Deep Learning for (Structural) Individual Heterogenity

Sanjog Misra

Booth School Of Business
University of Chicago

DSE 2022 @ MIT

Structural Heterogeneity

- A fundamental building block of decisions in economics and marketing is the construct of individual heterogeneity.
- Accounting for such heterogeneity is relevant for a number of tasks
  - Accuracy (Bias)
  - Inference (Variance)
  - Decisions, Policy Design and Evaluation, Targeting, Segmentation and Personalization
- This presentation is about the **practice** of estimating and using heterogeneity measures in structrual economic models.

Messages

Deep Learning for Individual Heterogneiety*

(*This paper, builds on...)

Deep Neural Networks for Estimation and Inference

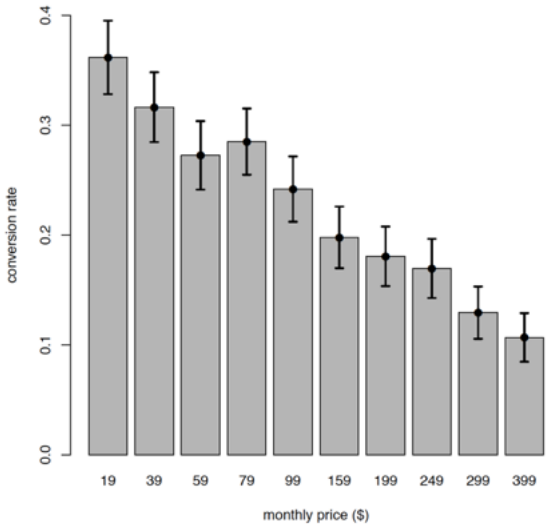
(Econometrica 2021) [https://arxiv.org/abs/1809.09953]

(both with with Max Farrell and Tengyuan Liang)


Agenda

- Deep Learning Heterogeneity: The Idea
- A general Framework
- Applications

Motivating Structure: Pricing Experiment



Pricing Experiment

- Let's assume that we are interested in learning a demand function

$$y = f(p)$$

- So that we can find optimal prices

$$p^* = \arg\max_p (p - c) f(p)$$

- We decide to use Random forests to obtain $\widehat{f}(p)$
- Plugging in and optimizing gives us...

$$p^* =$$

Pricing Experiment

- Let's assume that we are interested in learning a demand function

$$y = f(p)$$

- So that we can find optimal prices

$$p^* = \arg\max_p (p - c) f(p)$$

- We decide to use Random forests to obtain $\widehat{f}(p)$
- Plugging in and optimizing gives us...

$$p^* = \infty$$

The typical structural model.

$$\ell(\boldsymbol{Y}, \boldsymbol{T}, \boldsymbol{\theta}_i)$$

A parametric per-observation loss function for a structural model.

The typical cheat.

$$\ell(\boldsymbol{Y}, \boldsymbol{T}, \boldsymbol{\theta})$$

The easiest way to deal with heterogeneity is to ignore it.

The not-so new idea.

$$\ell(\boldsymbol{Y}, \boldsymbol{T}, \boldsymbol{\theta}(\boldsymbol{X}))$$

We suggest projecting $\boldsymbol{\theta}_i$ on $\boldsymbol{x}_i$. Old idea, used quite often.

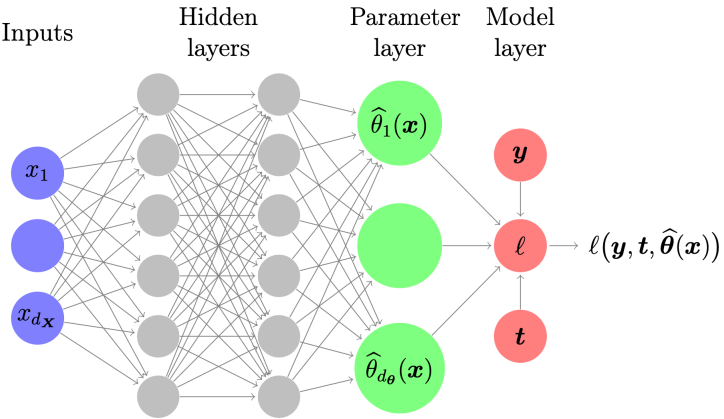
Pretty much the same idea as interaction except that θ is a function..

The new idea.

$$\ell(\boldsymbol{Y}, \boldsymbol{T}, \boldsymbol{\theta}_{\mathrm{DNN}}(\boldsymbol{X}))$$

More specifically, we are proposing that the parameter functions be treated as a **Deep Neural Network** (DNN).

The Framework



The key idea is that the deep learning is targeted towards the parameters. The model is still parameteric and defined by economic structure.

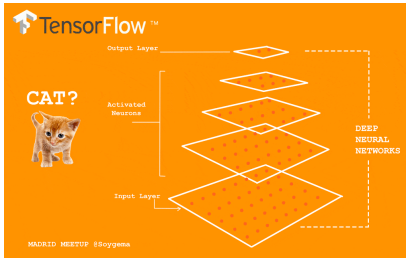
The estimator.

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta} \in \mathcal{F}_{\mathrm{DNN}}} \frac{1}{n} \sum_i \ell\left(\boldsymbol{y}_i, \boldsymbol{t}_i, \boldsymbol{\theta}(\boldsymbol{x}_i)\right)$$

Implementing an estimator is then relatively strightforward.

We can use standard tools such as **Tensorflow** or **Torch**.

Why DNNs?



Universal Approximators

Can approximate any smooth function.

Scalability

Infrastructure allows for true scalability (up or down!)

Structural Compatibility

Global estimators that can be embedded into a structural model.

Interpretation maintained!

Structural Compatibility

- Consider a utility function

$$U_i = \alpha + \beta t_i + \varepsilon_i$$

- Where t_i is some treatment (for our example, say a targeted price)
- Then a "structural" choice with heterogeneity and the usual EVTII error gives

$$\mathbb{P}\left(y_i = 1 | \mathbf{x}_i, t_i\right) = \frac{\exp\left(\alpha_i + \beta_i t_i\right)}{1 + \exp\left(\alpha_i + \beta_i t_i\right)}$$

- Changing this to

$$\mathbb{P}\left(y_i = 1 | \mathbf{x}_i, t_i\right) = \frac{\exp\left(\alpha_{\mathsf{DNN}}\left(\mathbf{x}_i\right) + \beta_{\mathsf{DNN}}\left(\mathbf{x}_i\right) t_i\right)}{1 + \exp\left(\alpha_{\mathsf{DNN}}\left(\mathbf{x}_i\right) + \beta_{\mathsf{DNN}}\left(\mathbf{x}_i\right) t_i\right)}$$

- retains the structural interpretation *completely*.
  - $\beta\left(x\right)$ is still the price effect!
  - Can still use usual tricks for
    - WTP, Elasticity, Surplus, ...

Pieces of the puzzle...

$$\left[\beta_i - \widehat{\beta}_{\text{DNN}}\left(\boldsymbol{x}_i\right)\right] = \underbrace{\left[\beta_i - \beta\left(\boldsymbol{x}_i\right)\right]}_{\text{approximation}} + \underbrace{\left[\beta\left(\boldsymbol{x}_i\right) - \beta_{\text{DNN}}\left(\boldsymbol{x}_i\right)\right]}_{\text{bias}} + \underbrace{\left[\beta_{\text{DNN}}\left(\boldsymbol{x}_i\right) - \widehat{\beta}_{\text{DNN}}\left(\boldsymbol{x}_i\right)\right]}_{\text{variance}}$$

Is this approximation a good idea?

$$\beta(\boldsymbol{x}_i) = \beta(\quad)$$

Data is aplenty. Really.

Any unobserved heterogeneity will have to be orthogonal to *all* observed heterogeneity.

What about the other pieces?

$$\underbrace{\left|\theta_i - \theta\left(\boldsymbol{x}_i\right)\right|}_{\text{approximation}} \leq \underbrace{\epsilon \text{ [as information}(\boldsymbol{x}_i) \to \infty]}_{\text{assumption}}$$

$$\underbrace{\left|\theta\left(\boldsymbol{x}_i\right) - \theta_{\mathrm{DNN}}\left(\boldsymbol{x}_i\right)\right|}_{\text{bias}} = \ ?$$

$$\underbrace{\left|\theta_{\mathrm{DNN}}\left(\boldsymbol{x}_i\right) - \widehat{\theta}_{DNN}\left(\boldsymbol{x}_i\right)\right|}_{\text{variance}} = \ ?$$

We got this.

$$\underbrace{\left|\theta_i - \theta\left(\boldsymbol{x}_i\right)\right|}_{\text{approximation}} \leq \underbrace{\epsilon\ [\text{as information}(\boldsymbol{x}_i) \to \infty]}_{\text{assumption}}$$

$$\underbrace{\left|\theta\left(\boldsymbol{x}_i\right) - \theta_{\text{DNN}}\left(\boldsymbol{x}_i\right)\right|}_{\text{bias}} \leq \underbrace{\epsilon\ [\text{as complexity (DNN)} \to \infty]}_{\text{proof}}$$

$$\underbrace{\left|\theta_{\text{DNN}}\left(\boldsymbol{x}_i\right) - \widehat{\theta}_{\text{DNN}}\left(\boldsymbol{x}_i\right)\right|}_{\text{variance}} \leq \underbrace{\epsilon\,[\text{as sample size}(n) \to \infty]}_{\text{proof}}$$

Result

$$\|\widehat{\boldsymbol{\theta}}_{\mathrm{DNN}} - \boldsymbol{\theta}\|_2^2 = O_{\mathbb{P}}\left(n^{-p/(p+d_C)} \log^8 n\right)$$

The rates are fast enough for us to do useful things with the estimators.

Inference Objects

$$\boldsymbol{\mu}_0 = \mathbb{E}\Big[\boldsymbol{H}(\boldsymbol{X}, \boldsymbol{\theta}_0(\boldsymbol{X}); \boldsymbol{t}^*)\Big]$$

We usually estimate structural models because we wish to compute some economic objects.

These could me measures, counterfactuals or policy interventions.

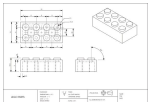
Generic Influence function

$$\boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{t}, \boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{\Lambda}) = \boldsymbol{H}\left(\boldsymbol{x}, \boldsymbol{\theta}(\boldsymbol{x}); \boldsymbol{t}^*\right) - \boldsymbol{H}_{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{\theta}(\boldsymbol{x}); \boldsymbol{t}^*)\boldsymbol{\Lambda}(\boldsymbol{x})^{-1}\boldsymbol{\ell}_{\boldsymbol{\theta}}(\boldsymbol{w}, \boldsymbol{\theta}(\boldsymbol{x})).$$

where

$$\boldsymbol{\Lambda}(\boldsymbol{x}) = \mathbb{E}[\boldsymbol{\ell}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{y}, \boldsymbol{t}, \boldsymbol{\theta}(\boldsymbol{x})) \mid \boldsymbol{X} = \boldsymbol{x}]$$

Applies to any smooth function and can be computed *automatically* using standard AD technology!



Applications

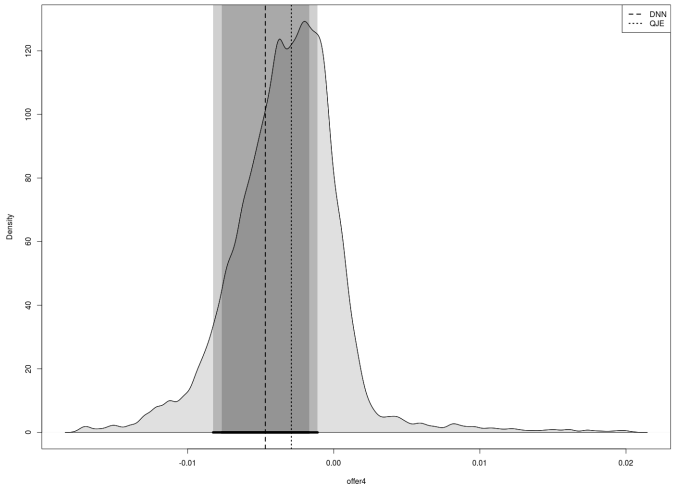
Content of Advertising

- **Bertrand, Karlan, Mullainathan, Shafir, Zinman (QJE, 2010)**:
  - Advertising for shortish-term loans in South Africa
  - Original questions: does advertising content matter? How much?
  - $Y = \{0, 1\}$ Applied for a loan (scalar in this case)
  - Policy/Treatment $\boldsymbol{T} = \mathbf{12}$ ad characteristics, randomly assigned (based on $\boldsymbol{x}$)
  - $T_1 =$ interest rate, directly compute valuation
  - Other qualities (photos, tables, uses) can be then valued
  - $\boldsymbol{X} = 11$ individual characteristics, some discrete
  - Economic Model for relationship of $\boldsymbol{T} \leftrightarrow Y$, enriched w/ heterogeneity in $\boldsymbol{X}$

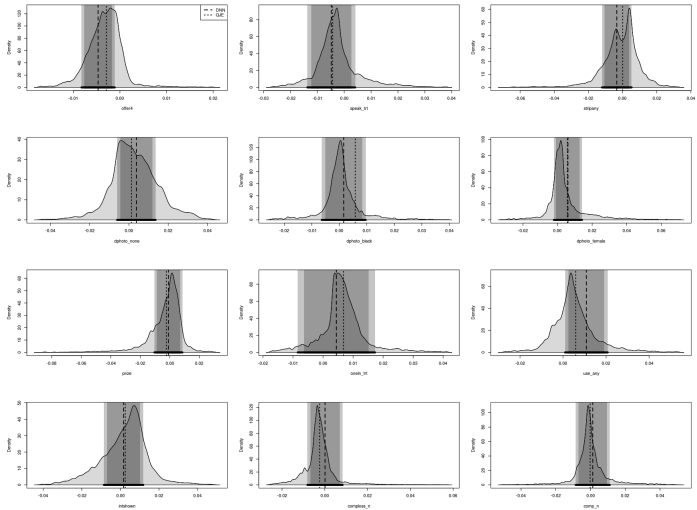
Marginal Effects Table

Variable	QJE	DNN ME	95% CI		$\widehat{Pr}\left(\hat{\beta}(x) > 0\right)$	Coef. of Variation
Interest rate offer	-0.0029	-0.0047	-0.0083	-0.0011	0.1337	1.0211
We speak your language	-0.0043	-0.0048	-0.0137	0.0041	0.2533	2.0542
Special rate for you	0.0001	-0.0034	-0.0120	0.0053	0.5001	4.4506
No photo	0.0013	0.0038	-0.0060	0.0136	0.5723	3.4931
Black photo	0.0058	0.0016	-0.0064	0.0096	0.5402	5.1348
Female photo	0.0057	0.0060	-0.0021	0.0141	0.6820	2.3375
Cell phone raffle	-0.0023	-0.0009	-0.0104	0.0085	0.4812	17.0059
Example loan shown	0.0068	0.0044	-0.0084	0.0173	0.8631	1.9379
No loan use mentioned	0.0059	0.0108	0.0009	0.0207	0.7499	1.0936
Interest rate shown	0.0025	0.0017	-0.0085	0.0119	0.6289	7.9903
Loss comparison	-0.0024	0.0001	-0.0081	0.0083	0.2342	89.3606
Competitors rate shown	-0.0002	0.0013	-0.0085	0.0111	0.4107	9.6790

Offer Rate Coefficient



Same for the Other Estimates



Beyond Marginal Effects: Optimal Offers

- Assume the firm wishes to maximize profits:

$$\max_{r=\text{rate}} \pi(r) := \max_{r=\text{rate}} L\left[rG\left(r\right)\right]\left[1 - D\left(r\right)\right]$$

  - $L =$ expected dollar loan amount, normalize to 1 (doesn't impact the rate)
  - $r =$ the interest rate offered
  - $G(r) =$ the probability of acceptance (depends on \alpha(\bx_{i}), \btheta(\bx_{i}))
  - $\left[1 - D\left(r\right)\right] =$ probability of non-default on the loan

- Then it is straightforward to show that

$$\frac{\partial \pi}{\partial r} = \left(r\dot{G}\left(r\right)\beta_r + G\left(r\right)\right)\left[1 - D\left(r\right)\right] - rG\left(r\right)\dot{D}\left(r\right)\delta = 0$$

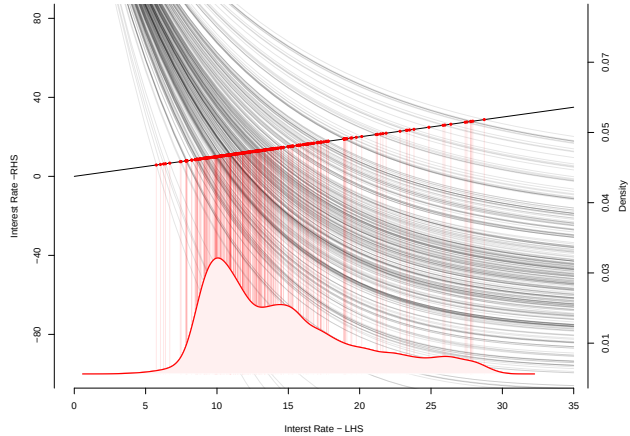
Beyond Marginal Effects: Optimal Offers

- There will a unique fixed point since the denominator of the RHS is decreasing in r for $\beta_r < 0$ (remember those 13%?) and$\delta > 0$
- Therefore

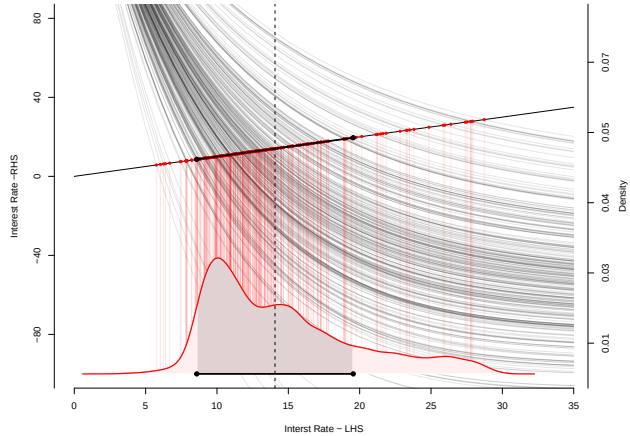
$$r^* = \frac{1 + r^* \left(1 - G\left(r^*\right)\right) \beta_r}{D\left(r^*\right) \delta}$$

- Even if we don't have it in closed form, r^* is a smooth function of $\boldsymbol{\theta}(\boldsymbol{x})$
- We can do inference on any $\boldsymbol{\mu}_0 = \mathbb{E}[\boldsymbol{H}(\boldsymbol{X}, \boldsymbol{\theta}, r^*)]$
- Impossible without
  1. Taking heterogeneity seriously
  2. Our new methods and ideas

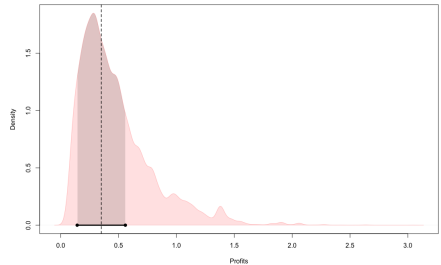
Optimal Offers



Optimal Offers



Profit from Personalized Offers



- Inference is automatic because \pi is just another **H** function:

$$\mu = \mathbb{E}[\pi(r^*)] = \mathbb{E}[\pi(r^*(\boldsymbol{\theta}(\boldsymbol{x})))]$$

- We cannot write the IF down in closed form, but we can evaluate it:
- $\widehat{\mu} = \$0.35$, with a 95% confidence interval of $(\$0.1421, \$0.5586)$
- Incremental 5.7% in expected profits over the optimal (uniform) interest rate


SNAP

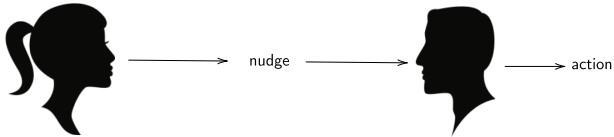
Application: SNAP Recertification

- SNAP participants often fail to re-certify at the 6 month mark (over 50%!)
- Simply need to fill out a form.
- Note: No interview at the 6-month mark in CA.

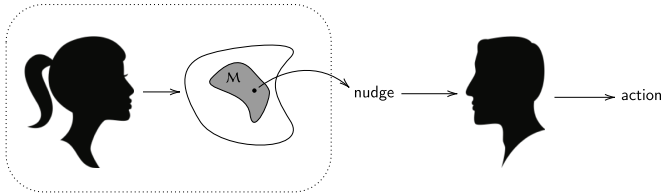
a [nudge] message

Hi [name], GetCalFresh here! To keep getting CalFresh, fill out your Semi-Annual Report (SAR 7) as soon as you can. It's required even if you have no changes to report. Do it online at: [link]

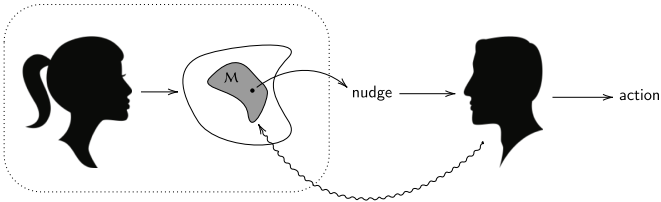
Message Architect



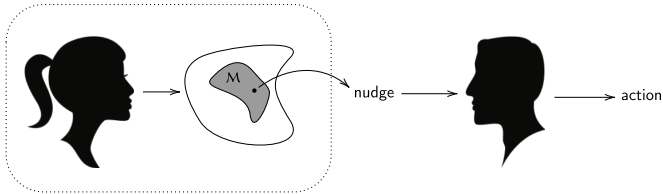
Designing Context



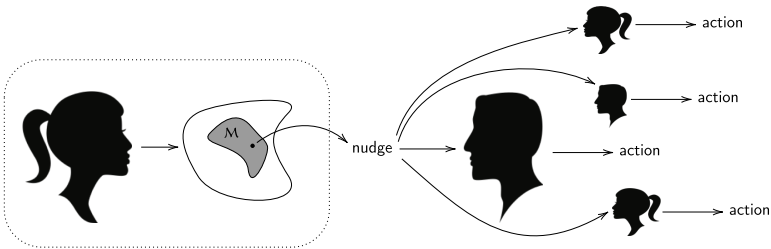
Evidence/Beliefs



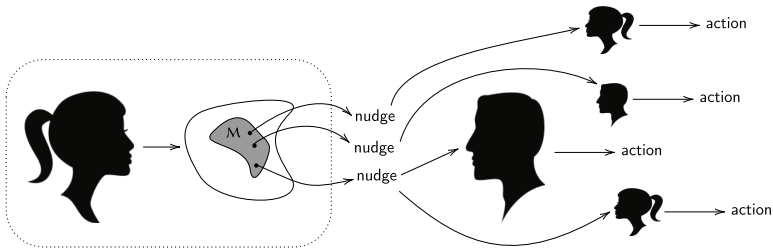
Messages



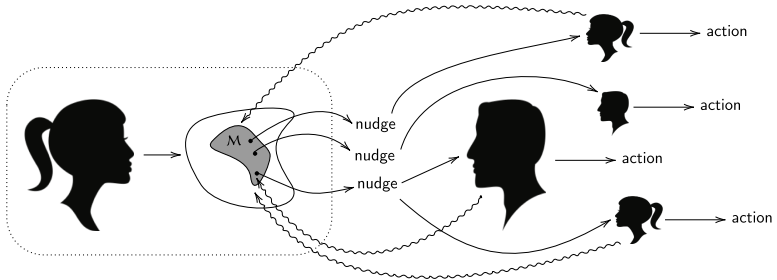
Heterogeneity



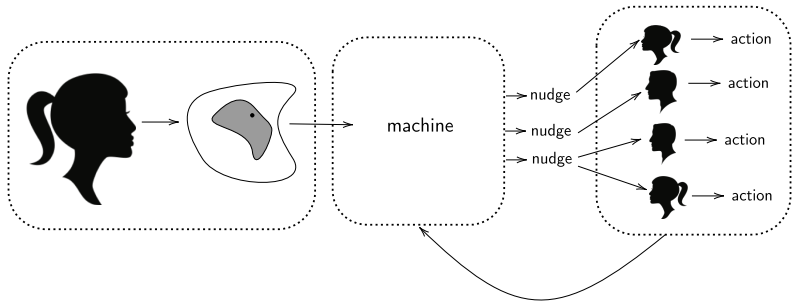
Personalization



Heterogeneity and beliefs



Algorithmic Messages



Algorithm

1. Featurization
   - Construct a mapping from "textual" content to numeric features that is reversable
2. Parameterization
   - Construct a model that related features to choices
3. Estimation
   - Learn parameters of this model
4. Optimization
   - Use parameters to find the optimal message

Control Message (Code for America)

Hi [name], GetCalFresh here! To keep getting CalFresh, fill out your Semi-Annual Report (SAR 7) as soon as you can. It's required even if you have no changes to report. Do it online at: [link]

- Better than status quo by around 13%.(communicated to me)

Featurizing messages

- We borrow ideas from Context free grammars (Chomsky)
- The architect chooses the grammar: $G = \langle \mathcal{F}, V, S, R \rangle$
  - $\mathcal{F}$ is a finite set of terminal symbols (the content)
  - V is a finite set of variables (nonterminals)
  - S are start symbols
  - R is a finite relation $V \to (V \cup \mathcal{F})^*$ (a set of rules)
- Note that

$$\mathbf{F} \in \boldsymbol{\mathcal{F}} = \bigcup_{k \in K} \mathcal{F}_k$$

- with

$$m \Leftrightarrow \mathbf{F} \equiv \{F_1, F_2, \ldots, F_K\}$$

- and

$$F_k \in \mathcal{F}_k$$

- We can then use the optimal $\mathbf{F}^*$ in conjunction with the grammar to define nudges.

Featurizing messages

- Context free grammars (Chomsky)

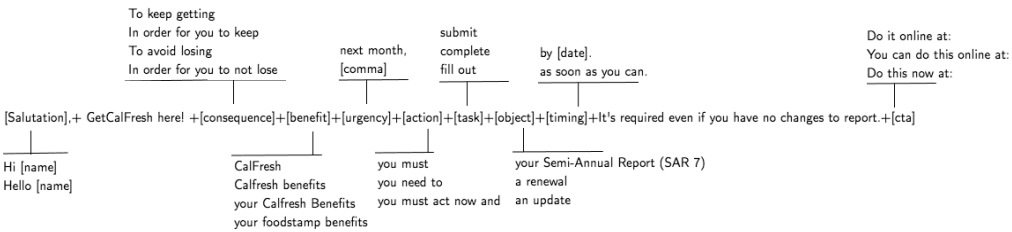
Featurizing messages

- Consider the simple example

Featurizing messages

- Messages have structure or grammar (defined by architect)
- We exploit this grammar to construct features that are modular.
- These feature-sets are bijective to the message
  - 101 : John got on an airplane to Boston
  - 010 : Mary caught a ride to Chicago
- Think conjoint with a bit more structure

Featurization



Model of Behavior

- We will assume the types are captured by a vector of characteristics $(\mathbf{x}_i)$
- We will assume that participants undertake an action (e.g. open emails) if the expected utility/value of doing so is positive.
- And under assumptions some assumptions...

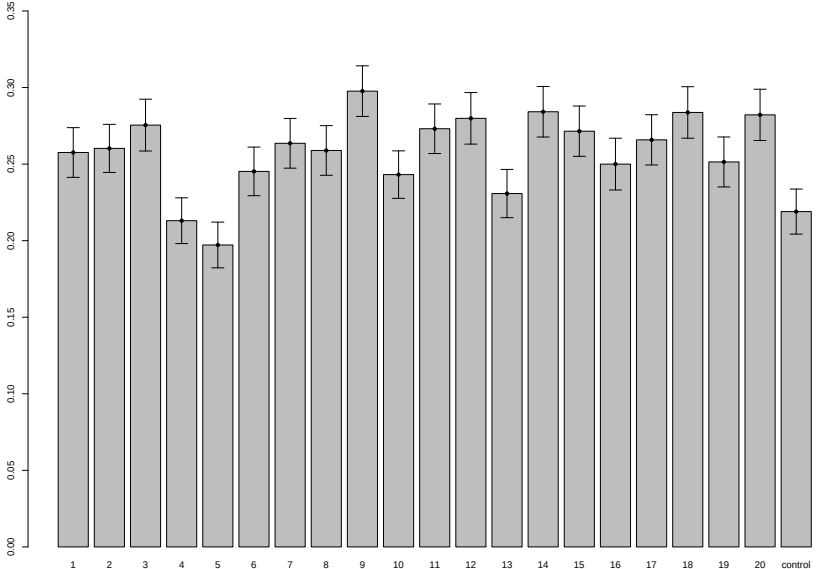
$$P\left(\text{response} = 1 | \mathbf{F}, \mathbf{x}_i\right) = \frac{1}{1 + \exp\left(-\left(\alpha\left(\mathbf{x}_i\right) + \mathbf{F}'\beta\left(\mathbf{x}_i\right)\right)\right)}$$

  - where the β are the sensitivity of a person with characteristics $\mathbf{x}_i$ to message features $\mathbf{F}$

Experiment: Design and Data

- Q: Data? A: Experiments
- The first experiment was conducted in October, 2020
  - Message Space: $\mathcal{M} = \prod_{k=1}^{K} |\mathcal{F}_k| = \mathbf{13824}$
  - Design: $\mathcal{D} = 1 + \sum_{k=1}^{K} (|\mathcal{F}_k| - 1) = \mathbf{19}$
- We chose to test 20 messages ($N \approx 732$ each) + control
  - Chosen by Federov exchange algorithm.
- Total sample was $N = 15380$

Experiment: Results



Results

Feature	$\mathbb{E}[\beta(x)]$	SE	t-stat	95%CI (lower)	95%CI (upper)
Intercept	-1.2903	0.0856	-15.071	-1.4581	-1.1225
Salutation: "Hi [name],"	–				
Salutation: "Hello [name],"	0.0593	0.0360	1.6453	-0.0113	0.1299
Consequence: "To keep getting "	–				
Consequence: "In order for you to keep "	0.1852	0.0447	4.1396	0.0975	0.273
Consequence: "To avoid losing "	0.1643	0.0499	3.2927	0.0665	0.2621
Consequence: "In order for you to not lose "	0.0480	0.0511	0.941	-0.0520	0.1481
Benefit: "CalFresh"	–				
Benefit: "CalFresh benefits"	-0.0069	0.0504	-0.1368	-0.1056	0.0918
Benefit: "your CalFresh benefits"	0.1012	0.0504	2.0073	0.0024	0.2000
Benefit: "your food stamp benefits"	-0.0411	0.0506	-0.8126	-0.1402	0.0580
Urgency: "next month,"	–				
Urgency: ","	-0.1392	0.0360	-3.8677	-0.2097	-0.0686
Action: " you must"	–				
Action: " you need to"	-0.0164	0.0450	-0.3636	-0.1046	0.0719
Action: " you must act now and"	-0.0266	0.0487	-0.5458	-0.1220	0.0688
Action: ""	-0.019	0.0531	-0.3587	-0.1231	0.0850
Task: "submit "	–				
Task: "complete "	0.0925	0.0452	2.0485	0.0040	0.1810
Task: "fill out "	0.1497	0.0449	3.3362	0.0617	0.2376
Object: "your Semi-Annual Report (SAR 7) "	–				

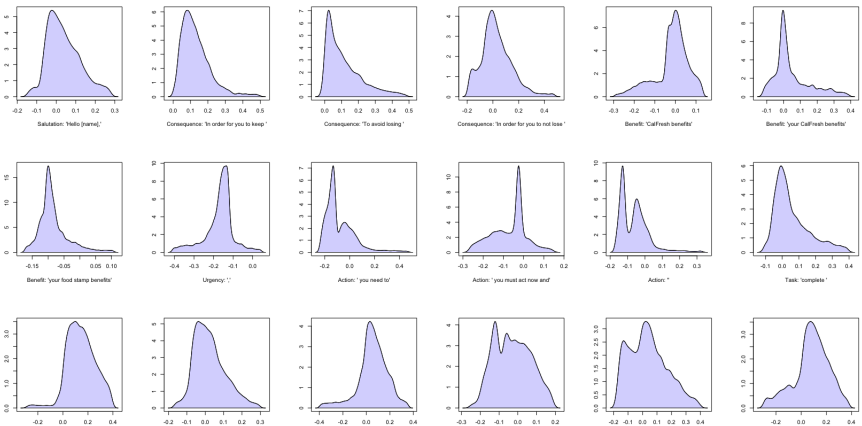
Results [control]

Feature	$\mathbb{E}[\beta(x)]$	SE	t-stat	95%CI (lower)	95%CI (upper)
Intercept	-1.2903	0.0856	-15.071	-1.4581	-1.1225
Salutation: "Hi [name],"	–				
Salutation: "Hello [name],"	0.0593	0.0360	1.6453	-0.0113	0.1299
Consequence: "To keep getting "	–				
Consequence: "In order for you to keep "	0.1852	0.0447	4.1396	0.0975	0.273
Consequence: "To avoid losing "	0.1643	0.0499	3.2927	0.0665	0.2621
Consequence: "In order for you to not lose "	0.0480	0.0511	0.941	-0.0520	0.1481
Benefit: "CalFresh"	–				
Benefit: "CalFresh benefits"	-0.0069	0.0504	-0.1368	-0.1056	0.0918
Benefit: "your CalFresh benefits"	0.1012	0.0504	2.0073	0.0024	0.2000
Benefit: "your food stamp benefits"	-0.0411	0.0506	-0.8126	-0.1402	0.0580
Urgency: "next month,"	–				
Urgency: ","	-0.1392	0.0360	-3.8677	-0.2097	-0.0686
Action: " you must"	–				
Action: " you need to"	-0.0164	0.0450	-0.3636	-0.1046	0.0719
Action: " you must act now and"	-0.0266	0.0487	-0.5458	-0.1220	0.0688
Action: ""	-0.019	0.0531	-0.3587	-0.1231	0.0850
Task: "submit "	–				
Task: "complete "	0.0925	0.0452	2.0485	0.0040	0.1810
Task: "fill out "	0.1497	0.0449	3.3362	0.0617	0.2376
Object: "your Semi-Annual Report (SAR 7) "	–				

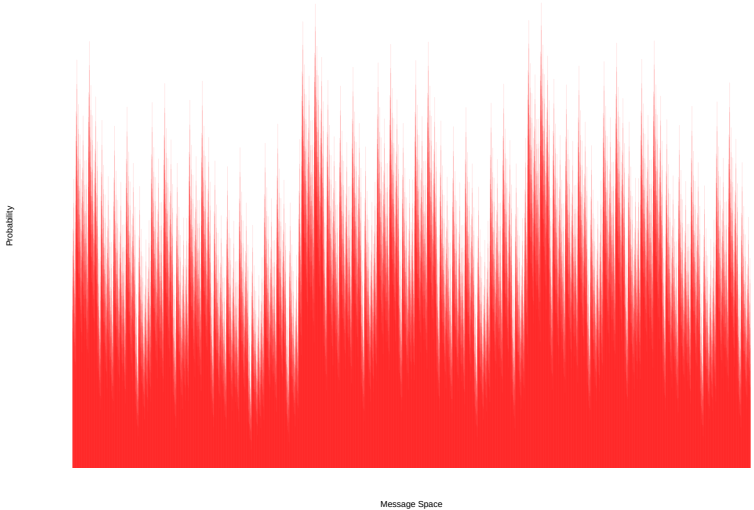
Optimal Uniform

Feature	$\mathbb{E}[\beta(x)]$	SE	t-stat	95%CI (lower)	95%CI (upper)
Intercept	-1.2903	0.0856	-15.071	-1.4581	-1.1225
Salutation: "Hi [name],"	–				
Salutation: "Hello [name],"	0.0593	0.0360	1.6453	-0.0113	0.1299
Consequence: "To keep getting "	–				
Consequence: "In order for you to keep "	0.1852	0.0447	4.1396	0.0975	0.273
Consequence: "To avoid losing "	0.1643	0.0499	3.2927	0.0665	0.2621
Consequence: "In order for you to not lose "	0.0480	0.0511	0.941	-0.0520	0.1481
Benefit: "CalFresh"	–				
Benefit: "CalFresh benefits"	-0.0069	0.0504	-0.1368	-0.1056	0.0918
Benefit: "your CalFresh benefits"	0.1012	0.0504	2.0073	0.0024	0.2000
Benefit: "your food stamp benefits"	-0.0411	0.0506	-0.8126	-0.1402	0.0580
Urgency: "next month,"	–				
Urgency: ","	-0.1392	0.0360	-3.8677	-0.2097	-0.0686
Action: " you must"	–				
Action: " you need to"	-0.0164	0.0450	-0.3636	-0.1046	0.0719
Action: " you must act now and"	-0.0266	0.0487	-0.5458	-0.1220	0.0688
Action: ""	-0.019	0.0531	-0.3587	-0.1231	0.0850
Task: "submit "	–				
Task: "complete "	0.0925	0.0452	2.0485	0.0040	0.1810
Task: "fill out "	0.1497	0.0449	3.3362	0.0617	0.2376
Object: "your Semi-Annual Report (SAR 7) "	–				

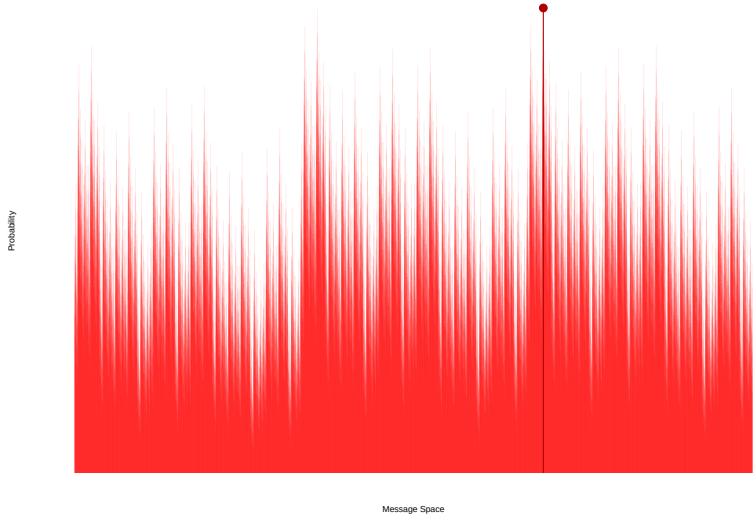
Heterogeneity



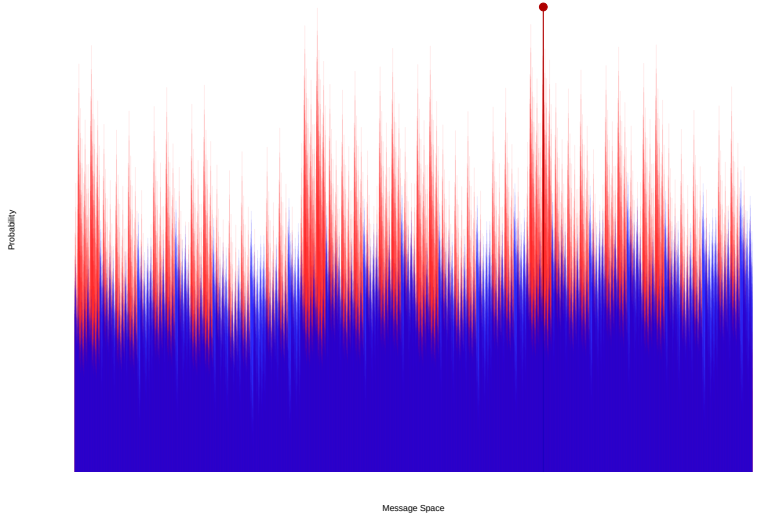
Heterogeneity



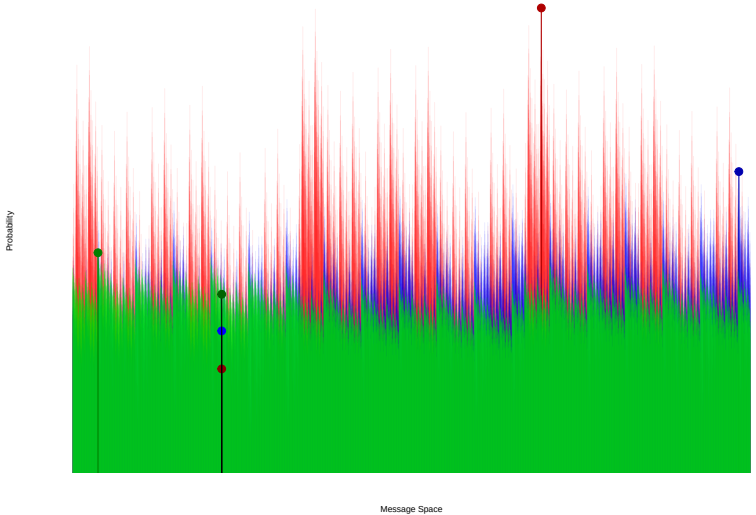
Heterogeneity



Heterogeneity



Heterogeneity



November Validation Experiment

- We implemented a second experiment ($N = 7773$) in November 2020.(*)
- We allocated individuals to four arms
  - Algorithmic (50%)
  - Control (10%)
  - Random (30%)
  - Optimal (10%)
- Messages were constructed based on the parameters estimated.
  - 252 Unique algorithmic messages

Predictions

	(Raw) Prediction
Algorithmic	0.4654
Control	0.2373
Optimal	0.3557
Random	0.2678

- Predictions are aggressive
- Do not account for noise/uncertainty
- Election day!!!
- Distribution of recipients not in our control.
- Doomsday Validation.

Validation

	(Raw) Prediction	Actual
Algorithmic	0.4654	0.2756
Control	0.2373	0.2346
Optimal	0.3557	0.2593
Random	0.2678	0.2559

- **Validated improvement from control to algorithmic is 17.47%!**

Valuing Algorithmic Nudges

- **Idea:** Mullainathan and Shafir (2013) articulate a "bandwidth tax" that suggests that poor individuals more likely to fail to undertake high value activities.
- Estimator using validation (November)

$$\Delta = \left[\frac{\sum_{i \in T_{\text{alg}}} \widehat{B}_i \times 1\left[t_i = \text{alg}\right]}{\Pr\left(t_i = \text{alg}\right)} \right] - \left[\frac{\sum_{i \in T_{\text{control}}} \widehat{B}_i \times 1\left[t_i = \text{control}\right]}{\Pr\left(t_i = \text{control}\right)} \right]$$

- $\widehat{B}_i$ taken as max SNAP benefit conditional on family size.

$$\$\Delta = \$193,503.8/\text{month}$$

- Assuming they stay on SNAP for next **6 months** and using **12 cohorts** we get

$$\$\Delta_{\textbf{year}} \approx \$13.93 \text{ million}$$

- **For** $(\textbf{Income} = 0)$

$$\$\Delta_{\textbf{year}} \approx \$7.09 \textbf{ million}$$

Valuing Algorithmic Nudges

- To scale this consider that there are 441,000,000 SNAP recipients in the us anually
- So the total incremental disbursement from algorithmic nudges will be

$$\$\Delta_{\text{year}} \approx \$12.39 \textbf{ Billion}$$

- Thats about 11% of disbursements!

Summary

- We have proposed a simple, scalable approach to inforporating flexible heterogeneity in structural models.
- The framework provides *automatic* inference for a large class of economic constructs.
- The implementation uses standard software and a package is forthcoming.