MACHINE LEARNING IN ECONOMETRICS ‡

Double/Debiased/Neyman Machine Learning of Treatment Effects[†]

By VICTOR CHERNOZHUKOV, DENIS CHETVERIKOV, MERT DEMIRER, ESTHER DUFLO, CHRISTIAN HANSEN, AND WHITNEY NEWEY*

Chernozhukov et al. (2016) provide a generic double/debiased machine learning (ML) approach for obtaining valid inferential statements about focal parameters, using Neyman-orthogonal scores and cross-fitting, in settings where nuisance parameters are estimated using ML methods. In this note, we illustrate the application of this method in the context of estimating average treatment effects (ATE) and average treatment effects on the treated (ATTE) using observational data. Empirical illustrations and code are available as supplementary material to this paper, and more general discussion and references to the existing literature are available in Chernozhukov et al. (2016).

I. Scores for Average Treatment Effects

We consider estimation of ATE and ATTE under the unconfoundedness assumption of Rosenbaum and Rubin (1983). We consider the case where treatment effects are fully heterogeneous and the treatment variable, D, is binary, $D \in \{0, 1\}$. We let Y denote the outcome variable of interest and Z denote a set of control variables. We then model random vector (Y, D, Z) as

$$(1) \quad Y = g_0(D, Z) + \zeta, \qquad E[\zeta \mid Z, D] = 0,$$

$$(2) \quad D = m_0(Z) + \nu, \qquad E[\nu \mid Z] = 0.$$

Since D is not additively separable, this model allows for very general heterogeneity in treatment effects. Common target parameters θ_0 in this model are the ATE,

$$\theta_0 = E[g_0(1, Z) - g_0(0, Z)],$$

and the ATTE,

$$\theta_0 = E[g_0(1, Z) - g_0(0, Z) | D = 1].$$

The confounding factors Z affect the treatment variable D via the propensity score, $m_0(Z) := E[D|Z]$, and the outcome variable via the function $g_0(D, Z)$. Both of these functions are unknown and potentially complicated, and we consider estimating these functions via the use of ML methods.

We proceed to set up moment conditions with scores that obey a type of orthogonality with respect to nuisance functions. Specifically, we make use of scores $\psi(W, \theta, \eta)$ that satisfy the identification condition

(3) $E\psi(W, \theta_0, \eta_0) = 0,$

and the *Neyman orthogonality condition*

(4) $\partial_\eta E\psi(W, \theta_0, \eta)|_{\eta=\eta_0} = 0,$

where $W = (Y, D, Z)$, θ_0 is the parameter of interest, η denotes nuisance functions with population value η_0 and $\partial_\eta f|_{\eta=\eta_0}$ denotes the derivative of f with respect to η (the Gateaux derivative operator).

Using moment conditions that satisfy (4) to construct estimators and inference procedures that are robust to small mistakes in nuisance parameters has a long history in statistics, e.g., Neyman (1959). Using moment conditions that satisfy (4) is also crucial to developing valid inference procedures for θ_0 after using ML methods to produce estimators $\hat{\eta}$ as discussed, e.g., in Chernozhukov, Hansen, and Spindler (2015). In practice, estimation of θ_0 will be based on the empirical analog of (3) with η_0 replaced by $\hat{\eta}_0$, and the Neyman orthogonality condition (4) ensures sufficient insensitivity to this replacement that high-quality inference for θ_0 may be obtained. The second *critical ingredient*, that enables the use of a wide array of modern ML estimators is *data splitting*, as discussed in the next section.

Neyman-orthogonal scores are readily available for both the ATE and ATTE one can use the doubly robust/efficient scores of Robins and Rotnitzky (1995) and Hahn (1998) which are automatically Neyman orthogonal. For estimating the ATE, we employ

(5) $\psi(W, \theta, \eta) := g(1, Z) - g(0, Z)$

$$+ \frac{D(Y - g(1, Z))}{m(Z)}$$

$$- \frac{(1-D)(Y - g(0, Z))}{1 - m(Z)} - \theta,$$

with

$$\eta(Z) := (g(0, Z), g(1, Z), m(Z))',$$

$$\eta_0(Z) := (g_0(0, Z), g_0(1, Z), m_0(Z))',$$

where $\eta(Z)$ is the nuisance parameter with true value denoted by $\eta_0(Z)$ consisting of P-square integrable functions, for P defined

in Assumption 1, mapping the support of Z to $\mathbb{R} \times \mathbb{R} \times (\varepsilon, 1 - \varepsilon)$ where $\varepsilon > 0$ is a constant. For estimation of ATTE, we use the score

(6) $\psi(W, \theta, \eta) := \dfrac{D(Y - g(0, Z))}{m}$

$$- \frac{m(Z)(1 - D)(Y - g(0, Z))}{(1 - m(Z))m} - \theta\frac{D}{m},$$

with

$$\eta(Z) := (g(0, Z), g(1, Z), m(Z), m)',$$

$$\eta_0(Z) := (g_0(0, Z), g_0(1, Z), m_0(Z), E[D])',$$

where again $\eta(Z)$ is the nuisance parameter with true value denoted by $\eta_0(Z)$ consisting of three P-square integrable functions, for P defined in Assumption 1, mapping the support of Z to $\mathbb{R} \times \mathbb{R} \times (\varepsilon, 1 - \varepsilon)$ and a constant $m \in (\varepsilon, 1 - \varepsilon)$. The respective scores for ATE and ATTE obey the identification condition (3) and the Neyman orthogonality property (4). Note that all semi-parametrically efficient scores share the orthogonality property (4), but not all orthogonal scores are efficient. In some problems, we may use inefficient orthogonal scores to have more robustness. Moreover, the use of efficient scores could be considerably refined using the targeted maximum likelihood approach of Scharfstein et al. (1999) and van der Laan and Rubin (2006) in many contexts.

II. Algorithm and Result

We describe the estimator of θ_0 using random sample $(W_i)_{i=1}^N$. The algorithm makes use of a form of sample splitting, which we call cross-fitting. It builds on the ideas in, e.g., Angrist and Krueger (1995). The use of sample-splitting is a *crucial ingredient* to the approach that helps avoid overfitting which can easily result from the application of complex, flexible methods such as boosted linear and tree models, random forests, and various ensemble and hybrid ML methods.

A. *Algorithm: Estimation by K-fold Cross-Fitting*

Step 1: Let K be a fixed integer. Form a K-fold random partition of $\{1, \ldots, N\}$ by dividing it into equal parts $(I_k)_{k=1}^K$ each of size

$n := N/K$, assuming that N is a multiple of K. For each set I_k, let I_k^c denote all observation indices that are not in I_k.

Step 2: Construct K estimators

$$\check{\theta}_0(I_k, I_k^c), \quad k = 1, \ldots, K,$$

that employ the machine learning estimators

$$\hat{\eta}_0(I_k^c) = \left(\hat{g}_0(0, Z; I_k^c), \hat{g}_0(1, Z; I_k^c), \right.$$

$$\left. \hat{m}_0(Z; I_k^c), \frac{1}{N-n} \sum_{i \in I_k^c} D_i \right)',$$

of the nuisance parameters

$$\eta_0(Z) = (g_0(0, Z), g_0(1, Z), m_0(Z), E[D])',$$

and where each estimator $\check{\theta}_0(I_k, I_k^c)$ is defined as the root θ of

$$\frac{1}{n} \sum_{i \in I_k} \psi(W, \theta, \hat{\eta}_0(I_k^c)) = 0,$$

for the score ψ defined in (5) for the ATE and in (6) for the ATTE.

Step 3: Average the K estimators to obtain the final estimator:

$$(7) \qquad \tilde{\theta}_0 = \frac{1}{K} \sum_{k=1}^{K} \check{\theta}_0(I_k, I_k^c).$$

An approximate standard error for this estimator is $\hat{\sigma}/\sqrt{N}$, where

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{N} \hat{\psi}_i^2,$$

$\hat{\psi}_i := \psi(W_i, \tilde{\theta}_0, \hat{\eta}_0(I_{k(i)}^c))$, and $k(i) := \{k \in \{1, \ldots, K\} : i \in I_k\}$. An approximate $(1 - \alpha) \times 100$ percent confidence interval is

$$CI_n := [\tilde{\theta}_0 \pm \Phi^{-1}(1 - \alpha/2)\hat{\sigma}/\sqrt{N}].$$

We now state a formal result that provides the asymptotic properties of $\tilde{\theta}_0$. Let $(\delta_n)_{n=1}^{\infty}$ and $(\Delta_n)_{n=1}^{\infty}$ be sequences of positive constants approaching 0. Let c, ε, C, and $q > 4$ be fixed positive constants, and let K be a fixed integer.

ASSUMPTION 1: *Let $\mathcal{P}$ be the set of probability distributions P for (Y, D, Z) such that (i) equations (1)–(2) hold, with $D \in \{0, 1\}$; (ii) the following conditions on moments hold for all N and $d \in \{0, 1\}$: $\|g(d, Z)\|_{P,q} \leq C$, $\|Y\|_{P,q} \leq C$, $P(\varepsilon \leq m_0(Z) \leq 1 - \varepsilon) = 1$, $P(E_P[\zeta^2|Z] \leq C) = 1, \|\zeta\|_{P,2} \geq c$ and $\|\nu\|_{P,2} \geq c$; and (iii) the ML estimators of the nuisance parameters based upon a random subset I_k^c of $\{1, \ldots, N\}$ of size $N - n$, obey the following conditions for all $N \geq 2K$ and $d \in \{0, 1\}$: $\| \hat{g}_0(d, Z; I_k^c) - g_0(d, Z) \|_{P,2} \cdot \| \hat{m}_0(Z; I_k^c) - m_0(Z) \|_{P,2} \leq \delta_n n^{-1/2}, \| \hat{g}_0(d, Z; I_k^c) - g_0(d, Z) \|_{P,2} + \| \hat{m}_0(Z; I_k^c) - m_0(Z) \|_{P,2} \leq \delta_n$, and $P(\varepsilon \leq \hat{m}_0(Z; I_k^c) \leq 1 - \varepsilon) = 1$, with P_P-probability no less than $1 - \Delta_n$.*

The assumption on the rate of estimating the nuisance parameters is a non-primitive condition. These rates of convergence are available for most often used ML methods and are case-specific, so we do not restate conditions that are needed to reach these rates. The conditions are not the tightest possible but are chosen for simplicity.

THEOREM 1: *Suppose that the ATE, $\theta_0 = E_P[g_0(1, Z) - g_0(0, Z)]$, is the target parameter and we use the estimator $\tilde{\theta}_0$ and other notations defined above. Alternatively, suppose that the ATTE, $\theta_0 = E_P[g_0(1, Z) - g_0(0, Z) | D = 1]$, is the target parameter and we use the estimator $\tilde{\theta}_0$ and other notations above. Consider the set $\mathcal{P}$ of probability distributions P defined in Assumption 1. Then, uniformly in $P \in \mathcal{P}$, the estimator $\tilde{\theta}_0$ concentrates around θ_0 with the rate $1/\sqrt{N}$ and is approximately unbiased and normally distributed:*

$$\sigma^{-1}\sqrt{N}(\tilde{\theta}_0 - \theta_0) \rightsquigarrow N(0, 1),$$

$$\sigma^2 = E_P[\psi^2(W, \theta_0, \eta_0(Z))],$$

and the result continues to hold if σ^2 is replaced by $\hat{\sigma}^2$. Moreover, confidence regions based upon $\tilde{\theta}_0$ have uniform asymptotic validity:

$$\sup_{P \in \mathcal{P}} |P(\theta_0 \in CI_n) - (1 - \alpha)| \to 0.$$

The scores ψ are the efficient scores, so both estimators are asymptotically efficient, in the

sense of reaching the semi-parametric efficiency bound of Hahn (1998).

The proof, given in the online Appendix, relies on the application of Chebyshev inequality and the central limit theorem.

III. Accounting for Uncertainty Due to Sample-Splitting

The method outlined in this note relies on subsampling to form auxiliary samples for estimating nuisance functions and main samples for estimating the parameter of interest. The specific sample partition has no impact on estimation results asymptotically but may be important in finite samples. Specifically, the dependence of the estimator on the particular split creates an additional source of variation. Incorporating a measure of this additional source of variation into estimated standard errors of parameters of interest may be important for quantifying the true uncertainty of the parameter estimates.

Hence we suggest making a slight modification to the asymptotically valid estimation procedure detailed in Section II. Specifically, we propose repeating the main estimation procedure S times, for a large number S, repartitioning the data in each replication $s = 1, \ldots, S$. Within each partition, we then obtain an estimate of the parameter of interest, $\tilde{\theta}_0^s$. Rather than report point estimates and interval estimates based on a single replication, we may then report estimates that incorporate information from the distribution of the individual estimates obtained from the S different data partitions.

For point estimation, two natural quantities that could be reported are the sample average and the sample median of the estimates obtained across the S replications, $\tilde{\theta}_0^{\text{Mean}}$ and $\tilde{\theta}_0^{\text{Median}}$. Both of these reduce the sensitivity of the estimate for θ_0 to particular splits. $\tilde{\theta}_0^{\text{Mean}}$ could be strongly affected by any extreme point estimates obtained in the different random partitions of the data, and $\tilde{\theta}_0^{\text{Median}}$ is obviously much more robust. We note that asymptotically the specific random partition is irrelevant, and $\tilde{\theta}_0^{\text{Mean}}$ and $\tilde{\theta}_0^{\text{Median}}$ should be close to each other.

To quantify and incorporate the variation introduced by sample splitting, one might also compute standard errors that add an element to capture the spread of the estimates obtained across the S different sets of partitions. For $\tilde{\theta}_0^{\text{Mean}}$, we propose adding an element that captures the spread of the estimated $\tilde{\theta}_0^s$ around $\tilde{\theta}_0^{\text{Mean}}$. Specifically, we suggest

$$\hat{\sigma}^{\text{Mean}} = \sqrt{\frac{1}{S}\sum_{s=1}^{S}\left(\hat{\sigma}_s^2 + \left(\tilde{\theta}_0^s - \frac{1}{S}\sum_{j=1}^{S}\tilde{\theta}_0^j\right)^2\right)},$$

where $\hat{\sigma}_s$ is defined as in Section II. The second term in this formula takes into account the variation due to sample splitting which is added to a usual estimate of sampling uncertainty. Using this estimated standard error obviously results in more conservative inference than relying on the $\hat{\sigma}_s$ alone. We adopt a similar formulation for $\tilde{\theta}_0^{\text{Median}}$. Specifically, we propose a median deviation defined as

$$\hat{\sigma}^{\text{Median}} = \text{median}\left\{\sqrt{\hat{\sigma}_i^2 + \left(\hat{\theta}_i - \hat{\theta}^{\text{Median}}\right)^2}\right\}_{i=1}^{S}.$$

This standard error is more robust to outliers than $\hat{\sigma}^{\text{Mean}}$.

REFERENCES

Angrist, Joshua D., and Alan B. Krueger. 1995. "Split-Sample Instrumental Variables Estimates of the Return to Schooling." *Journal of Business and Economic Statistics* 13 (2): 225–35.

Chernozhukov, Victor, Christian Hansen, and Martin Spindler. 2015. "Valid Post-Selection and Post-Regularization Inference: An Elementary, General Approach." *Annual Review of Economics* 7: 649–88.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, and anad Whitney Newey. 2016. "Double Machine Learning for Treatment and Causal Parameters." https://arxiv.org/abs/1608.00060 (accessed March 21, 2017).

Hahn, Jinyong. 1998. "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects." *Econometrica* 66 (2): 315–31.

Neyman, J. 1959. "Optimal Asymptotic Tests of Composite Statistical Hypotheses." In *Probability and Statistics: The Harald Cramer Volume*, edited by Ulf Grenander, 213–34. Stockholm: Almqvist and Wiksell.

Robins, James M., and Andrea Rotnitzky. 1995. "Semiparametric Efficiency in Multivariate Regression Models with Missing Data." *Journal of the American Statistical Association* 90 (429): 122–29.

Rosenbaum, Paul R., and Donald Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70 (1): 41–55.

Scharfstein, Daniel O., Andrea Rotnitzky, and James M. Robins. 1999. "Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models: Rejoiner." *Journal of the American Statistical Association* 94 (448): 1135–146.

Van der Laan, M. J., and D. Rubin. 2006. "Targeted Maximum Likelihood Learning." *International Journal of Biostatistics* 2 (1).

This article has been cited by:

1. Jorge M. Uribe. 2025. Investment in intangible assets and economic complexity. *Research Policy* **54**:1, 105133. [Crossref]

2. Fuyang Jiang, Hussain Kazmi. 2025. What-if: A causal machine learning approach to control-oriented modelling for building thermal dynamics. *Applied Energy* **377**, 124550. [Crossref]

3. Jie Yuan, Shucheng Liu. 2024. A double machine learning model for measuring the impact of the Made in China 2025 strategy on green economic growth. *Scientific Reports* **14**:1. . [Crossref]

4. Daijiro Kabata, Elizabeth A. Stuart, Ayumi Shintani. 2024. Prognostic score-based model averaging approach for propensity score estimation. *BMC Medical Research Methodology* **24**:1. . [Crossref]

5. Yaming Guo, Meng Li, Keqiang Li, Huiping Li, Yunxuan Li. 2024. Unraveling the determinants of traffic incident duration: A causal investigation using the framework of causal forests with debiased machine learning. *Accident Analysis & Prevention* **208**, 107806. [Crossref]

6. Martin Kroczek, Philipp Kugler. 2024. Heterogeneous effects of monetary and non-monetary job characteristics on job attractiveness in nursing. *Labour Economics* **91**, 102632. [Crossref]

7. Kosuke Inoue, Motohiko Adomi, Orestis Efthimiou, Toshiaki Komura, Kenji Omae, Akira Onishi, Yusuke Tsutsumi, Tomoko Fujii, Naoki Kondo, Toshi A. Furukawa. 2024. Machine learning approaches to evaluate heterogeneous treatment effects in randomized controlled trials: a scoping review. *Journal of Clinical Epidemiology* **176**, 111538. [Crossref]

8. Yong Qiu, Ziyang Guo, Junlong Li, Biao Li, Dan Qiao. 2024. Impact of controlling shareholders' equity pledges on firms' labour investment efficiency. *Finance Research Letters* **70**, 106408. [Crossref]

9. Silu Pang, Guihong Hua. 2024. How does digital tax administration affect R&D manipulation? Evidence from dual machine learning. *Technological Forecasting and Social Change* **208**, 123691. [Crossref]

10. Lucas Kook, Sorawit Saengkyongam, Anton Rask Lundborg, Torsten Hothorn, Jonas Peters. 2024. Model-Based Causal Feature Selection for General Response Types. *Journal of the American Statistical Association* **4**, 1-12. [Crossref]

11. Dmitry Arkhangelsky, Guido Imbens. 2024. Causal models for longitudinal and panel data: a survey. *The Econometrics Journal* **27**:3, C1-C61. [Crossref]

12. Alexis H. Villacis, Syed Badruddoza, Ashok K. Mishra. 2024. A machine learning-based exploration of resilience and food security. *Applied Economic Perspectives and Policy* **1**. . [Crossref]

13. Dmitry Arkhangelsky, Guido W Imbens. 2024. Fixed Effects and the Generalized Mundlak Estimator. *Review of Economic Studies* **91**:5, 2545-2571. [Crossref]

14. Cheng Hsiao, Qiankun Zhou. 2024. Statistical inference for the low dimensional parameters of linear regression models in the presence of high-dimensional data: An orthogonal projection approach. *Journal of Econometrics* **33**, 105851. [Crossref]

15. Zaid Hattab, Silvia Moler-Zapata, Edel Doherty, Zia Sadique, Padmanabhan Ramnarayan, Stephen O'Neill. 2024. Exploring Heterogeneity in the Cost-Effectiveness of High-Flow Nasal Cannula Therapy in Acutely Ill Children—Insights From the Step-Up First-line Support for Assistance in Breathing in Children Trial Using a Machine Learning Method. *Value in Health* **8**. . [Crossref]

16. Hadi Movaghari, Serafeim Tsoukas, Evangelos Vagenas-Nanos. 2024. Corporate cash policy and double machine learning. *International Journal of Finance & Economics* **59**. . [Crossref]

17. Philippe Boileau, Ning Leng, Nima S Hejazi, Mark van der Laan, Sandrine Dudoit. 2024. A nonparametric framework for treatment effect modifier discovery in high dimensions. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **62**. . [Crossref]

18. Xuhang Shen, Jiaxuan Zhang. 2024. Are female leaders hiring more female workers? Evidence from developing countries. *Applied Economics* **62**, 1-17. [Crossref]

19. Laura Derksen, Jason T Kerwin, Natalia Ordaz Reynoso, Olivier Sterck. 2024. Healthcare Appointments as Commitment Devices. *The Economic Journal* **88**. . [Crossref]

20. Alex Luedtke, Incheoul Chung. 2024. One-step estimation of differentiable Hilbert-valued parameters. *The Annals of Statistics* **52**:4. . [Crossref]

21. Hongxiang Qiu, Eric Tchetgen Tchetgen, Edgar Dobriban. 2024. Efficient and multiply robust risk estimation under general forms of dataset shift. *The Annals of Statistics* **52**:4. . [Crossref]

22. Alexei Alexandrov, Russell Pittman, Olga Ukhaneva. 2024. Double marginalization in the pricing of complements: The case of US freight railroads. *Journal of Economics & Management Strategy* **113**. . [Crossref]

23. Anna Baiardi, Andrea A. Naghi. 2024. The effect of plough agriculture on gender roles: A machine learning approach. *Journal of Applied Econometrics* **27**. . [Crossref]

24. Qijia He, Fei Gao, Oliver Dukes, Sinead Delany-Moretlwe, Bo Zhang. 2024. Generalizing the Intention-to-Treat Effect of an Active Control from Historical Placebo-Controlled Trials: A Case Study of the Efficacy of Daily Oral TDF/FTC in the HPTN 084 Study. *Journal of the American Statistical Association* 1-15. [Crossref]

25. Ryunosuke Goto, Kosuke Inoue, Itsuki Osawa, Katherine Baicker, Scott L Fleming, Yusuke Tsugawa. 2024. Machine learning for detection of heterogeneous effects of Medicaid coverage on depression. *American Journal of Epidemiology* **193**:7, 951-958. [Crossref]

26. Zaid Hattab, Edel Doherty, Zia Sadique, Padmanabhan Ramnarayan, Stephen O'Neill. 2024. Exploring Heterogeneity in Cost-Effectiveness Using Machine Learning Methods. *Medical Care* **62**:7, 449-457. [Crossref]

27. Haowen Wang, Xinyan Ye, Yikang Wang, Yangze Zhou, Zhiyi Zhang, Longhan Zhang, Jing Jiang, Yiteng Zhai. Uplift Modeling based on Graph Neural Network Combined with Causal Knowledge 1487-1492. [Crossref]

28. Anna Baiardi, Andrea A Naghi. 2024. The value added of machine learning to causal inference: evidence from revisited studies. *The Econometrics Journal* **27**:2, 213-234. [Crossref]

29. Dominik S. Meier. 2024. Compassion for All: Real-World Online Donations Contradict Compassion Fade. *Nonprofit and Voluntary Sector Quarterly* **32**. . [Crossref]

30. Shucheng Liu, Jie Yuan. 2024. Can government digitalization promote the urban–rural equalization of basic public services? Evidence from double machine learning. *Applied Economics* 1-16. [Crossref]

31. Koichiro Shiba, Kosuke Inoue. 2024. Harnessing causal forests for epidemiologic research: key considerations. *American Journal of Epidemiology* **193**:6, 813-818. [Crossref]

32. Ziwei Mei, Zhentao Shi. 2024. On LASSO for high dimensional predictive regression. *Journal of Econometrics* **242**:2, 105809. [Crossref]

33. Jialuo Liu, Dong Xi. 2024. Covariate adjustment and estimation of difference in proportions in randomized clinical trials. *Pharmaceutical Statistics* **118**. . [Crossref]

34. Prodosh Eugene Simlai. 2024. Risk Characterization of Firms with ESG Attributes Using a Supervised Machine Learning Method. *Journal of Risk and Financial Management* **17**:5, 211. [Crossref]

35. Paolo Libenzio Brignoli, Alessandro Varacca, Cornelis Gardebroek, Paolo Sckokai. 2024. Machine learning to predict grains futures prices. *Agricultural Economics* **55**:3, 479-497. [Crossref]

36. Nicholas Larsen, Jonathan Stallrich, Srijan Sengupta, Alex Deng, Ron Kohavi, Nathaniel T. Stevens. 2024. Statistical Challenges in Online Controlled Experiments: A Review of A/B Testing Methodology. *The American Statistician* **78**:2, 135-149. [Crossref]

37. Manjari Das, Edward H. Kennedy, Nicholas P. Jewell. 2024. Doubly Robust Capture-Recapture Methods for Estimating Population Size. *Journal of the American Statistical Association* **119**:546, 1309-1321. [Crossref]

38. Qixian Zhong, Jane-Ling Wang. 2024. Neural Networks for Partially Linear Quantile Regression. *Journal of Business & Economic Statistics* **42**:2, 603-614. [Crossref]

39. Esfandiar Maasoumi, Jianqiu Wang, Zhuo Wang, Ke Wu. 2024. Identifying factors via automatic debiased machine learning. *Journal of Applied Econometrics* **39**:3, 438-461. [Crossref]

40. Chen Liang, Yili Hong, Bin Gu. 2024. Monitoring and the Cold Start Problem in Digital Platforms: Theory and Evidence from Online Labor Markets. *Information Systems Research* **3**. . [Crossref]

41. Sijia Li, Wei Pan, Paul Siu Fai Yip, Jing Wang, Wenwei Zhou, Tingshao Zhu. 2024. Uncovering the heterogeneous effects of depression on suicide risk conditioned by linguistic features: A double machine learning approach. *Computers in Human Behavior* **152**, 108080. [Crossref]

42. Vignesh Yoganathan, Victoria-Sophie Osburg. 2024. The mind in the machine: Estimating mind perception's effect on user satisfaction with voice-based conversational agents. *Journal of Business Research* **175**, 114573. [Crossref]

43. Xiaohui Liu, Sean Qian, Hock-Hai Teo, Wei Ma. 2024. Estimating and Mitigating the Congestion Effect of Curbside Pick-ups and Drop-Offs: A Causal Inference Approach. *Transportation Science* **58**:2, 355-376. [Crossref]

44. Susan Athey, Guido W. Imbens, Jonas Metzger, Evan Munro. 2024. Using Wasserstein Generative Adversarial Networks for the design of Monte Carlo simulations. *Journal of Econometrics* **240**:2, 105076. [Crossref]

45. Alberto Abadie, Jiaying Gu, Shu Shen. 2024. Instrumental variable estimation with first-stage heterogeneity. *Journal of Econometrics* **240**:2, 105425. [Crossref]

46. Myoung-jae Lee, Chirok Han. 2024. Ordinary least squares and instrumental-variables estimators for any outcome and heterogeneity. *The Stata Journal: Promoting communications on statistics and Stata* **24**:1, 72-92. [Crossref]

47. Sarah E Robertson, Jon A Steingrimsson, Nina R Joyce, Elizabeth A Stuart, Issa J Dahabreh. 2024. Estimating Subgroup Effects in Generalizability and Transportability Analyses. *American Journal of Epidemiology* **193**:1, 149-158. [Crossref]

48. Yi Ding, Yingying Li, Rui Song. 2024. Statistical Learning for Individualized Asset Allocation. *Journal of the American Statistical Association* **119**:545, 639-649. [Crossref]

49. Iván Díaz, Katherine L. Hoffman, Nima S. Hejazi. 2024. Causal survival analysis under competing risks using longitudinal modified treatment policies. *Lifetime Data Analysis* **30**:1, 213-236. [Crossref]

50. Shiva Madadkhani, Svetlana Ikonnikova. 2024. Toward high-resolution projection of electricity prices: A machine learning approach to quantifying the effects of high fuel and CO2 prices. *Energy Economics* **129**, 107241. [Crossref]

51. Shourya Mehra, Shivom Aggarwal, Pritha Mitra. Predictive Analytics and AI-Driven Strategies for Enhanced Cash Flow Forecasting 296-315. [Crossref]

52. George Panagopoulos, Daniele Malitesta, Fragkiskos D. Malliaros, Jun Pang. Uplift Modeling Under Limited Supervision 127-144. [Crossref]

53. Samir Bhatt, Neil Ferguson, Seth Flaxman, Axel Gandy, Swapnil Mishra, James A Scott. 2023. Semi-mechanistic Bayesian modelling of COVID-19 with renewal processes. *Journal of the Royal Statistical Society Series A: Statistics in Society* **186**:4, 601-615. [Crossref]

54. Daijiro Kabata, Mototsugu Shintani. 2023. Variable selection in double/debiased machine learning for causal inference: an outcome-adaptive approach. *Communications in Statistics - Simulation and Computation* **52**:12, 5880-5893. [Crossref]

55. Henry V. Burton, Jack W. Baker. 2023. Evaluating the effectiveness of ground motion intensity measures through the lens of causal inference. *Earthquake Engineering & Structural Dynamics* **52**:15, 4842-4864. [Crossref]

56. Stefan Buijsman. 2023. Causal scientific explanations from machine learning. *Synthese* **202**:6. . [Crossref]

57. Theodoros Evgeniou, Ahmed Guecioueur, Rodolfo Prieto. 2023. Uncovering Sparsity and Heterogeneity in Firm-Level Return Predictability Using Machine Learning. *Journal of Financial and Quantitative Analysis* **58**:8, 3384-3419. [Crossref]

58. Zhen Li, Jie Chen, Eric Laber, Fang Liu, Richard Baumgartner. 2023. Optimal Treatment Regimes: A Review and Empirical Comparison. *International Statistical Review* **91**:3, 427-463. [Crossref]

59. Fuyang Jiang, Johan Driesen, Hussain Kazmi. On the causality of data-driven building thermal models 454-457. [Crossref]

60. Daijiro Kabata, Mototsugu Shintani. 2023. On propensity score misspecification in double/debiased machine learning for causal inference: ensemble and stratified approaches. *Communications in Statistics - Simulation and Computation* **43**, 1-11. [Crossref]

61. Daniel Hoang, Kevin Wiegratz. 2023. Machine learning methods in finance: Recent applications and prospects. *European Financial Management* **29**:5, 1657-1701. [Crossref]

62. Ruoxuan Xiong, Allison Koenecke, Michael Powell, Zhu Shen, Joshua T. Vogelstein, Susan Athey. 2023. Federated causal inference in heterogeneous observational data. *Statistics in Medicine* **42**:24, 4418-4439. [Crossref]

63. Kan Chen, Qishuo Yin, Qi Long. 2023. Covariate-Balancing-Aware Interpretable Deep Learning Models for Treatment Effect Estimation. *Statistics in Biosciences* **34**. . [Crossref]

64. Denis Agniel, Boris P Hejblum, Rodolphe Thiébaut, Layla Parast. 2023. Doubly robust evaluation of high-dimensional surrogate markers. *Biostatistics* **24**:4, 985-999. [Crossref]

65. Philippe Boileau, Nina Ting Qi, Mark J van der Laan, Sandrine Dudoit, Ning Leng. 2023. A flexible approach for predictive biomarker discovery. *Biostatistics* **24**:4, 1085-1105. [Crossref]

66. Oliver Rojas Conde, Siddharth Misra, Rui Liu. Causal Inference for the Characterization of Microseismic Events Induced by Hydraulic Fracturing . [Crossref]

67. Oliver Rojas Conde, Siddharth Misra, Rui Liu. Use of Double Machine Learning for Microseismic Data Analysis: What Factors Influence Microseismic Events? . [Crossref]

68. Jules Gazeaud, Nausheen Khan, Eric Mvukiyehe, Olivier Sterck. 2023. With or without him? Experimental evidence on cash grants and gender-sensitive trainings in Tunisia. *Journal of Development Economics* **165**, 103169. [Crossref]

69. Byung-Kwan Lee, Junho Kim, Yong Man Ro. Mitigating Adversarial Vulnerability through Causal Parameter Estimation by Adversarial Double Machine Learning 4476-4486. [Crossref]

70. Aparup Khatua, Wolfgang Nejdl. Why do we Hate Migrants? 1-10. [Crossref]

71. Xiqian Wang, Yong Bian, Qin Zhang. 2023. The effect of cooking fuel choice on the elderly's well-being: Evidence from two non-parametric methods. *Energy Economics* **125**, 106826. [Crossref]

72. Waverly Wei, Maya Petersen, Mark J van der Laan, Zeyu Zheng, Chong Wu, Jingshen Wang. 2023. Efficient Targeted Learning of Heterogeneous Treatment Effects for Multiple Subgroups. *Biometrics* **79**:3, 1934-1946. [Crossref]

73. Igor Shuryak. 2023. Analysis of causal effects of 137Cs deposition on 137Cs concentrations in trees after the Fukushima accident using machine learning. *Journal of Environmental Radioactivity* **264**, 107205. [Crossref]

74. Jingcun Cao, Pradeep Chintagunta, Shibo Li. 2023. From Free to Paid: Monetizing a Non-Advertising-Based App. *Journal of Marketing Research* **60**:4, 707-727. [Crossref]

75. Chengchun Shi, Xiaoyu Wang, Shikai Luo, Hongtu Zhu, Jieping Ye, Rui Song. 2023. Dynamic Causal Effects Evaluation in A/B Testing with a Reinforcement Learning Framework. *Journal of the American Statistical Association* **118**:543, 2059-2071. [Crossref]

76. Kimia Kamal, Bilal Farooq. 2023. Debiased Machine Learning for Estimating the Causal Effect of Urban Traffic on Pedestrian Crossing Behavior. *Transportation Research Record: Journal of the Transportation Research Board* **2677**:7, 196-208. [Crossref]

77. Kevin Munger. 2023. Temporal validity as meta-science. *Research & Politics* **10**:3. . [Crossref]

78. Luke Guerdan, Amanda Coston, Kenneth Holstein, Zhiwei Steven Wu. Counterfactual Prediction Under Outcome Measurement Error 1584-1598. [Crossref]

79. Gareth Liu-Evans, Shalini Mitra. 2023. Formal sector enforcement and welfare. *International Tax and Public Finance* **30**:3, 706-728. [Crossref]

80. Ashkan Ertefaie, Nima S. Hejazi, Mark J. van der Laan. 2023. Nonparametric Inverse-Probability-Weighted Estimators Based on the Highly Adaptive Lasso. *Biometrics* **79**:2, 1029-1041. [Crossref]

81. Myoung-jae Lee, Goeun Lee, Jin-young Choi. 2023. Linear Probability Model Revisited: Why It Works and How It Should Be Specified. *Sociological Methods & Research* **188**. . [Crossref]

82. Fan Li, Peng Ding, Fabrizia Mealli. 2023. Bayesian causal inference: a critical review. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **381**:2247. . [Crossref]

83. Ting Ye, Marlena Bannick, Yanyao Yi, Jun Shao. 2023. Robust variance estimation for covariate-adjusted unconditional treatment effect in randomized clinical trials with binary outcomes. *Statistical Theory and Related Fields* **7**:2, 159-163. [Crossref]

84. Jong Hee Park, Soichiro Yamauchi. 2023. Change-Point Detection and Regularization in Time Series Cross-Sectional Data Analysis. *Political Analysis* **31**:2, 257-277. [Crossref]

85. Peter S. Lee, Ishita Chakraborty, Shrabastee Banerjee. Artificial Intelligence Applications to Customer Feedback Research: A Review 169-190. [Crossref]

86. Ricardo P. Masini, Marcelo C. Medeiros, Eduardo F. Mendes. 2023. Machine learning advances for time series forecasting. *Journal of Economic Surveys* **37**:1, 76-111. [Crossref]

87. Sriram Somanchi, Ahmed Abbasi, Ken Kelley, David Dobolyi, Ted Tao Yuan. 2023. Examining User Heterogeneity in Digital Experiments. *ACM Transactions on Information Systems* **73**. . [Crossref]

88. Kosuke Inoue, Teresa E. Seeman, Tamara Horwich, Matthew J. Budoff, Karol E. Watson. 2023. Heterogeneity in the Association Between the Presence of Coronary Artery Calcium and Cardiovascular Events: A Machine-Learning Approach in the MESA Study. *Circulation* **147**:2, 132-141. [Crossref]

89. Si Fang, Gibran Hemani, Tom G Richardson, Tom R Gaunt, George Davey Smith. 2023. Evaluating and implementing block jackknife resampling Mendelian randomization to mitigate bias induced by overlapping samples. *Human Molecular Genetics* **32**:2, 192-203. [Crossref]

90. Stijn Vansteelandt. 2023. Comment on" Estimands and their Estimators for Clinical Trials Impacted by the COVID-19 Pandemic: A Report from the NISS Ingram Olkin Forum Series on Unplanned Clinical Trial Disruptions". *Statistics in Biopharmaceutical Research* **15**:1, 112-115. [Crossref]

91. Caiwei Zhang, Zemin Zheng. 2023. Inference of subgroup-level treatment effects via generic causal tree in observational studies. *JUSTC* **53**:11, 1102. [Crossref]

92. Alexander Betts, Maria Flinder Stierna, Naohiko Omata, Olivier Sterck. 2023. Refugees welcome? Inter-group interaction and host community attitude formation. *World Development* **161**, 106088. [Crossref]

93. Sebastian Krantz. 2023. Mapping Africa's Infrastructure Potential with Geospatial Big Data, Causal ML, and XAI. *SSRN Electronic Journal* **66**. . [Crossref]

94. Kwok Ping Tsang, Zichao Yang. 2023. Agree to Disagree: Measuring Hidden Dissents in FOMC Meetings. *SSRN Electronic Journal* **142**. . [Crossref]

95. David M. A. Murphy. 2023. Alcohol, Labor, and Agricultural Productivity. *SSRN Electronic Journal* **21**. . [Crossref]

96. Miquel Oliu-Barton, Bary S. R. Pradelski, Nicolas Woloszko, Lionel Guetta-Jeanrenaud, Philippe Aghion, Patrick Artus, Arnaud Fontanet, Philippe Martin, Guntram B. Wolff. 2022. The effect of COVID certificates on vaccine uptake, health outcomes, and the economy. *Nature Communications* **13**:1. . [Crossref]

97. Francesco Decarolis, Cristina Giorgiantonio. 2022. Corruption red flags in public procurement: new evidence from Italian calls for tenders. *EPJ Data Science* **11**:1. . [Crossref]

98. Dmitry Arkhangelsky, Guido W Imbens. 2022. Doubly robust identification for causal panel data models. *The Econometrics Journal* **25**:3, 649-674. [Crossref]

99. Liqun Diao, Richard J. Cook. 2022. Nested doubly robust estimating equations for causal analysis with an incomplete effect modifier. *Canadian Journal of Statistics* **50**:3, 776-794. [Crossref]

100. Marcos Delprato, Alessia Frola, Germán Antequera. 2022. Indigenous and non-Indigenous proficiency gaps for out-of-school and in-school populations: A machine learning approach. *International Journal of Educational Development* **93**, 102631. [Crossref]

101. Edvard Bakhitov, Amandeep Singh. Causal Gradient Boosting: Boosted Instrumental Variable Regression 604-605. [Crossref]

102. Oliver Hines, Oliver Dukes, Karla Diaz-Ordaz, Stijn Vansteelandt. 2022. Demystifying Statistical Learning Based on Efficient Influence Functions. *The American Statistician* **76**:3, 292-304. [Crossref]

103. Zenobia T. Chan, Sophie Meunier. 2022. Behind the screen: Understanding national support for a foreign investment screening mechanism in the European Union. *The Review of International Organizations* **17**:3, 513-541. [Crossref]

104. Johannes Jakubik, Stefan Feuerriegel. 2022. Data-driven allocation of development aid toward sustainable development goals: Evidence from HIV/AIDS. *Production and Operations Management* **31**:6, 2739-2756. [Crossref]

105. Yuqian Zhang, Jelena Bradic. 2022. High-dimensional semi-supervised learning: in search of optimal inference of the mean. *Biometrika* **109**:2, 387-403. [Crossref]

106. Alexander Buchholz, Vito Bellini, Giuseppe Di Benedetto, Yannik Stein, Matteo Ruffini, Fabian Moerchen. Fair Effect Attribution in Parallel Online Experiments 77-83. [Crossref]

107. Elliott Ash, Daniel L. Chen, Sergio Galletta. 2022. Measuring Judicial Sentiment: Methods and Application to US Circuit Courts. *Economica* **89**:354, 362-376. [Crossref]

108. Helmut Wasserbacher, Martin Spindler. 2022. Machine learning for financial forecasting, planning and analysis: recent developments and pitfalls. *Digital Finance* **4**:1, 63-88. [Crossref]

109. Hongning Wang, Sanjun Ma. 2022. Preventing crimes against public health with artificial intelligence and machine learning capabilities. *Socio-Economic Planning Sciences* **80**, 101043. [Crossref]

110. Yiqi Cao, Qiao Kang, Baiyu Zhang, Zhiwen Zhu, Guihua Dong, Qinhong Cai, Kenneth Lee, Bing Chen. 2022. Machine learning-aided causal inference for unraveling chemical dispersant and salinity effects on crude oil biodegradation. *Bioresource Technology* **345**, 126468. [Crossref]

111. Yasufumi Gon, Daijiro Kabata, Hideki Mochizuki. 2022. Association between kidney function and intracerebral hematoma volume. *Journal of Clinical Neuroscience* **96**, 101-106. [Crossref]

112. ANDREAS FUSTER, PAUL GOLDSMITH-PINKHAM, TARUN RAMADORAI, ANSGAR WALTHER. 2022. Predictably Unequal? The Effects of Machine Learning on Credit Markets. *The Journal of Finance* **77**:1, 5-47. [Crossref]

113. Alexander Betts, Maria Flinder Stierna, Naohiko Omata, Olivier Sterck. Social Cohesion and Refugee-Host Interactions: Evidence from East Africa **5**, . [Crossref]

114. Qingliang Fan, Yu-Chin Hsu, Robert P. Lieli, Yichong Zhang. 2022. Estimation of Conditional Average Treatment Effects With High-Dimensional Data. *Journal of Business & Economic Statistics* **40**:1, 313-327. [Crossref]

115. Nima Hejazi, Kara Rudolph, Iván Díaz. 2022. medoutcon: Nonparametric efficient causal mediation analysis with machine learning in R. *Journal of Open Source Software* **7**:69, 3979. [Crossref]

116. Yuanzun Zhao, Tao Jin, Qian Lin. 2022. A Dynamic Allocation-Execution Decision Method Based on In-App Customer Lifetime Value Prediction. *SSRN Electronic Journal* **12**. . [Crossref]

117. Peter Lee, Ishita Chakraborty, Shrabastee Banerjee. 2022. AI Applications to Customer Feedback Research: A Review. *SSRN Electronic Journal* **1**. . [Crossref]

118. Victor Chernozhukov, Whitney K. Newey, Rahul Singh. 2022. Automatic Debiased Machine Learning of Causal and Structural Effects. *Econometrica* **90**:3, 967-1027. [Crossref]

119. Guido W. Imbens. 2022. Causality in Econometrics: Choice vs Chance. *Econometrica* **90**:6, 2541-2566. [Crossref]

120. Yuya Sasaki, Takuya Ura, Yichong Zhang. 2022. Unconditional quantile regression with high-dimensional data. *Quantitative Economics* **13**:3, 955-978. [Crossref]

121. Mehmet Canayaz. 2022. Financial Consequences of the Belt and Road Initiative. *SSRN Electronic Journal* **119**. . [Crossref]

122. Mohammed Alyakoob, Mohammad Saifur Rahman. 2022. Market Design Choices, Racial Discrimination, and Equitable Micro-Entrepreneurship in Digital Marketplaces. *SSRN Electronic Journal* **86**. . [Crossref]

123. Hao Jiang, Peng Qi, Jingying Zhou, Jack Zhou, Sharath Rao. A Short Survey on Forest Based Heterogeneous Treatment Effect Estimation Methods: Meta-learners and Specific Models 3006-3012. [Crossref]

124. Li-Jeng Huang, Shyh-Haur Chen, Duc-Hien Le. Fourier-Bessel series neural networks for classification 1-7. [Crossref]

125. Hongxiang Qiu, Alex Luedtke, Marco Carone. 2021. Universal sieve-based strategies for efficient estimation using machine learning tools. *Bernoulli* **27**:4. . [Crossref]

126. Yuqi Deng, Sudeeksha Murari. When a Voice Assistant Asks for Feedback: An Empirical Study on Customer Experience with A/B Testing and Causal Inference Methods 183-191. [Crossref]

127. Matthew C. Harding, Carlos Lamarche. 2021. Small Steps with Big Data: Using Machine Learning in Energy and Environmental Economics. *Annual Review of Resource Economics* **13**:1, 469-488. [Crossref]

128. Prodosh E. Simlai. 2021. Predicting owner-occupied housing values using machine learning: an empirical investigation of California census tracts data. *Journal of Property Research* **38**:4, 305-336. [Crossref]

129. Matthew J. Vowels, Necati Cihan Camgoz, Richard Bowden. Targeted VAE: Variational and Targeted Learning for Causal Inference 132-141. [Crossref]

130. Carl Bonander. 2021. A (Flexible) Synthetic Control Method for Count Data and Other Nonnegative Outcomes. *Epidemiology* **32**:5, 653-660. [Crossref]

131. Vasilis Syrgkanis, Greg Lewis, Miruna Oprescu, Maggie Hei, Keith Battocchi, Eleanor Dillon, Jing Pan, Yifeng Wu, Paul Lo, Huigang Chen, Totte Harinen, Jeong-Yoon Lee. Causal Inference and Machine Learning in Practice with EconML and CausalML 4072-4073. [Crossref]

132. Carl Bonander, Mikael Svensson. 2021. Using causal forests to assess heterogeneity in cost-effectiveness analysis. *Health Economics* **30**:8, 1818-1832. [Crossref]

133. Oliver Hines, Stijn Vansteelandt, Karla Diaz-Ordaz. 2021. Robust Inference for Mediated Effects in Partially Linear Models. *Psychometrika* **86**:2, 595-618. [Crossref]

134. Xin Huang, Hesen Li, Yihua Gu, Ivan S. F. Chan. 2021. Predictive Biomarker Identification for Biopharmaceutical Development. *Statistics in Biopharmaceutical Research* **13**:2, 239-247. [Crossref]

135. Michael C Knaus, Michael Lechner, Anthony Strittmatter. 2021. Machine learning estimation of heterogeneous causal effects: Empirical Monte Carlo evidence. *The Econometrics Journal* **24**:1, 134-161. [Crossref]

136. Yanbo Xu, Divyat Mahajan, Liz Manrao, Amit Sharma, Emre Kıcıman. Split-Treatment Analysis to Rank Heterogeneous Causal Effects for Prospective Interventions 409-417. [Crossref]

137. Yibo Guo. Signal Processing Based on Machine Learning Optical Communication 141-147. [Crossref]

138. Jushan Bai, Sung Hoon Choi, Yuan Liao. 2021. Feasible generalized least squares for panel data with cross-sectional and serial correlations. *Empirical Economics* **60**:1, 309-326. [Crossref]

139. Kathy Baylis, Thomas Heckelei, Hugo Storm. Machine learning in agricultural economics 4551-4612. [Crossref]

140. Lei Zhu, Zhuo Yu, Hui Zhan. 2021. Impact of Industrial Agglomeration on Regional Economy in a Simulated Intelligent Environment Based on Machine Learning. *IEEE Access* **9**, 20695-20702. [Crossref]

141. Ziwei CONG, Jia Liu, Puneet Manchanda. 2021. The Role of 'Live' in Livestreaming Markets: Evidence Using Orthogonal Random Forest. *SSRN Electronic Journal* **47**. . [Crossref]

142. Ruoxuan Xiong, Allison Koenecke, Michael Powell, Zhu Shen, Joshua T. Vogelstein, Susan Carleton Athey. 2021. Federated Causal Inference in Heterogeneous Observational Data. *SSRN Electronic Journal* **61**. . [Crossref]

143. Jungjun Choi, Xiye Yang. 2021. Convolution of Kernels and Recursive Bias Correction. *SSRN Electronic Journal* **4**. . [Crossref]

144. Andrea Naghi. 2021. Finite Sample Evaluation of Causal Machine Learning Methods: Guidelines for the Applied Researcher. *SSRN Electronic Journal* **113**. . [Crossref]

145. Vaagn Minasyan, Diana Mirzoyan, Alexander Sakhnov. 2021. Estimation of Average Treatment Effect on Residuals: Bias Derivation. *SSRN Electronic Journal* **2**. . [Crossref]

146. Kathryn N. Vasilaky, Aurélie Patricia Harou, Katherine Alfredo, Ishita Singh Kapur. 2021. What Works for Water Conservation? Evidence from a Field Experiment in India. *SSRN Electronic Journal* **29**. . [Crossref]

147. Mehmet Canayaz, Ilja Kantorovitch, Roxana Mihet. 2021. Privacy Laws and Value of Personal Data. *SSRN Electronic Journal* **54**. . [Crossref]

148. Judith Favereau. Bibliographie 263-283. [Crossref]

149. Guido W. Imbens. 2020. Potential Outcome and Directed Acyclic Graph Approaches to Causality: Relevance for Empirical Practice in Economics. *Journal of Economic Literature* **58**:4, 1129-1179. [Abstract] [View PDF article] [PDF with links]

150. Jong Hee Park, Byung Koo Kim. 2020. Why your neighbor matters: Positions in preferential trade agreement networks and export growth in global value chains. *Economics & Politics* **32**:3, 381-410. [Crossref]

151. Andres Algaba, David Ardia, Keven Bluteau, Samuel Borms, Kris Boudt. 2020. ECONOMETRICS MEETS SENTIMENT: AN OVERVIEW OF METHODOLOGY AND APPLICATIONS. *Journal of Economic Surveys* **34**:3, 512-547. [Crossref]

152. Adi Lin, Jie Lu, Junyu Xuan, Fujin Zhu, Guangquan Zhang. 2020. A Causal Dirichlet Mixture Model for Causal Inference from Observational Data. *ACM Transactions on Intelligent Systems and Technology* **11**:3, 1-29. [Crossref]

153. Hugo Storm, Kathy Baylis, Thomas Heckelei. 2020. Machine learning in agricultural and applied economics. *European Review of Agricultural Economics* **47**:3, 849-892. [Crossref]

154. Duncan Simester, Artem Timoshenko, Spyros I. Zoumpoulis. 2020. Targeting Prospective Customers: Robustness of Machine-Learning Methods to Typical Data Challenges. *Management Science* **66**:6, 2495-2522. [Crossref]

155. Marco Avella-Medina. Robust Methods for High-Dimensional Regression and Covariance Matrix Estimation 625-653. [Crossref]

156. Zenobia Chan, Sophie Meunier-Aitsahalia. 2020. Behind the Screen: Understanding National Support for a Foreign Investment Screening Mechanism in the European Union. *SSRN Electronic Journal* **91**. . [Crossref]

157. K. John McConnell, Stephan Lindner. 2019. Estimating treatment effects with machine learning. *Health Services Research* **54**:6, 1273-1282. [Crossref]

158. Adi Lin, Jie Lu, Junyu Xuan, Fujin Zhu, Guangquan Zhang. One-Stage Deep Instrumental Variable Method for Causal Inference from Observational Data 419-428. [Crossref]

159. Liangjun Su, Takuya Ura, Yichong Zhang. 2019. Non-separable models with high-dimensional data. *Journal of Econometrics* **212**:2, 646-677. [Crossref]

160. Susan Athey, Guido W. Imbens. 2019. Machine Learning Methods That Economists Should Know About. *Annual Review of Economics* **11**:1, 685-725. [Crossref]

161. Mario Molina, Filiz Garip. 2019. Machine Learning for Sociology. *Annual Review of Sociology* **45**:1, 27-45. [Crossref]

162. Zhuan Pei, Jörn-Steffen Pischke, Hannes Schwandt. 2019. Poorly Measured Confounders are More Useful on the Left than on the Right. *Journal of Business & Economic Statistics* **37**:2, 205-216. [Crossref]

163. Andres Algaba, David Ardia, Keven Bluteau, Samuel Borms, Kris Boudt. 2019. Econometrics Meets Sentiment: An Overview of Methodology and Applications. *SSRN Electronic Journal* **53**. . [Crossref]

164. Nicolaj Mühlbach. 2019. Tree-Based Methods: Consequences of Moving the US Embassy. *SSRN Electronic Journal* **10**. . [Crossref]

165. Sergio Galletta, Elliott Ash, Daniel L. Chen. 2019. Do Judicial Sentiments Affect Social Attitudes?. *SSRN Electronic Journal* **13**. . [Crossref]

166. Jungjun Choi, Xiye Yang. 2019. Bias Correction and Robust Inference in Semiparametric Models. *SSRN Electronic Journal* **62**. . [Crossref]

167. Mateus Souza. 2019. Predictive Counterfactuals for Event Studies with Staggered Adoption: Recovering Heterogeneous Effects from a Residential Energy Efficiency Program. *SSRN Electronic Journal* **105**. . [Crossref]

168. David M. A. Murphy, Dries Roobroeck, David R. Lee. 2019. Show and Tell: Causal Impacts of Field Days on Farmer Learning for Organic Inputs in Kenya. *SSRN Electronic Journal* **28**. . [Crossref]

169. Jingbo Wang, Yufeng Huang. 2019. Nonparametric Estimation of Price Elasticities: A Heterogeneous Treatment Effect Approach. *SSRN Electronic Journal* **47**. . [Crossref]

170. Anna Grodecka-Messi, Isaiah Hull. 2019. The Impact of Local Taxes and Public Services on Property Values. *SSRN Electronic Journal* **355**. . [Crossref]

171. Ahmed M. Alaa, Mihaela van der Schaar. 2018. Bayesian Nonparametric Causal Inference: Information Rates and Learning Algorithms. *IEEE Journal of Selected Topics in Signal Processing* **12**:5, 1031-1046. [Crossref]

172. Yinchu Zhu, Jelena Bradic. 2018. Significance testing in non-sparse high-dimensional linear models. *Electronic Journal of Statistics* **12**:2. . [Crossref]

173. Alexei Alexandrov, Russell Pittman, Olga Ukhaneva. 2018. Pricing of Complements in the U.S. Freight Railroads: Cournot Versus Coase. *SSRN Electronic Journal* **113**. . [Crossref]

174. Yingying Fan, Jinchi Lv, Jingbo Wang. 2018. DNN: A Two-Scale Distributional Tale of Heterogeneous Treatment Effect Inference. *SSRN Electronic Journal* **74**. . [Crossref]

175. Xiye Yang. 2018. Semiparametric Estimation in Continuous-Time: Asymptotics for Integrated Volatility Functionals With Small and Large Bandwidths. *SSRN Electronic Journal* **112**. . [Crossref]

176. Maria Cuellar, Edward Kennedy. 2018. A Nonparametric Projection-Based Estimator for the Probability of Causation, with Application to Water Sanitation in Kenya. *SSRN Electronic Journal* **43**. . [Crossref]

177. Andreas Fuster, Paul Goldsmith-Pinkham, Tarun Ramadorai, Ansgar Walther. 2017. Predictably Unequal? The Effects of Machine Learning on Credit Markets. *SSRN Electronic Journal* **110**. . [Crossref]

178. Monica Andini, Emanuele Ciani, Guido de Blasio, Alessio D'Ignazio, Viola Salvestrini. 2017. Targeting Policy-Compliers with Machine Learning: An Application to a Tax Rebate Programme in Italy. *SSRN Electronic Journal* **113**. . [Crossref]

179. Chen Liang, Yili Hong, Bin Gu. 2016. Effects of IT-Enabled Monitoring on Labor Contracting in Online Platforms: Evidence from a Natural Experiment. *SSRN Electronic Journal* **72**. . [Crossref]

180. Chen Liang, Yili Hong, Bin Gu. 2016. Effects of IT-Enabled Monitoring on Labor Contracting in Online Platforms: Evidence from a Natural Experiment. *SSRN Electronic Journal* **72**. . [Crossref]