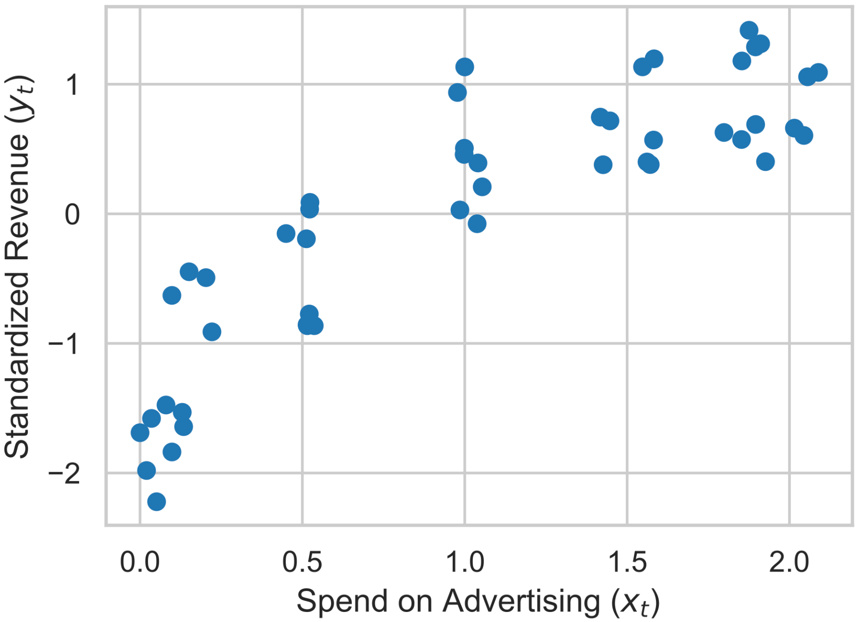
$$ln(\hat{y}_t) = l_t + s_t + \sum_{p=1}^{P} ln(x_{t,p})\beta_{t,p}$$

$$\beta_{t,p} = \sum_j w_j(t) \cdot b_{j,p},$$

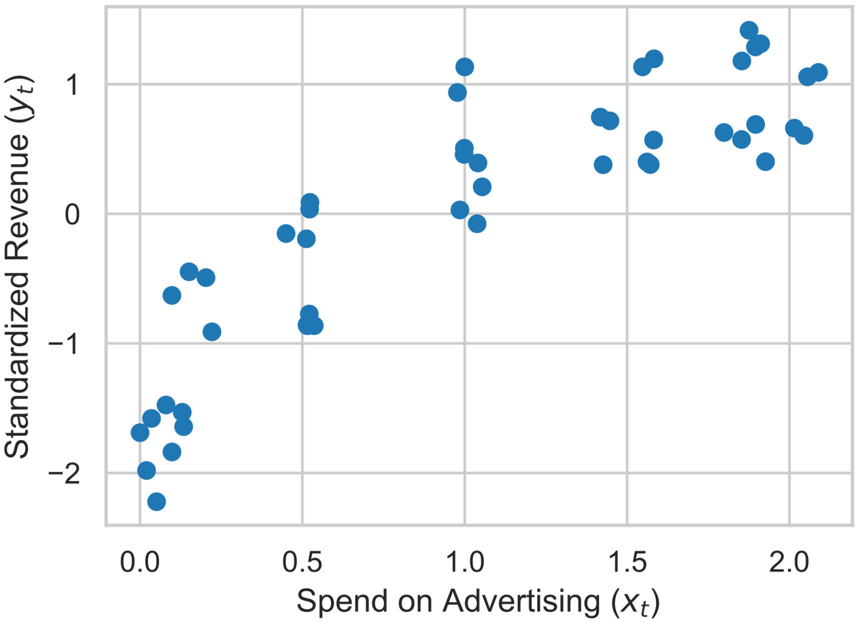
$$w_j(t) = k(t, t_j) / \sum_{i=1}^{J} k(t, t_i),$$



1
0
−1
−2
0.0
0.5
1.0
1.5
2.0
Standardized Revenue (y_t)
Spend on Advertising (x_t)



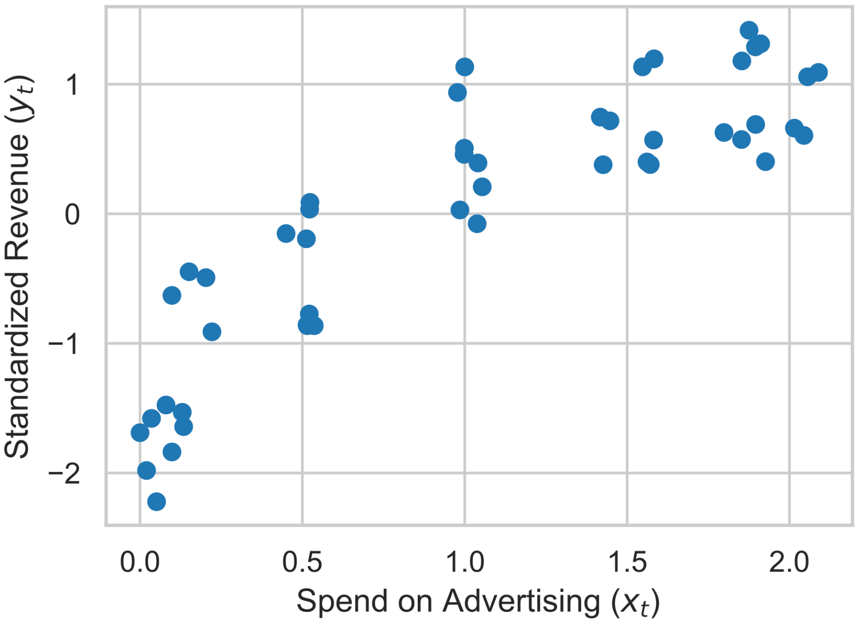
Surely this is nonlinear, right...?



Surely this is nonlinear, right...?

True data generating process (DGP):

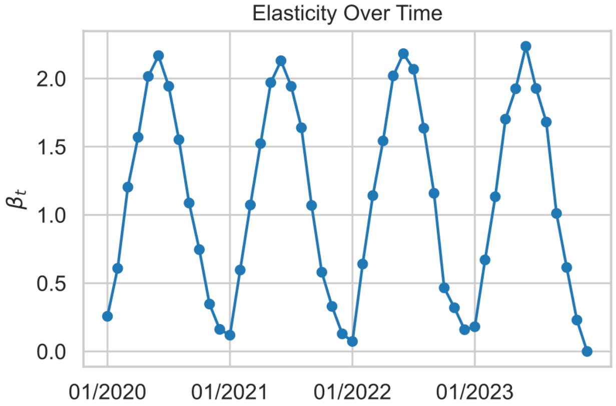
$$y_t = \beta_t x_t + \epsilon_t$$

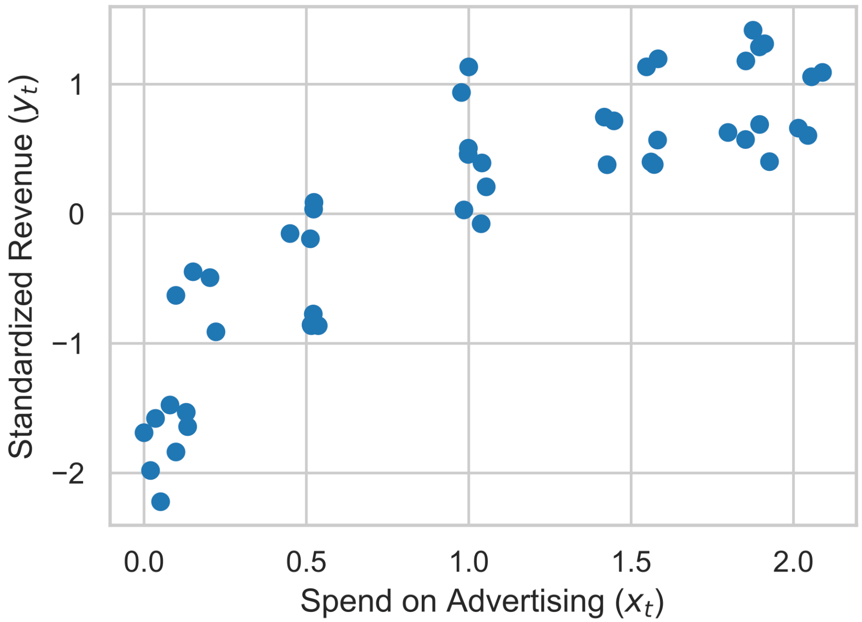


Surely this is nonlinear, right...?

True data generating process (DGP):

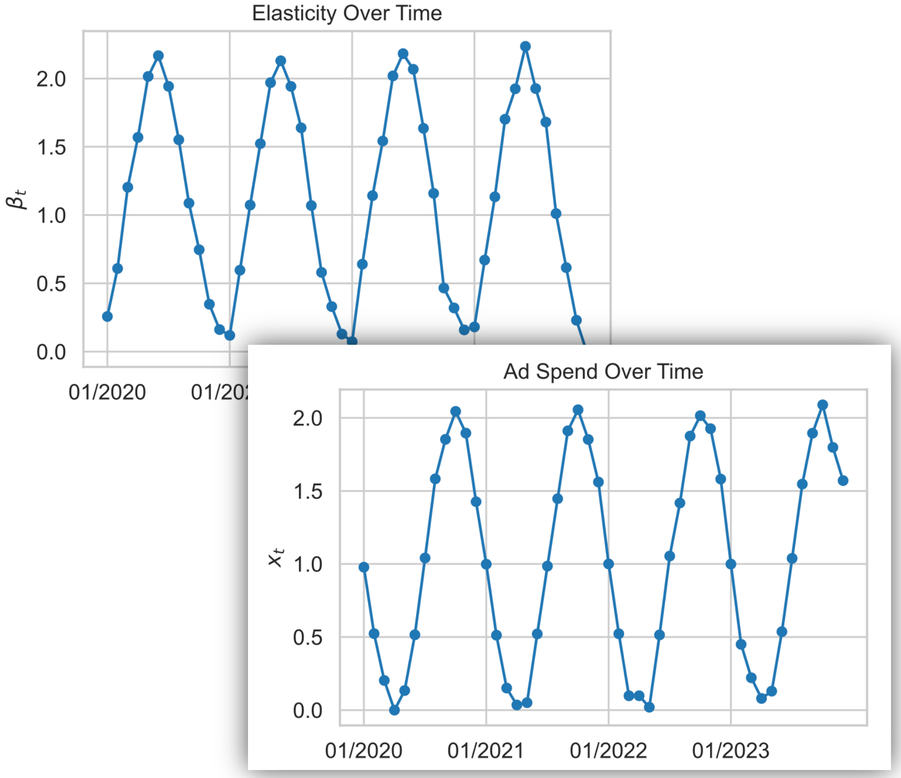
$$y_t = \beta_t x_t + \epsilon_t$$





True data generating process (DGP):

$$y_t = \beta_t x_t + \epsilon_t$$



Surely this is nonlinear, right...?

When can nonlinear and time-varying effects be identified in marketing mix models?

When can nonlinear and time-varying effects be identified in marketing mix models?

1. Introduce a **flexible, nonparametric framework** to estimate both types of effects

When can nonlinear and time-varying effects be identified in marketing mix models?

1. Introduce a **flexible, nonparametric framework** to estimate both types of effects

2. Through both theory and simulations, we show that, under realistic conditions, **the two models <u>will</u> be conflated**

When can nonlinear and time-varying effects be identified in marketing mix models?

1. Introduce a **flexible, nonparametric framework** to estimate both types of effects
2. Through both theory and simulations, we show that, under realistic conditions, **the two models <u>will</u> be conflated**
3. Problems are exacerbated under common managerial practices, like **autoregressive decision-making**

When can nonlinear and time-varying effects be identified in marketing mix models?

1. Introduce a **flexible, nonparametric framework** to estimate both types of effects
2. Through both theory and simulations, we show that, under realistic conditions, **the two models will be conflated**
3. Problems are exacerbated under common managerial practices, like **autoregressive decision-making**
4. Similarly fitting models can have **fundamentally different implications** in terms of optimal decision-making

(A little) Math

When can dynamic approximate nonlinear?

When can dynamic approximate nonlinear?

$$y_t = \sum_j \beta_{jt} x_{jt} + \epsilon_t$$

$$y_t = \sum_j f_j(x_{jt}) + \epsilon_t$$

When can dynamic approximate nonlinear?

$$y_t = \sum_j \beta_{jt} x_{jt} + \epsilon_t$$

$$y_t = \sum_j f_j(x_{jt}) + \epsilon_t$$

- First, a simple point: if $\beta_t = y_t / x_t$, the dynamic model will fit perfectly!

When can dynamic approximate nonlinear?

$$y_t = \sum_j \beta_{jt} x_{jt} + \epsilon_t \qquad\qquad y_t = \sum_j f_j(x_{jt}) + \epsilon_t$$

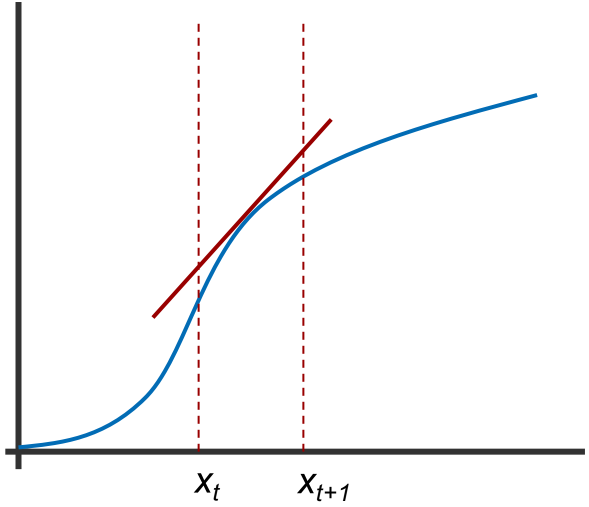
- First, a simple point: if $\beta_t = y_t / x_t$, the dynamic model will fit perfectly!
- A more nuanced answer: in any local region of x, a local linear approximation will work pretty well

When can dynamic approximate nonlinear?

$$y_t = \sum_j \beta_{jt} x_{jt} + \epsilon_t$$

$$y_t = \sum_j f_j(x_{jt}) + \epsilon_t$$

- First, a simple point: if $\beta_t = y_t / x_t$, the dynamic model will fit perfectly!
- A more nuanced answer: in any local region of x, a local linear approximation will work pretty well

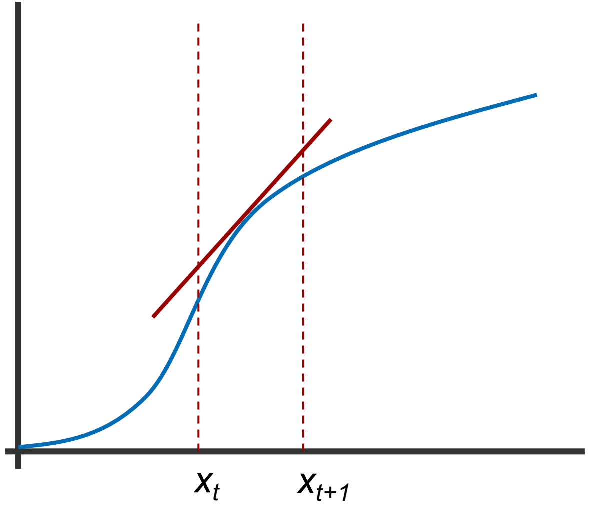


When can dynamic approximate nonlinear?

$$y_t = \sum_j \beta_{jt} x_{jt} + \epsilon_t$$

$$y_t = \sum_j f_j(x_{jt}) + \epsilon_t$$

- First, a simple point: if $\beta_t = y_t / x_t$, the dynamic model will fit perfectly!
- A more nuanced answer: in any local region of x, a local linear approximation will work pretty well
- When will β_t be smooth ("forecastable")? If *f* is smooth and x_t and x_{t+1} are close together!



When can nonlinear approximate dynamic?

$$y_t = \sum_j f_j(x_{jt}) + \epsilon_t$$

$$y_t = \sum_j \beta_{jt} x_{jt} + \epsilon_t$$

When can nonlinear approximate dynamic?

$$y_t = \sum_j f_j(x_{jt}) + \epsilon_t$$

$$y_t = \sum_j \beta_{jt} x_{jt} + \epsilon_t$$

- Additional notation: let $\beta_t = \mathcal{B}(t)$, $x_t = \mathcal{X}(t)$

When can nonlinear approximate dynamic?

$$y_t = \sum_j f_j(x_{jt}) + \epsilon_t \qquad\qquad y_t = \sum_j \beta_{jt} x_{jt} + \epsilon_t$$

- Additional notation: let $\beta_t = \mathcal{B}(t)$, $x_t = \mathcal{X}(t)$
- The nonlinear model can fully approximate the dynamic model when $\mathcal{B}(t) = g\big(\mathcal{X}(t)\big)$ for some function g

When can nonlinear approximate dynamic?

$$y_t = \sum_j f_j(x_{jt}) + \epsilon_t \qquad\qquad y_t = \sum_j \beta_{jt} x_{jt} + \epsilon_t$$

- Additional notation: let $\beta_t = \mathcal{B}(t)$, $x_t = \mathcal{X}(t)$
- The nonlinear model can fully approximate the dynamic model when $\mathcal{B}(t) = g\big(\mathcal{X}(t)\big)$ for some function g
- Simple case: $\mathcal{X}(t)$ is monotonic, in which case the nonlinear model is:
$$f(x) = \mathcal{B}\big(\mathcal{X}^{-1}(t)\big)\, x$$

When can nonlinear approximate dynamic?

$$y_t = \sum_j f_j(x_{jt}) + \epsilon_t \qquad\qquad y_t = \sum_j \beta_{jt} x_{jt} + \epsilon_t$$

- Additional notation: let $\beta_t = \mathcal{B}(t)$, $x_t = \mathcal{X}(t)$
- The nonlinear model can fully approximate the dynamic model when $\mathcal{B}(t) = g\big(\mathcal{X}(t)\big)$ for some function g
- Simple case: $\mathcal{X}(t)$ is monotonic, in which case the nonlinear model is:
$$f(x) = \mathcal{B}\big(\mathcal{X}^{-1}(t)\big)\, x$$
- Less simple case: if x and β both derive from a common parent process

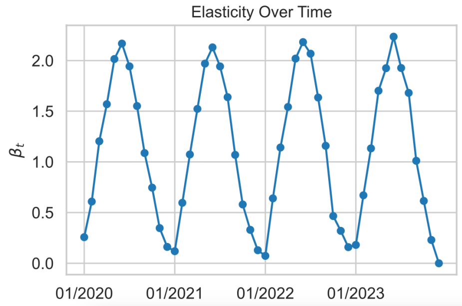
When can nonlinear approximate dynamic?

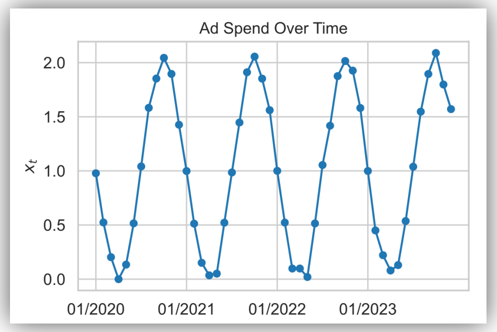
$$y_t = \sum_j f_j(x_{jt}) + \epsilon_t \qquad\qquad y_t = \sum_j \beta_{jt} x_{jt} + \epsilon_t$$

- Additional notation: let $\beta_t = \mathcal{B}(t)$, $x_t = \mathcal{X}(t)$
- The nonlinear model can fully approximate the dynamic model when $\mathcal{B}(t) = g\big(\mathcal{X}(t)\big)$ for some function g
- Simple case: $\mathcal{X}(t)$ is monotonic, in which case the nonlinear model is:

$$f(x) = \mathcal{B}\big(\mathcal{X}^{-1}(t)\big)\, x$$

- Less simple case: if x and β both derive from a common parent process





Simulations

When does conflation *actually* happen?

When does conflation *actually* happen?

Four types of simulations:

1. Flexible nonlinear response

$$y_t = f(x_t) + \varepsilon_t$$

2. Dynamic coefficients, linear (or log-linear) model

$$y_t = \beta(t)x_t + \varepsilon_t$$

3. Nonlinear response, parametric hill function

$$y_t = \mathrm{Hill}(x_t) + \varepsilon_t$$

4. Dynamic coefficients inherited from common parent

$$y_t = \beta(t)x(t) + \varepsilon_t, \;\; (\beta(t), x(t)) \sim \mathrm{Pa}(t)$$

When does conflation *actually* happen?

Four types of simulations:

1. Flexible nonlinear response

$$y_t = f(x_t) + \varepsilon_t$$

2. Dynamic coefficients, linear (or log-linear) model

$$y_t = \beta(t)x_t + \varepsilon_t$$

3. Nonlinear response, parametric hill function

$$y_t = \mathrm{Hill}(x_t) + \varepsilon_t$$

4. Dynamic coefficients inherited from common parent

$$y_t = \beta(t)x(t) + \varepsilon_t, \ (\beta(t), x(t)) \sim \mathrm{Pa}(t)$$

When does conflation *actually* happen?

Four types of simulations:

1. Flexible nonlinear response
$$y_t = \boxed{f(x_t)} + \varepsilon_t$$

2. Dynamic coefficients, linear (or log-linear) model
$$y_t = \boxed{\beta(t)} x_t + \varepsilon_t$$

3. Nonlinear response, parametric hill function
$$y_t = \mathrm{Hill}(x_t) + \varepsilon_t$$

4. Dynamic coefficients inherited from common parent
$$y_t = \beta(t)x(t) + \varepsilon_t, \; (\beta(t), x(t)) \sim \mathrm{Pa}(t)$$

When does conflation *actually* happen?

Four types of simulations:

1. Flexible nonlinear response

$$y_t = \boxed{f(x_t)} + \varepsilon_t$$

2. Dynamic coefficients, linear (or log-linear) model

$$y_t = \boxed{\beta(t)} x_t + \varepsilon_t$$

3. Nonlinear response, parametric hill function

$$y_t = \mathrm{Hill}(x_t) + \varepsilon_t$$

4. Dynamic coefficients inherited from common parent

$$y_t = \beta(t)x(t) + \varepsilon_t,\ (\beta(t), x(t)) \sim \mathrm{Pa}(t)$$

Gaussian processes

Two important levers:

- Smoothness
- Amplitude

Primer: Gaussian Processes

Gaussian processes: a Bayesian nonparametric approach to modeling unknown functions

Primer: Gaussian Processes

Gaussian processes: a Bayesian nonparametric approach to modeling unknown functions

$$f(x) \sim \mathcal{GP}(m(x), k(x, x'))$$

Primer: Gaussian Processes

Gaussian processes: a Bayesian nonparametric approach to modeling unknown functions

$$f(x) \sim \mathcal{GP}(m(x), k(x, x'))$$

$$\uparrow\downarrow$$

$$f(x_1, \ldots, x_N) \sim \mathcal{N}(m(x_1, \ldots, x_N), K), \text{ s.t. } K_{ij} = k(x_i, x_j)$$

Primer: Gaussian Processes

Gaussian processes: a Bayesian nonparametric approach to modeling unknown functions

$$f(x) \sim \mathcal{GP}(m(x), k(x, x'))$$

$$\Updownarrow$$

$$f(x_1, \ldots, x_N) \sim \mathcal{N}(m(x_1, \ldots, x_N), K), \text{ s.t. } K_{ij} = k(x_i, x_j)$$

$$k(x, x') = \eta^2 \exp\left\{-\frac{(x - x')^2}{2\rho^2}\right\}$$

Primer: Gaussian Processes

Gaussian processes: a Bayesian nonparametric approach to modeling unknown functions

$$f(x) \sim \mathcal{GP}(m(x), k(x, x'))$$

$$\Updownarrow$$

$$f(x_1, \ldots, x_N) \sim \mathcal{N}(m(x_1, \ldots, x_N), K), \text{ s.t. } K_{ij} = k(x_i, x_j)$$

Amplitude

$$k(x, x') = \boxed{\eta^2} \exp\left\{-\frac{(x - x')^2}{2\boxed{\rho^2}}\right\}$$

Lengthscale

Primer: Gaussian Processes

Gaussian processes: a Bayesian nonparametric approach to modeling unknown functions

$$f(x) \sim \mathcal{GP}(m(x), k(x, x'))$$

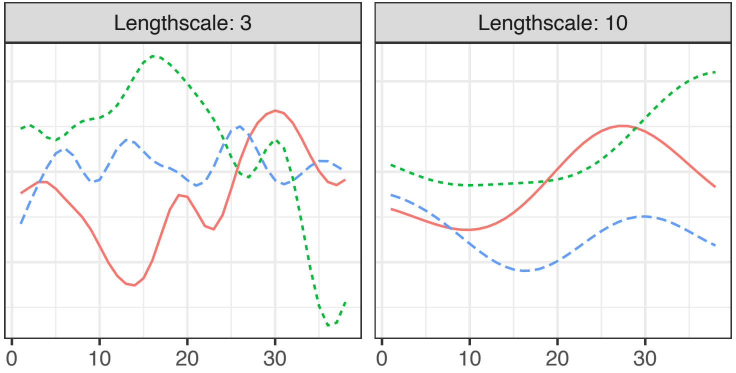
$$\Updownarrow$$

$$f(x_1, \ldots, x_N) \sim \mathcal{N}(m(x_1, \ldots, x_N), K), \text{ s.t. } K_{ij} = k(x_i, x_j)$$

Amplitude

$$k(x, x') = \boxed{\eta^2} \exp\left\{-\frac{(x - x')^2}{\boxed{2\rho^2}}\right\}$$

Lengthscale



When does conflation *actually* happen?

Four types of simulations:

1. Flexible nonlinear response

$$y_t = \boxed{f(x_t)} + \varepsilon_t$$

2. Dynamic coefficients, linear (or log-linear) model

$$y_t = \boxed{\beta(t)} x_t + \varepsilon_t$$

3. Nonlinear response, parametric hill function

$$y_t = \mathrm{Hill}(x_t) + \varepsilon_t$$

4. Dynamic coefficients inherited from common parent

$$y_t = \beta(t)x(t) + \varepsilon_t, \ (\beta(t), x(t)) \sim \mathrm{Pa}(t)$$

Gaussian processes

Two important levers:

- Smoothness
- Amplitude

When does conflation *actually* happen?

Four types of simulations:

1. Flexible nonlinear response
$$y_t = \boxed{f(x_t)} + \varepsilon_t$$
2. Dynamic coefficients, linear (or log-linear) model
$$y_t = \boxed{\beta(t)} x_t + \varepsilon_t$$
3. Nonlinear response, parametric hill function
4. Dynamic c

Gaussian processes
Two important levers:
- Smoothness
- Amplitude

Other manipulated features:
- Autoregressive coefficient in x
- Noise in x's autoregressive process
- Variance of the error term

Simulation Results

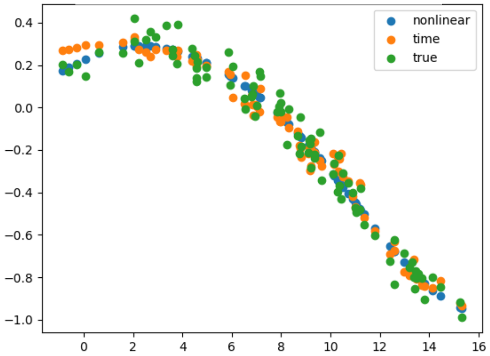
- For each simulation type, >300 settings, systematically varying the previously described factors, with 100 simulations per setting
- Fit both models (nonlinear and dynamic), measure conflation through validation RMSE

Examples

Nonlinear DGP

Smooth f, low noise

Data

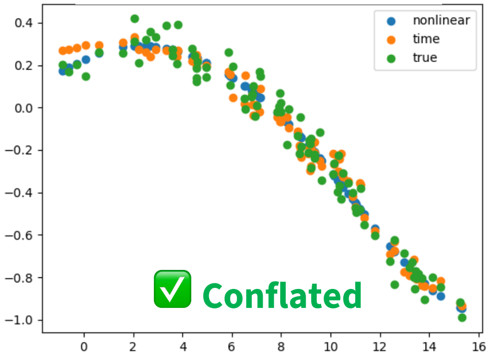


Examples

Nonlinear DGP

Smooth f, low noise

Data

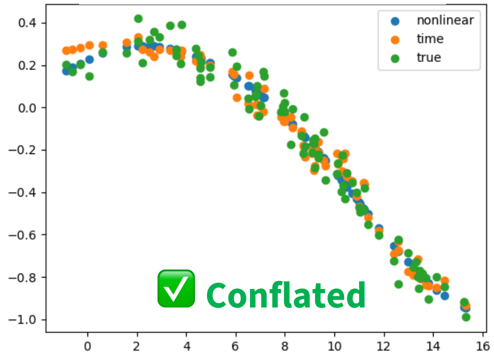


Examples

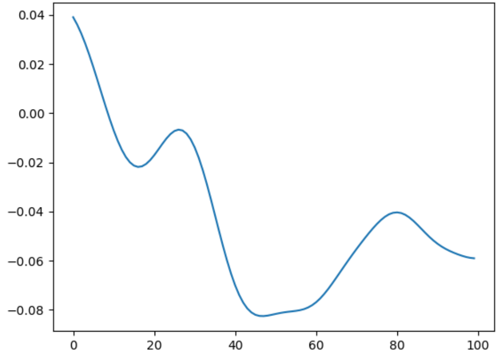
Nonlinear DGP

Smooth f, low noise

Data



Implied $\beta(t)$



Examples

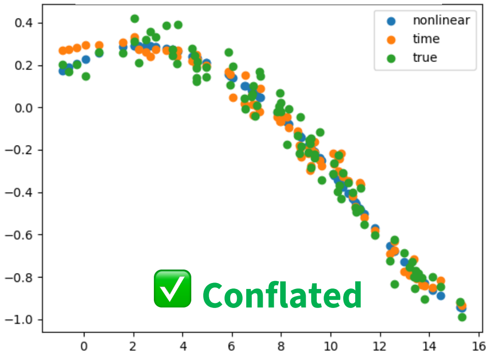
Nonlinear DGP

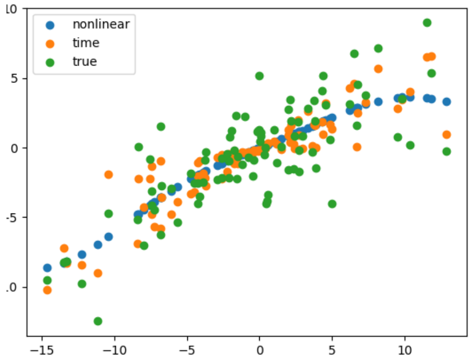
Smooth f, low noise

Dynamic DGP

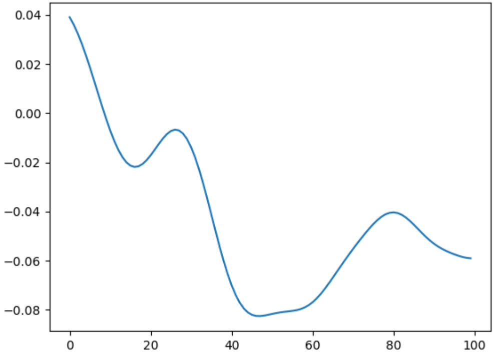
Smooth f, low noise

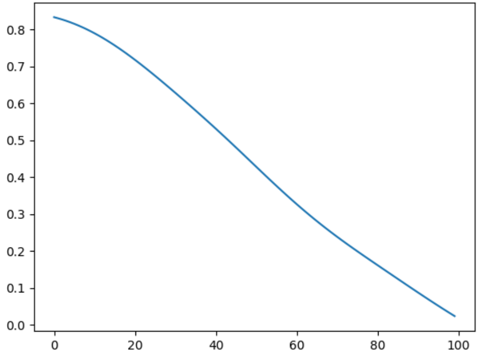
Data



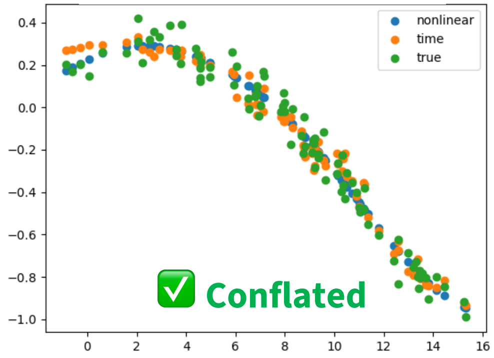
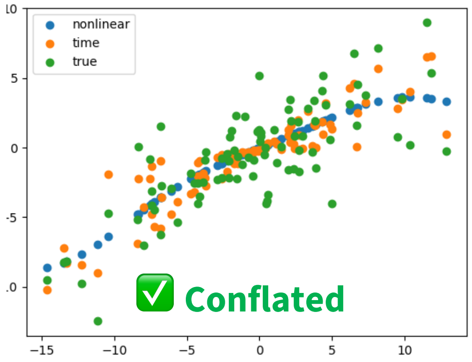
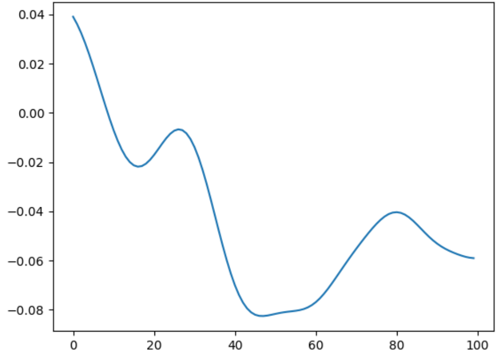
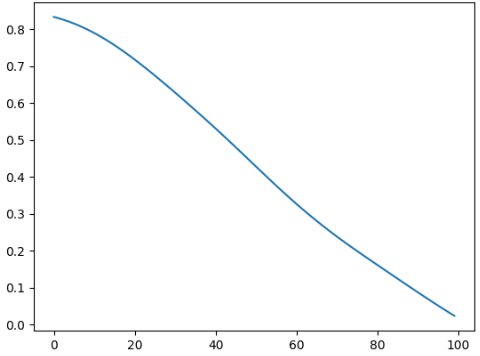


Implied
$\boldsymbol{\beta(t)}$

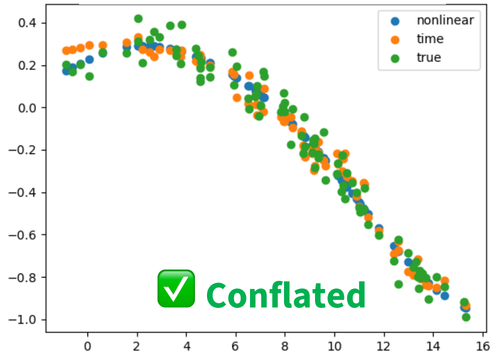
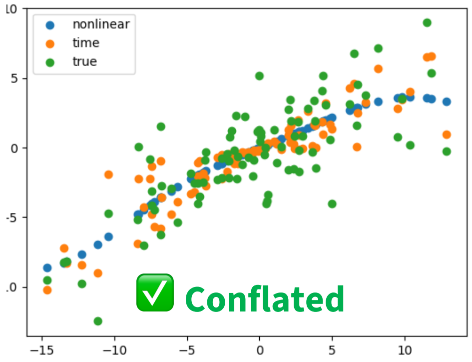
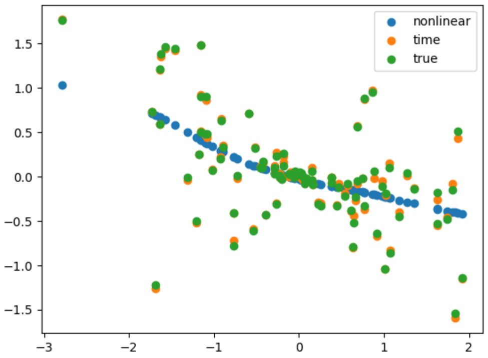
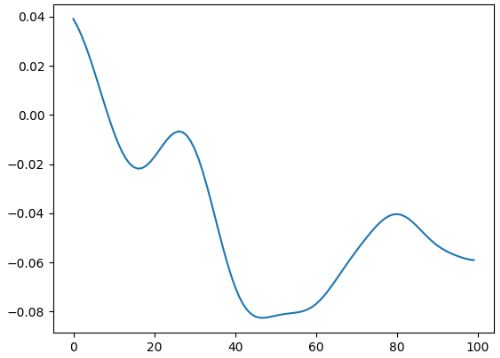
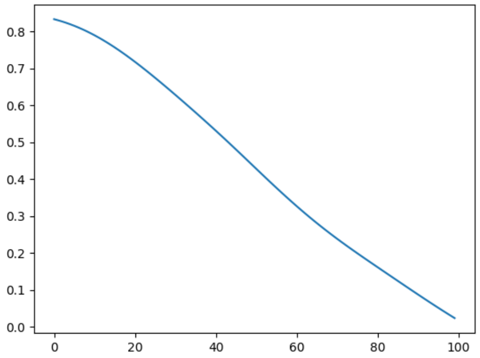
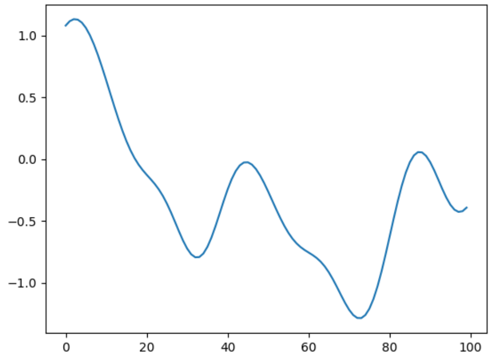




Examples

	Nonlinear DGP *Smooth f, low noise*	**Dynamic DGP** *Smooth f, low noise*
Data		
Implied $\boldsymbol{\beta(t)}$		

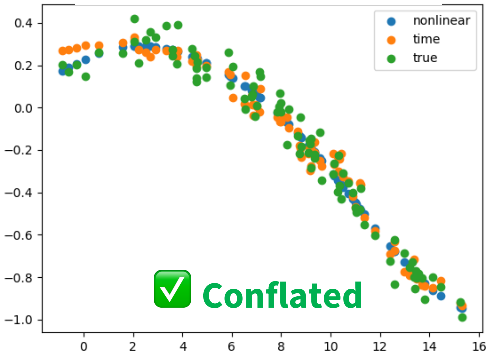
Examples

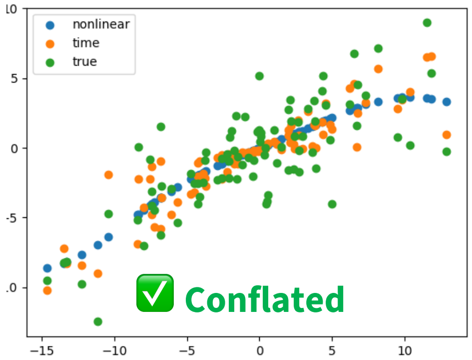
	Nonlinear DGP *Smooth f, low noise*	**Dynamic DGP** *Smooth f, low noise*	**Dynamic DGP** *Less smooth f, low noise*
Data			
Implied $\beta(t)$			

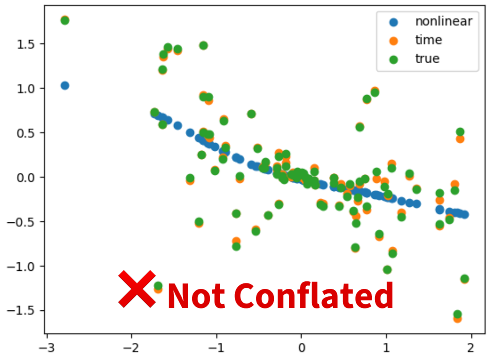
Examples

	Nonlinear DGP	**Dynamic DGP**	**Dynamic DGP**
	Smooth f, low noise	*Smooth f, low noise*	*Less smooth f, low noise*

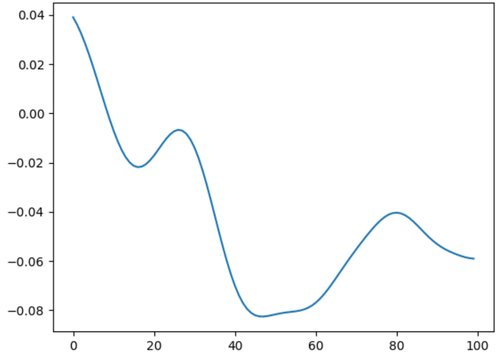
Data

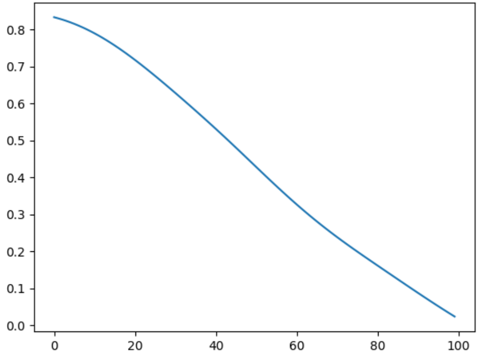


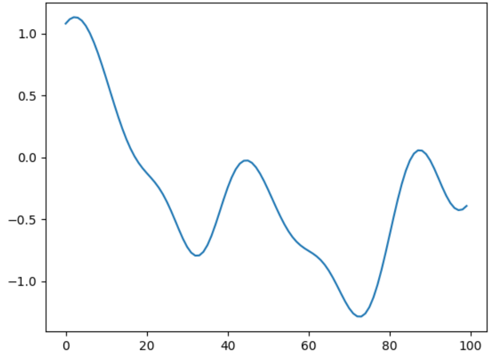




Implied $\boldsymbol{\beta(t)}$







Simulation

- Systematically vary the previously described factors, running 100 simulations per setting, and fit both models (nonlinear and dynamic)
- Any conflation = for at least one simulation of that setting, the wrong model had better validation RMSE
- Major conflation = conflation on over 25% of that setting's simulations

Simulation

- Systematically vary the previously described factors, running 100 simulations per setting, and fit both models (nonlinear and dynamic)
- Any conflation = for at least one simulation of that setting, the wrong model had better validation RMSE
- Major conflation = conflation on over 25% of that setting's simulations

	No Stock		High Stock	
True DGP	Any	Major	Any	Major
Nonlinear	81%	27%	85%	30%
Time-varying	91%	40%	99%	44%
Hill	83%	46%	92%	47%

Diving Deeper

DV: % Conflated Simulations

		Nonlinear DGP		Time-varying DGP	
Variable	Level	Coef	$P(>\|t\|)$	Coef	$P(>\|t\|)$
Amplitude, f:	Low (1)	-	-	-	
	Middle (2)	-0.24	0.65	0.23	0.62
	High (5)	-0.65	0.22	0.09	0.85
Smoothness, f:	Low (0.1)	-	-	-	
	Middle (0.5)	10.36	0.00	7.93	0.00
	High (1)	18.86	0.00	12.21	0.00
AR coef, x:	Low (0)	-	-	-	
	Middle (0.5)	-0.56	0.29	0.63	0.18
	High (1)	0.42	0.43	13.38	0.00
AR Variance, x:	Low (1)	-	-	-	
	Middle (5)	0.11	0.84	0.06	0.90
	High (10)	-0.47	0.37	0.36	0.44
Noise, y	Low (0.01)	-	-	-	
	Middle (0.1)	14.10	0.00	13.05	0.00
	High (0.2)	23.28	0.00	22.71	0.00
AdStock	Low (0)	-	-	-	
	Middle (0.3)	-0.14	0.79	0.83	0.08
	High (0.8)	1.59	0.00	2.49	0.00

Table 2: Simulation Results
DV = Percentage Conflation; Intercept omitted for clarity.

Diving Deeper

DV: % Conflated Simulations

Conflation more likely with…

- Noisier data (i.e., ϵ_t)

		Nonlinear DGP		Time-varying DGP	
Variable	Level	Coef	$P(>\|t\|)$	Coef	$P(>\|t\|)$
Amplitude, f:	Low (1)	-	-	-	
	Middle (2)	-0.24	0.65	0.23	0.62
	High (5)	-0.65	0.22	0.09	0.85
Smoothness, f:	Low (0.1)	-	-	-	
	Middle (0.5)	10.36	0.00	7.93	0.00
	High (1)	18.86	0.00	12.21	0.00
AR coef, x:	Low (0)	-	-	-	
	Middle (0.5)	-0.56	0.29	0.63	0.18
	High (1)	0.42	0.43	13.38	0.00
AR Variance, x:	Low (1)	-	-	-	
	Middle (5)	0.11	0.84	0.06	0.90
	High (10)	-0.47	0.37	0.36	0.44
Noise, y	Low (0.01)	-	-	-	
	Middle (0.1)	14.10	0.00	13.05	0.00
	High (0.2)	23.28	0.00	22.71	0.00
AdStock	Low (0)	-	-	-	
	Middle (0.3)	-0.14	0.79	0.83	0.08
	High (0.8)	1.59	0.00	2.49	0.00

Table 2: Simulation Results
DV = Percentage Conflation; Intercept omitted for clarity.

Diving Deeper

DV: % Conflated Simulations

Conflation more likely with...

- Noisier data (i.e., ϵ_t)
- Smoother functions

		Nonlinear DGP		Time-varying DGP	
Variable	Level	Coef	$P(>\|t\|)$	Coef	$P(>\|t\|)$
Amplitude, f:	Low (1)	-	-	-	
	Middle (2)	-0.24	0.65	0.23	0.62
	High (5)	-0.65	0.22	0.09	0.85
Smoothness, f:	Low (0.1)	-	-	-	
	Middle (0.5)	10.36	0.00	7.93	0.00
	High (1)	18.86	0.00	12.21	0.00
AR coef, x:	Low (0)	-	-	-	
	Middle (0.5)	-0.56	0.29	0.63	0.18
	High (1)	0.42	0.43	13.38	0.00
AR Variance, x:	Low (1)	-	-	-	
	Middle (5)	0.11	0.84	0.06	0.90
	High (10)	-0.47	0.37	0.36	0.44
Noise, y	Low (0.01)	-	-	-	
	Middle (0.1)	14.10	0.00	13.05	0.00
	High (0.2)	23.28	0.00	22.71	0.00
AdStock	Low (0)	-	-	-	
	Middle (0.3)	-0.14	0.79	0.83	0.08
	High (0.8)	1.59	0.00	2.49	0.00

Table 2: Simulation Results
DV = Percentage Conflation; Intercept omitted for clarity.

Diving Deeper

DV: % Conflated Simulations

Conflation more likely with…

- Noisier data (i.e., ϵ_t)
- Smoother functions
- High degree of autocorrelation

		Nonlinear DGP		Time-varying DGP	
Variable	Level	Coef	$P(>\lvert t \rvert)$	Coef	$P(>\lvert t \rvert)$
Amplitude, f:	Low (1)	-	-	-	
	Middle (2)	-0.24	0.65	0.23	0.62
	High (5)	-0.65	0.22	0.09	0.85
Smoothness, f:	Low (0.1)	-	-	-	
	Middle (0.5)	10.36	0.00	7.93	0.00
	High (1)	18.86	0.00	12.21	0.00
AR coef, x:	Low (0)	-	-	-	
	Middle (0.5)	-0.56	0.29	0.63	0.18
	High (1)	0.42	0.43	13.38	0.00
AR Variance, x:	Low (1)	-	-	-	
	Middle (5)	0.11	0.84	0.06	0.90
	High (10)	-0.47	0.37	0.36	0.44
Noise, y	Low (0.01)	-	-	-	
	Middle (0.1)	14.10	0.00	13.05	0.00
	High (0.2)	23.28	0.00	22.71	0.00
AdStock	Low (0)	-	-	-	
	Middle (0.3)	-0.14	0.79	0.83	0.08
	High (0.8)	1.59	0.00	2.49	0.00

Table 2: Simulation Results
DV = Percentage Conflation; Intercept omitted for clarity.

Diving Deeper

DV: % Conflated Simulations

Conflation more likely with...

- Noisier data (i.e., ϵ_t)
- Smoother functions
- High degree of autocorrelation

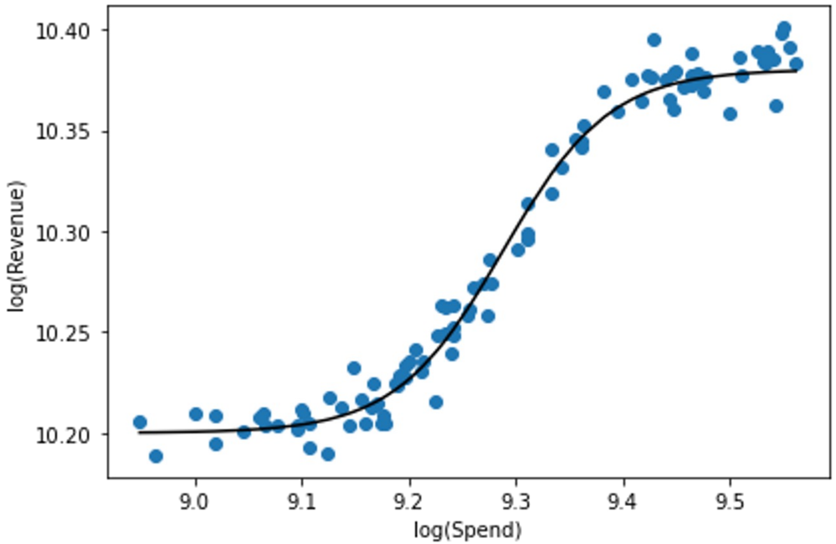
Decisions are often autocorrelated!

		Nonlinear DGP		Time-varying DGP	
Variable	Level	Coef	$P(>\|t\|)$	Coef	$P(>\|t\|)$
Amplitude, f:	Low (1)	-	-	-	
	Middle (2)	-0.24	0.65	0.23	0.62
	High (5)	-0.65	0.22	0.09	0.85
Smoothness, f:	Low (0.1)	-	-	-	
	Middle (0.5)	10.36	0.00	7.93	0.00
	High (1)	18.86	0.00	12.21	0.00
AR coef, x:	Low (0)	-	-	-	
	Middle (0.5)	-0.56	0.29	0.63	0.18
	High (1)	0.42	0.43	13.38	0.00
AR Variance, x:	Low (1)	-	-	-	
	Middle (5)	0.11	0.84	0.06	0.90
	High (10)	-0.47	0.37	0.36	0.44
Noise, y	Low (0.01)	-	-	-	
	Middle (0.1)	14.10	0.00	13.05	0.00
	High (0.2)	23.28	0.00	22.71	0.00
AdStock	Low (0)	-	-	-	
	Middle (0.3)	-0.14	0.79	0.83	0.08
	High (0.8)	1.59	0.00	2.49	0.00

Table 2: Simulation Results
DV = Percentage Conflation; Intercept omitted for clarity.

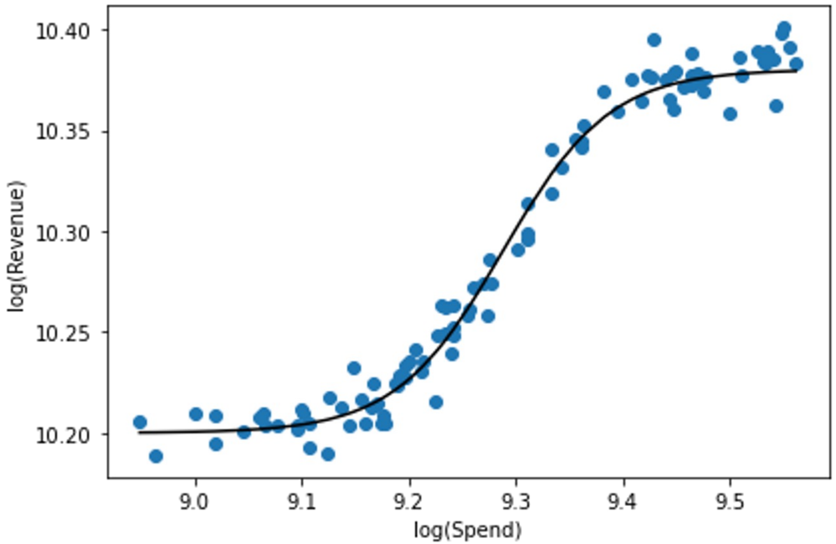
Implications

One last simulation…

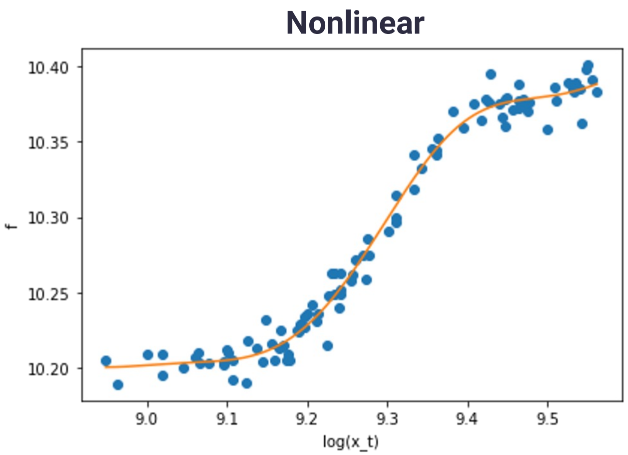


(or from roughly $8,000 to $14,000)

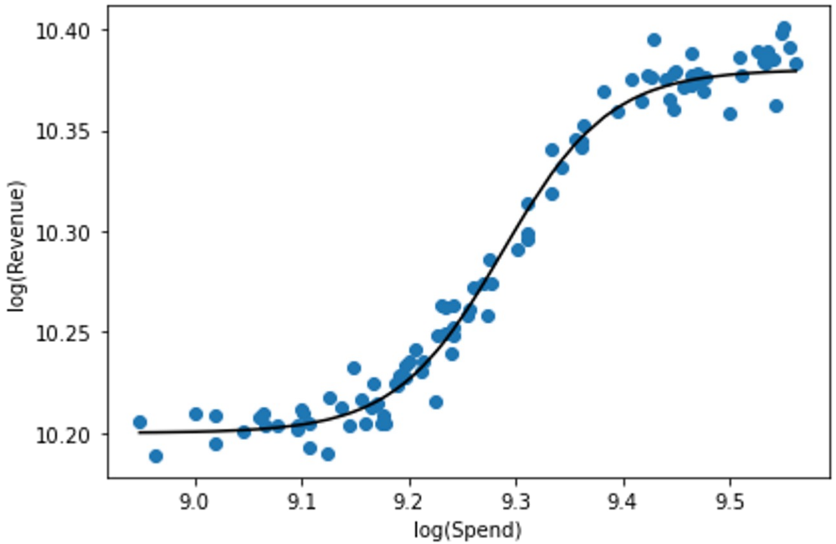
One last simulation…



(or from roughly $8,000 to $14,000)

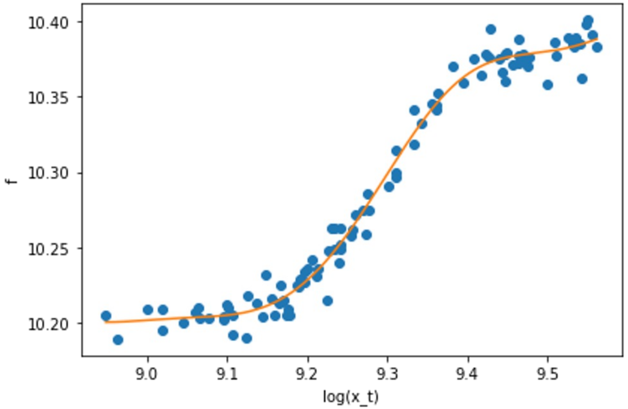


One last simulation…

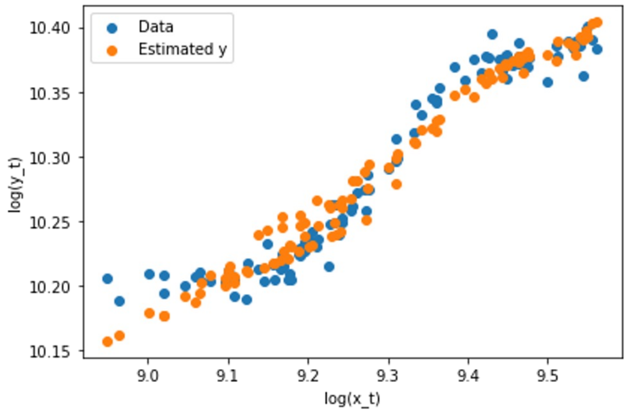


(or from roughly $8,000 to $14,000)

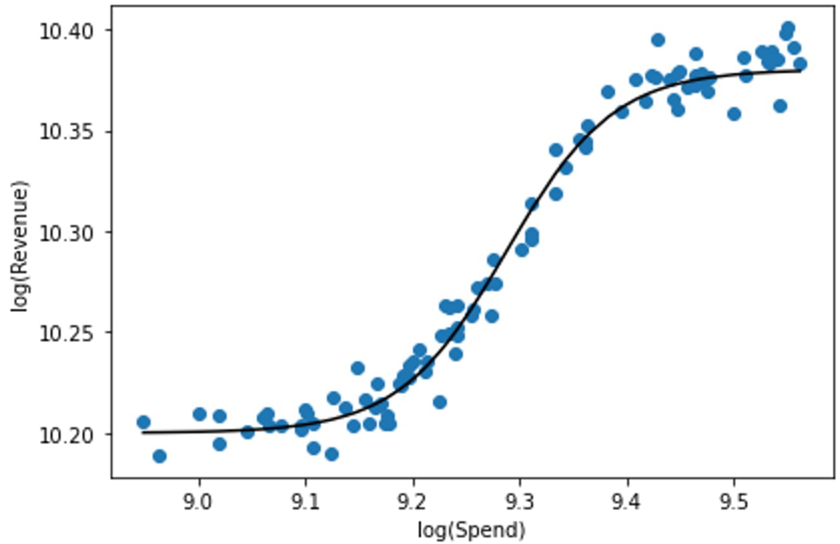
Nonlinear



Dynamic

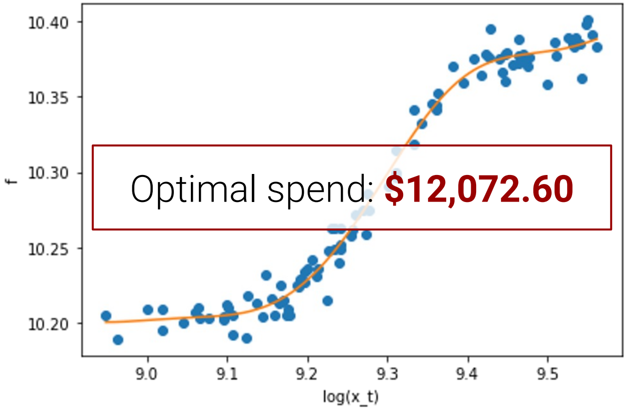


One last simulation…



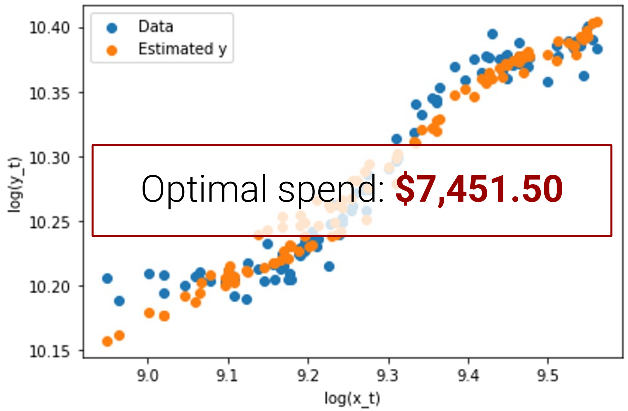
(or from roughly $8,000 to $14,000)

Nonlinear



Optimal spend: **$12,072.60**

Dynamic



Optimal spend: **$7,451.50**

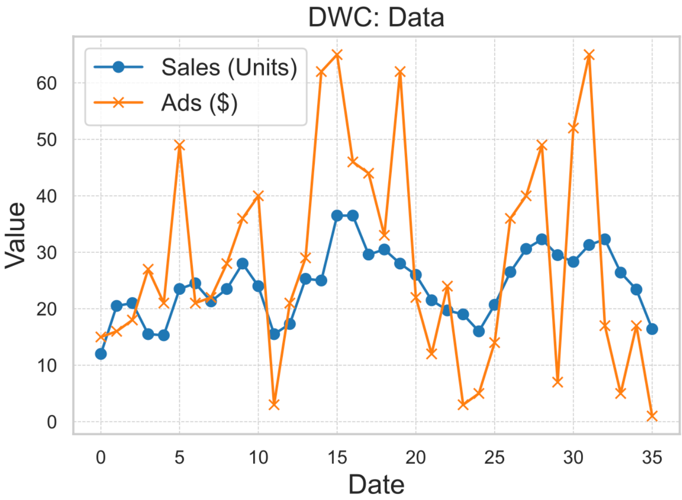
Real Data: Classic and Modern

- Classic Application 1: Dietary Weight Control (DWC)
  (Bass and Clark, 1972)

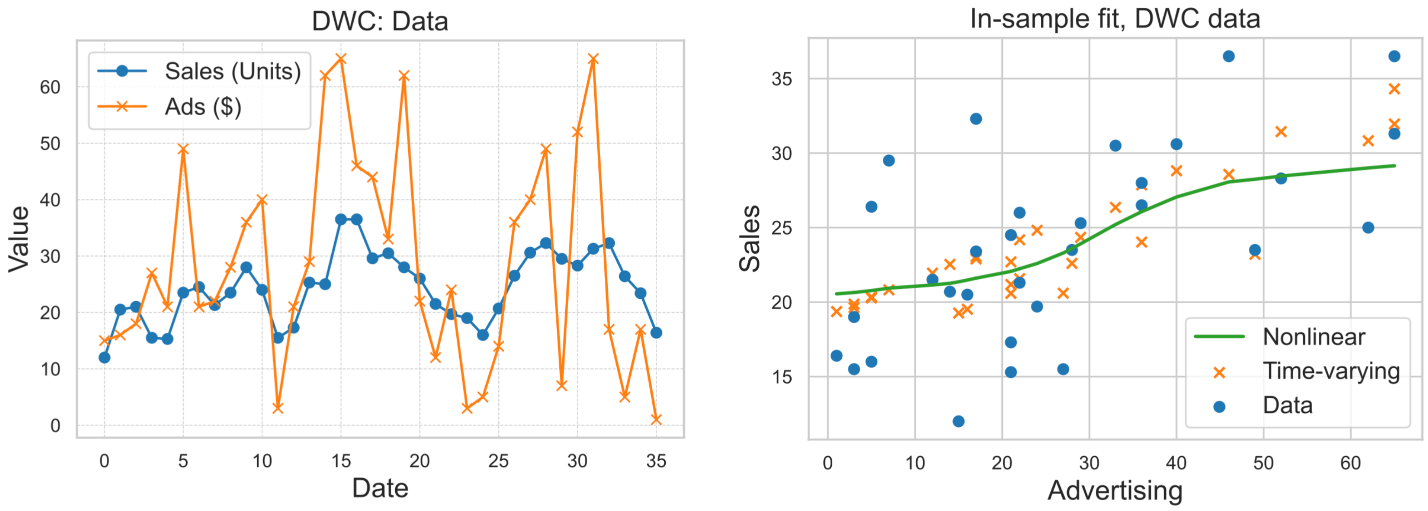
- Classic Application 2: Lydia Pinkham (LP)
  (Palda, 1964)

- Modern applications: MMM data from Nielsen

DWC

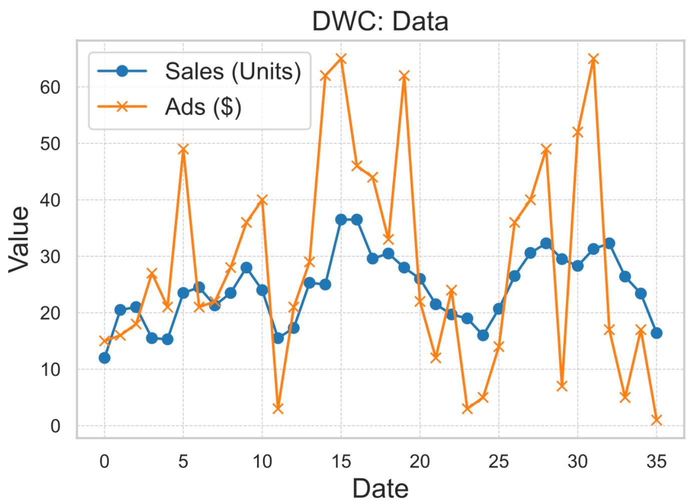

DWC: Data
Sales (Units)
Ads ($)
Value
Date
0
10
20
30
40
50
60
0
5
10
15
20
25
30
35

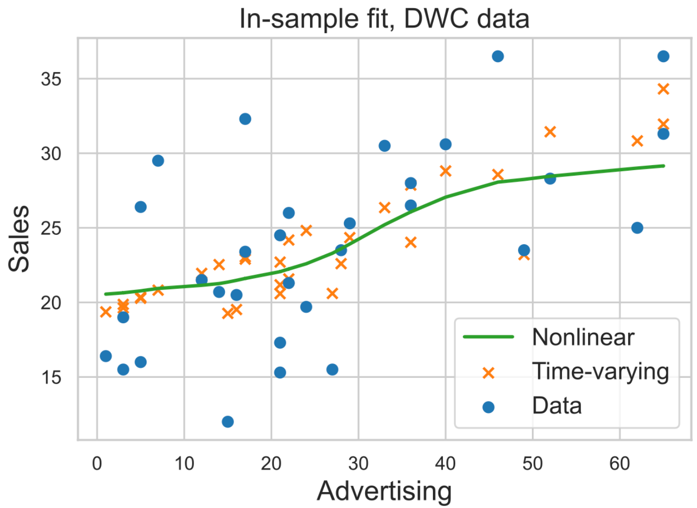
DWC



DWC: Data
Sales (Units)
Ads ($)
Value
Date
0
10
20
30
40
50
60
0
5
10
15
20
25
30
35
In-sample fit, DWC data
Nonlinear
Time-varying
Data
Sales
Advertising
15
20
25
30
35
0
10
20
30
40
50
60

DWC





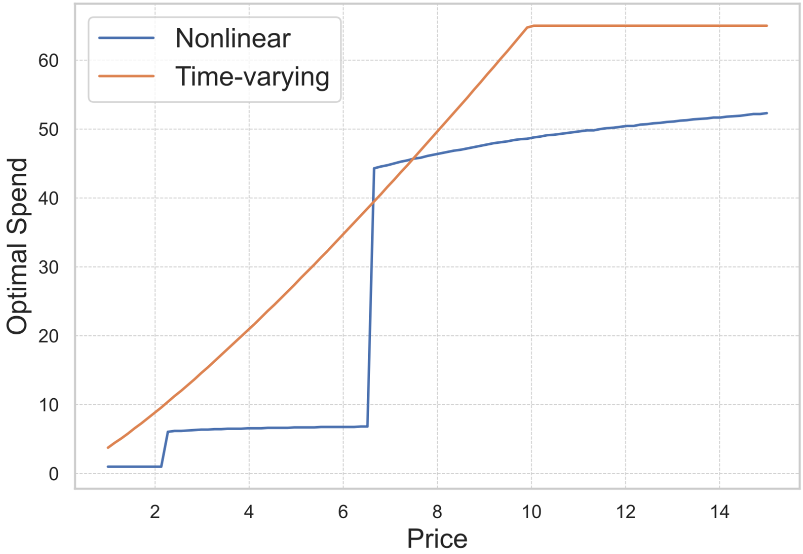
	Nonlinear	Time-varying	Hill	Log Time-varying
Dietary Weight Control	3.23 [2.22, 5.58]	2.97 [1.72, 5.09]	3.10 [2.59, 4.44]	2.55 [1.50, 4.13]

Spend Optimization

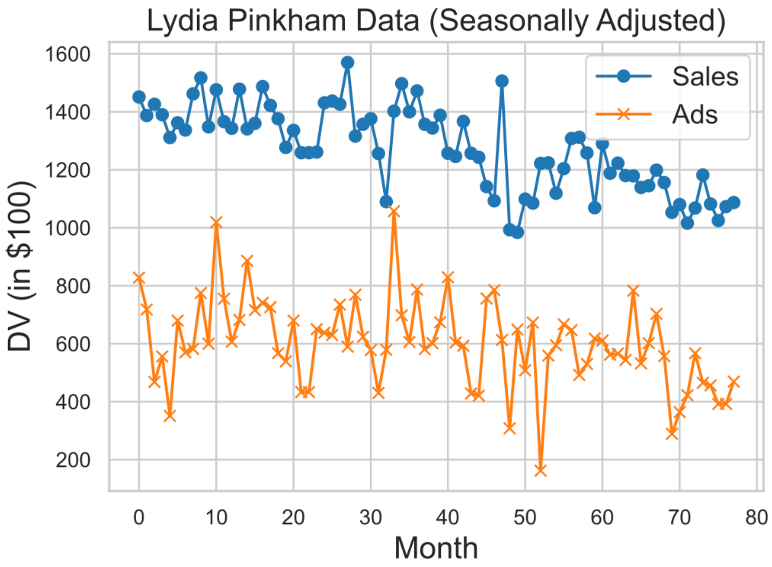
- DWC data: sales volume, but no prices
- Limitation = opportunity: optimal ad spend across a range of prices

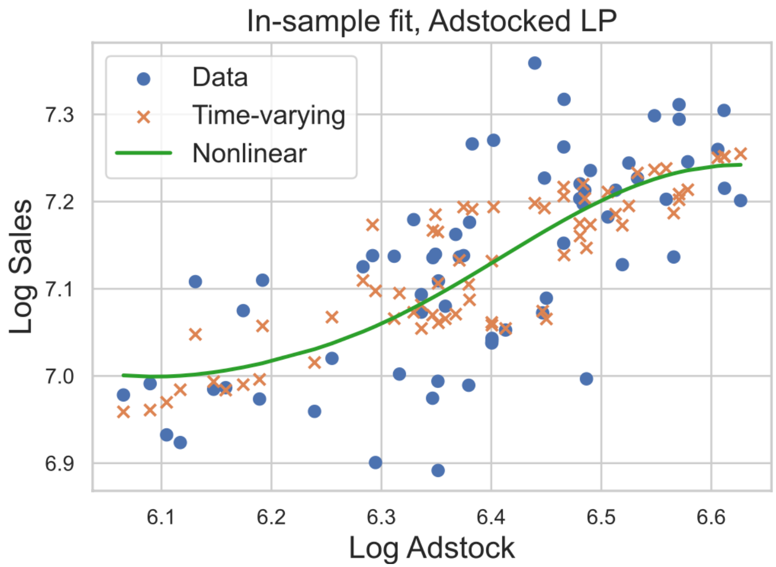
Spend Optimization

- DWC data: sales volume, but no prices
- Limitation = opportunity: optimal ad spend across a range of prices



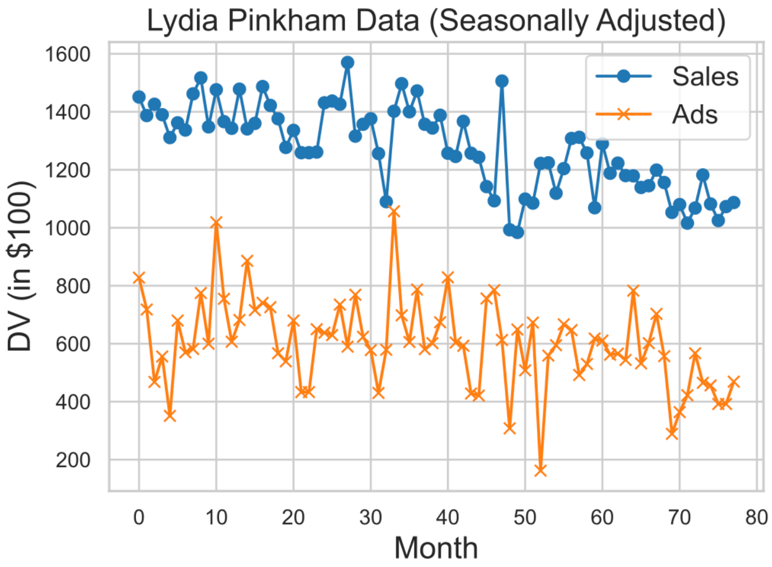
Lydia Pinkham



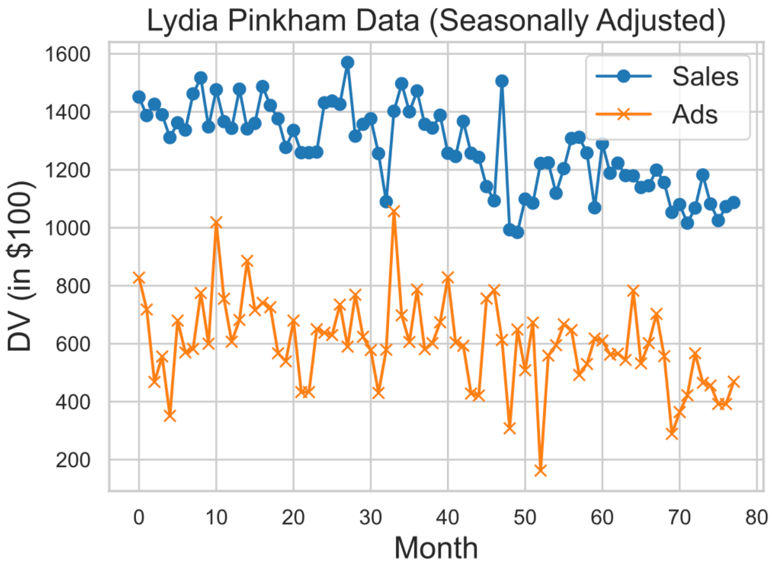


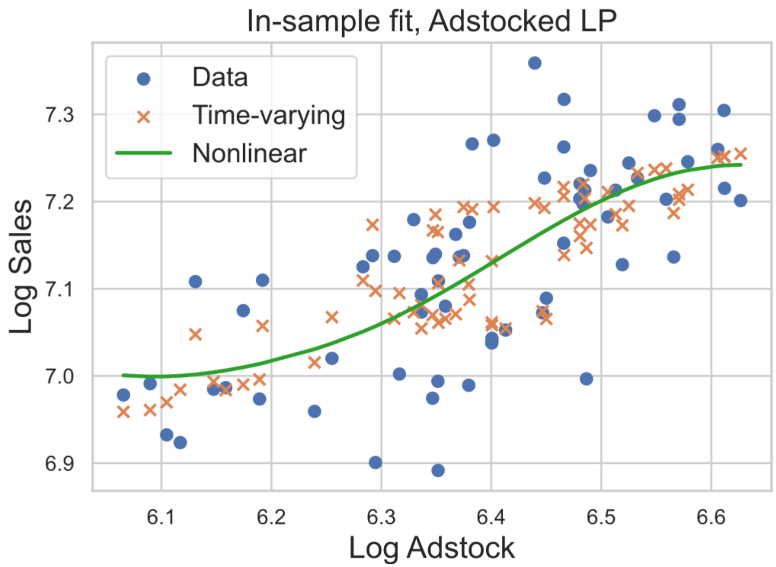
	Nonlinear	Time-varying	Hill	Log Time-varying
Lydia Pinkham	89.37	83.21	100.50	88.92
	[79.41, 104.88]	[75.44, 95.24]	[91.68, 114.97]	[83.23, 98.00]

Lydia Pinkham

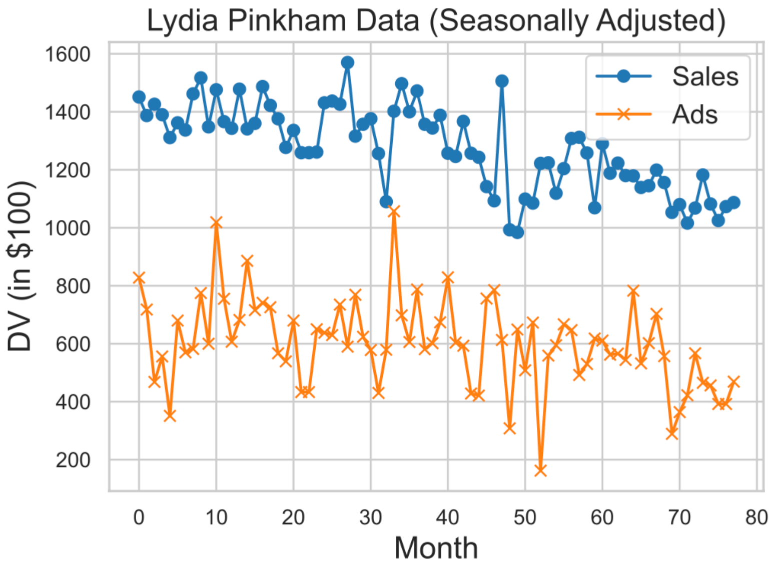


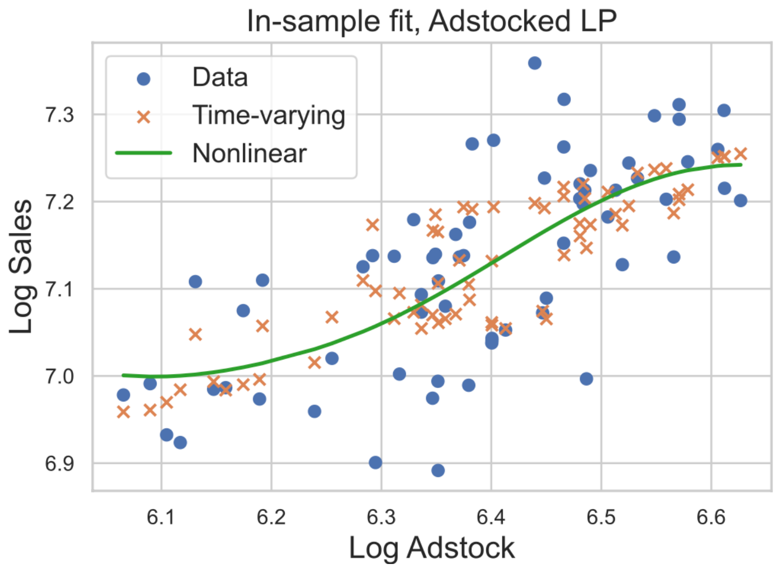
Lydia Pinkham





Lydia Pinkham





	Nonlinear	Time-varying	Hill	Log Time-varying
Lydia Pinkham	89.37	83.21	100.50	88.92
	[79.41, 104.88]	[75.44, 95.24]	[91.68, 114.97]	[83.23, 98.00]

Spend Optimization with Carryover

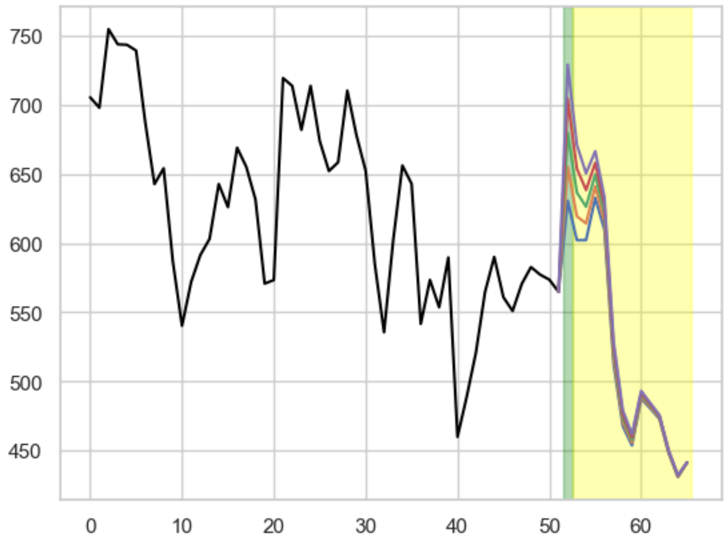
- Trickier: spending today affects AdStock for many future periods

Spend Optimization with Carryover

- Trickier: spending today affects AdStock for many future periods
- Follow Google's practice: for L carryover periods, determine optimal spend in a test window, predict revenue L periods out

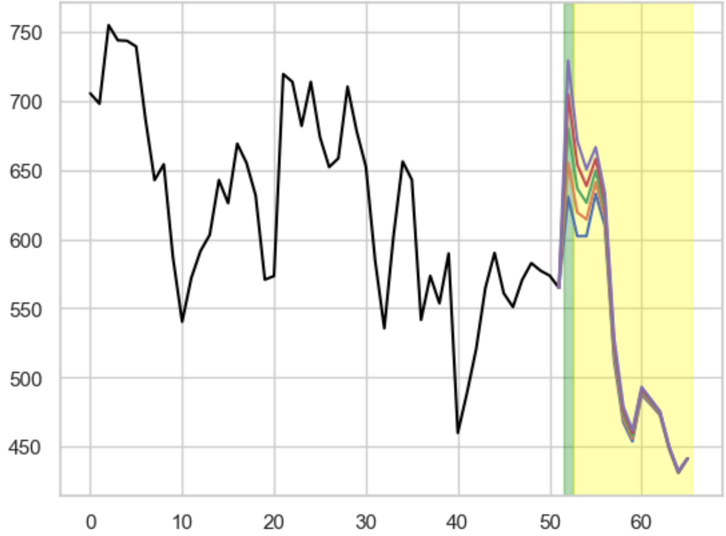
Spend Optimization with Carryover

- Trickier: spending today affects AdStock for many future periods
- Follow Google's practice: for *L* carryover periods, determine optimal spend in a test window, predict revenue *L* periods out



Spend Optimization with Carryover

- Trickier: spending today affects AdStock for many future periods
- Follow Google's practice: for *L* carryover periods, determine optimal spend in a test window, predict revenue *L* periods out
- For LP, optimal ad spend:
  - Nonlinear model: **$90,127**
  - Time-varying model: **$78,333**



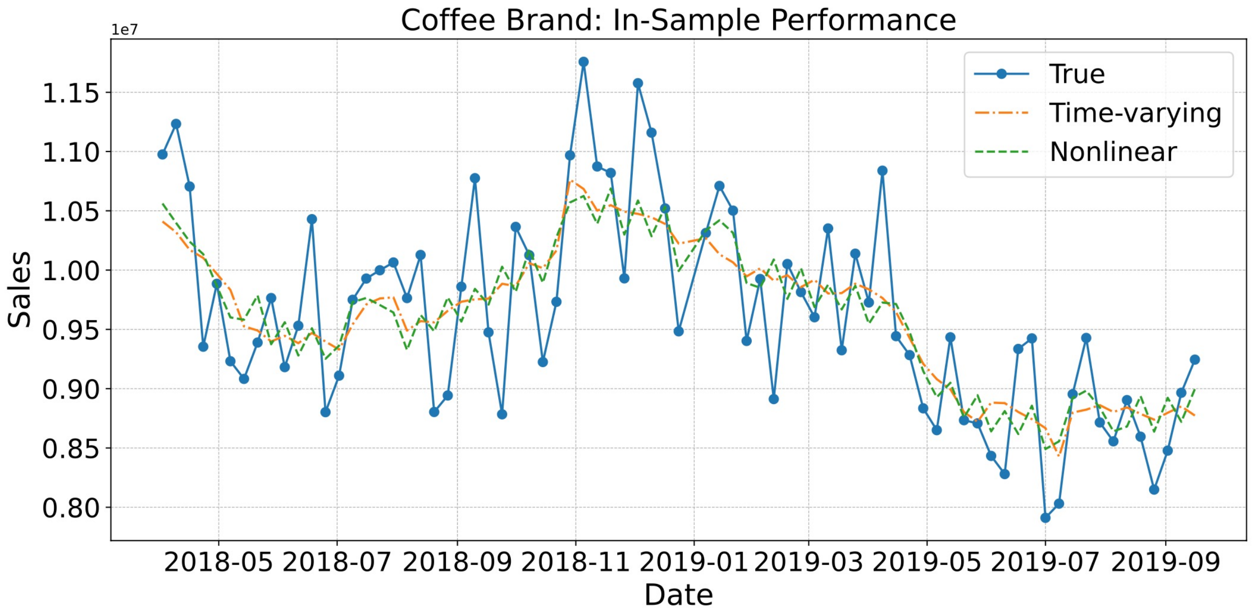
Modern MMM Data

- Again, clear conflation...

	Chocolate	Pet Food	Coffee	Beer
Nonlinear (GP)	368,766	48,438	722,076	443,817
	[207,634, 601,163]	[27,920, 83,249]	[509,527, 919,820]	[236,730, 911,732]
Time-varying (GP)	387,623	41,825	733,269	423,085
	[207,350, 671,246]	[24,631, 65,232]	[490,785, 994,303]	[207,914, 802,740]

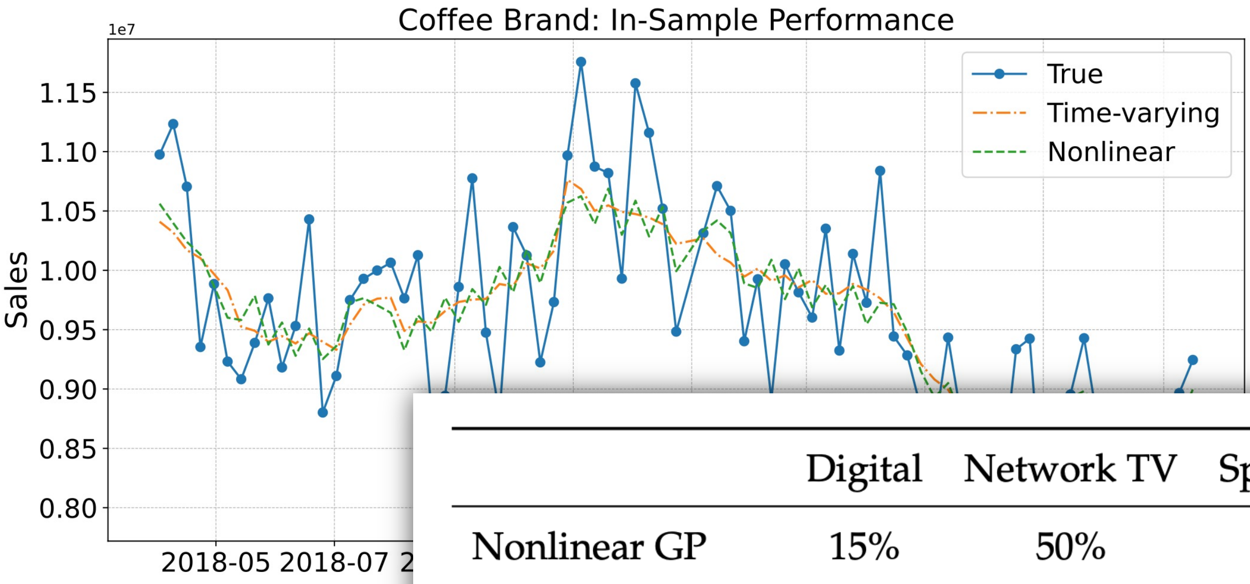
Modern MMM Data

- Case study: Coffee



Modern MMM Data

- Case study: Coffee



	Digital	Network TV	Spot TV	Conflation Cost
Nonlinear GP	15%	50%	35%	$227,000
Time-varying GP	15%	85%	0%	$61,000

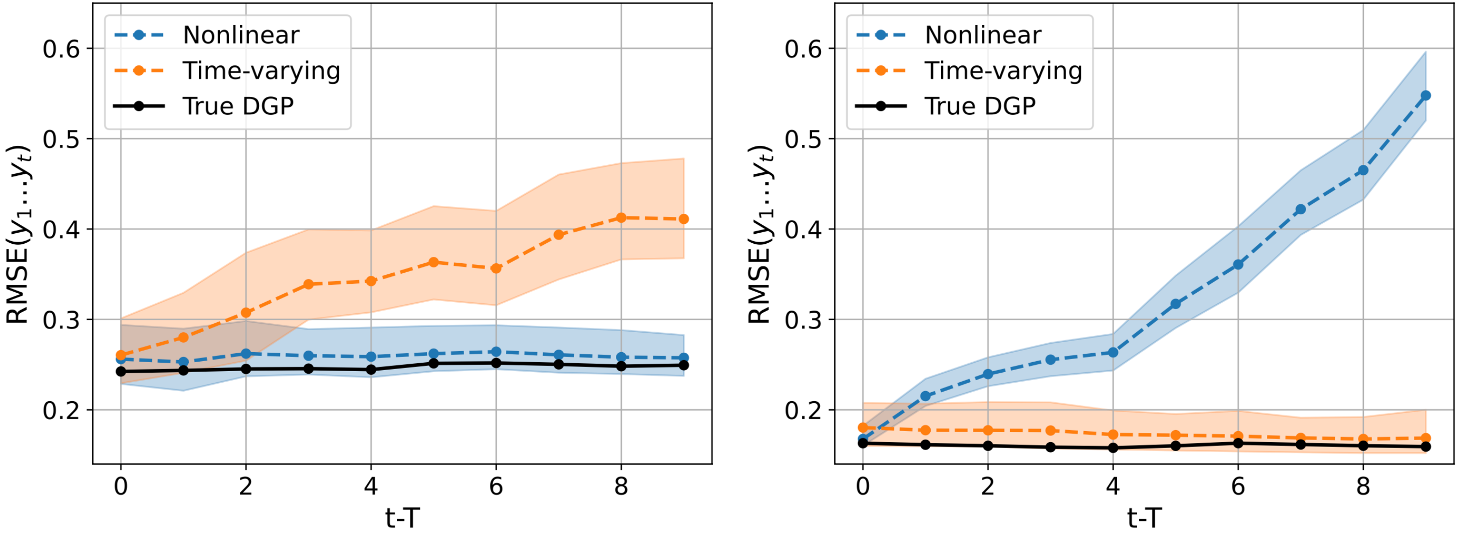
So is MMM doomed?

Test 1: Maximal Separation

- A type of “bump up” incrementality test
- Intuition: set spending next period to maximize the difference in predicted values across the two models

Test 1: Maximal Separation

- A type of “bump up” incrementality test
- Intuition: set spending next period to maximize the difference in predicted values across the two models

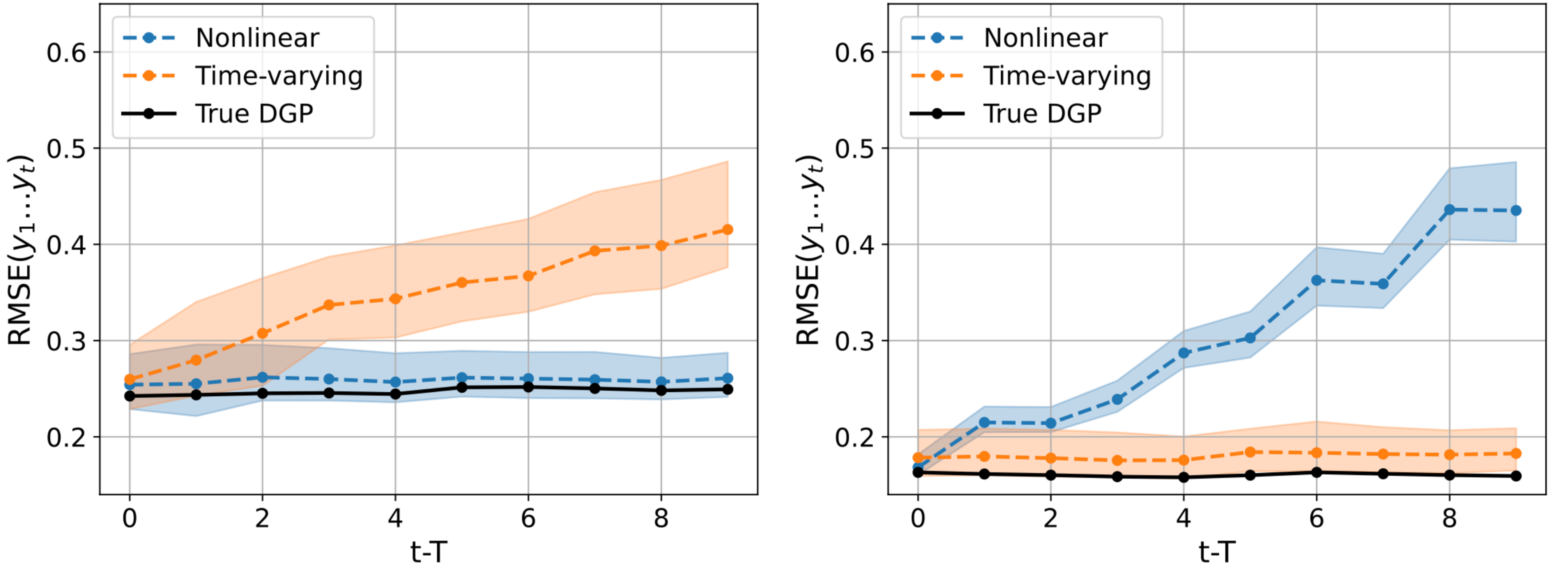


Test 2: See-saw

- Easier to implement
- Intuition: high-low spending over adjacent time periods

Test 2: See-saw

- Easier to implement
- Intuition: high-low spending over adjacent time periods



Conclusions

Current MMM Practice Might Be Flawed!

- We show that, under many common spending patterns, **time-varying and nonlinear effects cannot be disentangled**, *despite having different implications*
- This problem is potentially **very widespread**: increasing complexity in models, widespread practice of “model refreshes” to capture changing markets
- Our work both introduces a **framework for estimating** these types of models, and provides **solutions for understanding and preventing** conflation

Thanks!

Feedback or questions:

ryandew@wharton.upenn.edu

Working paper:

www.rtdew.com

You

generate an image of a marketing mix model paradise with some very happy computers and a cameo from the reverend thomas bayes