Economic AI

Matt Taddy – Microsoft and UChicago

Economic AI breaks complex systemic questions into sets of prediction tasks

What do economists do?

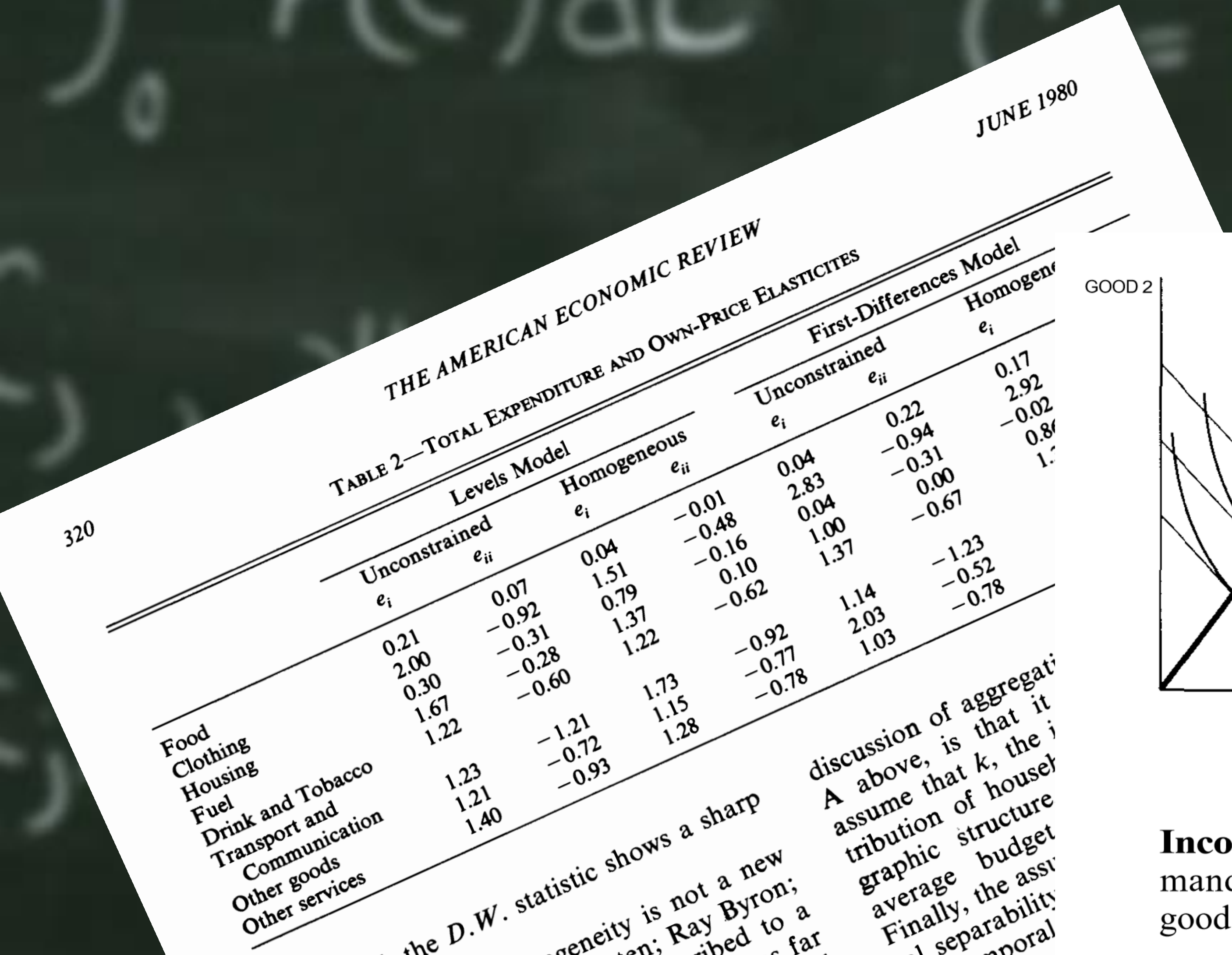
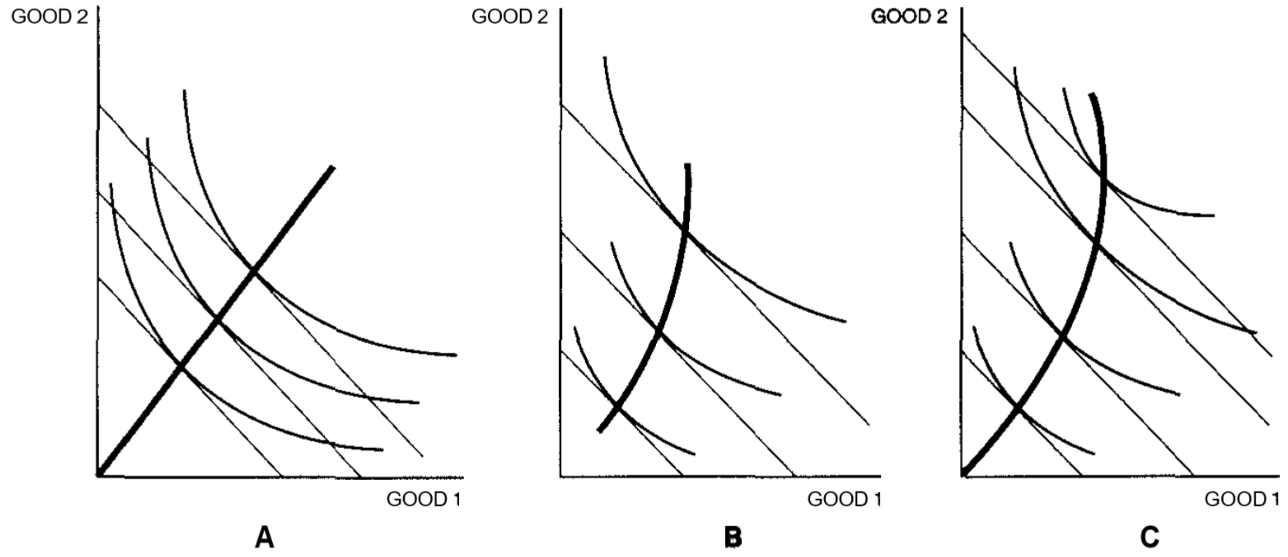


Table 2—Total Expenditure and Own-Price Elasticities

	Levels Model Unconstrained e_i	e_{ii}	Homogeneous e_i	e_{ii}	First-Differences Model Unconstrained e_i	e_{ii}	Homogene[illegible] e_i
Food	0.21	0.07	0.04	−0.01	0.04	0.22	0.17
Clothing	2.00	−0.92	1.51	−0.48	2.83	−0.94	2.92
Housing	0.30	−0.31	0.79	−0.16	0.04	−0.31	−0.02
Fuel	1.67	−0.28	1.37	0.10	1.00	0.00	[illegible]
Drink and Tobacco	1.22	−0.60	1.22	−0.62	1.37	−0.67	[illegible]
Transport and Communication	1.23	−1.21	1.73	−0.92	1.14	−1.23	
Other goods	1.21	−0.72	1.15	−0.77	2.03	−0.52	
Other services	1.40	−0.93	1.28	−0.78	1.03	−0.78	

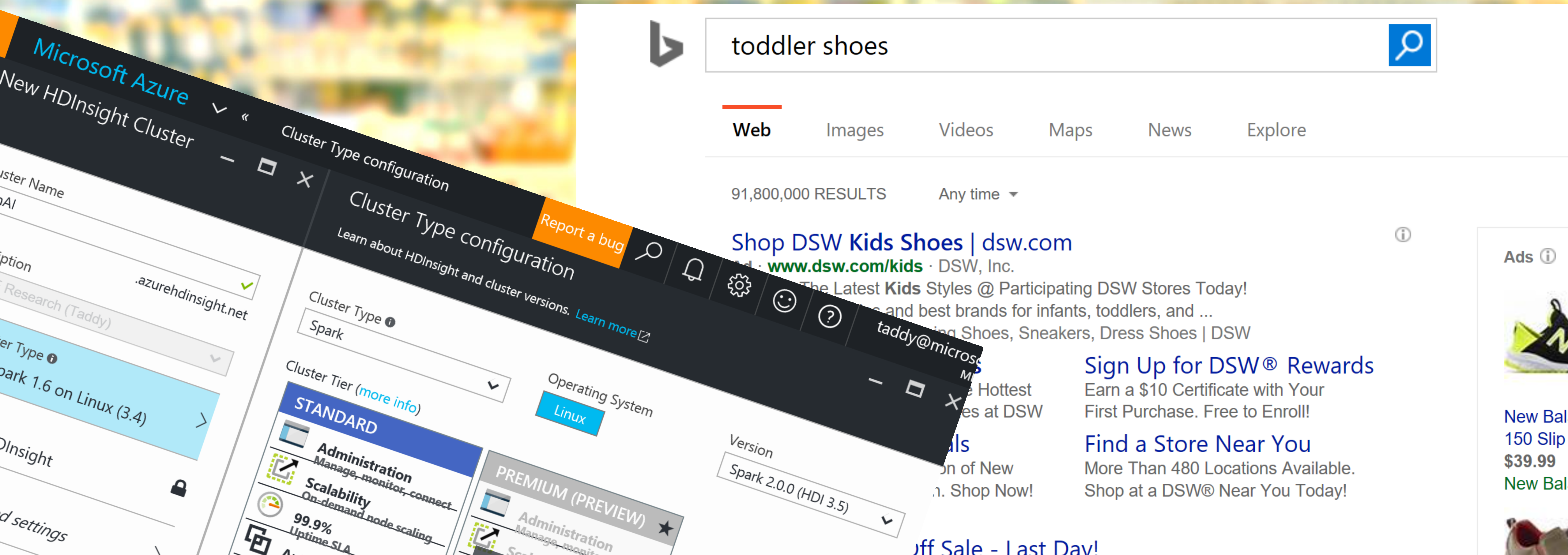
the *D.W.* statistic shows a sharp
geneity is not a new
ten; Ray Byron;
ibed to a
f far

discussion of aggregati
A above, is that it
assume that *k*, the i
tribution of househ
graphic structure
average budget
Finally, the assu
l separability
mporal



Income expansion paths. Panel **A** depicts unit elastic demands, in panel B good 2 is a luxury good, and in panel C, good 1 is an inferior good.

And what are they doing today?



Applied econometrics (via experimentation)

Example

- Question: what is the impact of sponsored search ads on revenue?
- Confound: revenue changes in time with other known and unknown factors
- Experiment: do an 'AB test', randomly turning off ads for certain users/markets

Limitations

- Expensive and politically difficult to run [big/long] experiments
- Design and analysis still requires high level of sophistication

Applied econometrics (mostly harmless version)

Example

- Question: what is the impact of going to charter school on college success?
- Confound: students who seek charter schools are different *to begin with*
- Experiment: compare students with high and low scores in enrollment lottery

Limitations

- Requires a high level of sophistication and a lot of luck
- Too cute: these natural experiments occur in special settings

Applied econometrics (may be hazardous)

Example

- Question: what is the sensitivity of consumers to prices (demand curve)?
- Confound: prices are set in response to consumer demand
- Experiment: compare transactions that match on observables (same demand info)

Limitations

- Results are very sensitive to the model specification
- Selection of the control variables is subjective and hugely labor intensive

Machine Learning can automate and accelerate tasks in these applied econometric workflows

Example: short-term price sensitivity

If I **drop** price 1%, by what % will quantity sold **increase?**
Ex. -3 ⇒ drop price 1%, quantity sold goes up 3%

Problem: both prices and sales respond to underlying demand
Need a causal effect of price on sales, not their co-movement

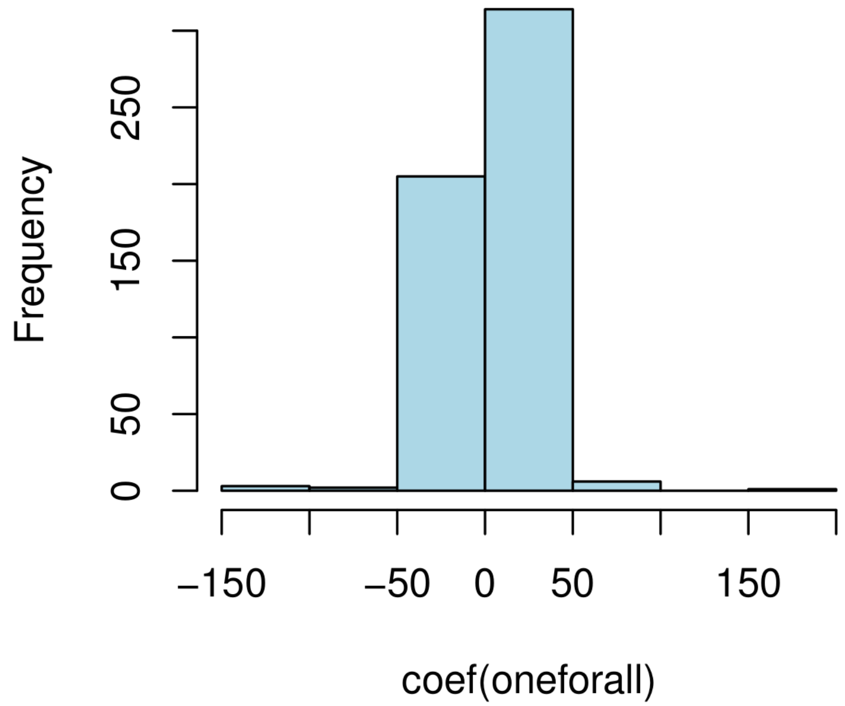
Beer Elasticity

```
> beer <- read.csv("smallbeer.csv",
+      colClasses=c(rep("factor",3),rep("numeric",2)))
> ( allforone <- lm(log(units) ~ log(price), data=beer) )

Call:
lm(formula = log(units) ~ log(price), data = beer)

Coefficients:
(Intercept)   log(price)
     1.3499      -0.2346

> oneforall <- lm(log(units) ~ log(price)*item, data=beer)
```



A single shared elasticity gives tiny -0.23

Separate elasticity for each gives wildly noisy zeros

Beer Elasticity

We need to group the products together using brand, pack, etc.
Quick: featurize the products from their text description.

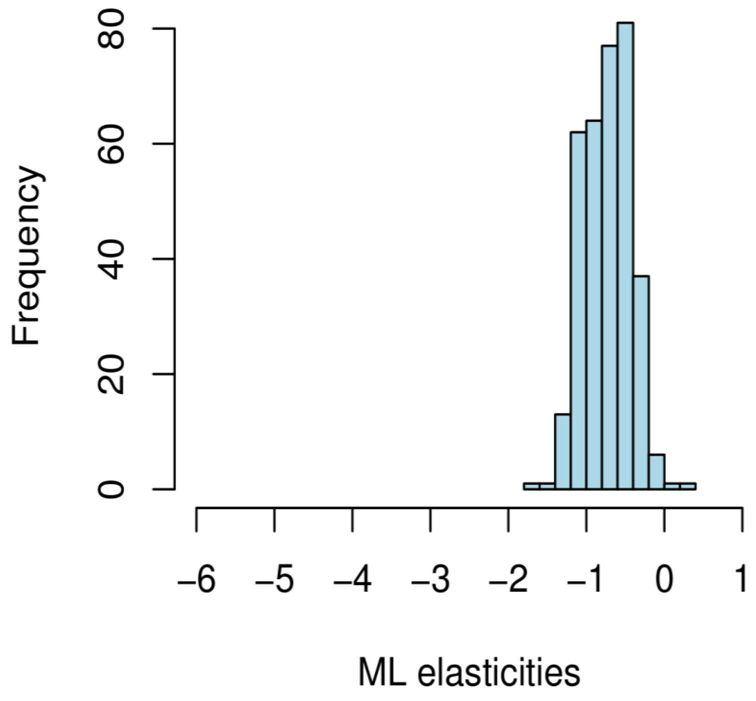
Say $w_{ik} = 1$ if word k is in description for beer i, then

$$\log y_i = \alpha_i + \delta_t + \boldsymbol{w}_i'\boldsymbol{\tau} + (\rho + \boldsymbol{w}_i'\boldsymbol{\gamma})\log p_i$$

Now we've got a large number of parameters. Just throw it all in a lasso?

Beer Elasticity

```
> xx <- cBind(xweek, xitem, xtext)
> xtreat <- cBind(1,xtext,xweek)
> dim(xx)
[1] 7969  827
> dim(xtreat)
[1] 7969  484
> naiveml <- gamlr(x=cBind(xtreat*log(beer$price), xx),
+                  y=log(beer$units),
+                  free=1, standardize=FALSE)
```



Now the elasticities are all unbelievably small

Our problem: this is not a pure prediction problem

Orthogonal Machine Learning

This naïve ML conflates two problems:

- selecting controls and predicting the response conditional upon controls.

Instead, Orthogonal ML

- Estimate nuisance functions that are orthogonal to γ in its conditional score.
- Then estimation for γ is robust to slow learning on these nuisance functions.

Our analysis is based on ideas in *Chernozhukov et al (2016) Double ML*

This in turn builds on BCH 2013/14, Newey 1994, and even Neyman 1979

Orthogonal ML for Pricing

Price sensitivity estimation breaks into two ML tasks:

1. Predict prices from the demand variables: $\boldsymbol{p} \sim \boldsymbol{x}$
2. Predict sales from the demand variables: $\boldsymbol{y} \sim \boldsymbol{x}$

Plus a final regression:

$$\left(\boldsymbol{y} - \widehat{\boldsymbol{y}}(\boldsymbol{x})\right) \sim \left(\boldsymbol{p} - \widehat{\boldsymbol{p}}(\boldsymbol{x})\right)$$

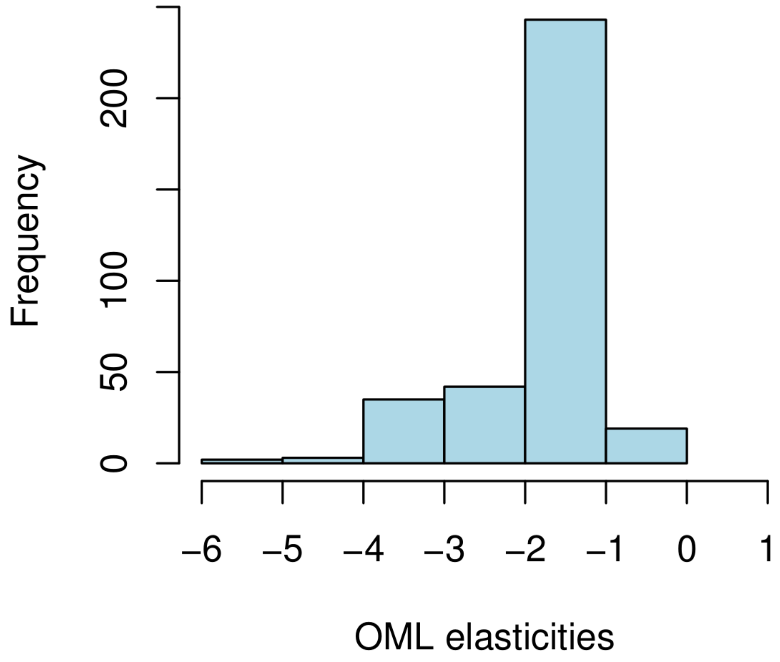
Estimated relationship is causal if $\boldsymbol{x}$ contains all demand info known to pricer

For inference you can data split: use one sample for 1-2, another for step 3

Orthogonal ML for Beer

For the final regression, interact price residuals with *text tokens* and *week*.

```
# OML steps 1-2
pfit <- gamlr(x=xx, y=log(beer$price), lmr=1e-5, standardize=FALSE)
qfit <- gamlr(x=xx, y=log(beer$units), lmr=1e-5, standardize=FALSE)
# Calculate residuals
lpr <- drop(log(beer$price) - predict(pfit, xx))
lqr <- drop(log(beer$units) - predict(qfit, xx))
# Run 3rd ML step to get gammas
ofit <- gamlr(x=(lpr*xtreat), y=lqr, standardize=FALSE, free=1)
gams <- coef(ofit)[-1,]
```



There's no ground truth,

but these elasticities are in the expected range

Orthogonal ML for Beer

The text encodes a natural hierarchy

- Many beers are `IPA` or `Cider` or `Draught`
- But individual brands also load; e.g., `Pyramid` or `Elysian`
- And we find technical terms:

4pk	6pk	12pk	24pk
-0.2	-0.4	0.0	0.3

Most price sensitive

```
> names(sort(el)[1:5])
[1] "GUINNESSS DRAUGHT 6PK BTL
[2] "GUINNESS DRAUGHT 4PK CAN
[3] "PYRAMID OUTBURST IMP IPA 6PK
[4] "ELYSIAN IMPORTAL IPA 6PK
[5] "PYRAMID OUTBURST IMP IPA 12PK
```

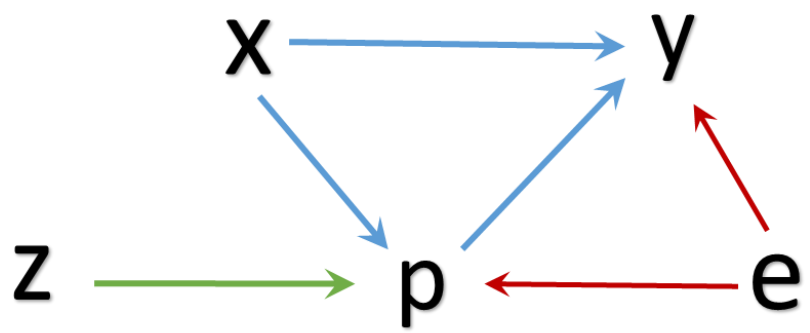
Least price sensitive

```
> names(sort(-el)[1:5])
[1] "2 TOWNS CRISP APPLE CIDER
[2] "2 TOWNS BAD APPLE CIDER
[3] "ATLAS BLKBRY APPLE CIDER
[4] "D'S WICKED BAKED APPLE CIDER
[5] "D'S WICKED GREEN APPLE CIDER
```

Econ + ML

This is what econometricians do: they break systems into measurable pieces

Another common example: Instrumental Variables



Ex. 2SLS: regress $p \approx z\tau$ then $y \approx (z\hat{\tau})\gamma$

Instead of OLS, we can break into two ML tasks

Deep IV

IV *exclusion structure* implies $\mathbb{E}[y|x,z] = \int g(p,x)dF(p|x,z)$

(e.g., Newey + Powell 2003)

Use arbitrary ML to learn $\hat{F}$, then solve

$$\min_{g \in G} \sum \left(y_i - \int g(p, x_i) d\hat{F}(p|x_i, z_i) \right)^2$$

See Hartford/Lewis/Leyton-Brown/Taddy ICML 2017

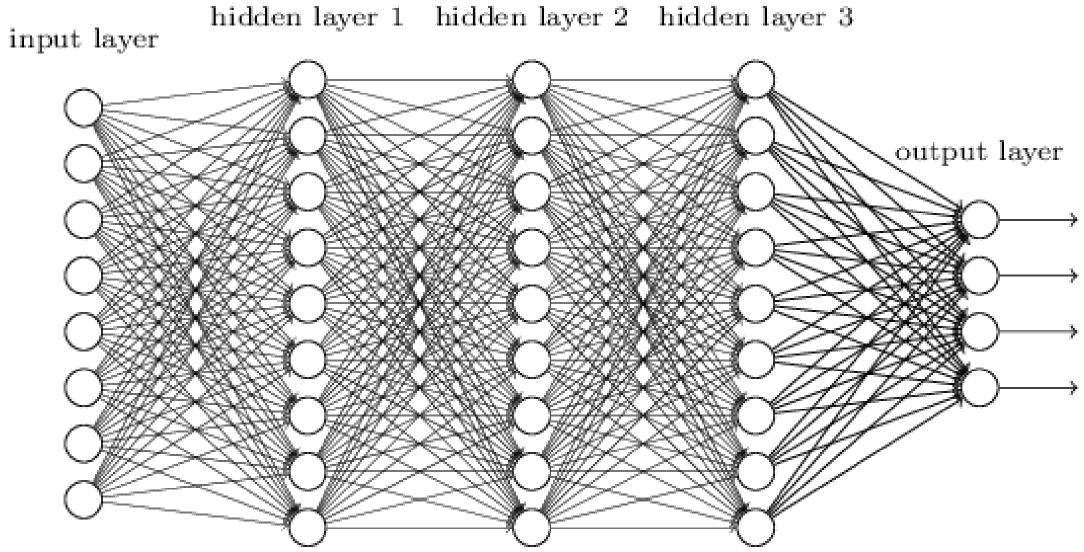
Deep Neural Networks

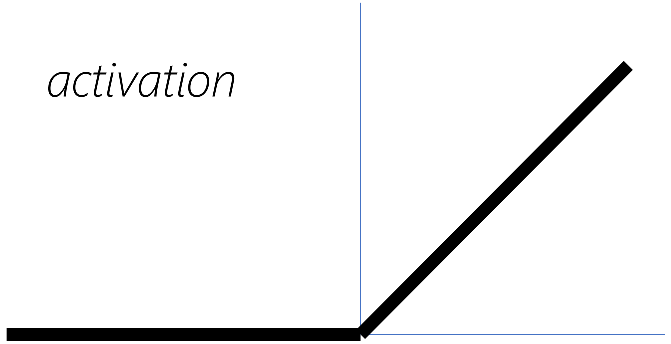
A massive number of parameters, mapping output of each layer to each node activation in the next layer

Fit with Stochastic Gradient Descent

Regularize with

- deviance penalties $\lambda||W||$
- dropout training (zeros in grad)





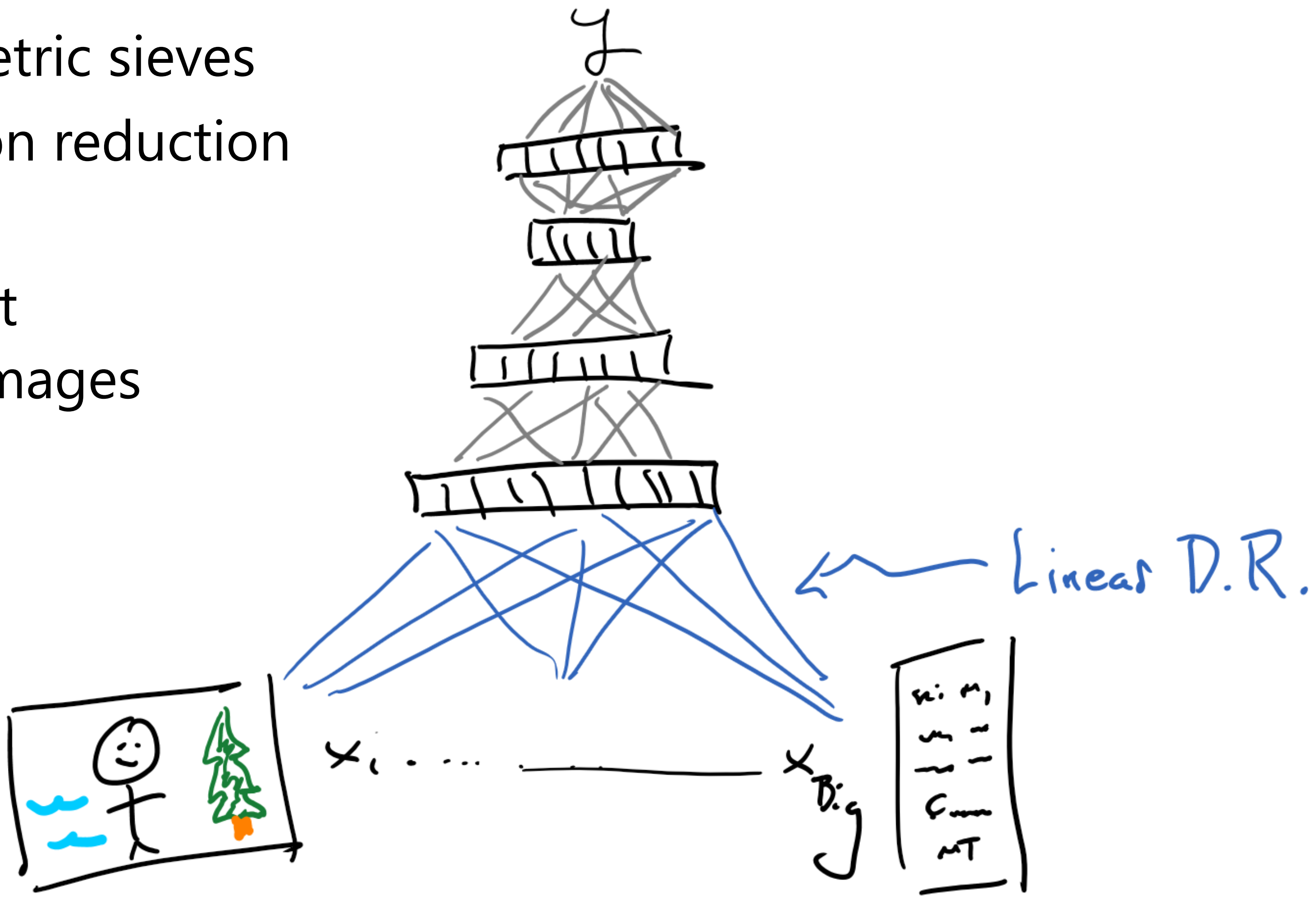
Deep Neural Networks

Deep nets are not nonparametric sieves

The 1st layer is a big dimension reduction

For example,

- word embedding for text
- matrix convolution for images



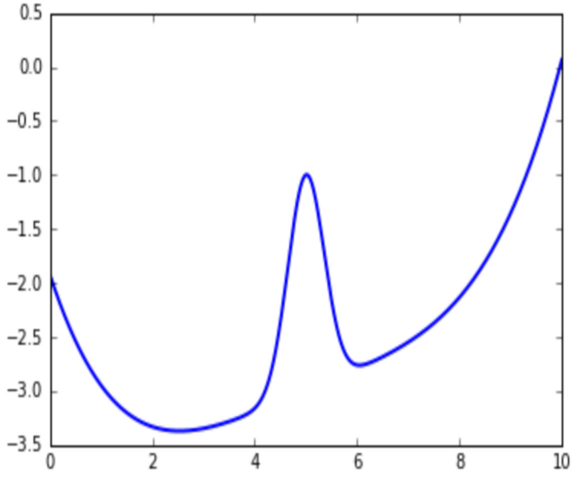
A pricing simulation

$$y = 100 + s\psi_t + (\psi_t - 2)p + e,$$

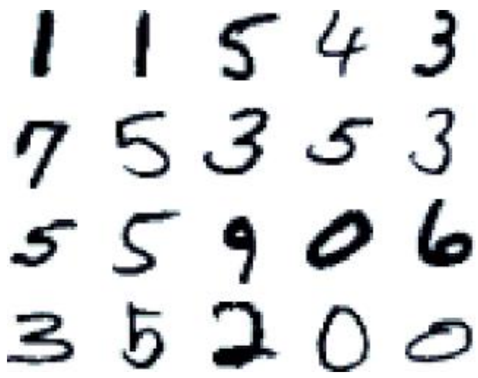
$$p = 25 + (z+3)\psi_t + v$$

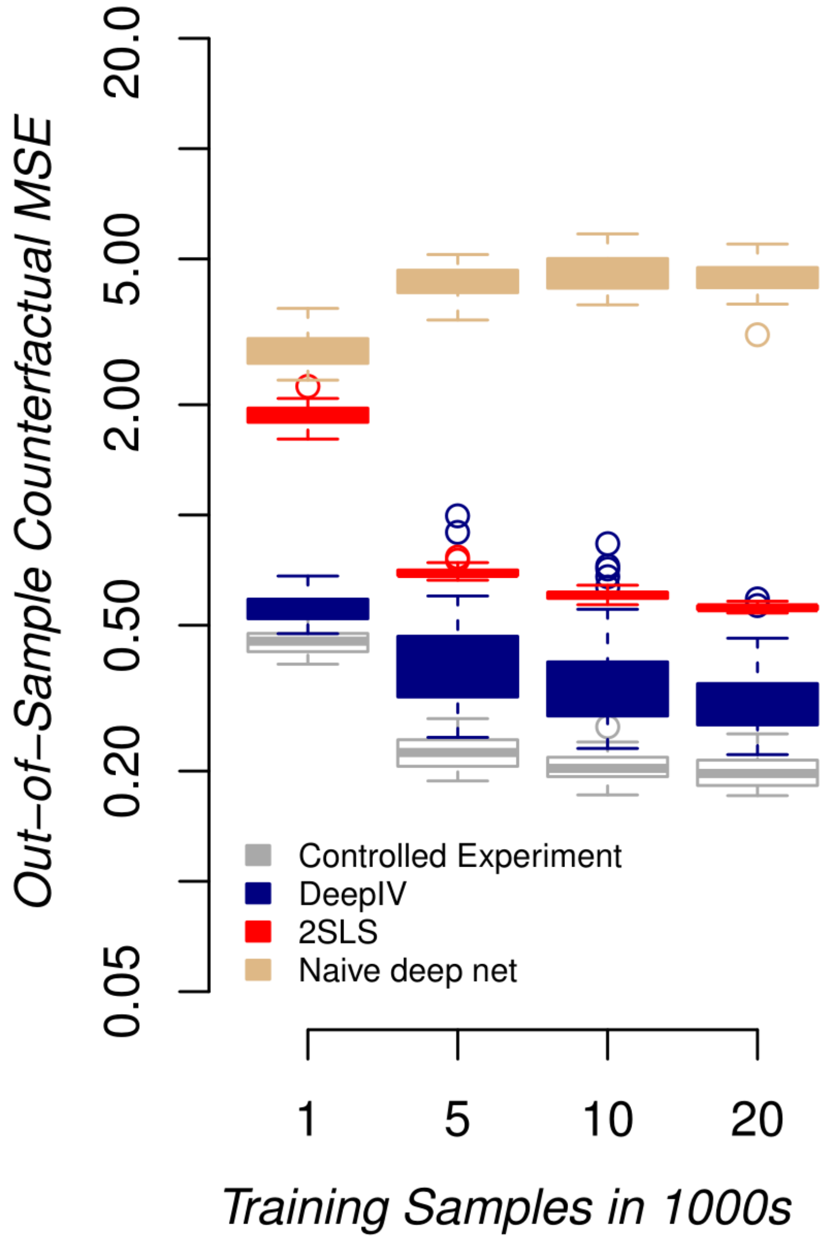
$$z,\ v \sim \mathrm{N}(0,1) \ \text{ and } \ e \sim \mathrm{N}(\rho v, 1-\rho^2),$$

Time-dependent ψ_t



Customer type 's'





Ads Application

Taken from Goldman and Rao (2014)

We have 74 mil click-rates over 4 hour increments for 10k search terms

Treatment: ad position 1-3

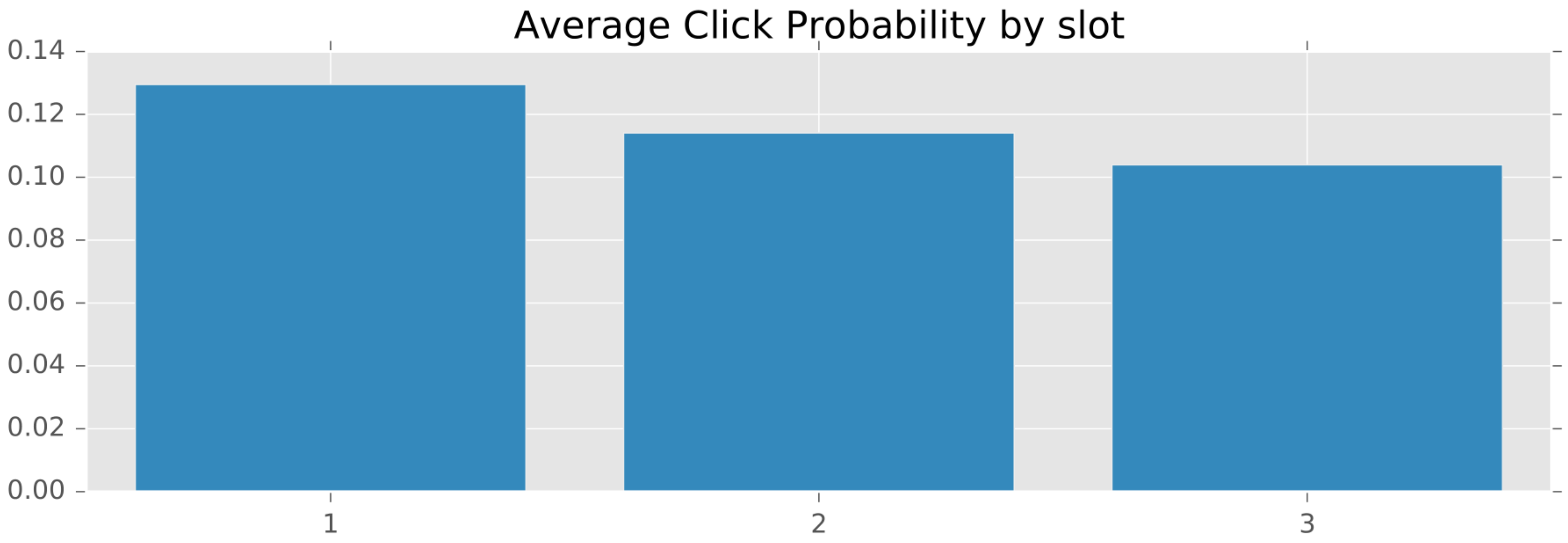
Instrument: background AB testing (bench of ~ 100 tests)

Covariates: advertiser id and ad properties, search text, time period

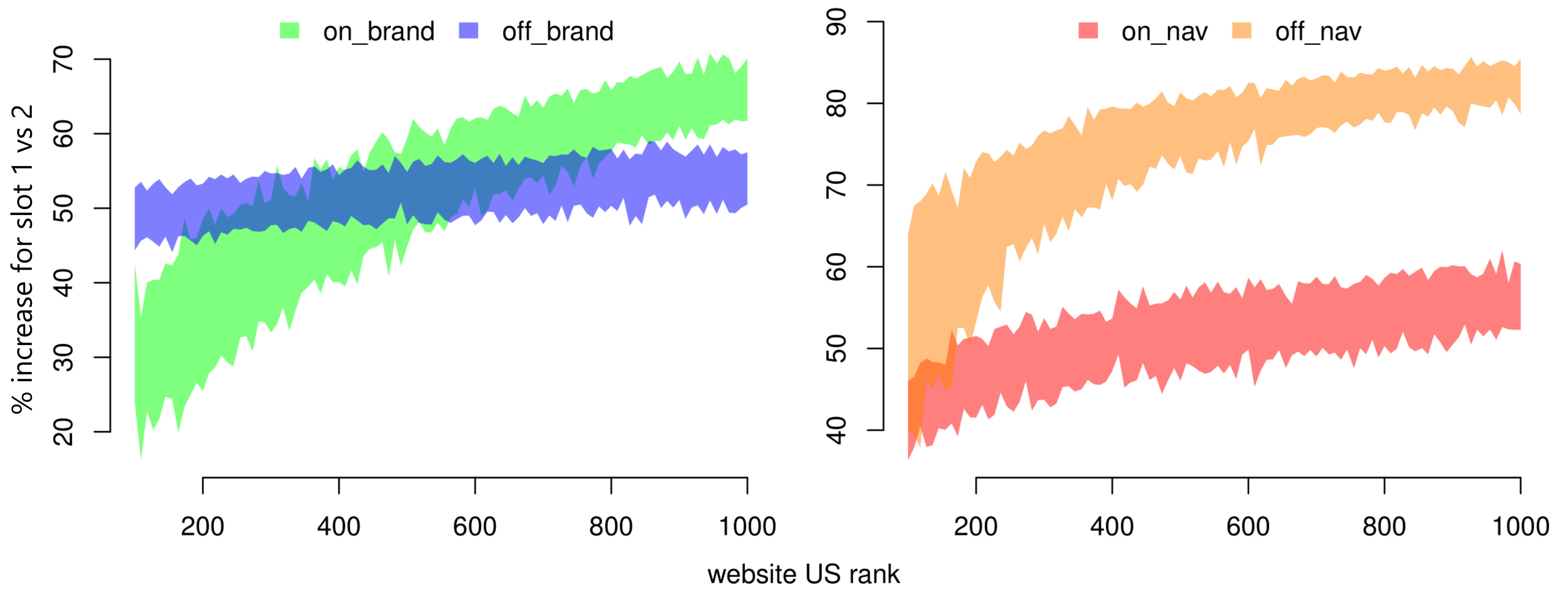
Ads in slot 1 are inherently different (better?) than those in slot 2

We need causal inference for the effect of position on clicks

Average Treatment Effects



Compare to observed click probabilities of 0.33, 0.1, and 0.05.



Fits heterogeneity across advertiser and search (automation of Goldman + Rao)

Economic AI

The ML doesn't create new economic insights or replace economists

It automates and accelerates subjective labor-intense measurement

Instruments are everywhere inside firms

If we push reinforcement learning there will be even more

Reduced form econometrics is low fruit; structural econometrics is next

Social Scientific AI

Deep learning revolution: good low-dev-cost off-the-shelf ML
As the tools become plug-n-play, teams get interdisciplinary
The next big gains in AI are coming from domain context

Use domain structure to break questions into ML problems
Don't re-learn things you already know with an AI baby