

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

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

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

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,

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

The confounding factors Z affect the treatment variable D via the propensity score, $m_0(Z) := E[D \mid Z]$, 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

[‡]*Discussants:* Panagiotis Toulis, University of Chicago; Matthew Harding, Duke University; Hai Wang, Singapore Management University.

*Chernozhukov: Massachusetts Institute of Technology, 50 Memorial Drive, Cambridge, MA 02142 (e-mail: vchern@mit.edu); Chetverikov: University of California Los Angeles, 315 Portola Plaza, Los Angeles, CA 90095 (e-mail: chetverikov@econ.ucla.edu); Demirer: Massachusetts Institute of Technology, 50 Memorial Drive, Cambridge, MA 02142 (e-mail: mdemirer@mit.edu); Duflo: Massachusetts Institute of Technology, 50 Memorial Drive, Cambridge, MA 02142 (e-mail: duflo@mit.edu); Hansen: University of Chicago, 5807 S. Woodlawn Avenue, Chicago, IL 60637 (e-mail: chansen1@chicagobooth.edu); Newey: Massachusetts Institute of Technology, 50 Memorial Drive, Cambridge, MA 02142 (e-mail: wnewey@mit.edu). This material is based upon work supported by the National Science Foundation under Grant No. 1558636.

[†]Go to <https://doi.org/10.1257/aer.p.20171038> to visit the article page for additional materials and author disclosure statement(s).

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

and the *Neyman orthogonality condition*

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

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

$$(5) \quad \psi(W, \theta, \eta) := g(1, Z) - g(0, Z) + \frac{D(Y - g(1, Z))}{m(Z)} - \frac{(1 - D)(Y - g(0, Z))}{1 - m(Z)} - \theta,$$

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

$$(6) \quad \psi(W, \theta, \eta) := \frac{D(Y - g(0, Z))}{m} - \frac{m(Z)(1 - D)(Y - g(0, Z))}{(1 - m(Z))m} - \theta \frac{D}{m},$$

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 $\{1, \dots, N\}$ 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

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

that employ the machine learning estimators

$$\hat{\eta}_0(I_k^c) = \left(\hat{g}_0(0, Z; I_k^c), \hat{g}_0(1, Z; I_k^c), \hat{m}_0(Z; I_k^c), \frac{1}{N-n} \sum_{i \in I_k^c} D_i \right)',$$

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

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

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:

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

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 $k(i) := \{k \in \{1, \dots, K\} : i \in I_k\}$. 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 $\{1, \dots, N\}$ 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:

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

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 $s = 1, \dots, S$. Within each partition, we then obtain an estimate of the parameter of interest, $\hat{\theta}_s^*$. 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, $\hat{\theta}_0^{\text{Mean}}$ and $\hat{\theta}_0^{\text{Median}}$. Both of these reduce the sensitivity of the estimate for θ_0 to particular splits. $\hat{\theta}_0^{\text{Mean}}$ could be strongly affected by any extreme point estimates obtained in the different random partitions of the data, and $\hat{\theta}_0^{\text{Median}}$ is obviously much more robust. We note that asymptotically the specific random partition is irrelevant, and $\hat{\theta}_0^{\text{Mean}}$ and $\hat{\theta}_0^{\text{Median}}$ 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 $\hat{\theta}_0^{\text{Mean}}$, we propose adding an element that captures the spread of the estimated $\hat{\theta}_0^s$ around $\hat{\theta}_0^{\text{Mean}}$. 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 $\hat{\theta}_0^{\text{Median}}$. 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: Rejoinder." *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. Yibo Guo. Signal Processing Based on Machine Learning Optical Communication 141-147. [[Crossref](#)]
2. 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](#)]
3. 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](#)]
4. 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](#)]
5. 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](#)]
6. Xin Huang, Hesen Li, Yihua Gu, Ivan S. F. Chan. 2020. Predictive Biomarker Identification for Biopharmaceutical Development. *Statistics in Biopharmaceutical Research* 1-9. [[Crossref](#)]
7. Maria Cuellar, Edward H. Kennedy. 2020. A non-parametric projection-based estimator for the probability of causation, with application to water sanitation in Kenya. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **183**:4, 1793-1818. [[Crossref](#)]
8. Qingliang Fan, Yu-Chin Hsu, Robert P. Lieli, Yichong Zhang. 2020. Estimation of Conditional Average Treatment Effects With High-Dimensional Data. *Journal of Business & Economic Statistics* **87**, 1-15. [[Crossref](#)]
9. 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](#)]
10. Hugo Storm, Kathy Baylis, Thomas Heckelei. 2020. Machine learning in agricultural and applied economics. *European Review of Agricultural Economics* **47**:3, 849-892. [[Crossref](#)]
11. Michael C Knaus, Michael Lechner, Anthony Strittmatter. 2020. Machine learning estimation of heterogeneous causal effects: Empirical Monte Carlo evidence. *The Econometrics Journal* **72**. . [[Crossref](#)]
12. 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](#)]
13. 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](#)]
14. Marco Avella-Medina. Robust Methods for High-Dimensional Regression and Covariance Matrix Estimation 625-653. [[Crossref](#)]
15. 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](#)]
16. K. John McConnell, Stephan Lindner. 2019. Estimating treatment effects with machine learning. *Health Services Research* **360**. . [[Crossref](#)]
17. Liangjun Su, Takuya Ura, Yichong Zhang. 2019. Non-separable models with high-dimensional data. *Journal of Econometrics* **212**:2, 646-677. [[Crossref](#)]
18. Susan Athey, Guido W. Imbens. 2019. Machine Learning Methods That Economists Should Know About. *Annual Review of Economics* **11**:1, 685-725. [[Crossref](#)]

19. Mario Molina, Filiz Garip. 2019. Machine Learning for Sociology. *Annual Review of Sociology* 45:1, 27-45. [[Crossref](#)]
20. 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](#)]
21. Jungjun Choi, Xiye Yang. 2019. Bias Correction and Robust Inference in Semiparametric Models. *SSRN Electronic Journal* . [[Crossref](#)]
22. Sergio Galletta, Elliott Ash, Daniel L. Chen. 2019. Do Judicial Sentiments Affect Social Attitudes?. *SSRN Electronic Journal* . [[Crossref](#)]
23. Nicolaj Mühlbach. 2019. Tree-Based Methods: Consequences of Moving the US Embassy. *SSRN Electronic Journal* . [[Crossref](#)]
24. Andres Algaba, David Ardia, Keven Bluteau, Samuel Borms, Kris Boudt. 2019. Econometrics Meets Sentiment: An Overview of Methodology and Applications. *SSRN Electronic Journal* . [[Crossref](#)]
25. Mateus Souza. 2019. Predictive Counterfactuals for Event Studies with Staggered Adoption: Recovering Heterogeneous Effects from a Residential Energy Efficiency Program. *SSRN Electronic Journal* . [[Crossref](#)]
26. Anna Grodecka-Messi, Isaiah Hull. 2019. The Impact of Local Taxes and Public Services on Property Values. *SSRN Electronic Journal* . [[Crossref](#)]
27. Alexei Alexandrov, Russell Pittman, Olga Ukhaneva. 2018. Pricing of Complements in the U.S. Freight Railroads: Cournot Versus Coase. *SSRN Electronic Journal* . [[Crossref](#)]
28. Yingying Fan, Jinchi Lv, Jingbo Wang. 2018. DNN: A Two-Scale Distributional Tale of Heterogeneous Treatment Effect Inference. *SSRN Electronic Journal* . [[Crossref](#)]
29. Xiye Yang. 2018. Semiparametric Estimation in Continuous-Time: Asymptotics for Integrated Volatility Functionals With Small and Large Bandwidths. *SSRN Electronic Journal* . [[Crossref](#)]
30. 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* . [[Crossref](#)]
31. 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* . [[Crossref](#)]
32. Andreas Fuster, Paul Goldsmith-Pinkham, Tarun Ramadorai, Ansgar Walther. 2017. Predictably Unequal? The Effects of Machine Learning on Credit Markets. *SSRN Electronic Journal* . [[Crossref](#)]