

Your MMM Is Broken: Identification of Nonlinear and Dynamic Effects in Marketing Mix Models

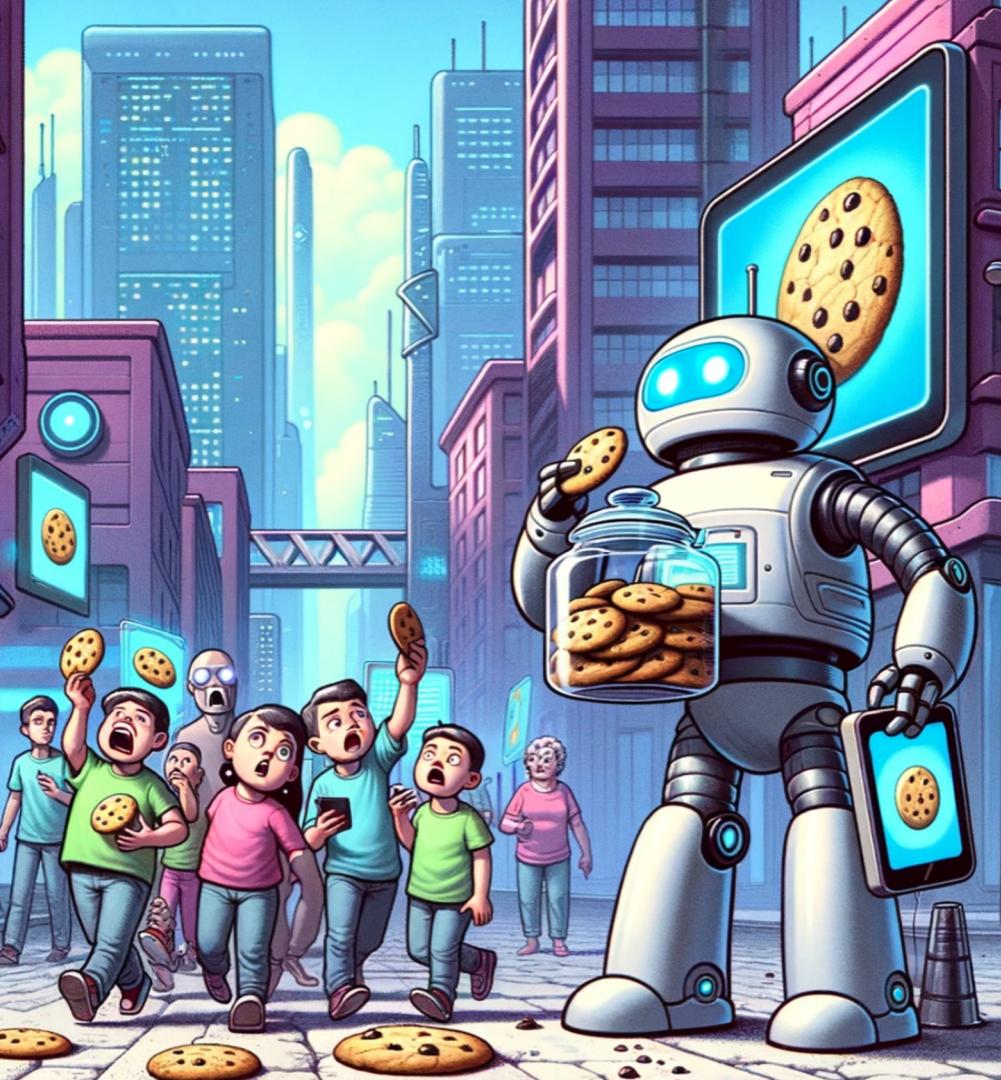
Ryan Dew

*The Wharton School
University of Pennsylvania*

Joint work with Anya Schetkina and Nicolas Padilla

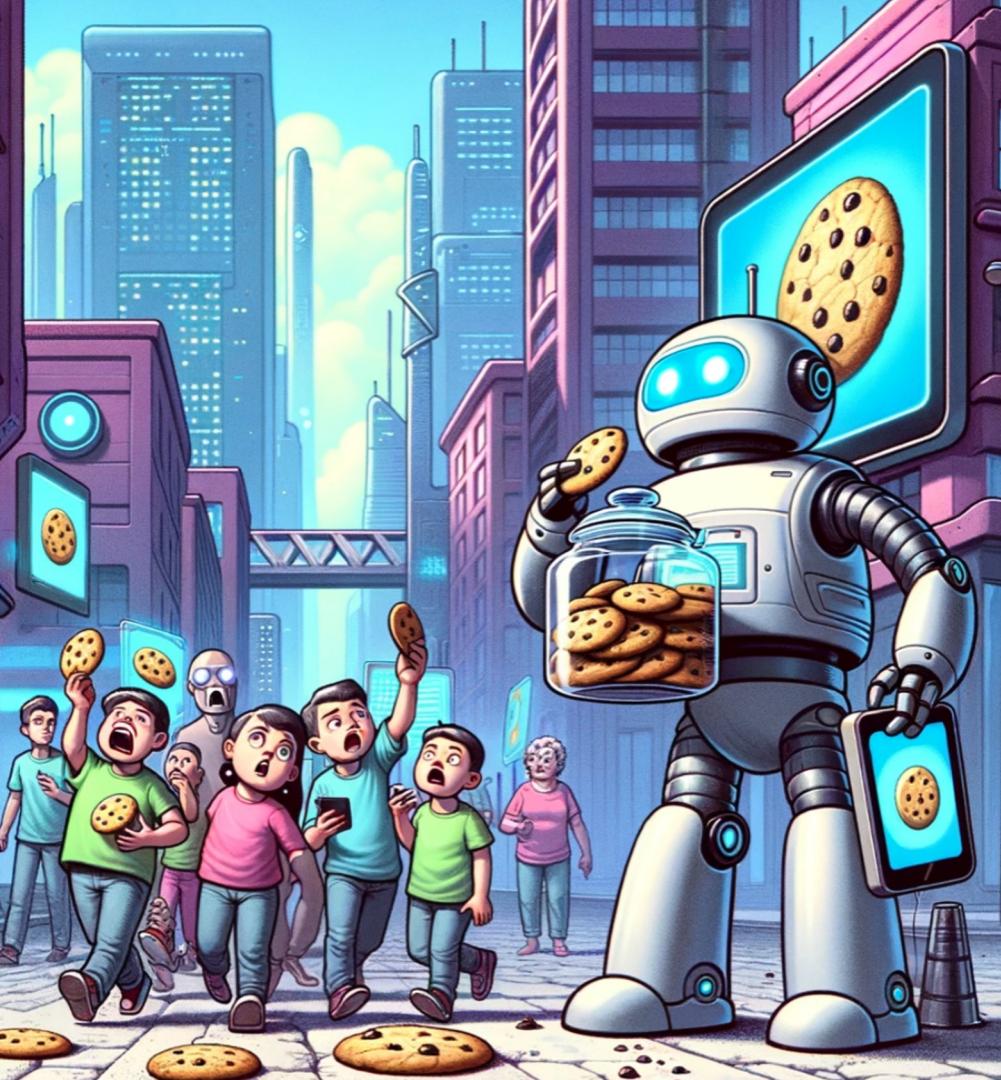
The Cookie-free Future

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



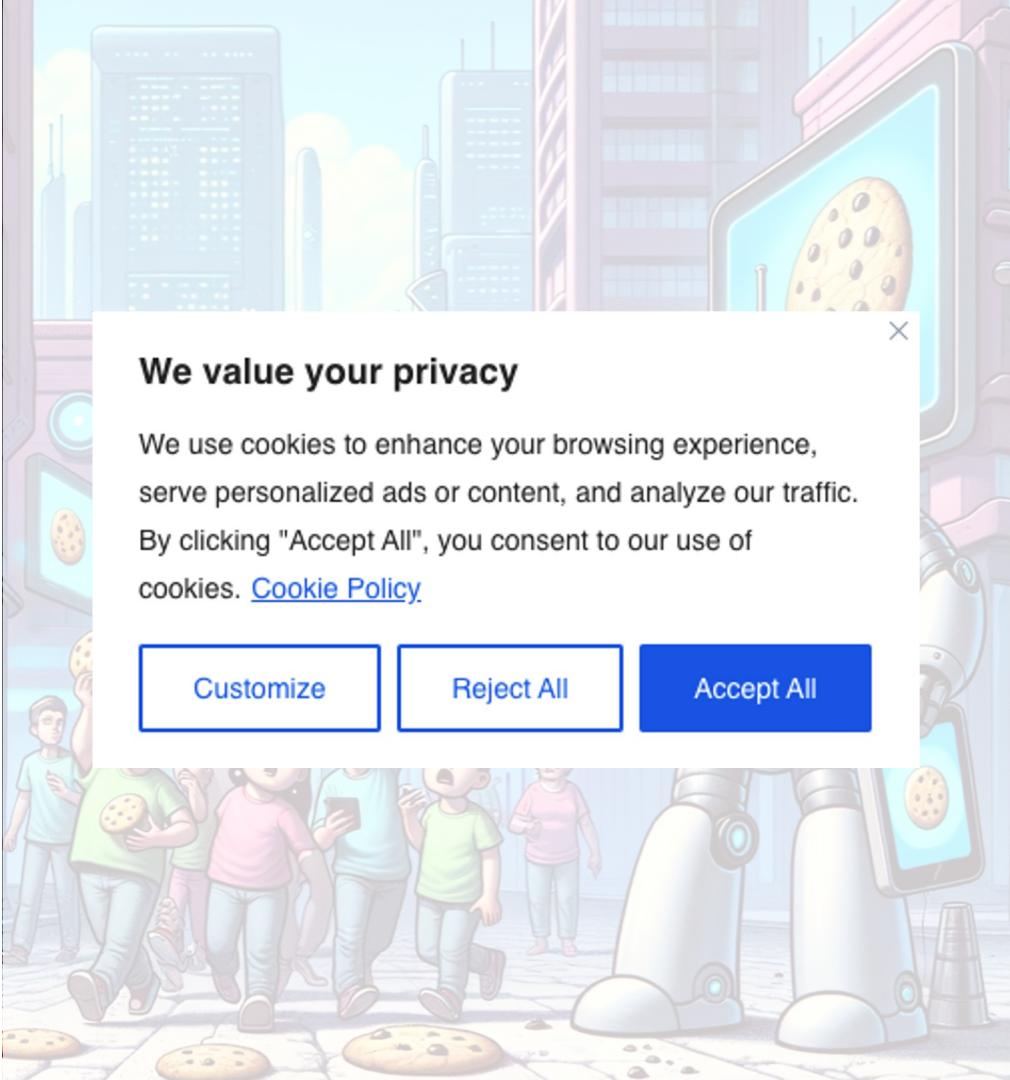
The Cookie-free Future

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



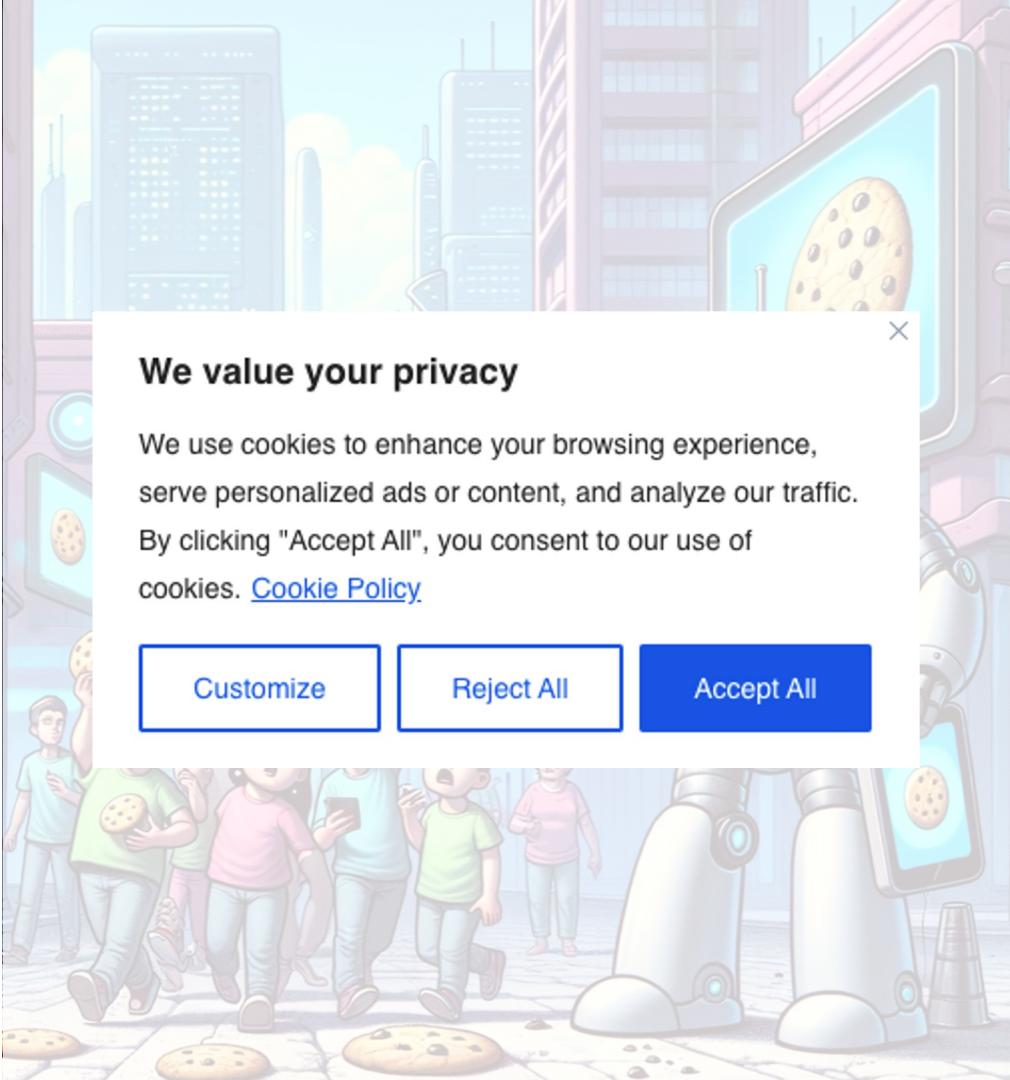
The Cookie-free Future

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

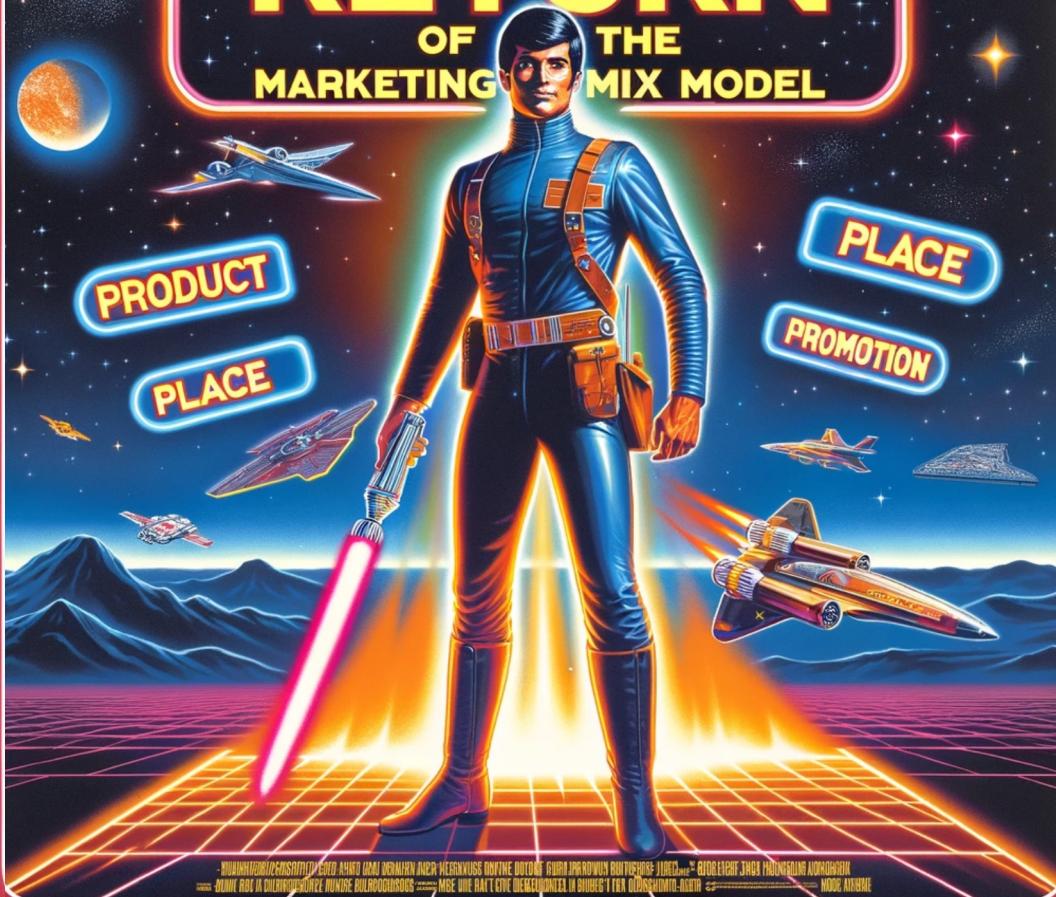


The Cookie-free Future

- 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



- HUANHEDZERNSDIT - GOL AHIO UMI DERNEN AER - HEGRKUSS DUVNE BOTON - GURJAPROWUN RUTTUSHEF - JUREL - BODGELEHT THOI YOUNGUNG NORGAGH
- BEUNE RUE IS DIERENGEDELE HUNDE BULBOORHOS - MEE UWE RAT' ETE DEMEDEUNES IN BUBB'S FRA UDERSHOMME - KUTU - MODE ARABIE

Statistics > Applications

[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.

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

Bayesian Time Varying Coefficients

Edwin Ng, Zhishi Wang, Athena Dai

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 variables 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.



Orbit

[Submitted on 7 Ju



Robyn

Documentation

Getting Started

Case Studies

Resources

Bayesian

Edwin Ng, Zhi

Both Bayesian and frequentist approaches have their own merits and generalizability. Robyn is built on a Bayesian framework, which provides a more principled model, equipped with prior distributions that can incorporate domain knowledge and coefficients are interpreted as posterior distributions. Robyn also provides methods to approximate the posterior distributions of the parameters, which are useful for traditional Marketing Mix Modeling (MMM) applications. Robyn is designed to be user friendly in terms of both the interface and the underlying statistical models.

Robyn

Robyn is an experimental, AI/ML-powered and open sourced Marketing Mix Modeling (MMM) package from Meta Marketing Science.

Getting Started

Statistics > Applications

[Submitted on 7 Ju

Bayesian

Edwin Ng, Zhi

Both Bayesian and frequentist methods have their own strengths and weaknesses. Robyn is a generalizable, open-source Python package that provides a unified interface for fitting Bayesian models, equipped with a variety of priors and samplers. It also provides a way to approximate coefficients and uncertainty intervals for parameters in a model. Robyn is designed to approximate coefficients and uncertainty intervals for parameters in a model. Robyn is designed to be used in both traditional MCMC and variational inference frameworks. Robyn is designed to be used in both traditional MCMC and variational inference frameworks. Robyn is designed to be used in both traditional MCMC and variational inference frameworks.

Robyn

Robyn is an experimental, AI-powered, open-source Python package for Marketing
Modeling (MMM) package

Getting Started



Case Studies

Resources

PyMC-Marketing

Statistics > Applications

[Submitted on 7 Ju

Bayesian

Edwin Ng, Zhi

Both Bayesian and frequentist approaches have their own strengths and weaknesses. Bayesian methods are more generalizable and can incorporate prior knowledge, while frequentist methods are more robust to model misspecification. Bayesian methods also provide a natural way to handle uncertainty and make predictions based on the posterior distribution of parameters. In contrast, frequentist methods often rely on asymptotic approximations and can be less transparent about the assumptions underlying their results.



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

Marketing

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)

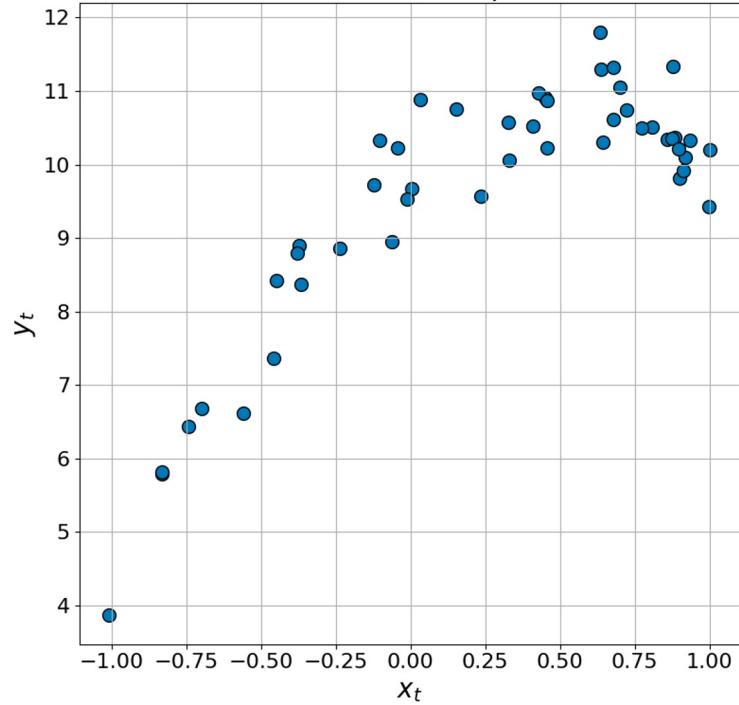
$$\begin{aligned}y_{g,t} = & \mu_t + \tau_g + \sum_{c=1}^C \gamma_{g,c} z_{g,t,c} \\& + \sum_{m=1}^M \beta_{g,m} HillAdstock \left(\left\{ x_{g,t-s,m} \right\}_{s=0}^L ; \alpha_m, ec_m, slope_m \right) \\& + \sum_{n=1}^N \beta_{g,n}^{(rf)} Adstock \left(\left\{ r_{g,t-s,n} \cdot Hill \left(f_{g,t-s,n};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

Uber's Orbit

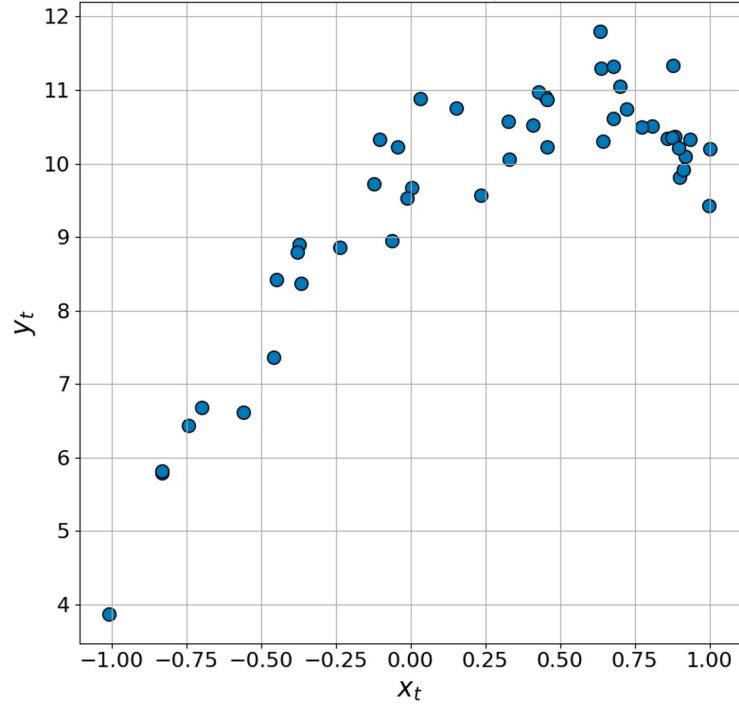
Ng et al., (2021)

$$\ln(\hat{y}_t) = l_t + s_t + \sum_{p=1}^P \ln(x_{t,p}) \beta_{t,p}$$
$$\beta_{t,p} = \sum_j w_j(t) \cdot b_{j,p},$$
$$w_j(t) = k(t, t_j) / \sum_{i=1}^J k(t, t_i),$$

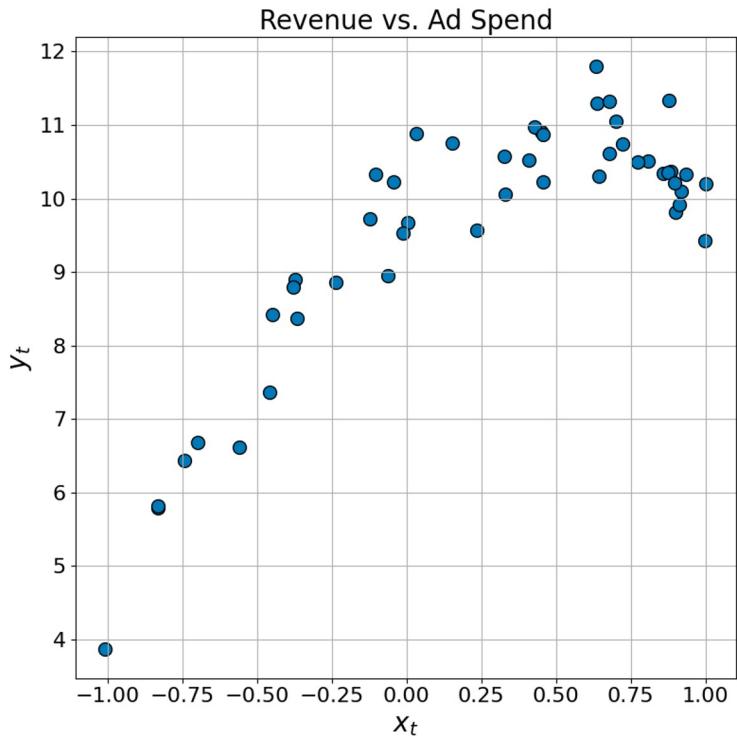
Revenue vs. Ad Spend



Revenue vs. Ad Spend



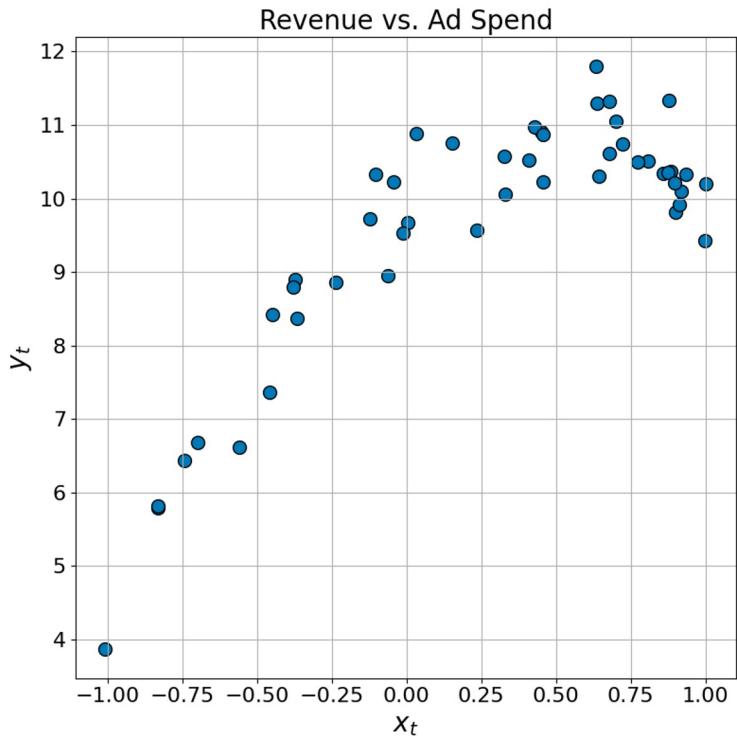
**Surely this is nonlinear,
right...?**



True data generating process (DGP):

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

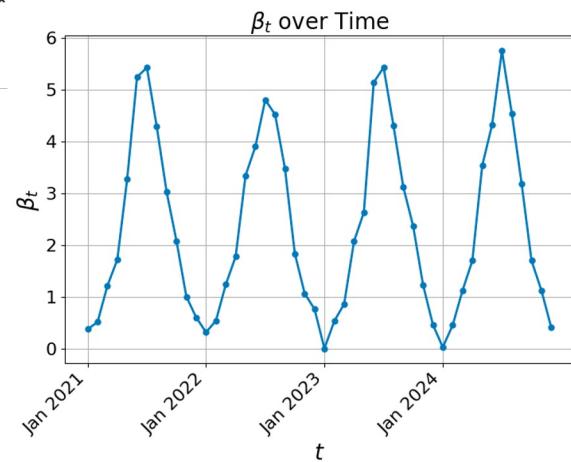
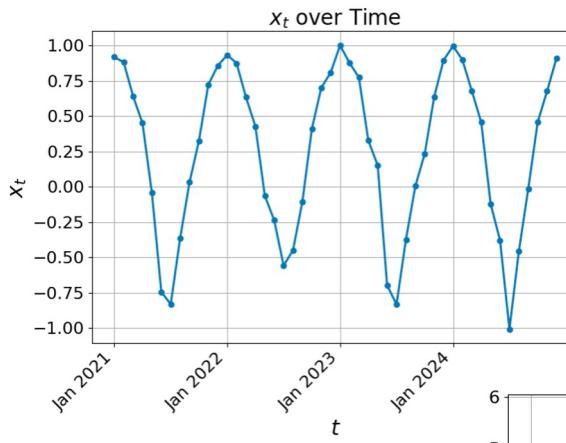
Surely this is nonlinear,
right...?



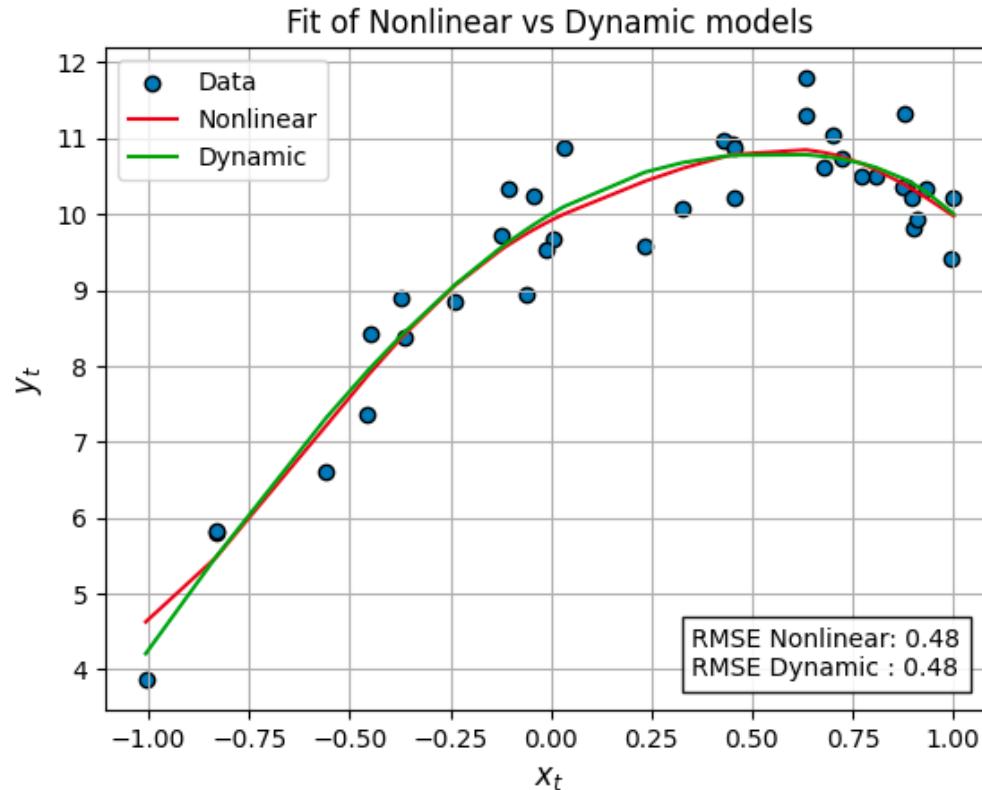
Surely this is nonlinear, right...?

True data generating process (DGP):

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



Identical Fit



**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 will 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 will 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?

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

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

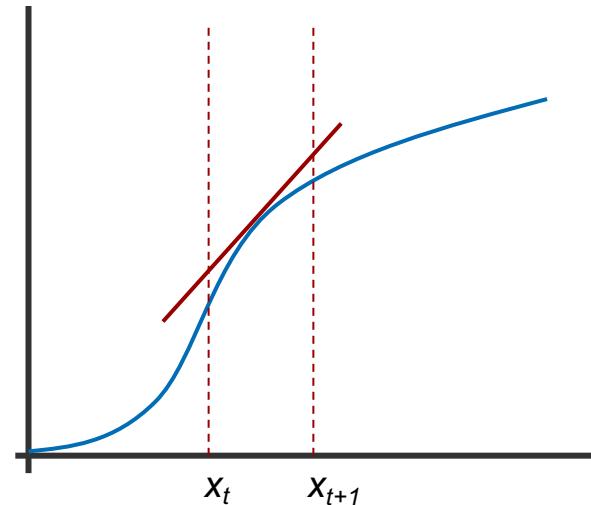
- First, a simple point: if $\beta_t = y_t/x_t$, the dynamic model will fit perfectly!
- A more nuanced answer: in any local region of x , a local linear approximation will work pretty well

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

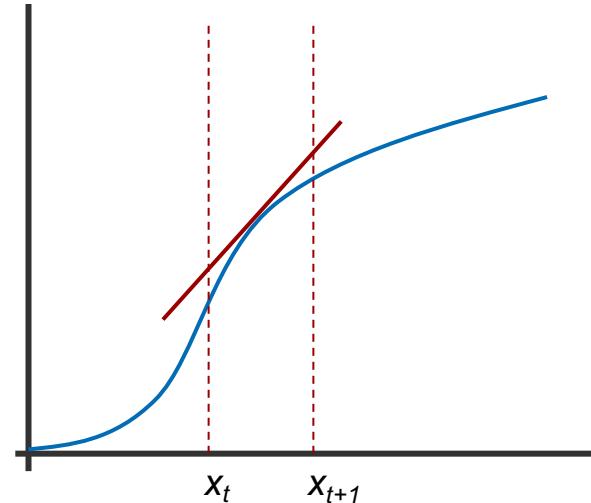


When can dynamic approximate nonlinear?

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

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

- 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?



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

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

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

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

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

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

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

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

When can nonlinear approximate dynamic?



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

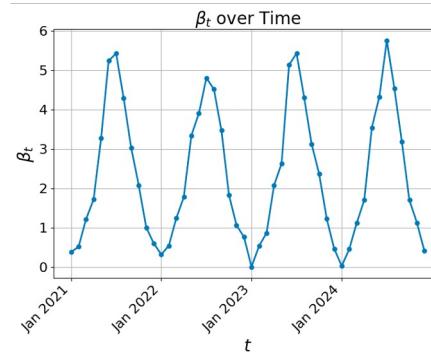
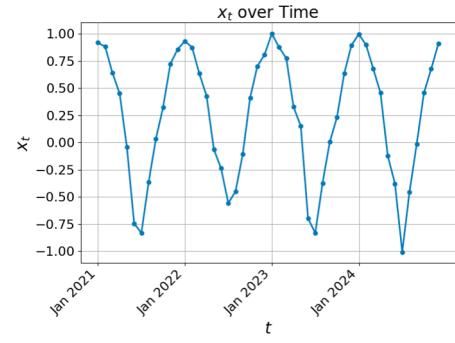


$$y_t = \sum_j \beta_{jt} x_{jt} + \epsilon_t$$

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

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

- 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

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

4. Dynamic coefficients inherited from common parent

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

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

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

4. Dynamic coefficients inherited from common parent

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

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

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

4. Dynamic coefficients inherited from common parent

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

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

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

4. Dynamic coefficients inherited from common parent

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

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

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



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

Primer: Gaussian Processes

Gaussian processes: a Bayesian nonparametric approach to modeling unknown functions

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

$\uparrow\downarrow$

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

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

Primer: Gaussian Processes

Gaussian processes: a Bayesian nonparametric approach to modeling unknown functions

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

↑↓

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

Amplitude

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

Lengthscale

Primer: Gaussian Processes

Gaussian processes: a Bayesian nonparametric approach to modeling unknown functions

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

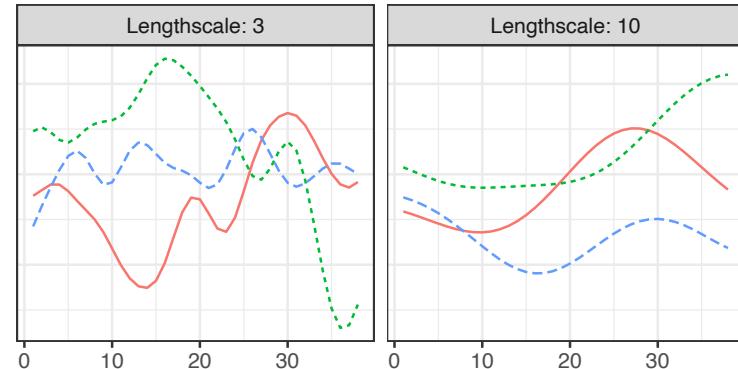


$$f(x_1, \dots, x_N) \sim \mathcal{N}(m(x_1, \dots, x_N), K), \text{ s.t. } K_{ij} = k(x_i, x_j)$$

Amplitude

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

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

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

4. Dynamic coefficients inherited from common parent

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

Gaussian processes

Two important levers:

- Smoothness
- Amplitude

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

Other manipulated features:

4. Dynamic coefficient
- Autoregressive coefficient in x
 - Noise in x 's autoregressive process
 - Variance of the error term

Gaussian processes

Two important levers:

- Smoothness
- Amplitude

Simulation Results

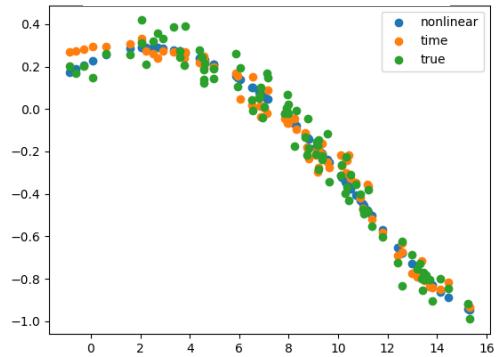
- For each simulation type, >300 settings, systematically varying the previously described factors, with 100 simulations per setting
- Fit both models (nonlinear and dynamic), measure conflation through validation RMSE

Examples

Nonlinear DGP

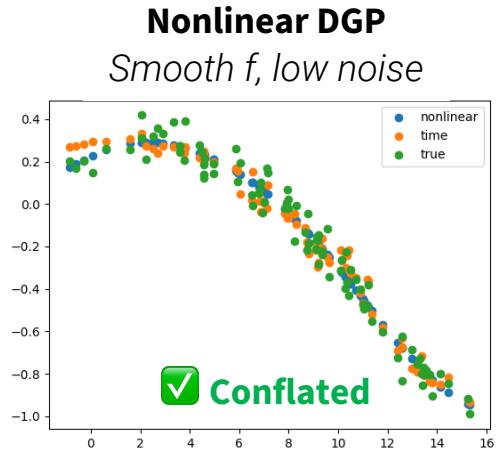
Smooth f , low noise

Data



Examples

Data

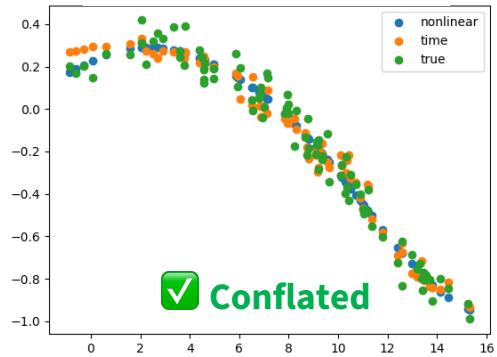


Examples

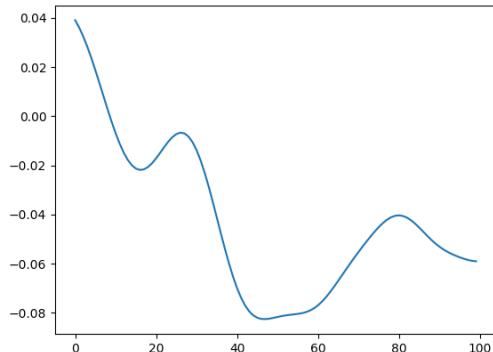
Nonlinear DGP

Smooth f , low noise

Data

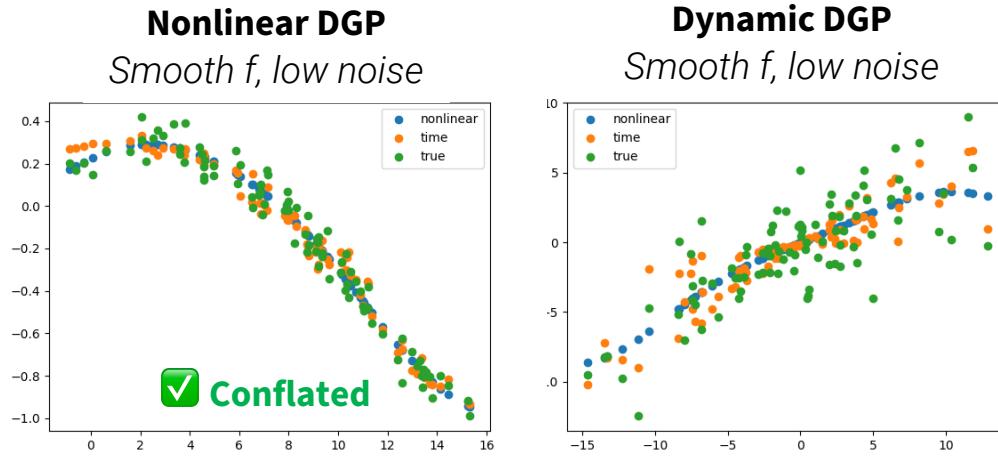


Implied
 $\beta(t)$

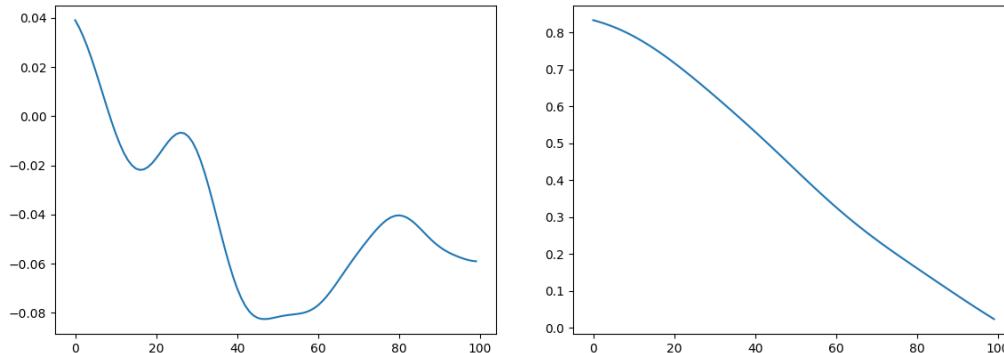


Examples

Data

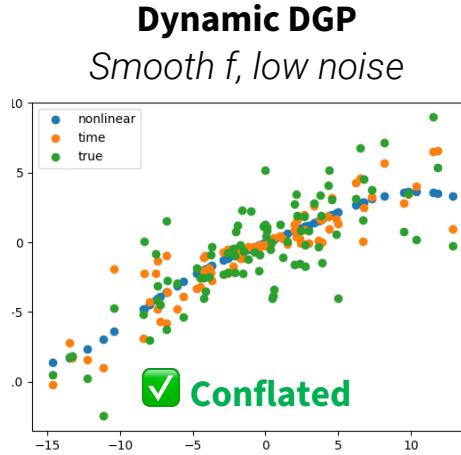
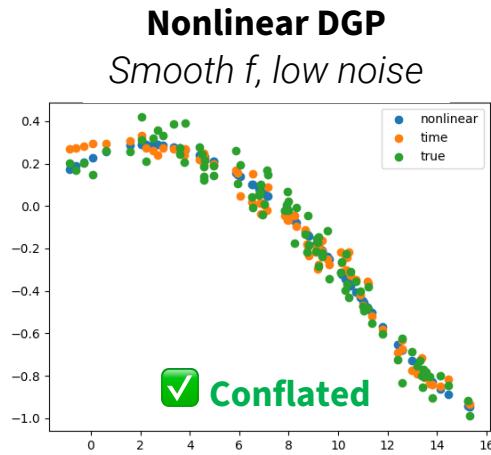


Implied
 $\beta(t)$

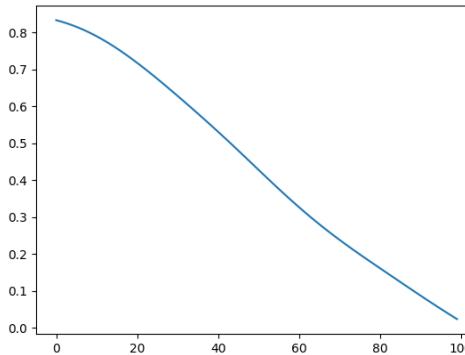
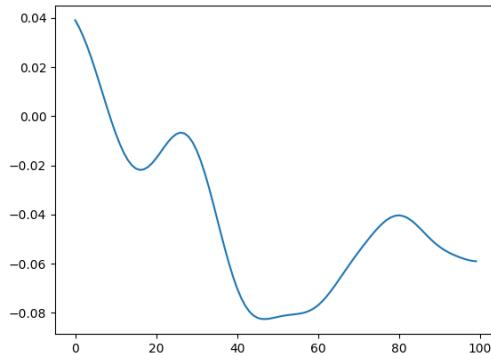


Examples

Data

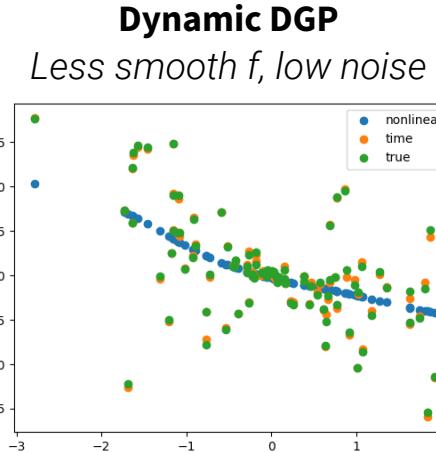
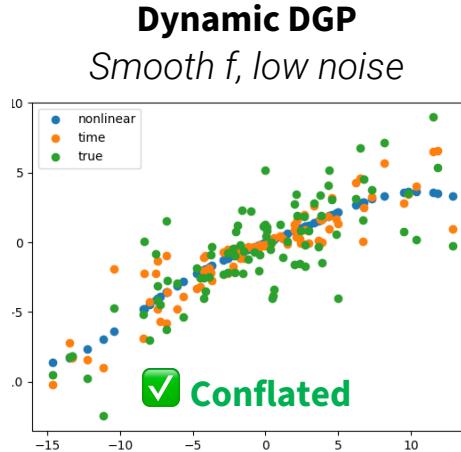
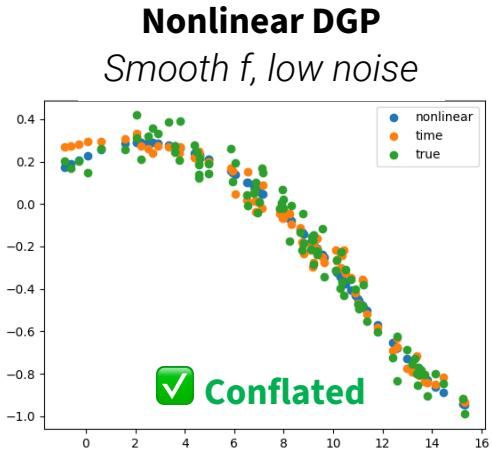


Implied
 $\beta(t)$

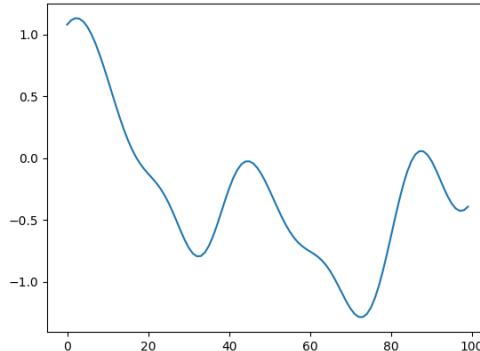
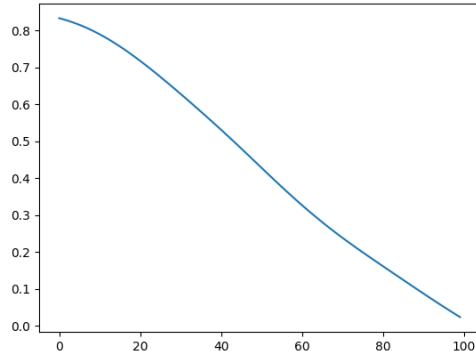
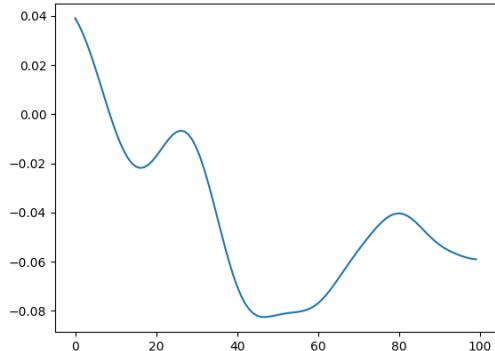


Examples

Data

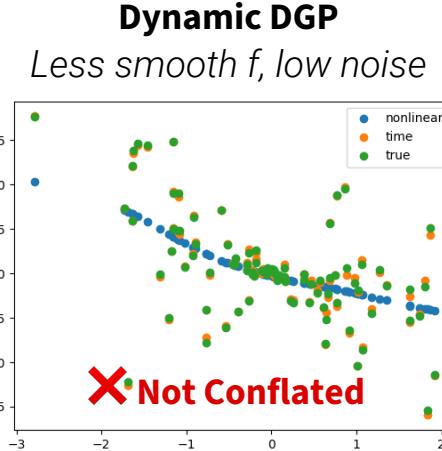
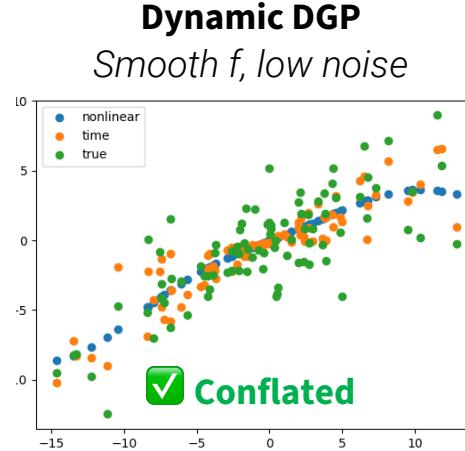
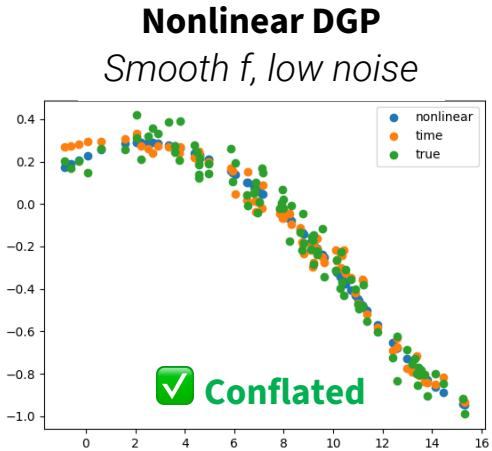


Implied
 $\beta(t)$

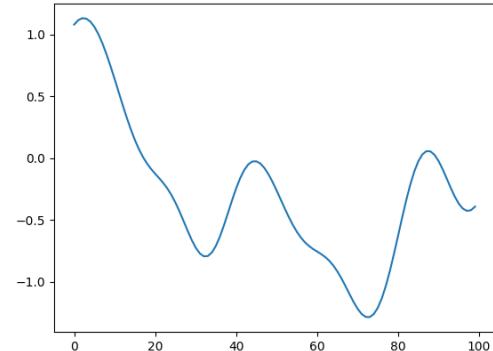
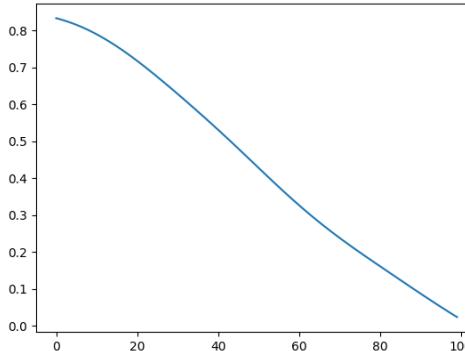
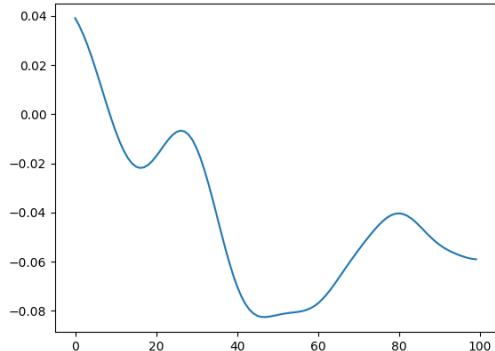


Examples

Data



Implied
 $\beta(t)$



Simulation Results

- Systematically vary the previously described factors, with 100 simulations per setting
- Fit both models (nonlinear and dynamic), measure conflation through validation RMSE

Simulation Results

- Systematically vary the previously described factors, with 100 simulations per setting
- Fit both models (nonlinear and dynamic), measure conflation through validation RMSE

Main result: huge prevalence of conflation

- Under the nonlinear DGP, **82%** exhibited some conflation, with **23% exhibiting major conflation** (*defined as >25% of simulations conflated*)
- Under the time-varying DGP, **80%** exhibited some conflation, with **27% exhibiting major conflation**

Diving Deeper

DV: % Conflated Simulations

Diving Deeper

DV: % Conflated Simulations

Variable	Level	Nonlinear DGP		Dynamic DGP	
		Coef	$P(> t)$	Coef	$P(> t)$
Amplitude, f :	Low	-	-	-	-
	Middle	-0.06	0.94	2.94	0.01
	High	-0.11	0.88	8.99	0.00
Smoothness, f :	Low	-	-	-	-
	Middle	9.99	0.00	3.44	0.00
	High	19.43	0.00	8.10	0.00
AR coef, x :	Low	-	-	-	-
	Middle	0.81	0.28	1.05	0.36
	High	2.66	0.00	4.77	0.00
Variance, x :	Low	-	-	-	-
	Middle	-0.16	0.83	5.80	0.00
	High	0.07	0.92	11.66	0.00
Noise, y	Low	-	-	-	-
	Middle	7.60	0.00	6.69	0.00
	High	14.81	0.00	20.25	0.00
	Very High	24.02	0.00	40.19	0.00

Table 1: Simulation Results: DV = Percentage Conflation; Intercept omitted for clarity.

Diving Deeper

DV: % Conflated Simulations

Conflation more likely with...

- Noisier data (i.e., ϵ_t)

Variable	Level	Nonlinear DGP		Dynamic DGP	
		Coef	$P(> t)$	Coef	$P(> t)$
Amplitude, f :	Low	-	-	-	-
	Middle	-0.06	0.94	2.94	0.01
	High	-0.11	0.88	8.99	0.00
Smoothness, f :	Low	-	-	-	-
	Middle	9.99	0.00	3.44	0.00
	High	19.43	0.00	8.10	0.00
AR coef, x :	Low	-	-	-	-
	Middle	0.81	0.28	1.05	0.36
	High	2.66	0.00	4.77	0.00
Variance, x :	Low	-	-	-	-
	Middle	-0.16	0.83	5.80	0.00
	High	0.07	0.92	11.66	0.00
Noise, y	Low	-	-	-	-
	Middle	7.60	0.00	6.69	0.00
	High	14.81	0.00	20.25	0.00
	Very High	24.02	0.00	40.19	0.00

Table 1: 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

Variable	Level	Nonlinear DGP		Dynamic DGP	
		Coef	$P(> t)$	Coef	$P(> t)$
Amplitude, f :	Low	-	-	-	-
	Middle	-0.06	0.94	2.94	0.01
	High	-0.11	0.88	8.99	0.00
Smoothness, f :	Low	-	-	-	-
	Middle	9.99	0.00	3.44	0.00
	High	19.43	0.00	8.10	0.00
AR coef, x :	Low	-	-	-	-
	Middle	0.81	0.28	1.05	0.36
	High	2.66	0.00	4.77	0.00
Variance, x :	Low	-	-	-	-
	Middle	-0.16	0.83	5.80	0.00
	High	0.07	0.92	11.66	0.00
Noise, y	Low	-	-	-	-
	Middle	7.60	0.00	6.69	0.00
	High	14.81	0.00	20.25	0.00
	Very High	24.02	0.00	40.19	0.00

Table 1: 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

Variable	Level	Nonlinear DGP		Dynamic DGP	
		Coef	$P(> t)$	Coef	$P(> t)$
Amplitude, f :	Low	-	-	-	-
	Middle	-0.06	0.94	2.94	0.01
	High	-0.11	0.88	8.99	0.00
Smoothness, f :	Low	-	-	-	-
	Middle	9.99	0.00	3.44	0.00
	High	19.43	0.00	8.10	0.00
AR coef, x :	Low	-	-	-	-
	Middle	0.81	0.28	1.05	0.36
	High	2.66	0.00	4.77	0.00
Variance, x :	Low	-	-	-	-
	Middle	-0.16	0.83	5.80	0.00
	High	0.07	0.92	11.66	0.00
Noise, y	Low	-	-	-	-
	Middle	7.60	0.00	6.69	0.00
	High	14.81	0.00	20.25	0.00
	Very High	24.02	0.00	40.19	0.00

Table 1: 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!

Variable	Level	Nonlinear DGP		Dynamic DGP	
		Coef	$P(> t)$	Coef	$P(> t)$
Amplitude, f :	Low	-	-	-	-
	Middle	-0.06	0.94	2.94	0.01
	High	-0.11	0.88	8.99	0.00
Smoothness, f :	Low	-	-	-	-
	Middle	9.99	0.00	3.44	0.00
	High	19.43	0.00	8.10	0.00
AR coef, x :	Low	-	-	-	-
	Middle	0.81	0.28	1.05	0.36
	High	2.66	0.00	4.77	0.00
Variance, x :	Low	-	-	-	-
	Middle	-0.16	0.83	5.80	0.00
	High	0.07	0.92	11.66	0.00
Noise, y	Low	-	-	-	-
	Middle	7.60	0.00	6.69	0.00
	High	14.81	0.00	20.25	0.00
	Very High	24.02	0.00	40.19	0.00

Table 1: Simulation Results: DV = Percentage Conflation; Intercept omitted for clarity.

Even worse with AdStock

Under the nonlinear DGP with AdStock, **93%** exhibited some conflation,
with **37% exhibiting major conflation**
(previously: 82% and 23%)

Even worse with AdStock

Under the nonlinear DGP with AdStock, **93%** exhibited some conflation,
with **37% exhibiting major conflation**

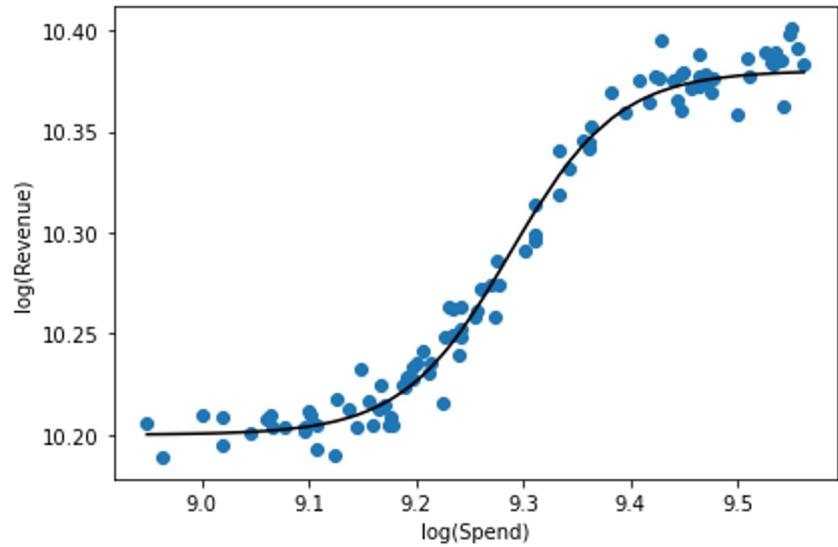
(previously: 82% and 23%)

Under the time-varying DGP with AdStock, **87%** exhibited some conflation,
with **52% exhibiting major conflation**

(previously: 80% and 27%)

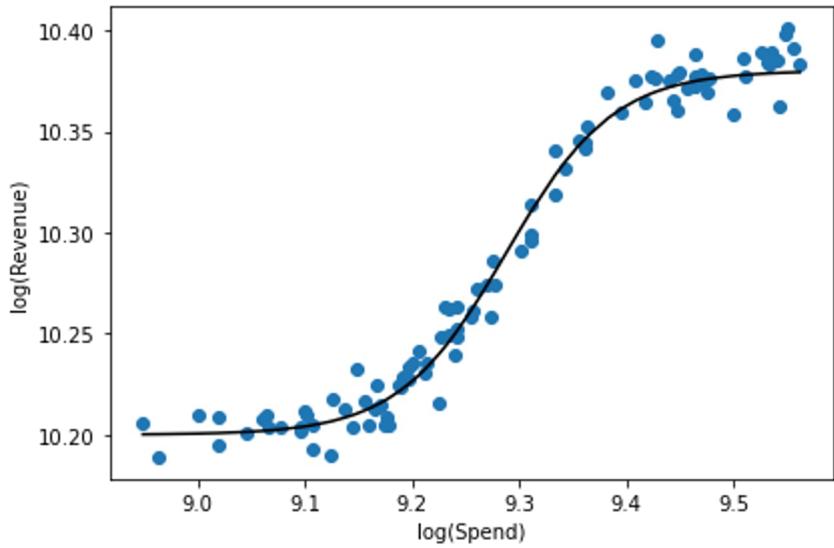
Implications

One last simulation...

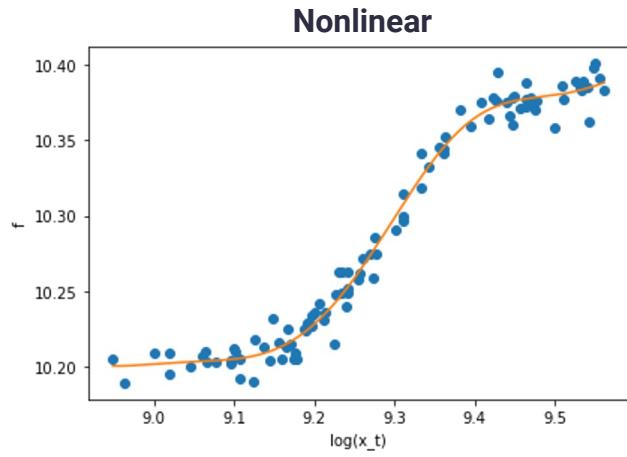


(or from roughly \$8,000 to \$14,000)

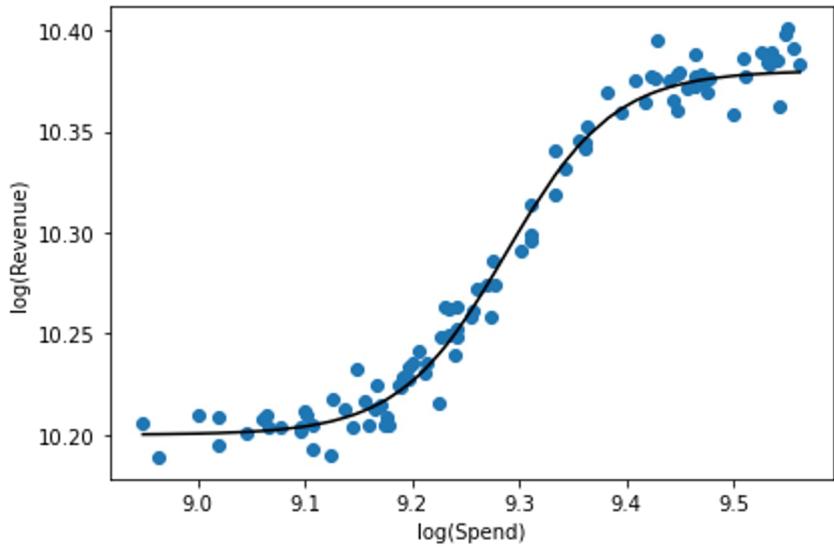
One last simulation...



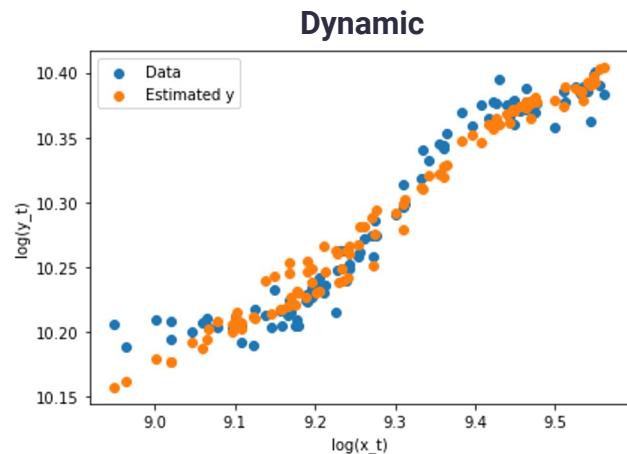
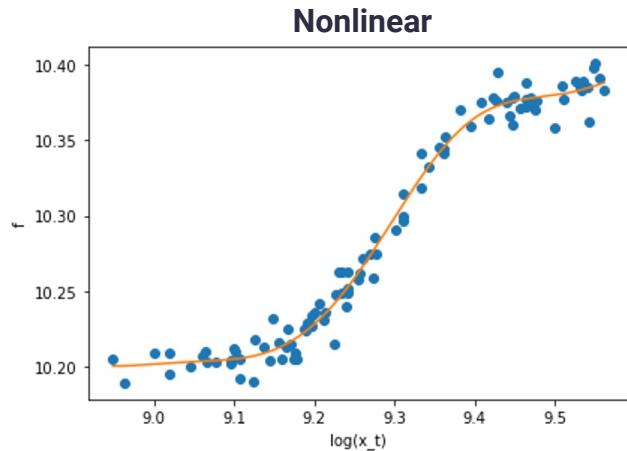
(or from roughly \$8,000 to \$14,000)



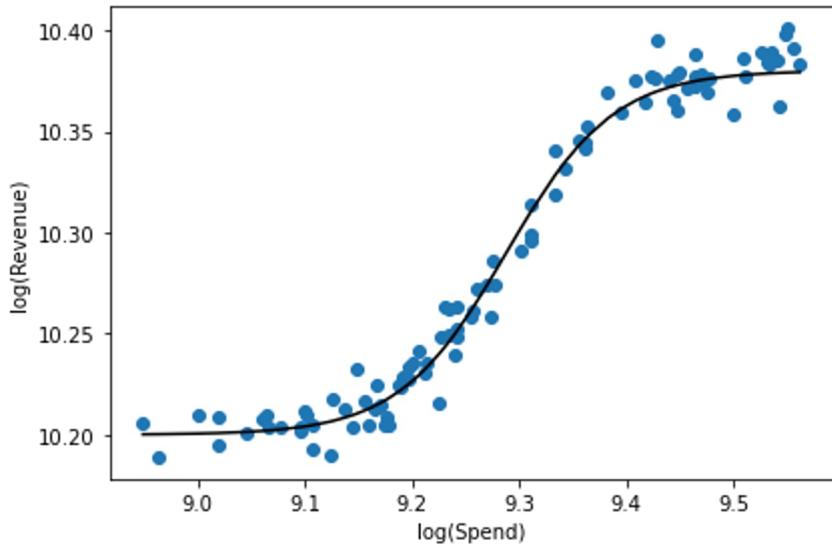
One last simulation...



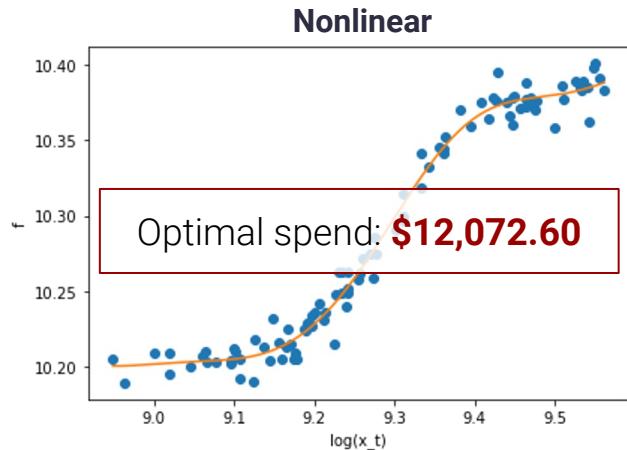
(or from roughly \$8,000 to \$14,000)



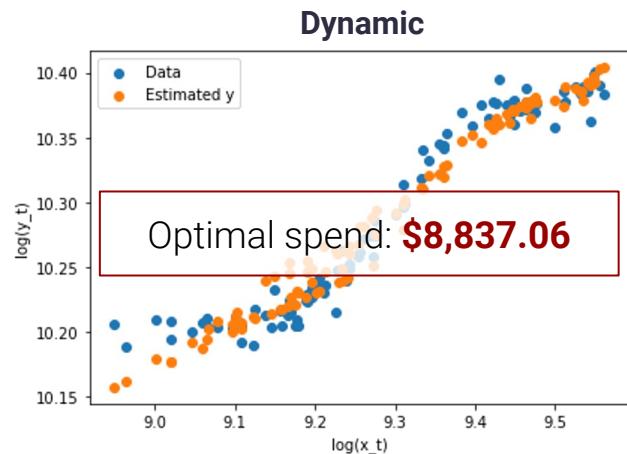
One last simulation...



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



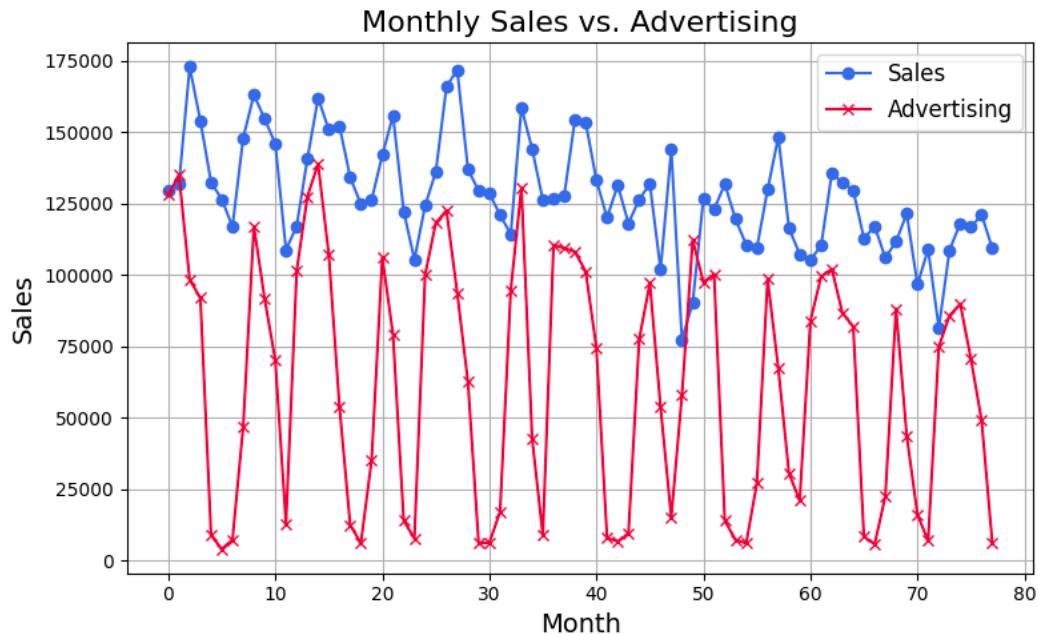
Optimal spend: **\$12,072.60**



Optimal spend: **\$8,837.06**

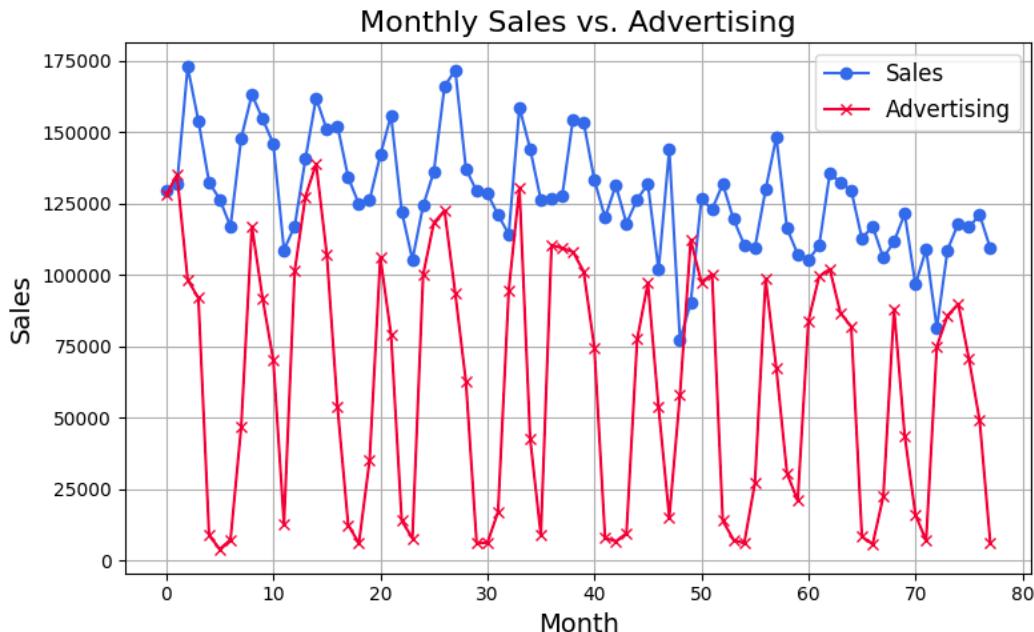
Application: The Classic Lydia Pinkham Dataset

Sales and advertising of Lydia Pinkham's herbal products, monthly, 1954-1960 ([Palda, 1964](#))



Application: The Classic Lydia Pinkham Dataset

Sales and advertising of Lydia Pinkham's herbal products, monthly, 1954-1960 ([Palda, 1964](#))



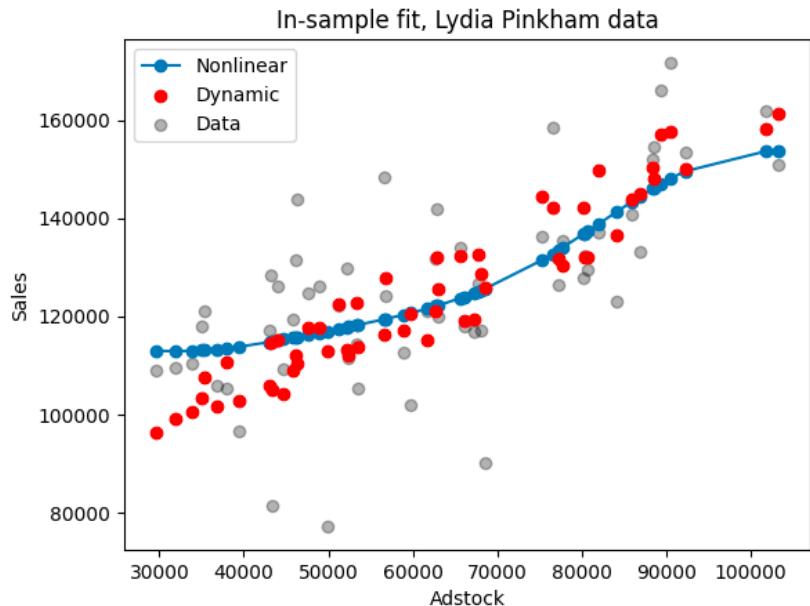
Lydia Pinkham
150 Tablets
★★★★★ 764
50+ bought in past month
\$21.99 (\$0.15/Count)
✓prime Two-Day
FREE delivery Thu, May 9
Add to cart

Conflation

Fit both nonlinear and dynamic models, using AdStock with carryover
 $\alpha = 0.7$ and $L = 13$ ([Bultez and Naert, 1979](#))

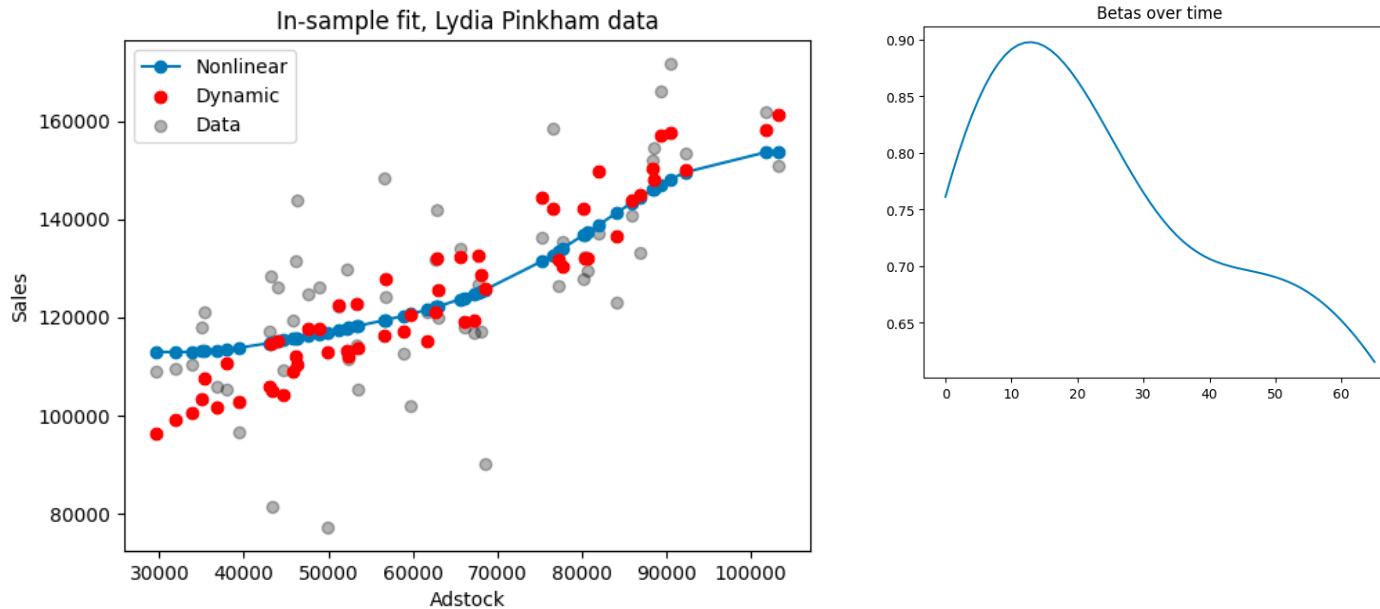
Conflation

Fit both nonlinear and dynamic models, using AdStock with carryover
 $\alpha = 0.7$ and $L = 13$ ([Bultez and Naert, 1979](#))



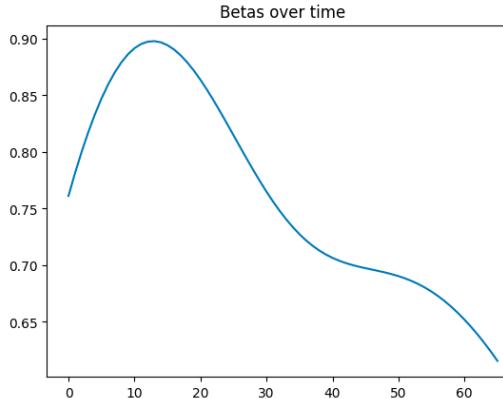
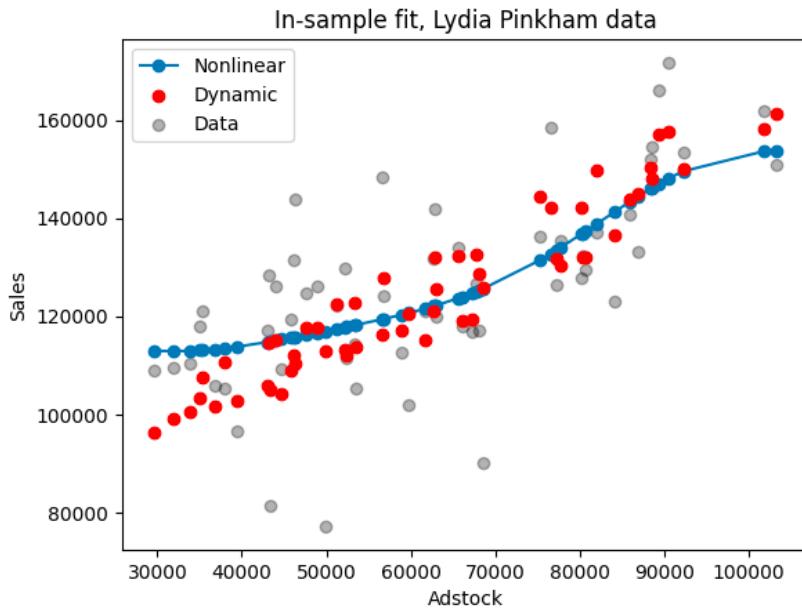
Conflation

Fit both nonlinear and dynamic models, using AdStock with carryover
 $\alpha = 0.7$ and $L = 13$ ([Bultez and Naert, 1979](#))



Conflation

Fit both nonlinear and dynamic models, using AdStock with carryover
 $\alpha = 0.7$ and $L = 13$ ([Bultez and Naert, 1979](#))



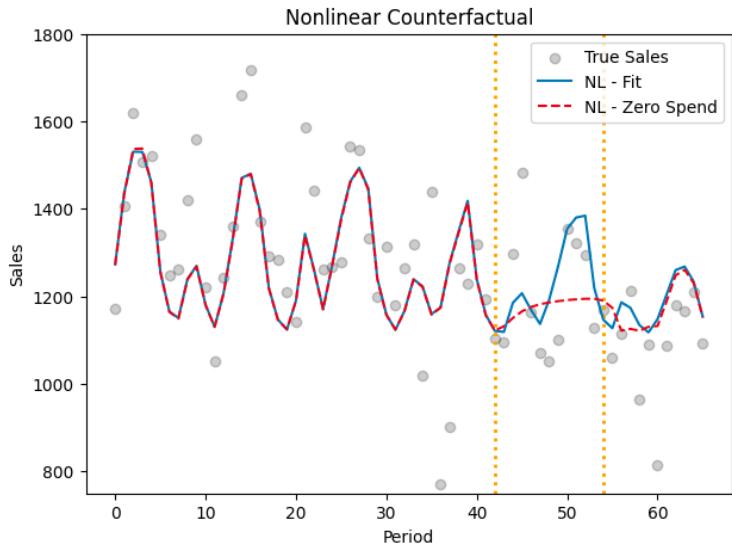
95% posterior predictive MSE intervals:
Nonlinear: [126.80, 174.78]
Dynamic: [93.31, 127.18]

Returns from Advertising

$$\text{Return}[T_1, T_2] = \sum_{t>T} \hat{y}_t(x_{[T_1, T_2]} = x_{\text{True}}) - \hat{y}_t(x_{[T_1, T_2]} = 0)$$

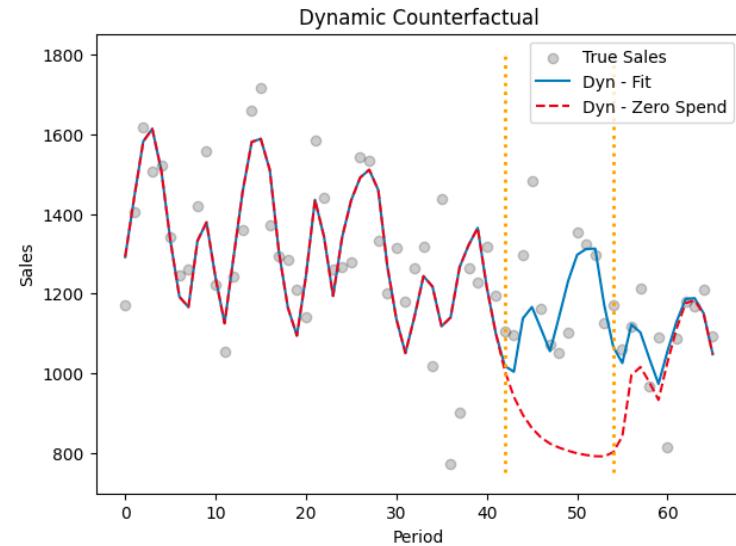
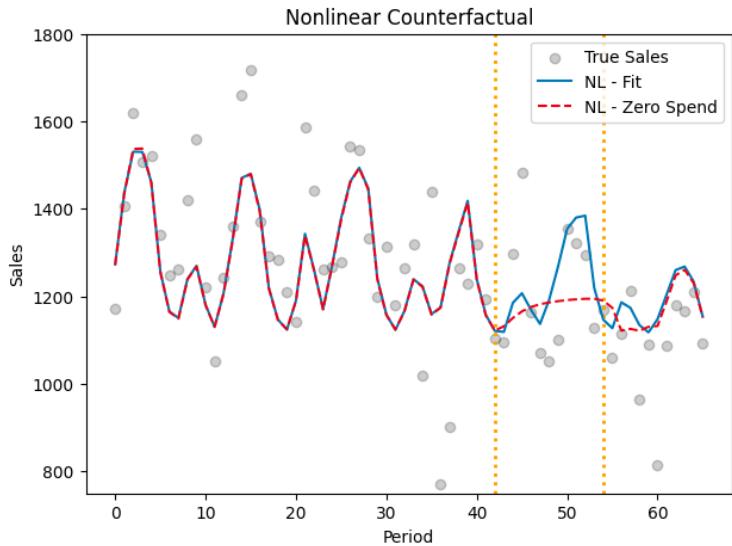
Returns from Advertising

$$\text{Return}[T_1, T_2] = \sum_{t>T} \hat{y}_t(x_{[T_1, T_2]} = x_{\text{True}}) - \hat{y}_t(x_{[T_1, T_2]} = 0)$$



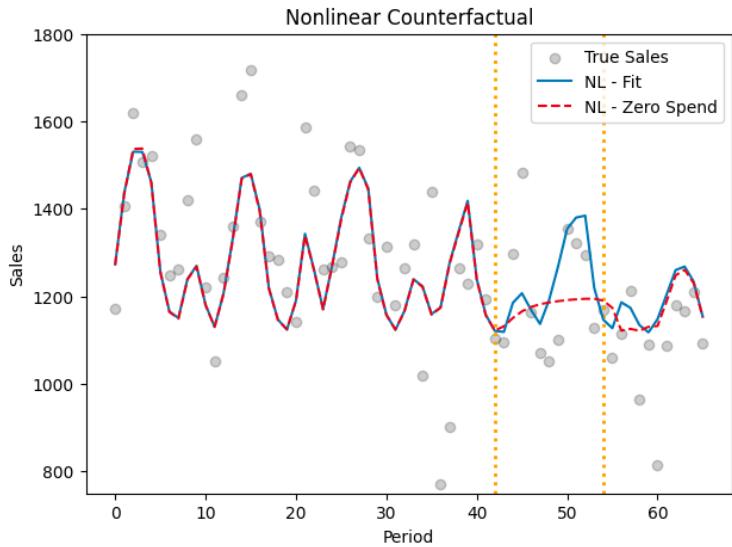
Returns from Advertising

$$\text{Return}[T_1, T_2] = \sum_{t>T} \hat{y}_t(x_{[T_1, T_2]} = x_{\text{True}}) - \hat{y}_t(x_{[T_1, T_2]} = 0)$$

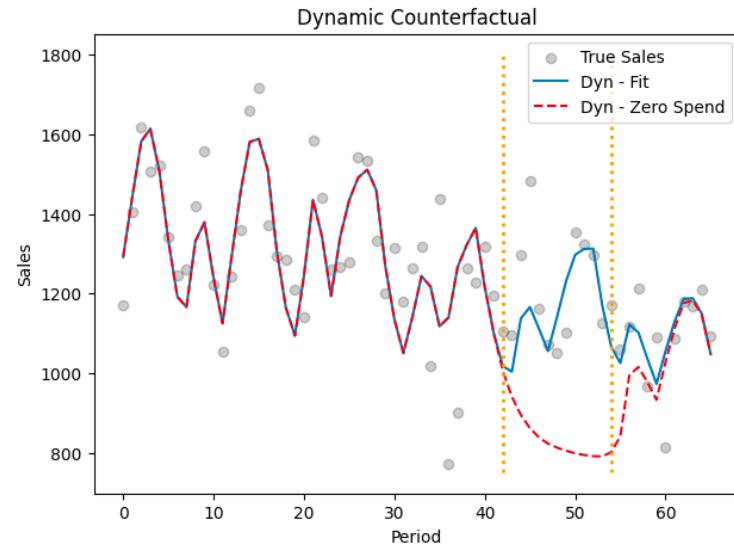


Returns from Advertising

$$\text{Return}[T_1, T_2] = \sum_{t>T} \hat{y}_t(x_{[T_1, T_2]} = x_{\text{True}}) - \hat{y}_t(x_{[T_1, T_2]} = 0)$$



Estimated revenue: **\$66,427**



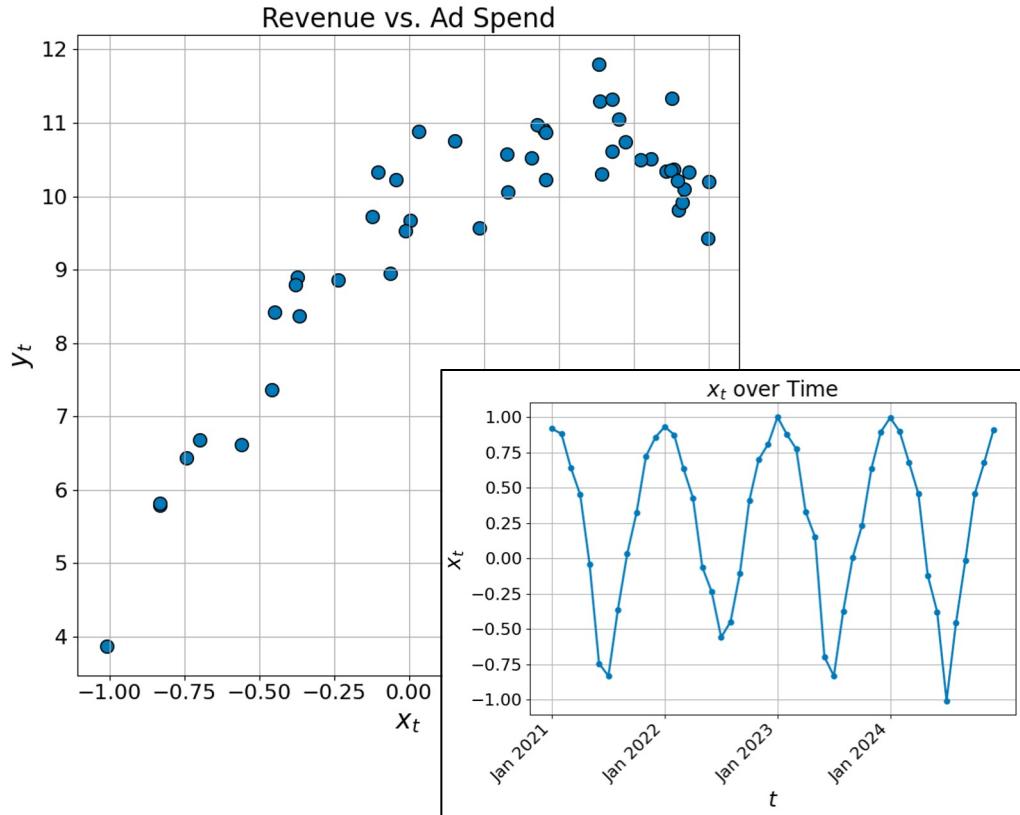
Estimated revenue: **\$468,832**

Is MMM doomed?

- In short: **no!**
- Careful ad planning can help disentangle the two stories

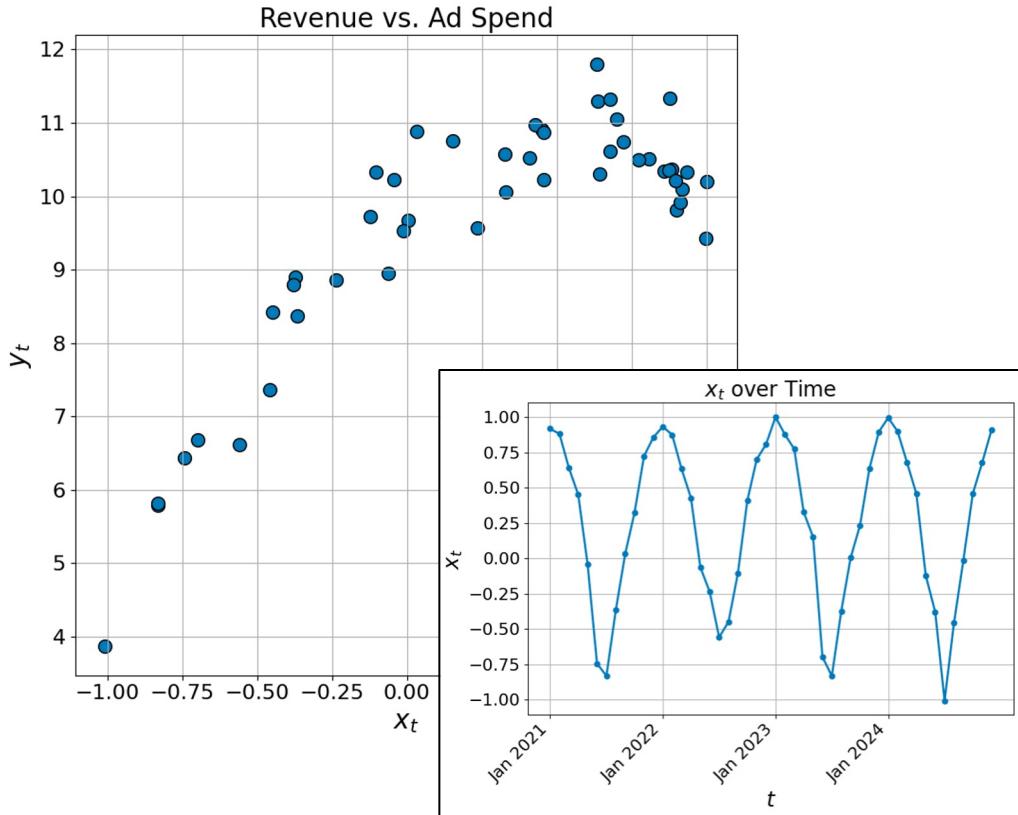
Is MMM doomed?

- In short: **no!**
- Careful ad planning can help disentangle the two stories



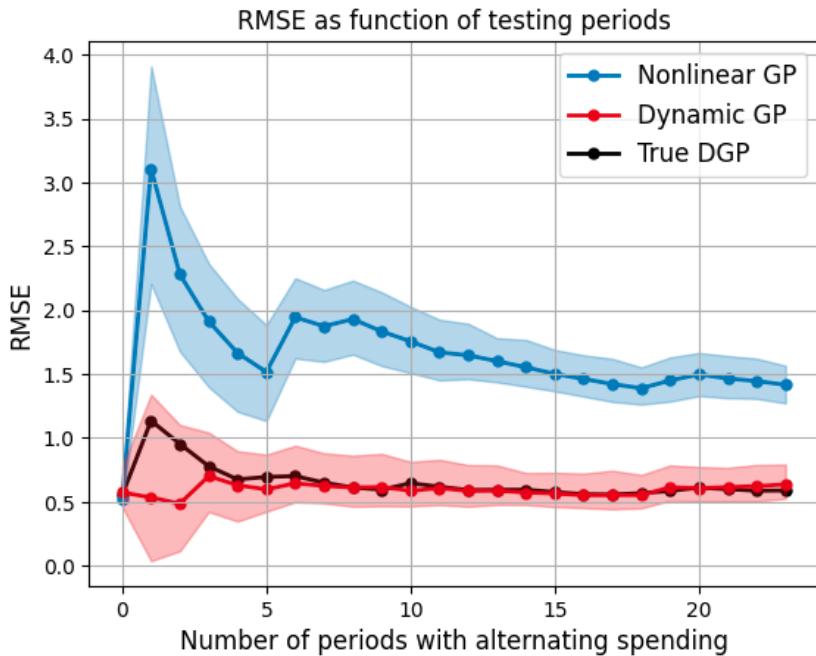
Is MMM doomed?

- In short: **no!**
- Careful ad planning can help disentangle the two stories
- One strategy that seems to work: **seesaw spending** (i.e., quick alternating x)



Is MMM doomed?

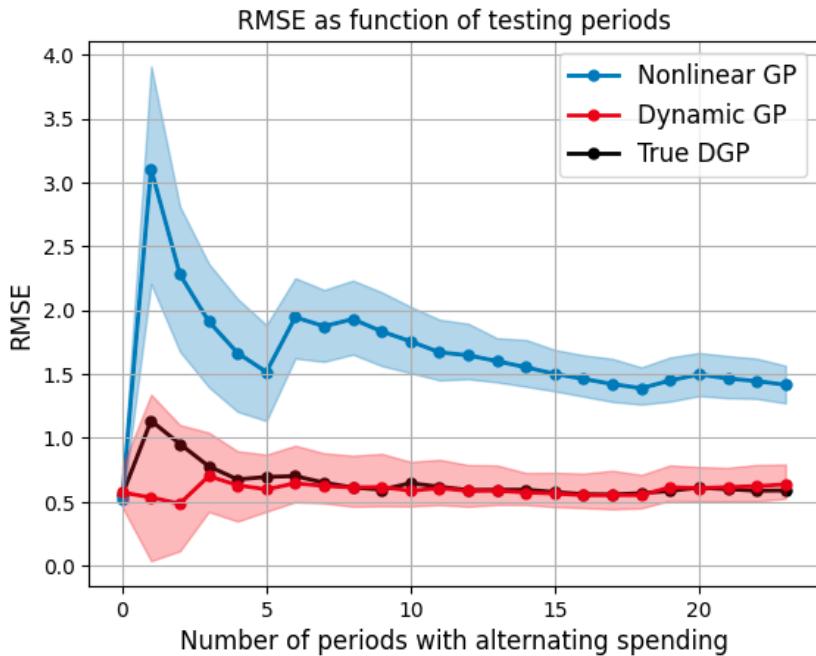
- In short: **no!**
- Careful ad planning can help disentangle the two stories
- One strategy that seems to work: **seesaw spending** (i.e., quick alternating x)



Is MMM doomed?

- In short: **no!**
- Careful ad planning can help disentangle the two stories
- One strategy that seems to work: **seesaw spending** (i.e., quick alternating x)

Ongoing work: optimal test design for model inference



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 soon:
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

