Today's Focus

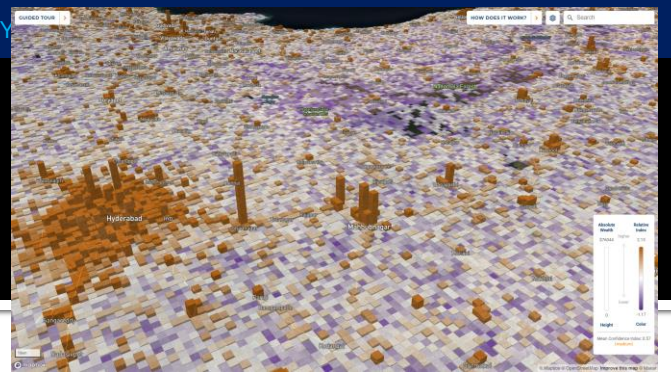
- How can ML help answer casual questions?
  - I'll highlight a few promising areas
  - But we're still figuring this out!

  - Great background: "Machine Learning Methods Economists Should Know About", by Susan Athey and Guido Imbens. 2019

Outline

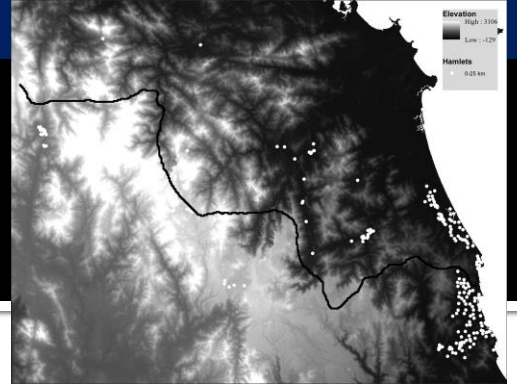
- **ML for measurement**
- Inference after selection
- Selecting among many controls
- Selecting among many instruments
- Heterogeneous treatment effects
- Other topics

ML for measurement



- One way ML can help answer causal questions is by making it possible (or easier) to observe a structural variable that was previously difficult to observe
- Version A: "Same question, same identification strategy, **new measurement strategy**"
  - e.g., "predicting poverty" lecture provides a method for measuring wealth in data-poor environments
    - Wealth is one of the most common/important "LHS variables"
    - Wealth is an important "RHS variable" or interaction/heterogeneity
      - Blumenstock et al., 2016. Airtime Transfers and Mobile Communications: Evidence in the Aftermath of Natural Disasters. *Journal of Development Economics*
  - Note: better measurement doesn't imply causality!!!

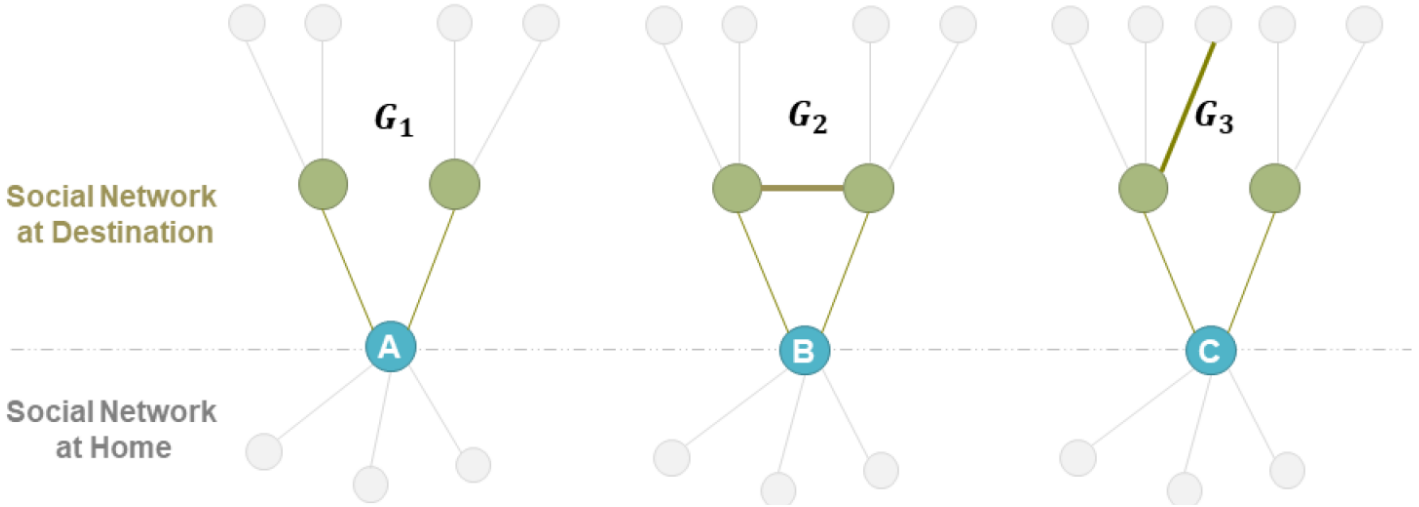
ML for measurement



- Version B: "Same question, **new identification strategy**, new measurement strategy"
  - Better measurement can facilitate new approaches to causal identification for long-standing questions
    - Spatial regression discontinuity
      - Donaldson, D., Storeygard, A., 2016. The View from Above: Applications of Satellite Data in Economics. *JEP*
      - Dell, M., Querubin, P., 2018. Nation building through foreign intervention: Evidence from discontinuities in military strategies. *QJE*
    - Event studies and temporal regression discontinuity
      - Xu, W., Li, Z., Cheng, C., Zheng, T., 2013. Data mining for unemployment rate prediction using search engine query data. *Service Oriented Computing and Applications*
      - Hall, A.S., 2018. Machine learning approaches to macroeconomic forecasting. *The Federal Reserve Bank of Kansas City Economic Review*

ML for measurement

- Version C: "**New question**, new identification strategy, new measurement strategy"
  - The combination of new measurement and new identification can make it possible to empirically analyze questions that were historically intractable
    - Blumenstock, J.E., Chi, G., Tan, X., 2020. Migration and the Value of Social Networks.



Outline

- ML for measurement
- **Inference after selection**
- Selecting among many controls
- Selecting among many instruments
- Heterogeneous treatment effects
- Other topics

ML for causal inference

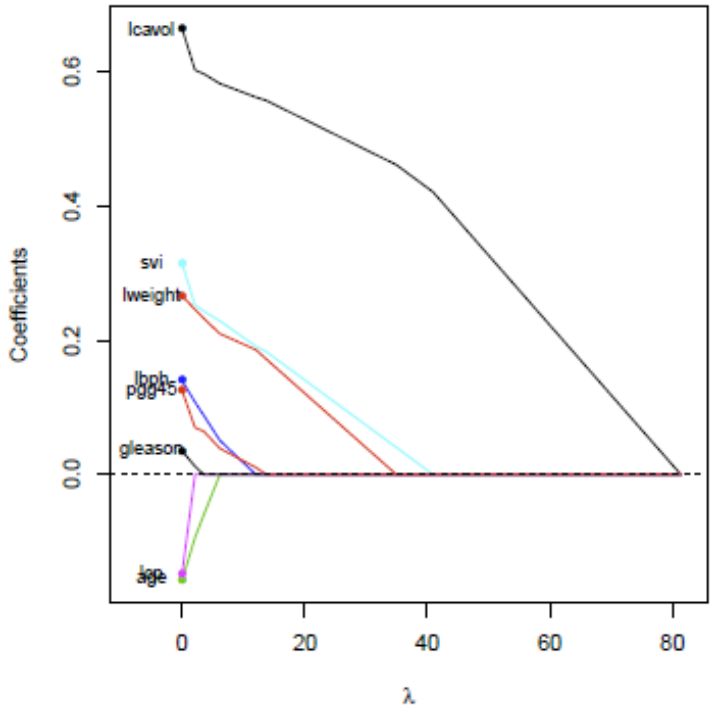
- How do econometricians describe ML?
  - The term "data mining" in a modern sense denotes a principled search for "true" predictive power that guards against false discovery and overfitting…
  - …does not erroneously equate in-sample fit to out-of-sample predictive ability…
  - and accurately accounts for using the same data to examine many different hypotheses or models.
    - Belloni, Chernozhukov, Hansen (2014 *JEP*)

Inference after selection

- Recall LASSO

$$J(\theta) = \frac{1}{2N}\sum_{i=1}^{N}\left(\theta_0 + \theta_1 X_i + \ \ldots + \theta_k X_i^k - Y_i\right)^2 + \lambda \sum_{j=1}^{k} |\theta_j|$$

- How to interpret the θ's?
  - Regularization shrinks them
  - Biased toward zero
  - Penalized by magnitude
    - Not just on zero vs. non-zero



Inference after selection

- LASSO
  - Good method for penalized estimation of the coefficients of a sparse linear model
  - Useful for prediction and forecasting
  - Provides an indication for which variables have a strong association with the particular outcome
  - Without additional steps, bad for inference!

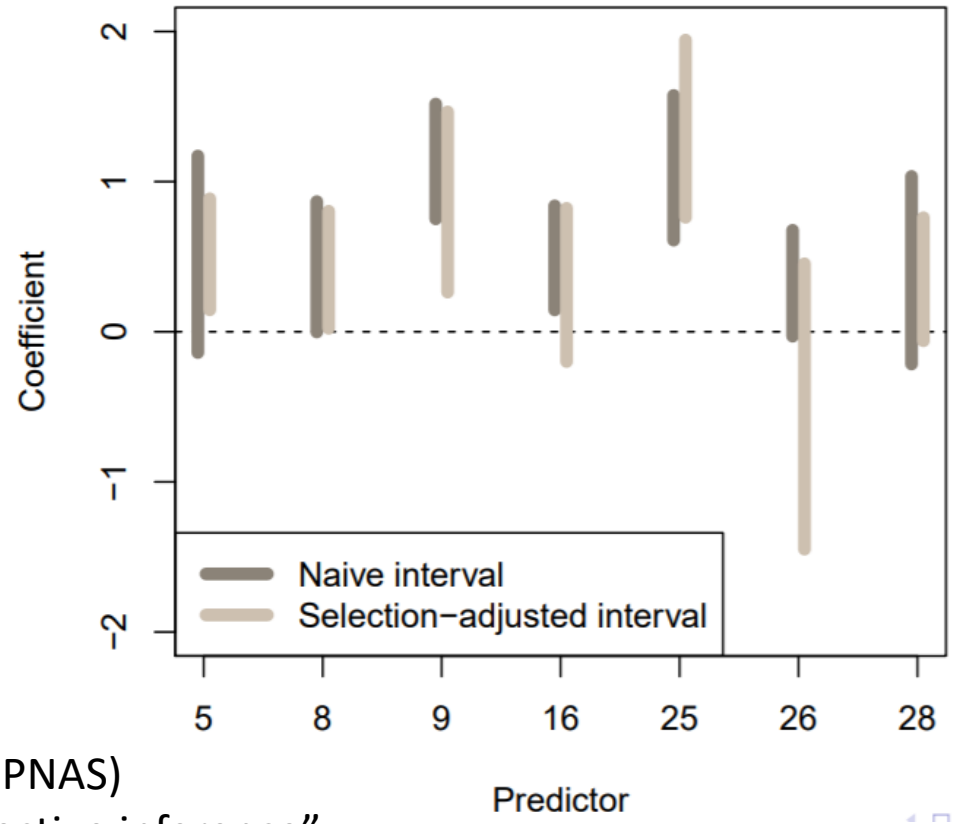
Inference after selection

- "Post LASSO"
  - What if I run LASSO for variable selection, and then run OLS on the remaining coefficients?
  - This is "cheating", since you've already looked at the data -- the resulting p-values and confidence intervals will no longer be valid
    - A "quiet scandal" in the statistical community – Leo Breiman
    - Similar to "p-hacking", but now done algorithmically!
    - Note: Under some conditions this may be okay, in particular, if "the set of variables selected by the lasso is deterministic and non-data dependent with high probability" – see Zhao et al. (2017), "In Defense of the Indefensible: A Very Naive Approach to High-Dimensional Inference"

Inference after selection

- Example

HIV data: mutations that predict response to a drug. Selection intervals for lasso with fixed tuning parameter λ.



Taylor and Tibshirani (2015 PNAS)
"Statistical learning and selective inference"

Inference after selection

- So… What to do?
  - **Data splitting**
  - Data carving
  - Randomized response
  - Exact procedures, polyhedral lemmas
  - Exponential family framework (MCMC)
  - …

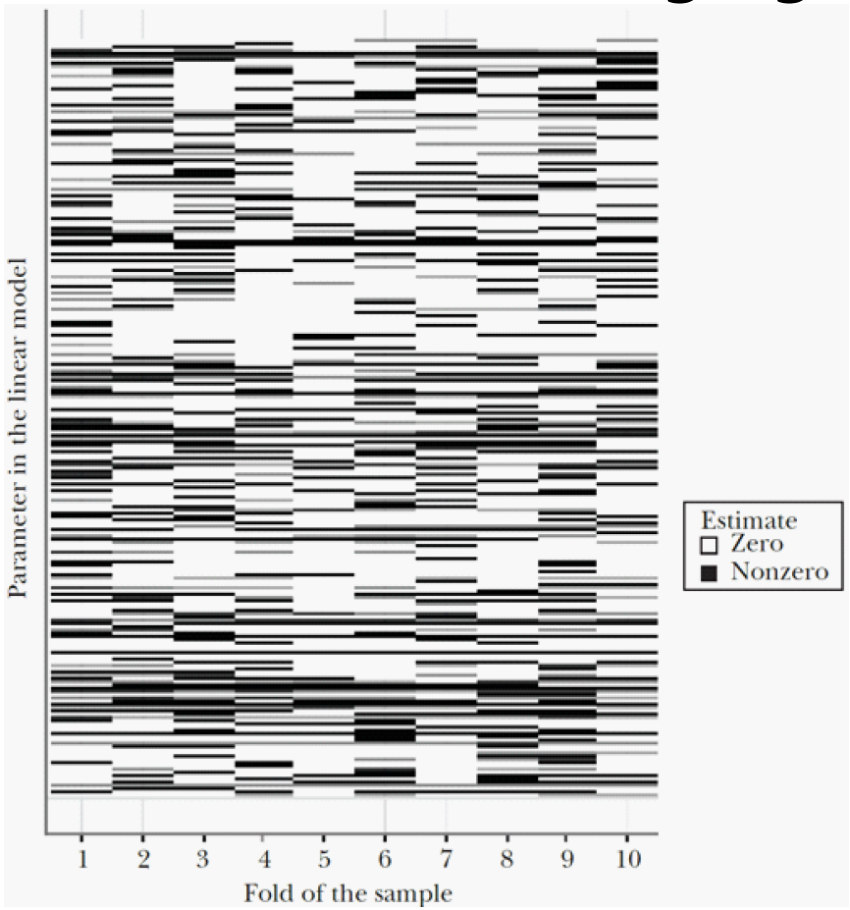
Inference after selection

- Split-sample "Post LASSO"
  1. Using half the data: Use LASSO to select **H** variables from **X**
  2. Using the other half : Run OLS of **Y** on **H**
     - This dates back to Cox, D. (1975). A note on data-splitting for the evaluation of significance levels. *Biometrika*
     - Good review article: Taylor, J. and Tibshirani, R. J. (2015). Statistical learning and selective inference. *PNAS*

  - Advantage: very easy and transparent
  - Disadvantages: you waste data, fitted model may be different between the two halves

Inference after selection

- Example of fitted model changing with folds



From Colin Cameron

Inference after selection

- Key citations:
  - Lee, Jason D., et al. "Exact post-selection inference, with application to the lasso." *The Annals of Statistics* 44.3 (2016): 907-927.
  - Taylor, Jonathan, and Robert J. Tibshirani. "Statistical learning and selective inference." *Proceedings of the National Academy of Sciences* 112.25 (2015): 7629-7634.

Outline

- Motivation
- ML for measurement
- Inference after selection
- **Selecting among many controls**
- Selecting among many instruments
- Heterogeneous treatment effects
- Other topics

Selecting among many controls

- "Selection on observables" design [recap]

$$Y_i = \alpha + \beta T_i + \gamma X_i + \epsilon_i$$

  - We are interested in unbiased estimates of β
  - We assume treatment assignment is exogenous after conditioning on control variables
  - "After controlling for X, T is "as good as random"
  - $E[\epsilon_i | T_i, X_i] = 0$

- Note:
  - This is a strong assumption, but very common

Selecting among many controls

$$Y_i = \alpha + \beta T_i + \gamma X_i + \epsilon_i$$

- Typically, in such models, $k<<N$
- What happens when $k>>N$?
  - Standard OLS will perfectly fit training data
  - LASSO can shrink β (as well as γ) - possibly to zero

Selecting among many controls

$$Y_i = \alpha + \beta T_i + \gamma X_i + \epsilon_i$$

- Why not just exclude β from LASSO penalty?
  - Use LASSO to select **H** variables from **X**
  - Then estimate $Y_i = \alpha + \beta T_i + \delta \boldsymbol{H_i} + \epsilon_i$
  - Issue: LASSO designed for *prediction*, not *inference*
  - Example: what if a variable in X is highly correlated with T?
    - From the standpoint of *prediction*, it should be dropped
    - Can create *omitted variable bias* if real relationship exists ($\gamma \neq 0$)

Selecting among many controls

- Option 1: "Regularized regression adjustment"
  - Assumes that T_i is randomly assigned
  - (Also requires other iid/Gaussian assumptions - see citations)

- Example (Ludwig et al 2018):
  - Use cross-validation to predict Y from X, call this $\hat{f}(X_i)$
  - Control for $\hat{f}(X_i)$: $Y_i = \alpha + \beta T_i + \gamma \hat{f}(X_i) + \epsilon_i$

Selecting among many controls

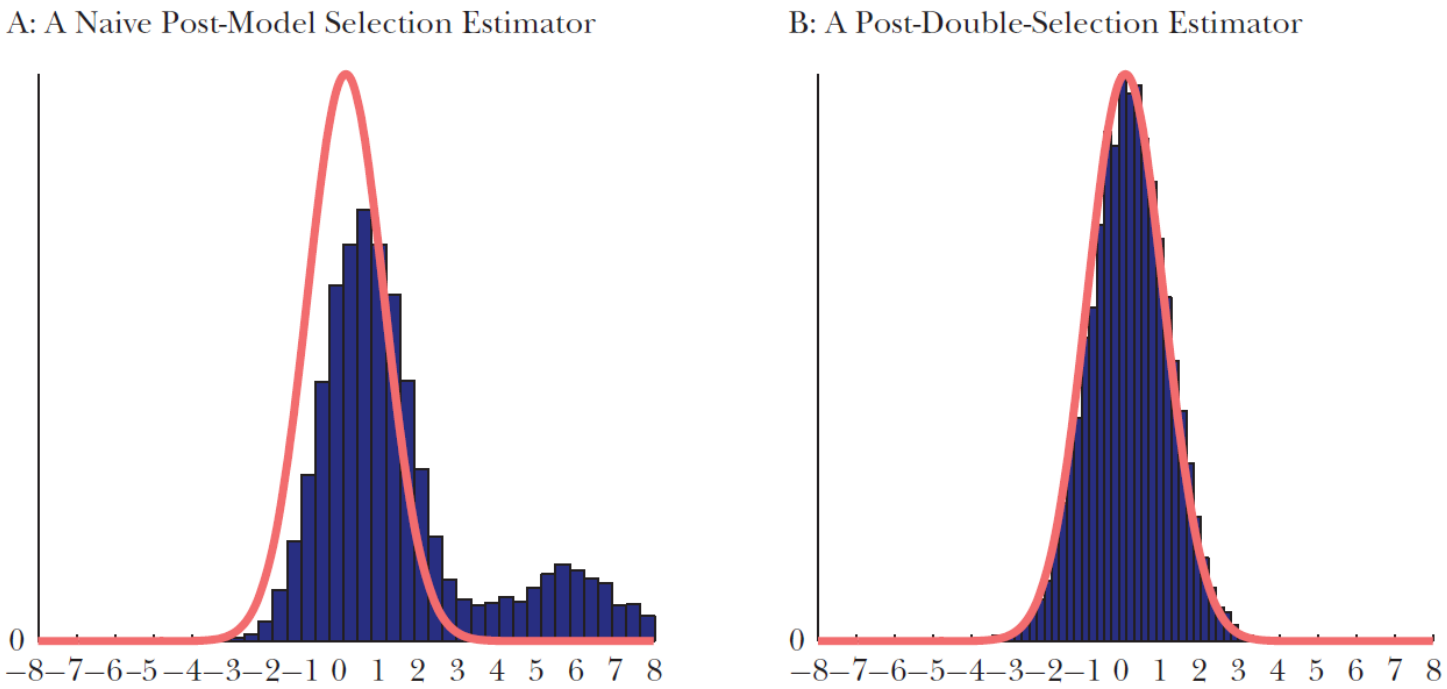
- Key Citations
  - Bloniarz, A., Liu, H., Zhang, C.-H., Sekhon, J.S., Yu, B., 2016. Lasso adjustments of treatment effect estimates in randomized experiments. PNAS 113, 7383–7390.
  - Ludwig, J., Mullainathan, S., Spiess, J., 2019. Augmenting pre-analysis plans with machine learning, in: AEA Papers and Proceedings. pp. 71–76.
  - Wager, S., Du, W., Taylor, J., Tibshirani, R.J., 2016. High-dimensional regression adjustments in randomized experiments. PNAS 113, 12673–12678.

Selecting among many controls

- Option 2: "Double Selection" (aka Post-double selection LASSO):
  1. Use LASSO to select H features from regressing Y on X
  2. Use LASSO to select K features from regressing T on X
     - If T is truly randomly assigned, K will be empty
     - In practice, K is almost never empty
  3. Regress Y on T and the union of H and K

- Key citation: Belloni, Chernozhukov, Hansen (2013 *ReStud*), "Inference onTreatment Effects after Selection amongst High- Dimensional Controls."

Selecting among many controls

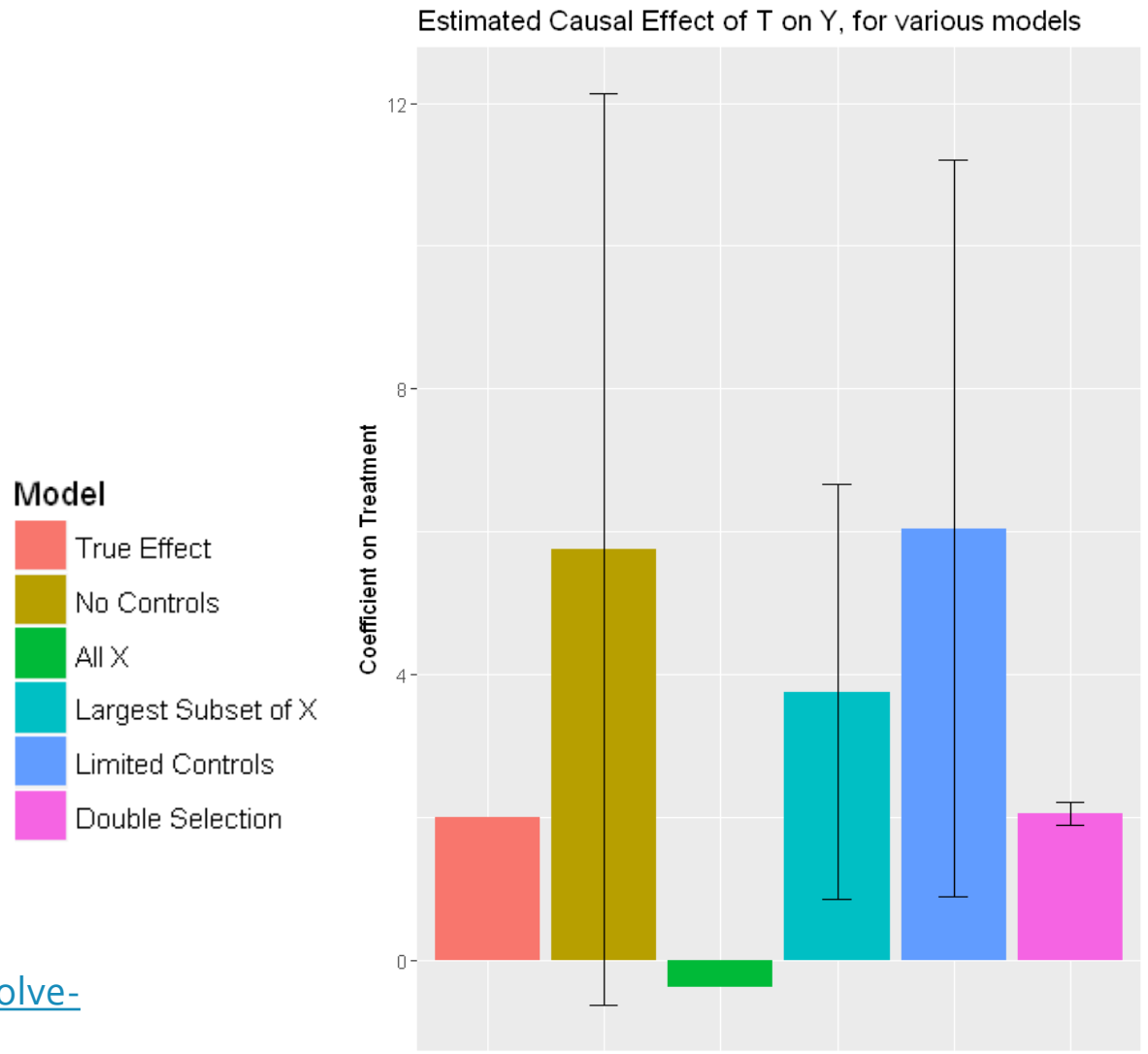


Source: Belloni, Chernozhukov, and Hansen (forthcoming).

Notes: The left panel shows the sampling distribution of the estimator of α based on the first naive procedure described in this section: applying LASSO to the equation $y_i = d_i + x_i' \theta_y + r_{yi} + \zeta_i$ while forcing the treatment variable to remain in the model by excluding α from the LASSO penalty. The right panel shows the sampling distribution of the "double selection" estimator (see text for details) as in Belloni, Chernozhukov, and Hansen (forthcoming). The distributions are given for centered and studentized quantities.

Selecting among many controls

- Simulation



https://medium.com/teconomics-blog/using-ml-to-resolve-experiments-faster-bd8053ff602e

Selecting among many controls

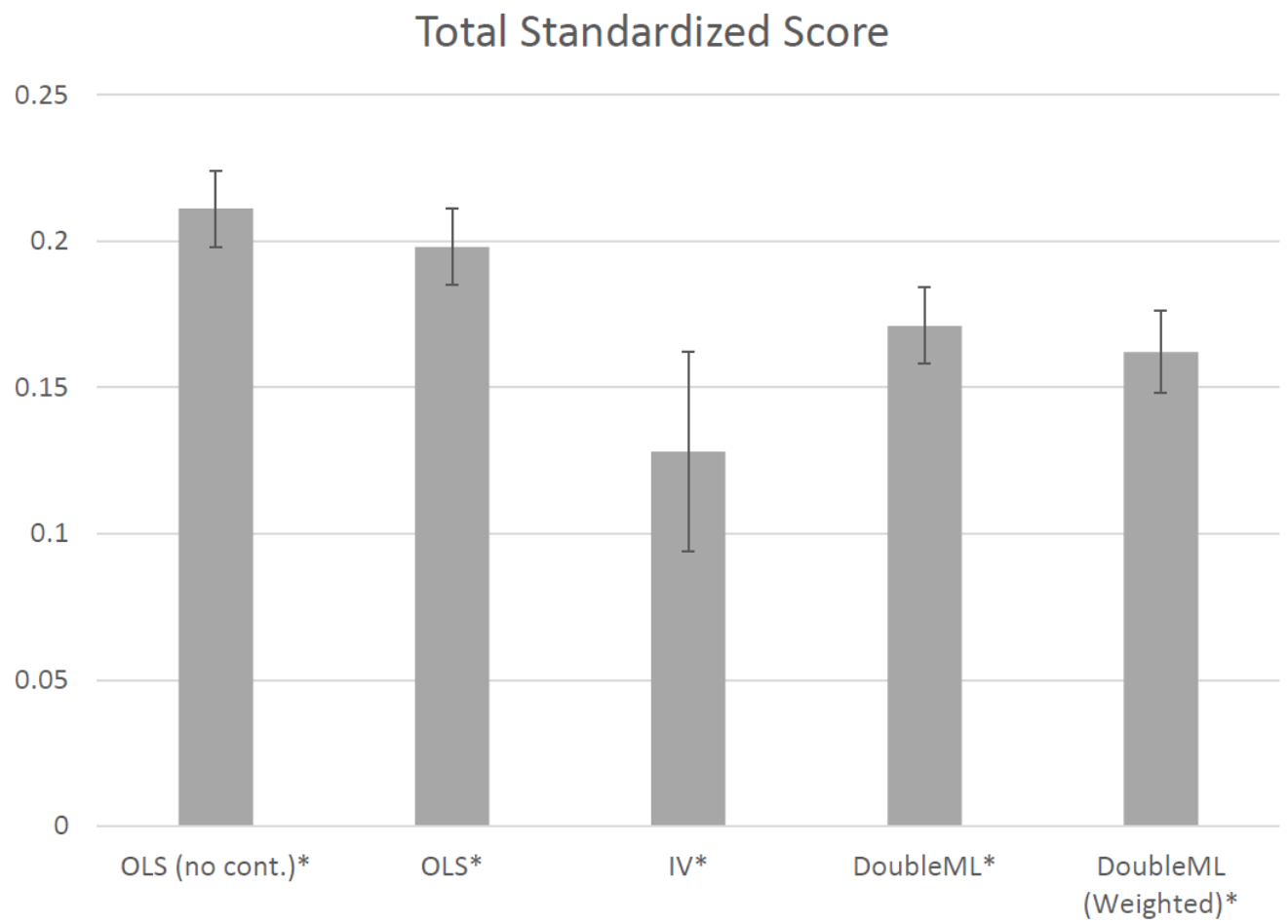
- Option 3: "Double Machine Learning"
  1. Start with double selection to obtain H and K
  2. Compute residuals Y* from regressing Y on H
  3. Compute residuals T* from regressing T on K
  4. Regress Y* on T*. This is the causal estimate

- Paper also introduces cross-validation
  - Key citation: Chernozhukov et al (2017 *Econometrics Journal*), "Double/debiased machine learning for treatment and structural parameters"

Selecting among many controls

- DML example: Duflo, Dupas, Kramer (2011 *AER*)
  - Effects of secondary schooling on labor market outcomes
    - RCT in Ghana provided secondary school scholarships to 682 randomly selected students
  - 2011 paper: scholarship is instrument for schooling
  - Re-analysis: Compares IV to OLS and DML

Selecting among many controls

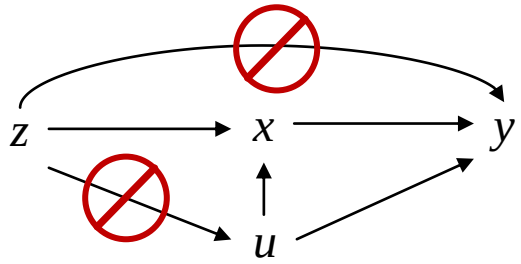


Outline

- Motivation
- ML for measurement
- Inference after selection
- Selecting among many controls
- **Selecting among many instruments**
- Heterogeneous treatment effects
- Other topics

Selecting among many instruments

- Recall instrumental variables:
  - You want to estimate: $Y_i = \alpha + \beta X_i + u_i$
  - Assume you have a valid instrument Z_i
    - Instrument relevance: $\mathrm{corr}(\mathrm{Z_i}, \mathrm{X}_i) \neq 0$
    - Instrument exogeneity: $\mathrm{corr}(\mathrm{Z_i}, \mathrm{u}_i) = 0$
  - Stage 1: $\mathrm{X_i} = b_0 + b_1 Z_i + \nu_i$
  - Stage 2: $\mathrm{Y_i} = \alpha + \beta \widehat{X}_i + u_i$



Selecting among many instruments

- Often, we struggle to find any instruments
- Sometimes, there are many potential instruments to choose between
  - Example: Tech firm running millions of randomized experiments, some of which might affect X
    - X = leaving a review; Y = subsequent purchases
- What to do?
  - Idea: **Use LASSO to select first stage instruments**

Selecting among many instruments

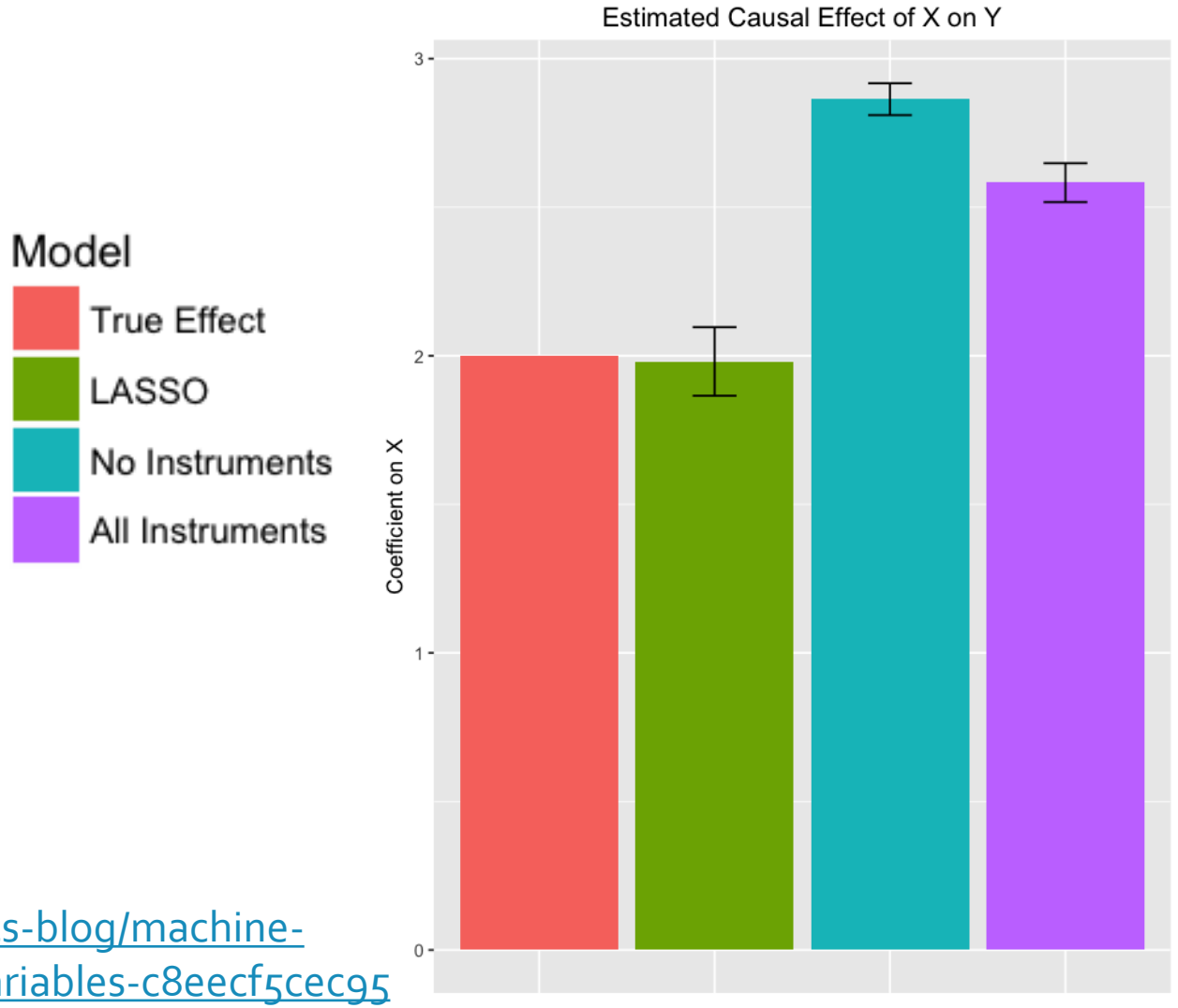
- Belloni et al (2012 ECTA)
  - Provides a set of formal conditions under which conventional inference from two-stage least squares, based on instruments selected by LASSO, is valid for learning about β
  - Intuition: The variable selection part of the problem is limited to the first-stage equation, which is a pure predictive relationship.
  - "The second-stage estimate is *orthogonal* or *immune* to variable selection errors where instruments… are mistakenly excluded from estimation."

Selecting among many instruments

- Belloni et al (2012 ECTA)
  1. Call M the full set of possible instruments. Run LASSO of X on M ➔ yields Z instruments
  2. Use standard IV of Y on X, using Z as instrument for X

Selecting among many instruments

- Simulation



https://medium.com/teconomics-blog/machine-learning-meets-instrumental-variables-c8eecf5cec95

Selecting among many instruments

- Key References
  - Belloni, Chen, Chernozhukov, and Hansen (2012 *Econometrica*), "Sparse models and methods for optimal instruments with an application to eminent domain".
  - Jason Hartford, Greg Lewis, Kevin Leyton-Brown, Matt Taddy (2017): "Deep IV: A Flexible Approach for Counterfactual Prediction." ICML.

Selecting among many instruments

- For another applied example, see:

Dissecting the Effect of Credit Supply on Trade: Evidence from Matched Credit-Export Data

Daniel Paravisini ✉, Veronica Rappoport, Philipp Schnabl, Daniel Wolfenzon

The Review of Economic Studies, Volume 82, Issue 1, January 2015, Pages 333–359,

- Idea: "credit supply" for a firm can depend on many different factors; they use LASSO to select relevant factors

Outline

- Motivation
- ML for measurement
- Inference after selection
- Selecting among many controls
- Selecting among many instruments
- **Heterogeneous treatment effects**
- Other topics

Heterogeneous treatment effects

- Recap of heterogeneous TE's (Lecture 4)
- How to simultaneously measure the effect of a treatment *T* **and** a non-experimental control variable *X* **and** a differential effect of treatment by control variable on an outcome *Y* in a regression setting?

$$Y_i = \alpha + \beta T_i + \gamma X_i + \boldsymbol{\delta}(\boldsymbol{T_i} * \boldsymbol{X_i}) + \epsilon_i$$

- When might this happen in practice?
  - Treatment effect is different for different types (X_i)
  - Some types of people respond to medicine, others don't
  - Some types of people respond to cookies, others don't

Heterogeneous treatment effects

- This begs the question: which "types" X_i are likely to exhibit heterogeneity?
- Typically don't to test all possible X_i (why?)
  - With many variables in X, data are sparse, overfitting likely
  - Biased confidence intervals / p-values
    - With *k* variables, this amounts to *k* statistical tests
    - At very least, would require multiple testing corrections
    - When *k* approaches *N*, need enormous/precise effects

Heterogeneous treatment effects

- Causal Tree (Athey and Imbens, 2016 *PNAS)*
  - Similar to a regression tree, with key differences:
  1. Split sample into two halves randomly: the "structure" sample and "estimation" sample
  2. Build regression tree on structure sample, *optimizing splits for treatment effect heterogeneity* (instead of purity of dependent variable)
     - More on this on next slide
  3. Use estimation sample to estimate treatment effects in each leaf
     - The estimate in each leaf is the average treatment effect for samples with those characteristics

Heterogeneous treatment effects

- Causal Tree (Athey and Imbens, 2016 *PNAS)*
  - What is the splitting criterion that optimizes for treatment effect heterogeneity?
  - In principle, we want to split to obtain more precise estimates of the average treatment effects
  - Problem: this typically requires that we observe true treatment effects, but of course we don't observe unit-level treatment effects
  - Instead, reweight Y by propensity score [details]:

$$Y_i^* = Y_i \cdot \frac{T_i - p(X)}{p(X) \cdot (1 - p(X))}$$

Heterogeneous treatment effects

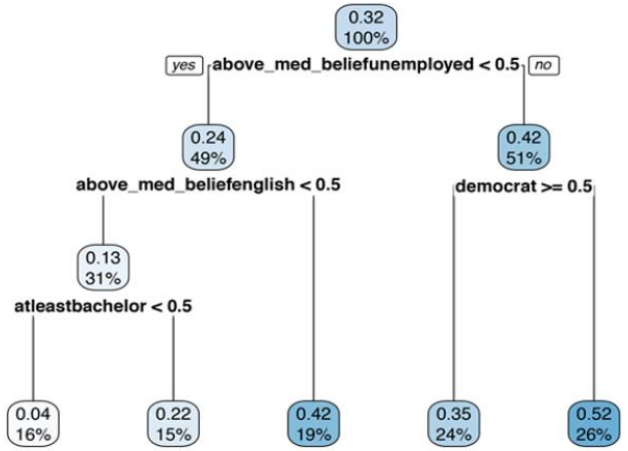
- Causal Tree (Athey and Imbens, 2016 *PNAS)*
  - Advantages:
    - Easy to implement and understand
    - A tree is "honest" if each training observation is only used to either estimate | either estimate decide where to split, or estimate local treatment effect, but not both
  - Disadvantages:
    - As with any single tree, the structure is data-dependent, sensitive to outliers, and somewhat arbitrary
    - The tree doesn't immediately answer the most common question: Which dimensions of heterogeneity are most important?

Example

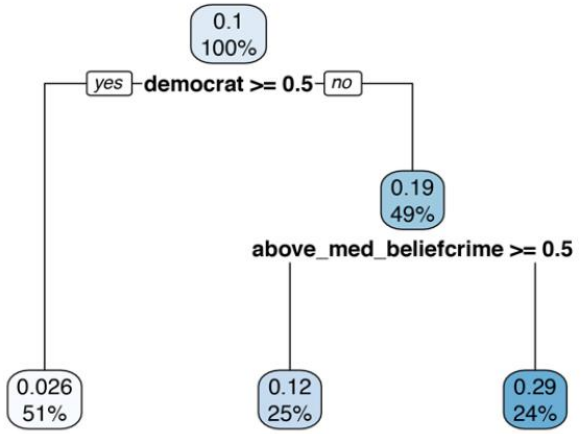
Does Information Change Attitudes Towards Immigrants?

Representative Evidence from Survey Experiments*

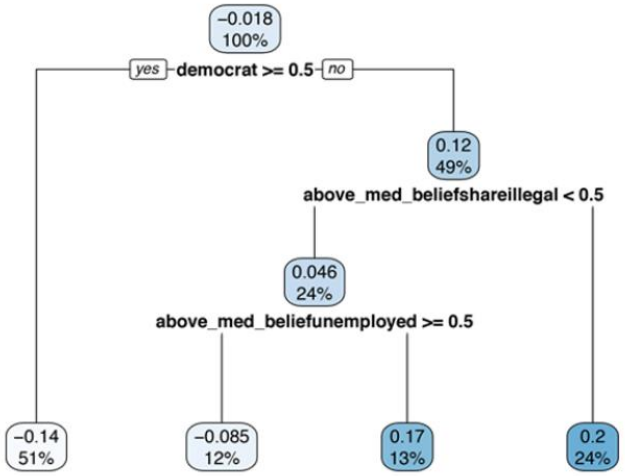
Alexis Grigorieff† Christopher Roth‡ Diego Ubfal§



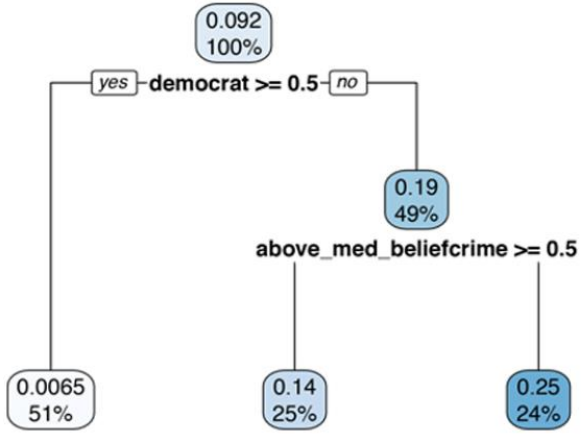
Panel A: Opinions 1



Panel B: Opinions 2



Panel C: Petition



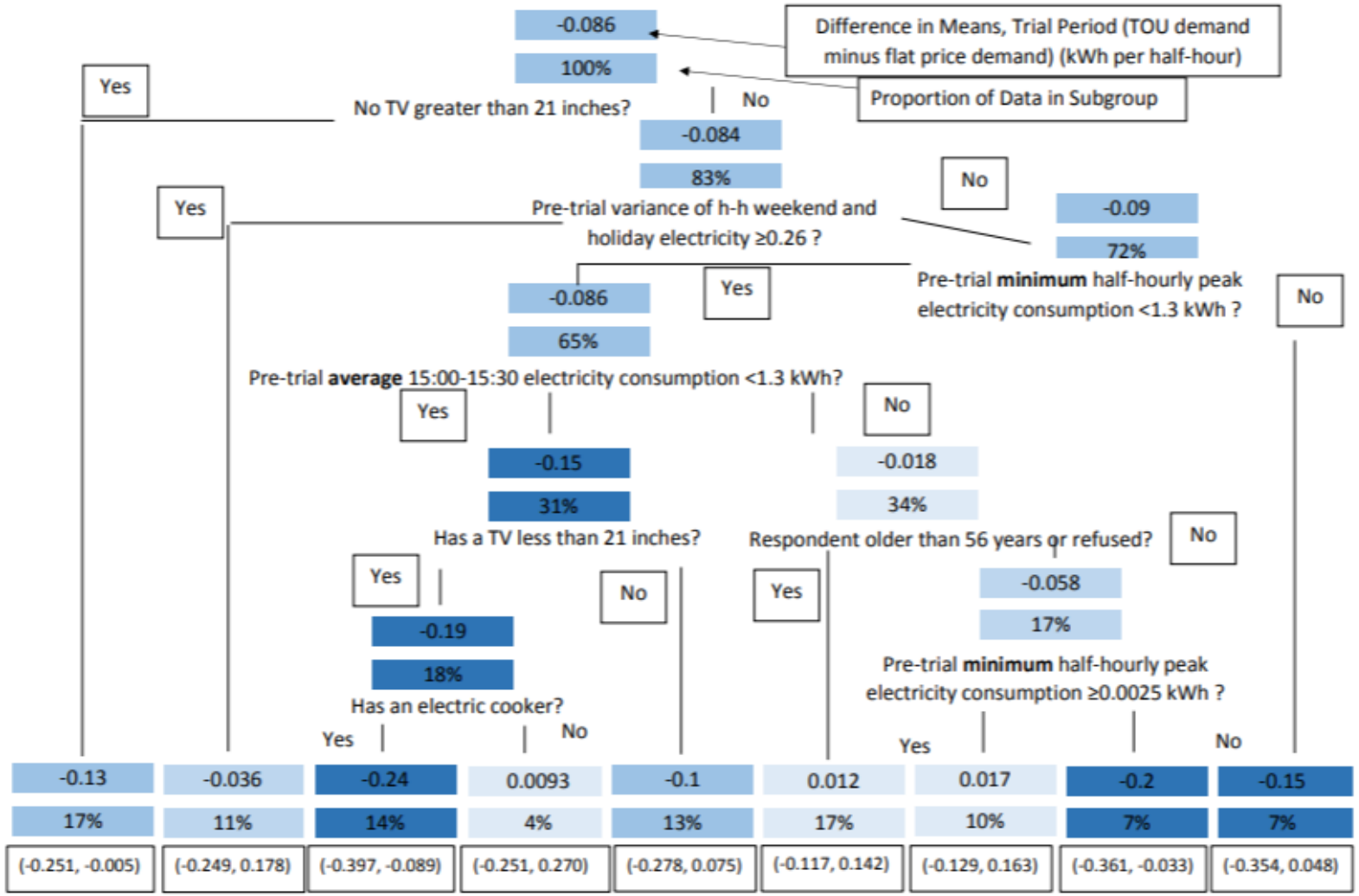
Panel D: Political Preferences

Causal Tree Estimation of Heterogeneous Household Response to Time-Of-Use Electricity Pricing Schemes

EPRG Working Paper 1906
Cambridge Working Paper in Economics 1865

Eoghan O'Neill and Melvyn Weeks

Example



(1) causalForest variable importance	(2)	(3) grf variable importance	(4)	(5) p-value	(1) causalForest variable importance	(2)	(3) grf variable importance	(4)	(5) p-value
attic insulated	0.04	*water instantly heated*	0	0.9	mean 13:00-13:30 usage	40.68	*number of freezers*	10.02	0.09
mean 01:00-01:30 usage	0.08	*number of washing machines*	0.17	0.95	mean 07:00-07:30 usage	40.77	mean h-h coef. of variation	10.51	1
mean 00:30-01:00 usage	0.1	*unheated, lack of money*	0.22	0.78	var. September peak usage	40.78	mean daytime usage	10.83	0.19
prop. elec. saving lightbulbs	0.14	*electric plugin heating*	0.25	0.27	*number of bedrooms*	40.9	variance night usage	11.03	0.98
mean 07:30-08:00 usage	0.16	*electric central heating*	0.34	0.95	mean 15:00-15:30 usage	41.1	mean 10:30-11:00 usage	11.33	0.94
mean usage - weekdays	0.86	*prop. double glazed windows*	0.42	1	*own or rent home*	41.21	mean 22:00-22:30 usage	11.4	0.76
mean 00:00-00:30 usage	1.3	*number of electric cookers*	0.52	1	mean 21:00-21:30 usage	41.43	mean 13:00-13:30 usage	11.82	0.86
variance daytime usage	1.37	*number of tumble dryers*	0.59	1	variance peak usage	42.06	var. night usage - weekdays	11.86	0.99
external walls insulated	1.51	*number of dishwashers*	0.73	1	mean 10:30-11:00 usage	42.13	mean 23:00-23:30 usage	12.09	0.93
mean 08:00-08:30 usage	1.8	*number of immersion heaters*	0.81	1	*number of small TVs*	42.9	mean 14:30-15:00 usage	12.11	0.8
mean 05:00-05:30 usage	1.89	*sex of respondent*	1.08	1	*type of home*	43.36	var. night usage - weekends	12.17	0.98
variance nonpeak usage	2.03	*type of cooker*	1.08	1	*electric central heating*	44.66	mean 21:30-22:00 usage	12.26	0.63
mean h-h coef. of variation	2.12	*attic insulated*	1.12	1	*education*	44.82	*number of laptop PCs*	12.56	0.19
lagging jacking	2.31	*own or rent home*	1.21	1	mean peak usage	45	mean 22:30-23:00 usage	12.7	0.81
mean 04:00-04:30 usage	2.33	*no. of elec. convector heaters*	1.22	1	mean 14:30-15:00 usage	45.12	mean 06:30-07:00 usage	12.72	0.97
mean 05:30-06:00 usage	2.53	*regular internet user*	1.24	1	mean night usage	45.36	mean daytime usage - weekends	12.78	0.1
mean daytime usage	2.56	*water pumped from elec. well*	1.4	1	*number of dishwashers*	45.38	mean 00:00-00:30 usage	13.66	0.89
mean 02:00-02:30 usage	2.81	*water immersion*	1.41	0.99	mean 12:30-13:00 usage	45.4	mean daytime usage - weekdays	14.16	0.12
no. of elec. convector heaters	3.19	*number of instant elec. showers*	1.47	1	*other internet users*	45.44	variance nonpeak usage	14.23	0.19
water pumped from elec. well	3.31	*other internet users*	1.48	0.61	mean daily max. usage	45.54	var. nonpeak usage - weekdays	15	0.26
mean 06:00-06:30 usage	3.4	*external walls insulated*	1.49	1	var. December peak usage	45.84	mean daily min. usage	15.84	0.9
number of desktop PCs	3.42	*number of hot tank elec. showers*	1.63	1	mean 10:00-10:30 usage	46.8	mean 10:00-10:30 usage	15.89	0.64
mean 03:30-04:00 usage	3.75	*water centrally heated*	2.12	0.98	*electric plugin heating*	46.85	mean 23:30-00:00 usage	15.9	0.78
min. half-hourly usage	3.86	*lagging jacking*	2.16	0.74	mean 12:00-12:30 usage	47.19	mean 07:30-08:00 usage	16.37	0.98
number of freezers	4.02	*age of home*	2.39	1	mean 21:30-22:00 usage	47.5	min. half-hourly usage	16.51	0.88
number of instant elec. showers	4.39	*has an energy rating*	2.85	0.6	mean 11:00-11:30 usage	48.6	variance daytime usage	16.58	0.14
variance of usage	4.91	*number of small TVs*	3.01	1	*lives alone*	48.65	mean lunchtime / mean day usage	16.61	1
number of big TVs	4.96	*number of games consoles*	3.29	0.85	mean 11:30-12:00 usage	50.24	mean 18:00-18:30 usage	16.82	0.34
number of games consoles	5.1	*lives alone*	3.39	0.82	mean 22:00-22:30 usage	50.95	var. daytime usage - weekdays	17.6	0.18
prop. double glazed windows	5.64	mean 02:30-03:00 usage	4.03	1	*unheated, lack of money*	51.56	mean 21:00-21:30 usage	17.61	0.26
max. half-hourly usage	5.73	*type of home*	4.06	1	var. October peak usage	51.7	mean 09:00-09:30 usage	18	0.69
mean 08:30-09:00 usage	6.29	*age of respondent*	4.25	1	*internet access*	52.08	variance of usage	18.14	0.05
var. usage - weekdays	6.52	*education*	4.26	1	*water centrally heated*	52.25	var. usage - weekdays	18.29	0.06
mean 03:00-03:30 usage	7.72	mean 12:00-12:30 usage	4.28	1	mean 16:30-17:00 usage	52.6	max. half-hourly usage	18.53	0.87
has an energy rating	8.97	*number of bedrooms*	4.52	0.96	mean 22:30-23:00 usage	52.7	mean 19:00-19:30 usage	18.67	0.21
mean 01:30-02:00 usage	9.69	*prop. elec. saving lightbulbs*	4.56	1	*type of cooker*	53.21	mean 19:30-20:00 usage	19.41	0.15
mean nonpeak usage	9.69	*internet access*	4.94	0.1	*water instantly heated*	53.31	mean 16:00-16:30 usage	19.46	0.44
mean of usage	10.08	mean 03:30-04:00 usage	4.96	1	*regular internet user*	53.37	mean 20:00-20:30 usage	20.3	0.08
number of laptop PCs	10.44	mean 06:00-06:30 usage	5.3	1	*water immersion*	54.64	mean 15:00-15:30 usage	21.12	0.28
mean 09:00-09:30 usage	12.29	mean 03:00-03:30 usage	5.4	1	mean 23:00-23:30 usage	54.82	var. usage - weekends	21.89	0.08
mean 02:30-03:00 usage	15.9	mean 00:30-01:00 usage	5.7	1	var. July peak usage	55.12	mean November peak usage	22.02	0.18
var. night usage - weekends	23.13	mean 05:30-06:00 usage	6.01	1	*number of electric cookers*	55.44	mean 18:30-19:00 usage	22.3	0.1
mean 04:30-05:00 usage	25.78	mean 04:30-05:00 usage	6.03	1	mean 23:30-00:00 usage	57.26	mean 08:00-08:30 usage	22.37	0.69
mean 16:00-16:30 usage	26.07	mean 01:30-02:00 usage	6.29	1	*number of immersion heaters*	57.53	mean 09:30-10:00 usage	23.8	0.37
mean 17:00-17:30 usage	27.02	mean 11:00-11:30 usage	6.46	1	mean night / mean day usage	59.33	var. daytime usage - weekends	23.94	0.06
mean daily min. usage	28.03	mean 04:00-04:30 usage	6.54	1	mean December peak usage	59.96	mean 16:30-17:00 usage	24.27	0.36
mean 17:30-18:00 usage	28.56	mean 05:00-05:30 usage	6.73	1	*social class*	61.34	var. November peak usage	25.8	0.3
mean 18:30-19:00 usage	28.87	*number of desktop PCs*	7.16	0.12	mean October peak usage	62.19	mean 15:30-16:00 usage	27.1	0.17
mean 18:00-18:30 usage	29.28	mean night usage - weekends	7.24	0.97	mean night usage - weekends	62.44	mean daily max. usage	27.36	0.04
variance night usage	29.51	*social class*	7.51	0.7	var. daytime usage - weekdays	62.5	mean 08:30-09:00 usage	30.15	0.33
mean July peak usage	29.74	*number of big TVs*	7.76	0.53	var. nonpeak usage - weekdays	64.71	mean peak usage - weekdays	33.35	0.03
mean 06:30-07:00 usage	30.99	mean 01:00-01:30 usage	7.91	1	*number of tumble dryers*	71	mean peak usage	34.52	0.01
mean 15:30-16:00 usage	31.99	*employment*	7.93	0.57	*age of respondent*	71.25	mean 20:30-21:00 usage	36.62	0.01
mean September peak usage	32.25	mean 11:30-12:00 usage	8.1	0.99	mean lunchtime / mean day usage	71.33	variance peak usage	40.62	0.01
mean 19:00-19:30 usage	32.69	mean 02:00-02:30 usage	8.1	0.98	*sex of respondent*	71.68	var. peak usage - weekdays	40.73	0.05
mean November peak usage	32.81	mean 12:30-13:00 usage	8.15	1	mean nonpeak usage - weekdays	74.69	var. December peak usage	40.75	0.1
var. August peak usage	33.2	mean night usage	8.2	0.88	*number of hot tank elec. showers*	75.34	mean September peak usage	47.41	0.02
number of washing machines	33.87	mean night usage - weekdays	8.88	0.91	mean usage - weekends	78.1	mean 17:00-17:30 usage	52.63	0.01
mean 20:00-20:30 usage	34.29	mean of usage	8.99	0.18	*employment*	78.23	mean December peak usage	53.13	0
mean 13:30-14:00 usage	35.56	mean night / mean day usage	9.09	1	var. daytime usage - weekends	80.36	mean July peak usage	53.33	0.03
mean 19:30-20:00 usage	35.69	mean nonpeak usage - weekdays	9.14	0.29	mean daytime usage - weekends	82.81	mean 17:30-18:00 usage	53.36	0
var. November peak usage	35.92	mean nonpeak usage	9.38	0.22	mean night usage - weekdays	85.08	mean August peak usage	54.18	0.01
mean 09:30-10:00 usage	36.4	mean 13:30-14:00 usage	9.41	0.95	var. peak usage - weekdays	87.83	var. July peak usage	55.36	0.1
mean 20:30-21:00 usage	36.68	mean 14:00-14:30 usage	9.41	0.92	var. usage - weekends	91.21	var. September peak usage	56.17	0.04
mean 14:00-14:30 usage	36.97	mean usage - weekdays	9.57	0.19	mean daytime usage - weekdays	94.64	mean October peak usage	64.96	0
mean August peak usage	39.59	mean 07:00-07:30 usage	9.91	1	var. night usage - weekdays	95.61	var. August peak usage	71.73	0
age of home	40.15	mean usage - weekends	10.02	0.23	mean peak usage - weekdays	100	var. October peak usage	100	0

Survey variables are in italics.

Table 7: Variable Importance results

Causal Tree Estimation of Heterogeneous Household Response to Time-Of-Use Electricity Pricing Schemes

EPRG Working Paper 1906

Cambridge Working Paper in Economics 1865

Eoghan O'Neill and Melvyn Weeks

Heterogeneous treatment effects

- "Generic" approach (Chernozhukov et al, 2018)
  - A challenge with forests is that it is difficult to describe the output, since the estimated function may be quite complex.
  - In this approach, "the key will be to give up on estimating all the possible heterogeneity but focus on a limited number of core features (is there heterogeneity? what are the characteristics of those with the largest treatment effect?)"

Heterogeneous treatment effects

- "Generic" approach focuses on "key features" of the CATE, rather than the CATE itself
  - **Best Linear Predictor** (BLP) of the CATE (conditional avg. treatment effect) based on the ML proxy predictor
  - **Sorted Group Average Treatment Effects** (GATES): average treatment effect by heterogeneity groups induced by the ML proxy predictor
  - **Classification Analysis** (CLAN): average characteristics of the most and least affected units defined in terms of the ML proxy predictor.

Heterogeneous treatment effects

- **Best Linear Predictor:** where is the heterogeneity?

TABLE 3. BLP of Microfinance Availability

	Elastic Net		Random Forest	
	ATE (β_1)	HET (β_2)	ATE (β_1)	HET (β_2)
Amount of Loans	1,163	0.238	1,180	0.390
	(545,1737)	(0.021,0.448)	(546,1770)	(0.037,0.779)
	[0.000]	[0.060]	[0.001]	[0.062]
Output	5,096	0.262	4,854	0.190
	(230,10027)	(0.084,0.431)	(-167,9982)	(-0.099,0.498)
	[0.079]	[0.008]	[0.116]	[0.385]
Profit	1,554	0.243	1,625	0.275
	(-1344,4388)	(0.079,0.416)	(-1332,4576)	(0.036,0.510)
	[0.584]	[0.008]	[0.577]	[0.045]
Consumption	-59.2	0.154	-58.5	0.183
	(-161.4,43.9)	(-0.054,0.382)	(-167.0, 45.9)	(-0.177,0.565)
	[0.513]	[0.270]	[0.494]	[0.617]

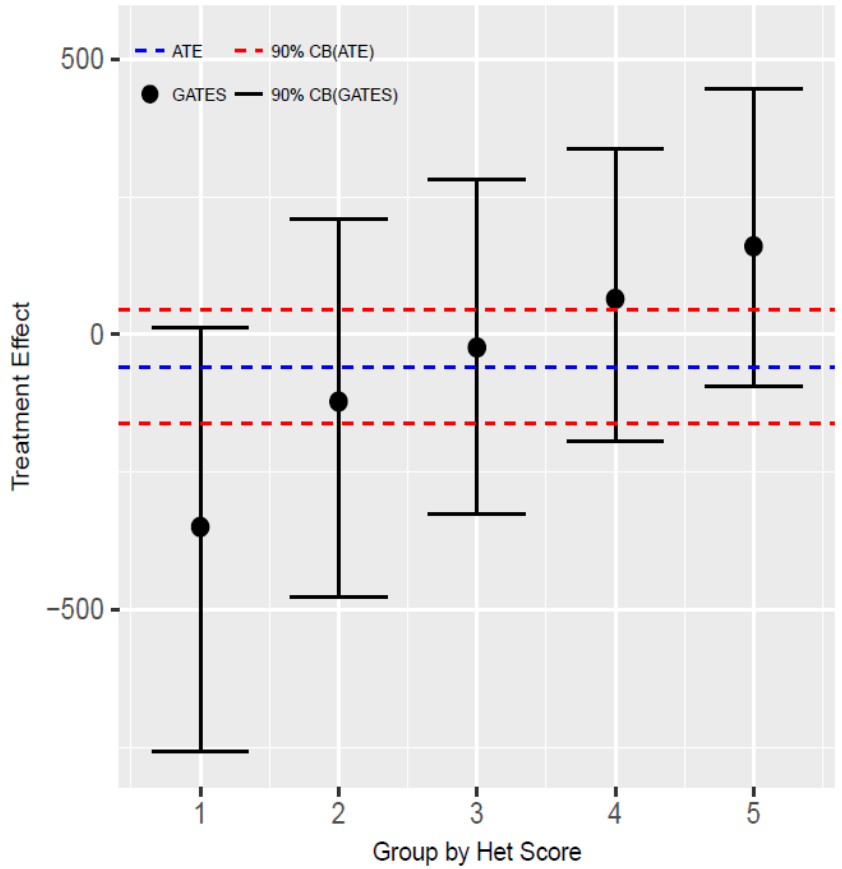
Notes: Medians over 100 splits. 90% confidence interval in parenthesis.
P-values for the hypothesis that the parameter is equal to zero in brackets.

"Reject null hypothesis of no heterogeneity in TE of Micro-finance on output"

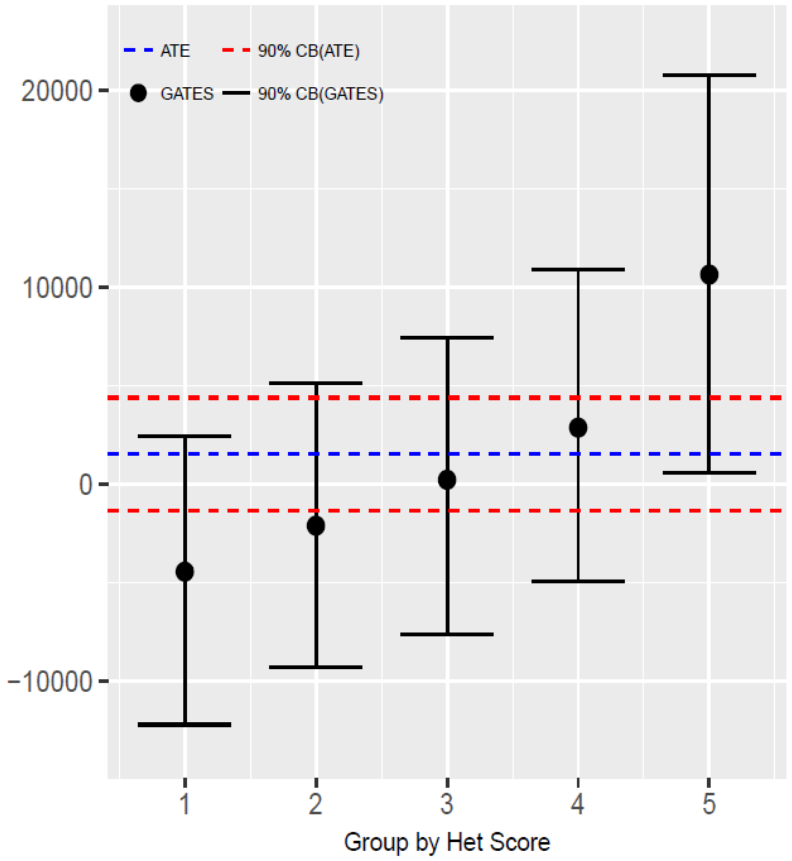
Heterogeneous treatment effects

- **Sorted Group Average Treatment Effects** (GATES):

Consumption



Total profit



Class

	Elastic Net		
	10% Most (δ_{10})	10% Least (δ_1)	Difference ($\delta_{10} - \delta_1$)
Amount of Loans			
Head Age	29.3 (26.3,32.4) -	35.2 (32.2,38.2) -	-6.6 (-10.9,-2.4) [0.004]
Non-agricultural self-emp.	0.199 (0.159,0.238) -	0.068 (0.030,0.108) -	0.123 (0.069,0.178) [0.000]
Borrowed from Any Source	0.144 (0.099,0.189) -	0.169 (0.124,0.212) -	-0.038 (-0.101,0.025) [0.448]
Output			
Head Age	36.280 (33.4,39.1) -	36.708 (33.6,39.6) -	-0.896 (-5.242,3.432) [1.000]
Non-agricultural self-emp.	0.275 (0.233,0.315) -	0.050 (0.007,0.093) -	0.226 (0.169,0.285) [0.000]
Borrowed from Any Source	0.193 (0.142,0.241) -	0.215 (0.167,0.262) -	-0.033 (-0.102,0.034) [0.687]
Profit			
Head Age	34.1 (31.2,37.0) -	40.4 (37.5,43.4) -	-6.5 (-10.7,-2.5) [0.003]
Non-agricultural self-emp.	0.181 (0.140,0.222) -	0.108 (0.068,0.149) -	0.082 (0.022,0.138) [0.014]
Borrowed from Any Source	0.180 (0.130,0.230) -	0.257 (0.207,0.307) -	-0.091 (-0.160,-0.022) [0.020]

Notes: Medians over 100 splits. 90% confidence interval in parenthesis.
P-values for the hypothesis that the parameter is equal to zero in brackets.

"Treatment effect on profits is larger for younger heads of household"

Other closely-related methods

- Causal Forests
  - Stefan Wager and Susan Athey. 2017. Estimation and inference of heterogeneous treatment effects using random forests. *JASA*
- SVM-based approach
  - Kosuke Imai, Marc Ratkovic, et al. 2013. Estimating treatment effect heterogeneity in randomized program evaluation. *Annals of Applied Statistics*
- Targeted Maximum Likelihood
  - Mark J van der Laan and Daniel Rubin. 2006. Targeted maximum likelihood learning. *The International Journal of Biostatistics*
- Meta-Learners
  - Künzel, S.R., Sekhon, J.S., Bickel, P.J., Yu, B., 2019. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the National Academy of Sciences*

Key Citations (Heterogeneous TE's)

- Athey, Susan and Guido W. Imbens (2016). "Recursive Partitioning for Heterogeneous Causal Effects". *Proceedings of the National Academy of Sciences*
- Chernozhukov, V., Demirer, M., Duflo, E., Fernández-Val, I., 2018. "Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments". NBER Working Paper No. 24678
- Wager, Stefan and Susan Athey (2018). "Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests". *Journal of the American Statistical Association*

Outline

- Motivation
- ML for measurement
- Inference after selection
- Selecting among many controls
- Selecting among many instruments
- Heterogeneous treatment effects
- **Other topics**

Other topics

- Also worth knowing about
  - ML to determine which outcomes are affected
    - Ludwig, J., Mullainathan, S., Spiess, J., 2019. "Machine-Learning Tests for Effects on Multiple Outcomes."
    - Ludwig, J., Mullainathan, S., Spiess, J., 2019. "Augmenting pre-analysis plans with machine learning"
  - Adaptive experimentation, ML to guide data collection and experimentation
    - If new units arrive over time, and we can adapt treatment choices, we can learn optimal treatment quickly
    - Bubeck, S. and Cesa-Bianchi, N. (2012). Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems. *Foundations and Trends in Machine Learning*

Other topics

- Also worth knowing about
  - Causal inference and casual/do-calculus
    - Pearl, J., 2009. Causality. Cambridge university press.
    - Pearl, J., Mackenzie, D., 2018. The Book of Why: The New Science of Cause and Effect. Basic Books
    - Pearl, J., 2019. The seven tools of causal inference, with reflections on machine learning. *Communications of the ACM*

Summary

- Contributions from causal inference
  - Identification and estimation of causal effects
  - Classical theory to yield asymptotically normal and centered confidence intervals
- Contributions from ML
  - Practical, high performance algorithms for personalized prediction and policy estimation
- Putting them together
  - Practical, high performance algorithms
  - Causal effects with valid confidence intervals

Good resources/repositories

- Key readings on bCourses
- Main innovators
  - Susan Athey, Alex Belloni, Victor Chernozhukov, Christian Hansen, Guido Imbens, Sendhil Mullainathan, Hal Varian
- Dario Sansone's list of resources
  - https://sites.google.com/view/dariosansone/resources/machine-learning
- Colin Cameron's list of resources
  - http://cameron.econ.ucdavis.edu/e240f/machinelearning.htm