# Comprehensive Causal Machine Learning*

Michael Lechner† Jana Mareckova

Swiss Institute for Empirical Economic Research
University of St.Gallen

February 14, 2025

Comments are very welcome.

**Abstract**

Uncovering causal effects in multiple treatment setting at various levels of granularity provides substantial value to decision makers. Comprehensive machine learning approaches to causal effect estimation allow to use a single causal machine learning approach for estimation and inference of causal mean effects for all levels of granularity. Focusing on selection-on-observables, this paper compares three such approaches, the modified causal forest (*mcf*), the generalized random forest (*grf*), and double machine learning (*dml*). It also compares the theoretical properties of the approaches and provides proven theoretical guarantees for the *mcf*. The findings indicate that *dml*-based methods excel for average treatment effects at the population level (ATE) and group level (GATE) with few groups, when selection into treatment is not too strong. However, for finer causal heterogeneity, explicitly outcome-centred forest-based approaches are superior. The *mcf* has three additional benefits: (i) It is the most robust estimator in cases when *dml*-based approaches underperform because of substantial selection into treatment; (ii) it is the best estimator for GATEs when the number of groups gets larger; and (iii), it is the only estimator that is internally consistent, in the sense that low-dimensional causal ATEs and GATEs are obtained as aggregates of finer-grained causal parameters.

**Keywords:** Causal machine learning, statistical learning, conditional average treatment effects, individualized treatment effects, multiple treatments, selection-on-observed-variables

**JEL classification:** C21, C87

**Correspondence to**: Michael Lechner or Jana Mareckova, Professors of Econometrics, Swiss Institute for Empirical Economic Research (SEW), University of St. Gallen, Switzerland, michael.lechner@unisg.ch, jana.mareckova@unisg.ch, `www.sew.unisg.ch`.

*Results from two research projects that were part of the National Research Programmes "Big Data" (NRP 75, `www.nrp75.ch`, grant number 407540_166999) and "Digital Transformation" (NRP 77, `www.nrp77.ch`, grant number 407740_187301) of the Swiss National Science Foundation (SNSF) were the basis of this paper. The theoretical part on the mcf (contained in Appendix A) is a revised version of the theoretical part of our unpublished "Modified Causal Forest" paper. We thank GPT-4 for some limited assistance in editing, and Phillip Heiler, Federica Mascolo, Fabian Muny, and Hannah Busshoff for helpful comments and suggestions. The paper was presented at research seminars at the Universities of Bern and the Collegio Carlo Alberto in Turino, at an invited session of the Annual Meeting of the German Economic Association in Berlin, at CompStat in Giessen, at American Causal Inference Conference in Seattle, at COMPIE in Amsterdam and at the Workshop on Causal Inference and Machine Learning in Groningen. We are grateful to participants for helpful remarks and interesting discussions.

†Michael Lechner is also affiliated with Örebro University, CEPR, London, CESIfo, Munich, IAB, Nuremberg, IZA, Bonn, and RWI, Essen.

# 1. Introduction

Machine learning (ML) has paved its way into academia and industry, impacting numerous fields from health care and finance to social sciences. At the core of the ML revolution is the remarkable predictive power of the methods. However, the growing debate around ML emphasizes that prediction does not imply causation. Going beyond mere predictive associations to identify cause-and-effect relationships is at the centre of most questions concerning the *effects* of policies, medical treatments, marketing campaigns, business decisions, etc. (see, e.g., Athey, 2017). The different focus of causal modelling calls for ML approaches that can reliably estimate causal effects, establishing causal machine learning (CML).

CML integrates principles of ML and causal inference. The causality literature provides conditions for identification and estimation of *effects*. Building on these conditions, causal problems can be transformed into specific prediction problems (see, e.g., Imbens and Wooldridge, 2009) for which ML methods are suitable (see, e.g., Hastie, Tibshirani, and Friedman, 2009, for an overview on ML methods). The flexibility of ML combined with careful use of data leads to versatile causal effect estimators that are also able to uncover effect heterogeneity (for an overview, see Athey and Imbens, 2017).

In recent years, researchers from different disciplines contributed to method development in CML. Although such multidisciplinary work on a common subject is an advantage from a scientific point of view, it raises the question which of these many proposed methods to use in any specific empirical study. In this paper, we focus on the relative performance of a specific class of estimators for a special empirical case, which is popular in the CML literature, namely (i) when causal effects are plausibly identified under unconfoundedness (i.e., selection-on-observables) in a multiple treatment setting, and (ii) the researcher is interested in aggregated average causal effects as well as their possibly fine-grained heterogeneity.

A key condition of any causal effect estimator to be attractive in applications is that it has acceptable proven statistical properties. Such theoretical guarantees and the implied ability to conduct inference are crucial, because, different to prediction settings, realisations from the true effects, the so-called 'ground truth', are unobservable. Therefore, they cannot be used to evaluate the performance of the specific estimation. Another important factor in practice is to keep the estimation of the many different causal parameters

sufficiently simple, for example, by using the same, or the same type of causal machine learner for all of them. This avoids the time-intensive task of tuning and monitoring many different estimators. Additionally, methods that can predict heterogeneous treatment effects on so far unseen data containing covariates only are also an advantage in empirical settings. We will call methods that (i) have available statistical inference, (ii) can estimate average and fine-grained effects using one type of ML framework, and (iii) can predict using covariates only, as Comprehensive Causal Machine Learners (CCMLs). Finally, it is advantageous to have *internal consistency* of the possibly many estimated effects, in the sense that effects at the higher aggregation levels are close to appropriately aggregated lower-level effects.[1] Internal consistency avoids scenarios in which conclusions across different aggregation levels might be contradictory.

In the light of these arguably desirable properties, we analyse three estimation principles that belong to the class of Comprehensive Causal Machine Learning. The methods are Debiased/Double Machine Learning (*dml*; Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins, 2018), the Generalized Random Forest (*grf*; Athey, Tibshirani, and Wager, 2019), and the Modified Causal Forest (*mcf*; Lechner, 2018). Only the latter fulfils the condition of internal consistency by construction.[2]

The intended contribution of this paper to the literature is two-fold: The first contribution is a large scale-simulation exercise to evaluate the finite-sample performance of all three CCMLs in many different settings to better understand which estimator may have advantages or disadvantages in certain situations and draw practical conclusions. The simulation assesses the CCMLs at three different levels of treatment effect granularity: (i) individual, (ii) group, and (iii) population level. Crucially, these simulations also cover the inference procedures that usually were absent from the small number of existing and far more limited simulation studies. The second contribution are the theoretical guarantees for the *mcf* and a comparison of the theoretical guarantees across the three different CCMLs. Additionally, similarities and differences between the three approaches are documented.

[1] A similar concept is known as *coherence* in hierarchical time series forecasting, where it led to development of methods that ensure coherent forecasts across collections of time series formed through aggregation. Athanasopoulos, Hyndman, Kourentzes, and Panagiotelis (2024) provide a review of such hierarchical time series methods.

[2] Within the classification of methods yielding coherent predictions for hierarchical time series, *mcf* would be classified as a single-level, bottom-up approach.

The results yield several practical recommendations: First, *dml*-based methods predominantly excel in estimation of average treatment effects (ATEs) or group treatment effects (GATEs) with fewer groups. Second, for finer causal heterogeneity, explicitly outcome-centred forest-based approaches are superior. Third, the *mcf* offers three additional benefits: (i) it is the most robust estimator even for the ATE in cases when *dml*-based approaches underperform because of substantial selection into treatment that needs to be corrected for; (ii) it outperforms both *dml* and *grf* in GATE estimation when the number of groups gets larger; (iii) it is the only estimator that is internally consistent.

The structure of the paper is as follows: Section 2 provides a review with a particular focus on the broad contribution of CML methods to the estimation of the ATE and more fine-grained conditional average treatment effects (CATEs) under unconfoundedness. After a general overview of CML methods, the section focuses on methods able to estimate effects at all levels of granularity and identifies the CCML methods. The section ends with an overview of CML simulation comparisons and applications. Section 3 introduces the notation of the potential outcome model, defines the parameters of interest, and states the identifying assumptions imposed under unconfoundedness. Section 4 discusses the three CCMLs. The large-scale simulation study is summarized in Section 5. Section 6 concludes. Appendix A provides proven theoretical guarantees for the *mcf*, its inference procedure, a comparison to the *grf*, as well as some new results for the *grf*. Implementational details of the simulation study are collected in Appendix B, while Appendix C holds its detailed results.

# 2. Review of causal machine learning estimation

## 2.1. Average treatment effect

For causal inference under the unconfoundedness assumption (see Rubin, 1974), the ATE represents one of the main parameters of interest capturing the overall impact of a treatment or an intervention. For ATE estimators, such as inverse probability weighting (IPW) and propensity score matching (see, e.g., Imbens, 2004), propensity score estimation plays a central role. Both estimators are consistent when a parametric model for the propensity score is correctly specified. Consequently, one of the first applications of ML methods in ATE estimation involved flexible propensity score estimation. Various ML ap-

proaches, such as Neural Networks, Support Vector Machines, Decision Trees (Westreich, Lessler, and Funk, 2010) or Boosting Models (McCaffrey, Ridgeway, and Morral, 2004; Westreich et al., 2010), have been proposed to improve the estimation of the propensity score.

However, even flexible ML methods may exhibit poor performance if they prioritize fitting the propensity score well instead of explicitly balancing the covariates across treatment groups, which is important to reduce the bias of the estimator for the causal parameter of interest. An estimator of the propensity score that explicitly aims to maximize covariate balancing is introduced in Imai and Ratkovic (2014). Alternatively, Graham, Pinto, and Egel (2012) improve IPW estimation by covariate balancing within the empirical likelihood framework, without explicit modelling of the propensity score.

Cannas and Arpino (2019) investigate the performance of matching and weighting estimators based on propensity scores obtained from Logistic Regression, Decision Tree, Bagging, Boosting, Random Forest, Neural Networks, and Naïve Bayes. Their simulation results indicate that Random Forests consistently outperform other methods, particularly in the context of IPW. Goller, Lechner, Moczall, and Wolff (2020) compare Random Forests, LASSO Logit Regression, and Probit estimation of the propensity score for a matching estimator in an active labour market policy setting, revealing that LASSO may yield more credible results than conventional propensity score estimation methods, but overall results were mixed.

Under the standard unconfoundedness assumptions, to be detailed in Section 3, ATE is alternatively identified as the expected difference between conditional expectations of the outcome variable for the treated and non-treated populations. This leads to regression-based estimation of the two conditional expectations. Assuming a correct specification of the outcome models, averaging their estimated differences yields a consistent ATE estimator. In the ML literature, tree-based methods such as Bayesian Additive Regression Trees (BART) (Hill, 2011) and Ensemble Methods (Austin, 2012) were introduced to obtain regression-based estimators of ATEs. However, issues like regularization bias (Athey, Imbens, and Wager, 2018) and slow convergence of ML methods render the outcome-based approach less popular.

Methods that combine propensity score-based and regression-based estimators to increase robustness to misspecification have gained popularity. Two related approaches pre-

vail in the literature: Double-robust (DR) estimators and Neyman-orthogonal scores. DR estimators utilize two nuisance functions, i.e., the propensity score and outcome models. They are consistent as long as at least one of the nuisance functions is correctly specified. Certain DR estimators were shown to be semi-parametrically efficient if both nuisance functions are correctly specified. Ideas to combine propensity score and outcome modelling in a double-robust way originally emerged in Robins, Rotnitzky, and Zhao (1994, 1995), leading to the Augmented Inverse Probability Weighting (AIPW) estimator for the ATE. Nevertheless, sensitivity to extreme propensity scores, as in the case of the IPW estimator, remains an issue.[3]

Approaches based on Neyman-orthogonality ensure that the moment conditions identifying the target parameters are not locally affected by small perturbances in the nuisance parameter estimates (Chernozhukov et al., 2018). Such an orthogonality property is leveraged, e.g., in Belloni, Chernozhukov, and Hansen (2014) for ATE estimation with binary treatment and in Farrell (2015) for multiple treatments. As Neyman-orthogonal scores mitigate first order bias arising from ML estimation of nuisance parameters (Bach, Kurz, Chernozhukov, Spindler, and Klaassen, 2024, p. 12), Neyman-orthogonality emerged as a key element for the Double Machine Learning (*dml*) framework introduced in Chernozhukov et al. (2018). *dml* is a generic framework for obtaining consistent estimators and valid inference for low-dimensional parameters. This framework combines Neyman-orthogonal scores with cross-fitting, a data partitioning technique that helps avoid further biases arising from the ML estimation of nuisance parameters. *dml* is well-suited for high-dimensional settings, as it operates under convergence rate conditions for nuisance parameter estimation that are achievable by many ML methods. Chernozhukov et al. (2018) apply the *dml* framework to ATE estimation using the DR score of Robins and Rotnitzky (1995)[4] and derive its statistical properties.

Another approach for ATE estimation is Targeted Minimum Loss Estimation (*tmle*), a general method that combines ML methods with a targeted updating step to reduce bias and variance of the estimated parameter of interest in semi-parametric and non-parametric models (van der Laan and Rubin, 2006). Among various *tmle* adaptations[5], cross-validated

[3] For potential benefits of trimming extreme propensity scores, see e.g., Huber, Lechner, and Wunsch (2013).

[4] Note that the DR score of Robins and Rotnitzky (1995) is Neyman-orthogonal.

[5] See van der Laan and Rose (2018) for a comprehensive overview.

*tmle*, *cv-tmle*, proposed by Zheng and van der Laan (2011), shares considerable similarities with *dml*. Both approaches require estimation of nuisance parameters that are combined in a way that yields a semi-parametrically efficient estimator of the ATE that is robust to small errors in nuisance parameter estimation under the same convergence rate conditions for nuisance parameter estimation and use of data splitting. Meanwhile *dml* utilizes the nuisance parameters in a Neyman-orthogonal score to mitigate the bias, *cv-tmle* combines them in an iterative estimation procedure addressing the bias and yielding quantities for each observation that can be averaged into an ATE. Despite this difference, *cv-tmle* and *dml* are asymptotically equivalent and can be expected to yield similar results in large samples under the same choice of methods estimating the nuisance parameters, as noted by Knaus (2021).

Automatic Debiased Machine Learning, introduced in Chernozhukov, Newey, and Singh (2022), offers another alternative for ATE estimation that can deal with bias stemming from using ML methods for estimation of nuisance parameters. The key idea is based on a "de-biased moment" utilizing an existence of a Riesz representer for parameters such as the ATE. The debiasing is automatic in the sense that it does not require any knowledge about the form of the Riesz representer. The Riesz representer is directly estimated within the procedure. In case of the ATE, the de-biased moment function corresponds to the DR moment function of Robins and Rotnitzky (1995). The estimation of the Riesz representer and nuisance parameters requires cross-fitting as in *dml*. The method is shown to yield a consistent and asymptotically normal estimator and provides a consistent asymptotic variance estimator.

## 2.2. Conditional average treatment effect

CML methods extend beyond ATE estimation and help uncover heterogeneity of the treatment effects through analysis of Conditional Average Treatment Effects (CATEs). CATEs quantify how the treatment affects an individual unit with specific low or large-dimensional characteristics. Prominent methods estimating CATEs include Causal Forests (Wager and Athey, 2018; Athey et al., 2019; Lechner, 2018), along with *dml*-based estimators yielding estimated components of the DR efficient score that are further regressed on the covariates as first proposed by van der Laan (2006). Alternative approaches include Meta-learners, which use multiple machine learning algorithms as "base learners" to es-

timate large-dimensional CATEs. The *X-learner*, introduced by Künzel, Sekhon, Bickel, and Yu (2019), is effective in cases with uneven treatment groups, while the *R-learner*, developed by Nie and Wager (2021), leverages the Robinson (1988) transformation to estimate heterogeneous causal effects. When interested in CATEs for specific subgroups, the above-mentioned quantities of the *tmle* iterative procedure can be averaged into a CATE, using the corresponding part of the sample for the estimation.[6] In case of a CATE for the finest granularity level, the *tmle* estimand depends on an observed outcome and cannot serve for a prediction on a new sample containing covariates only.[7]

The influential paper by Chernozhukov et al. (2018), averaging the estimated components of individual DR efficient scores of Robins and Rotnitzky (1995) to estimate ATE in a *dml* framework, spanned further methodological contributions, particularly in discovering heterogeneity along a chosen set of covariates, such as gender or age group. *dml*-based estimation of CATEs exploits the fact that the conditional expectation of the DR efficient score given such covariates identifies the respective CATEs. For low-dimensional sets of chosen covariates, OLS, Series, or Kernel Regressions of the estimated DR score components on the chosen set of covariates estimate the low-dimensional CATE and standard statistical inference applies, as shown in Semenova and Chernozhukov (2021), Zimmert and Lechner (2019), and Fan, Hsu, Lieli, and Zhang (2022). For a higher dimensionality of the set of the chosen covariates, Kennedy (2023) introduces the DR-learner that takes the components of a classic DR estimator of ATE and regresses them on all covariates to estimate a large-dimensional CATE and derives an upper bound on the DR-learner error relative to an oracle.

The foundations for heterogeneous treatment effect estimation by Causal Forests are laid in a seminal paper by Athey and Imbens (2016). This work introduces splitting rules for treatment effect estimation and honest data splitting, i.e. using different parts of data to build the tree and to estimate the parameters, to ensure valid inference for CATEs through tree-based methods. Subsequently, Wager and Athey (2018) develop a nonparametric Causal Forest algorithm based on honest trees and splitting rules maxi-

[6] Wei, Petersen, van der Laan, Zheng, Wu, and Wang (2023) note that the natural implementation of *tmle* "one-group-at-a-time" for CATE estimation with discrete groups might be computationally costly. As a remedy, they propose a *tmle* method that targets multi-subgroup treatment effects simultaneously.

[7] Based on Levy, van der Laan, Hubbard, and Pirracchio (2021) who introduces a *tmle* method for simultaneous estimation of ATE and variance of treatment effects, where estimated CATEs serve as nuisance parameters.

mizing heterogeneity in treatment effects across final leaves. The last step of the estimator of the possibly large-dimensional CATE is obtained by averaging individual tree estimates. Concerning its asymptotic guarantees, the Causal Forest is consistent and pointwise asymptotically normal under further regularity conditions. In a distinct but related approach, Athey et al. (2019) introduce the Generalized Random Forests (*grf*). This method is based on an honest forest with a gradient-based approximation of an estimator-specific splitting rule. The forest estimation determines importance weights of each observation for the parameters of interest. In other words, these weights determine a neighbourhood of observations that will contribute to the CATE estimation. Instead of estimating the CATE directly as an average across honest trees as in Wager and Athey (2018), the forest weights enter a set of local moment conditions identifying the CATE. This estimator is consistent and pointwise asymptotically Gaussian and generalizes beyond unconfoundedness.

Additionally, Lechner (2018) introduces a Modified Causal Forest (*mcf*) estimation procedure for ATEs and CATEs. In a multiple treatment framework, he proposes a weights-based inference procedure utilizing the weighted-outcome representation of forest estimator. *mcf* further differs from the other Causal Forests in using the mean squared error of the CATE directly to find the best split. Furthermore, it introduces a two-sample honesty, detailed in Subsection 4.3, for building the forest and the estimation of the effects. A simulation study in Lechner (2018) shows superior performance of *mcf* to the Causal Forest of Wager and Athey (2018). Bodory, Busshoff, and Lechner (2022) show that the *mcf* works in applications replicating results from several papers in different fields.

## 2.3. Comprehensive treatment effect estimation

Methods that (i) can estimate average and fine-grained effects using one type of ML framework, (ii) have available statistical inference for all the effects, and (iii) can predict using covariates only, are labelled as Comprehensive Causal Machine Learners (CCMLs). Some of the above-mentioned advances provide practitioners with methods able to estimate treatment effects across all different levels of granularity. Table 1 provides an overview of commonly used CML approaches for estimation of such treatment effects and evaluates them based on (i) available statistical inference, (ii) their ability to predict large-dimensional CATEs on new data (with covariate information only), and (iii) internal

consistency[8] of estimators. The overview reveals that *tmle* and Meta-learners do not have available statistical inference for all levels and that *tmle* cannot predict CATEs at the finest granularity level on a new data set that contains covariates only. Thus, these two methods are not considered to be CCMLs. On the other hand, *dml* and Causal Forests are approaches that fulfil the conditions for comprehensive treatment effect evaluation with available statistical inference and predictions for new data containing covariates only. For Causal Forests, there is a Bayesian alternative which will not be pursued further because its inference procedure is not based on repeated sampling inference.

Focusing on estimation at different levels of granularity via CCMLs, *dml*-based estimation operates via the estimated components of the DR efficient scores that are regressed on covariates to yield lower and higher-level aggregates of average treatment effects. Causal Forests, *grf* and *mcf*, provide in their first stage CATE estimates at the highest level of granularity. Using direct aggregation step of these CATEs, *mcf* leverages the weighted-outcome representation of forest predictions, while *grf* uses a variant of the AIPW score to yield higher level aggregations via a regression step. Table 1 summarizes the estimation at different levels of granularity for CCMLs and other methods.

Table 1: Properties of commonly used CML methods for treatment effect estimation

| | **Approach** | **Parameters** | | | **Statistical inference** | **CATE prediction depends only on covariates** | **Internal consistency** |
|---|---|---|---|---|---|---|---|
| | | *Dimension of CATE* | | *ATE* | | | |
| | | large | low | | | | |
| CCML | ***dml*** | Regress | Regress | Regress | Yes | Yes | No |
| CCML | ***grf*** | LocalGMM | Regress | Regress | Yes | Yes | No |
| CCML | ***mcf*** | MCF | Aggreg | Aggreg | Yes | Yes | Yes |
| | **Meta-L** | Meta-L | Aggreg* | Aggreg* | No** | Yes | Yes |
| | ***tmle*** | TMLE | TMLE | TMLE | Yes (low-dimen.) | No | Yes |
| | ***bcf*** | BCF | BayAggr | BayAggr | Bayesian | Yes | Yes |

Note: *dml*: Double / Debiased Machine Learning; *LocalGMM*: Local General Method of Moments; *mcf/MCF*: Modified Causal Forest; *bcf/BCF*: Bayesian Causal Forest; *tmle/TMLE*: Targeted Minimum Loss Estimation; *Meta-L*: Meta-learner.
*Regress*: Regression-based estimation; *Aggreg*: Aggregation of estimated IATEs; *BayAggr*: Bayesian Aggregation.
* Averaging of the finest large-dimensional CATEs for Meta-learners was implemented in, e.g., Salditt, Eckes, and Nestler (2024).
** However, see recent developments in constructing valid confidence intervals for Meta-learners based on conformal prediction in Alaa, Ahmad, and van der Laan (2024).

[8] An estimator is considered being internally consistent when effects at the higher aggregation levels are close to appropriately aggregated lower-level effects.

## 2.4. Simulations and empirical work

The literature on large scale simulation comparisons of finite sample properties of CML methods is very limited. Among notable contributions, Knaus, Lechner, and Strittmatter (2021) look at the finite-sample performance of selected machine learning estimators, covering *grf* and generic approaches that can be combined with any ML method, yielding large-dimensional CATE estimates based on pseudo-outcomes or modified covariates.[9] Higher level aggregates are simple averages of the corresponding CATEs. Using the Empirical Monte Carlo Study (EMCS) approach, many components of their simulations are based on real data. The results show that best-performing estimators at the large-dimensional CATE level produced most reliable estimates at higher aggregation levels, such as for the ATE, in terms of their mean squared error (MSE). In general, methods that utilized both outcome and treatment equations are among the best-performing methods, including *grf* with local centering. Caron, Baio, and Manolopoulou (2022) evaluate several Meta-learners in an EMCS based on health data. Their results highlight variability in performance that depends on the complexity of the data generating process. In a setting with a complex CATE, multitask learners designed to estimate both outcome equations jointly, such as Multitask Gaussian Process introduced in Alaa and van der Schaar (2017), perform best in terms of MSE. The *X-learner* performs the best in a setting with a simple CATE function and slightly unbalanced treatment groups. Another health-data-based EMCS of Wendling, Jung, Callahan, Schuler, Shah, and Gallego (2018) concentrates on large-dimensional CATE estimation in a common healthcare setting when outcomes are binary and rare. The compared methods include BART, *grf* with local centering, Causal Boosting, and Causal Multivariate Adaptive Regression Splines (MARS). The findings show that BART and Causal Boosting perform better across the scenarios, as evaluated by the root MSE in the conditional probability (risk) difference estimates. Except for one scenario involving highly heterogeneous treatment effects, the coverage rate for BART, *grf*, and causal MARS was close to its nominal level.[10]

[9] At the time the simulations of Knaus et al. (2021) were performed, the *mcf* was not yet available.

[10] To the best of our knowledge, this is the only study that also investigates the performance of the corresponding inference procedures. The simulation study in Wendling et al. (2018) differs from the simulation in Section 5 by using fixed covariates across replications. Despite this difference, their results are in line with the results in Section 5 confirming that high treatment heterogeneity negatively affects the coverage probability. Meanwhile DGPs with low treatment heterogeneity have coverage probability closer to nominal rates.

Empirical applications using CML for treatment estimation span across many fields and topics. The following non-exhaustive list illustrates the variety of applications ranging from active labour market policy (Davis and Heller, 2017; Bertrand, Crépon, Marguerie, and Premand, 2021; Pytka and Gulyas, 2021; Knaus, Lechner, and Strittmatter, 2022; Cockx, Lechner, and Bollens, 2023; Burlat, 2024), education (Knaus, 2021; Farbmacher, Kögel, and Spindler, 2021), social experiments (Athey and Wager, 2019; Strittmatter, 2023), energy use (Knittel and Stolper, 2021), natural resource rents (Hodler, Lechner, and Raschky, 2023), finance (Audrino, Chassot, Huang, Knaus, Lechner, and Ortega, 2024), and the dating market (Boller, Lechner, and Okasa, 2021) to medicine (Langenberger, Steinbeck, Schöner, Busse, Pross, and Kuklinski, 2023).

Regarding the comprehensive nature of the *dml* framework, Knaus (2022) demonstrates its flexibility in an active labour market programme evaluation setting. Rehill (2024) documents and evaluates some practices of using Causal Forests in applied work. Further illustrations of potential benefits of CML compared to traditional methods in empirical work is provided in Baiardi and Naghi (2024). They revisit estimates of average and heterogenous treatment effects from published observational studies and randomized control trials. Their results illustrate the potential of CML to capture non-linear confounding, its suitability in scenarios with large number of covariates relative to the sample size and its ability to discover treatment effect heterogeneity.

# 3. Causal framework

## 3.1. The potential outcome model

We use Rubin's potential outcome language (Rubin, 1974) to describe a multiple treatment model under unconfoundedness, also referred to as selection-on-observables or conditional independence (Imbens, 2000; Lechner, 2001). Let $D$ denote the treatment that may take a known number of $M$ different integer values from 0 to $M-1$. The (potential) outcome of interest that realises under treatment $d$ is denoted by $Y^d$. For each observation, we observe only the particular potential outcome corresponding to the observed treatment status, i.e., $y_i = \sum_{d=0}^{M-1} \underline{1}(d_i = d) y_i^d$, where $\underline{1}(\cdot)$ is the indicator function which

equals one if its argument is true.[11] There are two groups of variables to condition on, $\tilde{X}$ and $Z$. $\tilde{X}$ contains those covariates needed to correct for selection bias (confounders), while $Z$ contains variables that define (groups of) population members for which an average causal effect estimate is desired. $\tilde{X}$ and $Z$ may be discrete, continuous, or both. They may overlap in any way. The union of these two groups of variables is denoted by $X$, $X = (\tilde{X}, Z)$, $dim(X) = p$.[12]

Below, we investigate the following average causal effects:

$$
\begin{aligned}
&IATE(m, l; x) = E(Y^m - Y^l | X = x), \\
&GATE(m, l; z) = E(Y^m - Y^l | Z = z) = \int IATE(m, l; \tilde{x}, z) f_{\tilde{X}|Z=z}(\tilde{x}) d\tilde{x}, \\
&ATE(m, l) = E(Y^m - Y^l) = \int IATE(m, l; x) f_X(x) dx.
\end{aligned}
$$

The **I**ndividualized **A**verage **T**reatment **E**ffects (IATEs) measure the mean impact of treatment $m$ compared to treatment $l$ for units with covariates $x$. The IATEs represent the causal parameters at the finest aggregation level of the covariates. On the other extreme, the **A**verage **T**reatment **E**ffects (ATEs) represent the population averages. ATE is a classical parameter investigated in many empirical studies. The **G**roup **A**verage **T**reatment **E**ffect (GATE) parameters are in-between those two extremes with respect to their aggregation levels.[13] IATEs and GATEs are special cases of the already mentioned **C**onditional **A**verage **T**reatment **E**ffects (CATEs).

## 3.2. Identifying assumptions

The following set of assumptions identifies the causal effects discussed in the previous section (see Imbens, 2000; Lechner, 2001) (see Imbens, 2000, Lechner 2001):[14]

[11] If not obvious otherwise, capital letters denote random variables, and small letters their values. Small values subscripted by '$i$' denote the value of the respective variable of observation '$i$'.

[12] In all CCMLs $p$ is assumed to be finite.

[13] We presume that the analyst selects the variables $Z$ prior to estimation. However, the estimated IATEs may be analysed by methods picking $Z$ in a data-driven way to describe their dependence on certain covariates. See Section 6 in Lechner (2018) for more details. Note that Abrevaya, Hsu, and Lieli (2015) and Lee, Okui, and Whang (2017) introduce similar aggregated parameters that depend on a reduced conditioning set and discuss inference in their specific settings.

[14] To simplify the notation, we take the strongest form of these assumptions. Some parameters are identified under weaker conditions as well (for details see Imbens (2000, 2004), or Lechner (2001).

$$Y^d \perp\!\!\!\perp D|X = x, \qquad \text{(CIA)}$$

$$0 < P(D = d|X = x) = p_d(x), \quad \forall x \in \chi, \forall d \in \{0, \ldots, M-1\}; \qquad \text{(CS)}$$

$$Y = \sum_{d=0}^{M-1} \underline{1}(D = d)Y^d; \qquad \text{(Observation rule)}$$

The conditional independence assumption (CIA) holds if there are no covariates other than $X$ that jointly influence treatment and potential outcomes (for the values of $X$ that are in the support of interest, $\chi$). The common support (CS) assumption stipulates that for each value in $\chi$, there must be the possibility to observe all treatments. The stable-unit-treatment-value assumption (SUTVA, also referred to as observation rule or consistency condition) implies that the observed value of the treatment and the outcome does not depend on the treatment allocation of the other population members (ruling out spillover and treatment scale effects). Usually, to have an interesting interpretation of the effects, it is required that $X$ is not influenced by the treatment (exogeneity). In addition to these *identifying* assumptions, assume that a large random sample of size $N$ from the i.i.d. random variables $Y, X, D$, $(y_i, x_i, d_i)$, $i = 1, \ldots, N$, is available and that all necessary moments of these random variables exist.

If these assumptions hold, all IATEs are identified in the sense that they can be uniquely deduced from variables that have observable sample realisations (see Hurwicz, 1950):

$$\begin{aligned} IATE(m, l; x) &= E(Y^m - Y^l|X = x) \\ &= E(Y^m|X = x, D = m) - E(Y^l|X = x, D = l) \\ &= E(Y|X = x, D = m) - E(Y|X = x, D = l) \quad \forall x \in \chi, \forall m \neq l \in \{0, ..., M-1\}. \end{aligned}$$

Since the distributions used for aggregation, $f_{\tilde{X}|Z=z}(\tilde{x})$ and $f_X(x)$, relate to observable variables ($X$) only, they are identified as well (under standard regularity conditions). This in turn implies that the GATE and ATE parameters are identified.

# 4. Comprehensive approaches for estimation and inference

In this section, three comprehensive approaches for treatment effect estimation are presented. In the first two subsections, the underlying principles as well as the concrete

estimation and inference algorithms for Double Machine Learning (*dml)* and Generalized Random Forest (*grf)* are reviewed. The third subsection focuses on the Modified Causal Forest (*mcf)*. While the basic principles of the *mcf* are introduced in Lechner (2018), that paper does not contain explicit theoretical guarantees. Therefore, Appendix A.1 provides the proofs of consistency and asymptotic Gaussianity properties for IATEs, GATEs, and ATEs.

## 4.1. Double Machine Learning

*dml* estimation (Chernozhukov et al., 2018) is based on moment conditions with scores satisfying the identification condition as well as Neyman-orthogonality, yielding estimators that are robust to small estimation errors in nuisance parameters (as in Neyman, 1959). A Neyman-orthogonal score that identifies ATE, GATE, and IATE for treatments $m$ and $l$ under assumptions outlined in Subsection 3.2 is the DR score of Robins and Rotnitzky (1995):

$$\psi_{m,l}^{dml}(O;\theta_{m,l},\eta_{m,l}^{dml}(X)) = \Gamma_{m,l}^{dml}(O;\eta_{m,l}^{dml}(X)) - \theta_{m,l},$$
$$\Gamma_{m,l}^{dml}(O;\eta_{m,l}^{dml}(X)) = \mu_m(X) - \mu_l(X) + \frac{\underline{1}(D=m)(Y-\mu_m(X))}{p_m(X)} - \frac{\underline{1}(D=l)(Y-\mu_l(X))}{p_l(X)},$$

where $\mu_d(x) = E(Y|X=x, D=d)$, $\eta_{m,l}^{dml}(X) = (\mu_m(X), \mu_l(X), p_m(X), p_l(X))$ captures the nuisance parameters, and $\theta$ represents the treatment effect of interest. $O$ represents all observable variables, i.e., $O = (X, Y, D)$. Noting that the difference of the third and fourth term in this score has expectation zero conditional on $X$ at the true values of nuisance parameters, and letting $\eta^0$ and $\theta^0$ denote the true values of $\eta$ and $\theta$ respectively, $\psi_{m,l}^{dml}(O;\theta_{m,l},\eta_{m,l}^{dml}(X))$ identifies the different treatment effects:

$$\begin{aligned}
&E\left(\psi_{m,l}^{dml}(O;\theta_{m,l}^0,\eta_{m,l}^{dml,0})\right) = 0, && \theta_{m,l}^0 = ATE(m,l),\\
&E\left(\psi_{m,l}^{dml}(O;\theta_{m,l}^0(z),\eta_{m,l}^{dml,0})\,|Z=z\right) = 0, && \theta_{m,l}^0(z) = GATE(m,l;z),\\
&E\left(\psi_{m,l}^{dml}(O;\theta_{m,l}^0(x),\eta_{m,l}^{dml,0})\,|X=x\right) = 0, && \theta_{m,l}^0(x) = IATE(m,l;x), \text{ for all } x\in\chi.
\end{aligned}$$

Effectively, $ATE(m,l)$ is estimated as an average of the estimated component of the DR score, $\Gamma_{m,l}^{dml}\left(o_i;\hat{\eta}_{m,l}^{dml,-k(i)}(x_i)\right)$, across all observations, i.e.,

$$\widehat{ATE^{dml}}(m,l) = \hat{\theta}^{dml}_{m,l} = \frac{1}{N}\sum_{i=1}^{N} \Gamma^{dml}_{m,l}\left(o_i; \hat{\eta}^{dml,-k(i)}_{m,l}(x_i)\right).$$

This process involves the estimation of nuisance parameters, symbolized as $\hat{\eta}^{dml,-k(i)}_{m,l}(x_i)$, through $K$-fold cross-fitting. The latter ensures that for each observation $i$, the corresponding nuisance parameters are estimated without using that specific observation in their training data. Particularly, the elements of the vector $\hat{\eta}^{dml,-k(i)}_{m,l}(x_i)$ are estimated on $K-1$ folds not containing the observation $i$, which resides in the left-out fold $k$. This is indicated by the superscript $-k(i)$.

Neyman-orthogonal scores mitigate regularization bias in the estimation of the ATE, provided that the product of the convergence rates of the two nuisance parameter estimators within each treatment group is at least $\sqrt{N}$. This condition, which many ML methods achieve, allows for flexible estimation of nuisance parameters leading to estimators robust to small estimation errors in these nuisance parameters. The additional combination with cross-fitting is important to avoid overfitting when estimating the nuisance parameters and to subsequently guarantee $\sqrt{N}-$consistent estimation of the main parameters of interest. For technical details, refer to Chernozhukov et al. (2018). The variance estimator of the *dml* can be computed as the sample variance of the estimated DR scores.[15] The *dml* estimator asymptotically reaches the semiparametric efficiency bound of Hahn (1998).[16]

Building on the identification result, $GATE(m,l;z)$ can be estimated by regressing $\Gamma^{dml}_{m,l}\left(o_i; \hat{\eta}^{dml,-k(i)}_{m,l}(x_i)\right)$ on a low-dimensional vector $Z$ of pre-specified variables. Semenova and Chernozhukov (2021) provide statistical inference for the best linear predictor in this case, a method that is implemented in Section 5. Standard errors are estimated via heteroscedasticity-robust standard errors for asymptotically valid pointwise and uniform confidence bands. In general, the results in Semenova and Chernozhukov (2021) apply to linear projections onto a vector of basis functions. Under sufficient smoothness and richness of the basis functions, inference targets the $GATE(m,l;z)$ function. When basis functions are group indicators, it targets the corresponding group parameters. For continuous regressors, kernel regression offers a viable alternative, enabling the estimation of $GATE(m,l;z)$ and facilitating statistical inference, as demonstrated in Zimmert and Lechner (2019) and Fan et al. (2022).

[15] See Theorem 3.2 in Chernozhukov et al. (2018), and Theorem 1 and Remark 1 in Bach, Kurz, et al. (2024).

[16] See Theorem 5.1 in Chernozhukov et al. (2018).

In a similar fashion, a natural approach to obtain an estimator of $IATE(m,l;x)$ is by regressing $\Gamma_{m,l}^{dml}\left(o_i;\hat{\eta}_{m,l}^{dml,-k(i)}(x_i)\right)$ on the covariates.[17] Foster and Syrgkanis (2023) and Kennedy (2023) derive error bounds for this two-step procedure, referred to as DR-learner in Kennedy (2023). Particularly, the findings in Kennedy (2023) provide a theoretical foundation for the validity of inferential methods used in this context for linear smoothers satisfying the required stability condition and under further small-bias condition on $\Gamma_{m,l}^{dml}\left(o_i;\hat{\eta}_{m,l}^{dml,-k(i)}(x_i)\right)$. In practical applications, different techniques have been employed: Goller (2023) utilizes linear regression, while Knaus (2022) opts for a Random Forest approach.

## 4.2. Generalized Random Forests

*grf* (Athey et al., 2019, abbreviated as ATW19 in this section) extends the Random Forest into a nonparametric method estimating parameters identified by local moment conditions:

$$E\left(\psi^{grf}(O;\theta^0(x),\eta^{grf,0}(x))\,|X=x\right)=0 \quad \text{for all } x\in\chi,$$

where $\psi^{grf}(\cdot)$ is a score function identifying the true values of $\theta(x)$. As before, $\eta^{grf}(x)$ captures nuisance parameters, $\theta^0(x)$ and $\eta^{grf,0}(x)$ are the true values of $\theta(x)$ and $\eta^{grf}(x)$, and $O$ represents all observable variables. *grf* estimation of heterogeneous treatment effects is based on $\psi^{grf}(Y,\tilde{D},X;\theta(x),\eta_{IATE}^{grf}(x)) = \left(Y-\tilde{D}'\theta(x)-\eta_{IATE}^{grf}(x)\right)\left(1 \quad \tilde{D}'\right)'$, where $\tilde{D}$ is an $(M-1)\times 1$ vector containing dummy variables indicating whether treatment $d\in\{1,...,M-1\}$ was received, $\theta(x)$ is a parameter vector representing corresponding treatment effects (all treatment $d$ vs 0 pairs) for point $x$, and $\eta_{IATE}^{grf}(x)$ captures the intercept and all confounders in a single nuisance parameter.[18] Note that the score function in *dml* is based on a fully nonparametric DR score identifying treatment effects for all pairwise combinations of treatments, while the score function in *grf* stems from a partial linear model in which only treatment effects with respect to a reference treatment are identified. The reference treatment is the one left out from the vector $\tilde{D}$, here the control group with $d=0$. This is not too restrictive if effects that are conditioned on the treatment status are not of interest (as for IATE estimation), because the point estimates of the other treatment combinations can be obtained as $\theta_{m,l}(x)=\theta_{m,0}(x)-\theta_{l,0}(x)$.

[17] In Subsection 4.4 in Table 2, this step is called a smoothing step.

[18] This part is coined as an *intercept* term $c(x)$ in ATW19.

Building on the *local* generalized method of moments, *grf* estimates IATEs, $\theta(x)$, as

$$\left(\hat{\theta}^{grf}(x), \hat{\eta}^{grf}_{IATE}(x)\right) \in \underset{\theta(x),\eta(x)}{\arg\min} \left\| \sum_{i=1}^{N} w_i^{grf}(x) \psi^{grf}(o_i; \theta(x), \eta^{grf}_{IATE}(x)) \right\|_2,$$

where the weights, $w_i^{grf}(x)$, are obtained from an honest Random Forest whose gradient-based splitting rule is designed to maximize treatment heterogeneity in the daughter leaves. The required *Honesty* assumption involves a data partitioning strategy to prevent overfitting. The subsample drawn for each tree is further split into two halves: one for building the tree structure and the other one for estimating the parameters of interest – in this case, the tree weights which are subsequently aggregated into forest weights. As all observations alter between the two halves across trees when subsampled, we call this procedure 'one-sample honesty'. The forest weights, summing up to 1, can be seen as a measure of relevance of the observation $i$ for the estimation of $\theta(x)$. Given the forest weights $w_i^{grf}(x)$, the solution to the given optimization problem is:

$$\begin{aligned}\hat{\theta}^{grf}(x) &= \widehat{IATE}^{grf}(x) \\ &= \left(\sum_{i=1}^{N} w_i^{grf}(x) \left(\tilde{d}_i - \bar{\tilde{d}}_w\right) \left(\tilde{d}_i - \bar{\tilde{d}}_w\right)'\right)^{-1} \left(\sum_{i=1}^{N} w_i^{grf}(x) \left(\tilde{d}_i - \bar{\tilde{d}}_w\right) (y_i - \bar{y}_w)\right),\end{aligned}$$

where $\bar{\tilde{d}}_w = \sum_{i=1}^{N} w_i^{grf}(x)\tilde{d}_i$ and $\bar{y}_w = \sum_{i=1}^{N} w_i^{grf}(x) y_i$.[19]

As the splits are chosen to maximize treatment heterogeneity across daughter leaves, to mitigate confounding effects at early splits, ATW19 recommend to partial out effects of the covariates $X$ on the outcome $Y$ and treatment assignment $D$ as in Robinson (1988). This means building the forest using locally centred values $y_i^{cent} = y_i - \hat{\mu}^{(-i)}(x_i)$ and $\tilde{d}_i^{cent} = \tilde{d}_i - \hat{p}_d^{(-i)}(x_i)$ for all observations, where $\mu(x) = E(Y|X = x)$ and the superscript $(-i)$ denotes that the prediction did not use observation $i$. The best practice is to estimate $\mu(x)$ and $p_d(x)$, e.g., by $K$-fold cross-fitting as mentioned in ATW19. However, the available *grf* package performs local centring by subtracting out-of-bag predictions[20] obtained from regression forests specifically trained to predict the outcome variable $Y$ and the probabilities of treatment assignment, given the covariates $X$. Despite its computa-

[19] Compare with eq. (19) in ATW19.

[20] In ATW19 out-of-bag predictions are called leave-one-out predictions as the final prediction for observation $i$ averages predictions that are obtained from trees that did not use observation $i$ for splitting.

tional attractiveness, this implementation violates the requirement of complete separation of training and prediction data consistent with the theoretical results. Also, the simulation results in Appendix C.1 reveal the necessity of $K$-fold cross-fitting over out-of-bag predictions for local centring to remove bias.

The asymptotic results for $\hat{\theta}^{grf}(x)$ are based on their linear approximation motivated by the method of influence functions. ATW19 prove that the two are coupled and asymptotic properties of one apply to the other. Since the linear approximation can be interpreted as an honest Causal Forest estimator introduced in Wager and Athey (2018), their asymptotic result can be applied to establish asymptotic Gaussianity of $\hat{\theta}^{grf}(x)$, given that $\theta(x)$ and $\eta^{grf}_{IATE}(x)$ are consistently estimated.[21] The linear approximation is also leveraged to derive an estimator of the pointwise standard errors for $\hat{\theta}^{grf}(x)$. Derivation of the variance of the linear approximation yields two terms. One term can be in general consistently estimated by means of regressions and is a constant in case of $IATE(x)$. The other can be seen as an output of a regression forest with weights $w_i^{grf}(x)$ and the score function as outcome variable. ATW19 propose to estimate this term by a variant of a so-called bootstrap of little bags. This method was introduced in Sexton and Laake (2009). ATW19 motivate it by the observation that "*building confidence intervals via half-sampling – whereby evaluating an estimator on random halves of the training data to estimate its sampling error – is closely related to the bootstrap Efron (1982)*". A computationally efficient implementation to estimate within one forest, both the forest weights and the element of the variance by half-sampling, requires drawing for each small bag of trees a random half of the sample that is available for further subsampling. Honesty additionally requires the trees to be built on half of the subsample, and the weights $w_i^{grf}(x)$ to be computed on the half that was not used to build the tree. Thus, each $w_i^{grf}(x)$ is estimated on approximately a quarter of all the trees in the forest when the subsampling rate is close to 1.[22] Consistent estimation of the variance term yields asymptotically valid confidence intervals for the IATE at point $x$.

Like *dml*, *grf* offers comprehensive estimation of treatment effects at all levels of granularity. However, unlike *dml*, which estimates all effects in two steps by estimating the $\Gamma^{dml}_{m,l}\left(o_i; \hat{\eta}^{dml,-k(i)}_{m,l}(x_i)\right)$ first, and subsequently regressing them on a constant, group indi-

[21] This can be achieved, e.g., via honest forests as outlined in ATW19 in Theorem 3.

[22] Figure A.1 in Appendix A.3 captures the *grf* procedure graphically.

cators or regressors to obtain ATE, GATE and IATE estimators, *grf* takes a more direct approach. Specifically, *grf* estimates IATEs "directly", while ATE and GATEs are estimated through additional regressions. The *grf* package implements ATE estimation (treatment $m$ vs no treatment) by a variant of the AIPW estimator, plugging in the estimates from IATE estimation for the nuisance parameters:[23]

$$\widehat{ATE^{grf}}(m,0) = \hat{\theta}^{grf}_{m,0} = \frac{1}{N}\sum_{i=1}^{N} \Gamma^{grf}_{m,0}(o_i; \hat{\eta}^{grf}_{ATE(m,0)}(x_i)),$$

$$\Gamma^{grf}_{m,0}(o_i; \hat{\eta}^{grf}_{ATE(m,0)}(x_i)) = \hat{\theta}_{m,0} + \frac{\underline{1}(d_i = m)}{\hat{p}_m(x_i)}(y_i - \hat{\mu}_m(x_i)) - \frac{\underline{1}(d_i = 0)}{\hat{p}_0(x_i)}(y_i - \hat{\mu}_0(x_i)),$$

$$\text{where} \quad \hat{\mu}_0(x) = \hat{\mu}(x_i) - \sum_{d>0} \hat{p}_d(x_i)\hat{\theta}_{d,0}(x_i)$$

$$\text{and} \quad \hat{\mu}_m(x_i) = \hat{\mu}_0(x_i) + \hat{\theta}_{d,0}(x_i).$$

As default, the *grf* package estimates the nuisance parameters $\mu(x), p_1(x), \ldots, p_{M-1}(x)$ by out-of-bag predictions from Random Forests, and $\theta_{1,0}(x), \ldots, \theta_{M-1,0}(x)$ by a weighted local moment condition with forest-based weights. Derivations in Appendix A.4 show the equivalence of the *grf* score $\psi^{grf}_{m,0}(O; \theta_{m,0}, \eta^{grf}_{ATE(m,0)}(X)) = \Gamma^{grf}_{m,0}(O; \eta^{grf}_{ATE(m,0)}(X)) - \theta_{m,0}$ to the DR score of Robins and Rotnitzky (1995) for the ATE. It can therefore be expected that the *grf* estimator of the ATE is asymptotically efficient when all nuisance parameters are consistently estimated. Due to a different set of nuisance parameters, consistent estimation of the ATE is robust only to misspecifications of $\mu(x)$ and $\theta_{1,0}(x), \ldots, \theta_{M-1,0}(x)$ (see Appendix A.4). Consistent estimation of the ATE based on the above *grf* score therefore requires consistent estimation of the propensity scores. Appendix A.4 also shows Neyman-orthogonality of the *grf* score for the ATE, guaranteeing that the moment function is locally insensitive to errors in the nuisance parameter estimates. To obtain the GATE estimators, $\Gamma^{grf}_{m,0}(o_i; \hat{\eta}^{grf}_{ATE}(x_i))$ are regressed on $Z$, as in *dml*. Inference for the ATE and GATEs is conducted as in *dml*.

## 4.3. Modified Causal Forest

Lechner (2018, abbreviated as L18 in this section) builds on Wager and Athey (2018, abbreviated as WA18) and introduces the Modified Causal Forest estimator, *mcf*. One

[23] The equations for estimating the ATE are reconstructed from the *grf* R package (version 2.3.1). The derivation involves code implemented in the functions "get_scores.R" and "average_treatment_effect.R".

of the procedures proposed by WA18 builds trees by taking the outcome $Y$ as dependent variable and finding splits that maximize treatment effect heterogeneity. The tree estimator of an IATE at point $x$ is obtained from the final leaf containing $x$ by differencing the average outcomes of the treatment groups evaluated on the subsample that was not used to build the tree. The final forest estimator of the IATE at point $x$ is the average of all tree estimates. This procedure works well in an experimental design with heterogeneous treatment effects but may do poorly in the presence of confounding (see WA18). This observation led to further extensions addressing the confounding issue.

*grf* introduces local centring inside its Random Forest to remove the confounding effects by transforming the data when searching for splits that maximize effect heterogeneity. *mcf* takes a different approach. Motivated by the fact that in the presence of selection bias the difference of the outcome means of treated and controls within all leaves will not correspond to the means of the true effects, *mcf* proposes a splitting criterion targeting the minimisation of the MSE of IATE directly. In addition, a penalty is added penalizing squared differences between propensity scores. Taken together, the chosen split is predictive for $Y$, the differences of $Y$ across treatment groups, and the conditional treatment assignment probabilities.

The IATE is identified as the difference of two outcome regressions $\mu_d(x)$ in the different treatment groups. The splitting criterion of the *mcf* is therefore based on minimizing the sum of the expected MSEs at a given point $x$ over all unique treatment combinations, i.e., all IATEs relating to different treatment pairs are of equal importance:

$$\overline{MSE}_x = \sum_{m=1}^{M-1} \sum_{l=0}^{m-1} MSE\left(\widehat{IATE}(m,l;x)\right),$$
$$MSE\left(\widehat{IATE}(m,l;x)\right) = MSE^m(x) + MSE^l(x) - 2MCE^{m,l}(x),$$

where $MSE^d(x) = E\left(\hat{\mu}_d(x) - \mu_d^0(x)\right)$ and $MCE^{m,l}(x) = E\left(\hat{\mu}_m(x) - \mu_m^0(x)\right)\left(\hat{\mu}_l(x) - \mu_l^0(x)\right)$.

For finding splits based on this criterion, the $MSE^d$ and $MCE^{m,l}$ elements need to be estimated in the daughter leaves. Let $N^d_{S(x)}$ denote the number of observations with treatment value $d$ in a certain stratum (leaf) $S(x)$, defined by the values of the covariates $x$. Then a 'natural' choice to estimate the $MSE^d$'s in leaf $S(x)$ is:

$$\widehat{MSE^d_{S(x)}} = \frac{1}{N^d_{S(x)}} \sum_{i=1}^{N} \underline{1}(x_i \in S(x)) \underline{1}(d_i = d) \left(\hat{y}^d_{S(x)} - y_i\right)^2,$$

where $\hat{y}^d_{S(x)}$ is an average of the observed outcomes in treatment $d$ in leaf $S(x)$. To compute the $MCE$, the *mcf* uses the closest neighbour (in terms of similarity w.r.t. $x$) available in the other treatment (which is denoted by $\tilde{y}_{(i,d)}$ below).[24]

$$\widehat{MCE^{m,l}_{S(x)}} =$$
$$\frac{1}{N^m_{S(x)} + N^l_{S(x)}} \sum_{i=1}^{N} \underline{1}(x_i \in S(x)) \left(\underline{1}(d_i = m) + \underline{1}(d_i = l)\right) \left(\hat{y}^m_{S(x)} - \tilde{y}_{(i,m)}\right) \left(\hat{y}^l_{S(x)} - \tilde{y}_{(i,l)}\right),$$

$$\tilde{y}_{(i,d)} = \begin{cases} y_i, & \text{if } d_i = d, \\ \tilde{y}_{(i,d)}, & \text{if } d_i \neq d. \end{cases}$$

Analysing the estimator of $\overline{MSE}_x$ reveals that its minimization favours splits maximizing the differences of all $\hat{y}^d_{S(\cdot)}$ between the daughter leaves. Additionally, the splits favour large differences between $\hat{y}^m_{S(\cdot)}$ and $\hat{y}^l_{S(\cdot)}$ in the same leaf. In case of a binary treatment, this is equivalent to maximising treatment effect heterogeneity, as in WA18 and ATW19, under no selection into treatment. However, for multiple treatments, the *mcf* asymptotically targets treatment effect heterogeneity across daughter leaves while simultaneously increasing treatment outcome variability within leaves. Overlooking this dual focus may lead to splits that, while maximizing treatment effect heterogeneity, fail to minimise the sample MSE of parameters within leaves. Thus, *mcf* utilizes a splitting strategy that jointly considers all treatment combinations.

Although asymptotically confounding should be taken care of by the splitting rule derived above, the *mcf* adds a penalty term to the splitting criterion as an additional safeguard. As before, let denote $S(x')$ and $S(x'')$ denote the values of the covariates in the daughter leaves resulting from splitting some parent leaf. L18 proposes to add the following penalty to a combination of the two 'final' MSEs in the daughter leaves:

$$penalty(x', x'') = \lambda \left\{ 1 - \frac{1}{M} \sum_{d=0}^{M-1} \left(P(D = d|X \in S(x')) - P(D = d|X \in S(x''))\right)^2 \right\}, \lambda > 0.$$

The probabilities, local estimates of the propensity score, are computed as relative shares

[24] Implementational details on finding the closest neighbours for the MCE estimation are in Appendix B.3.1.

of the respective treatments in the potential daughter leaves. The penalty term is zero if the split leads to a perfect treatment prediction of the treatment probabilities. It reaches its maximum value, $\lambda$, when all probabilities are equal. Thus, the algorithm prefers a split that is also predictive for $P(D = d\,|X = x)$. Of course, the choice of the exact form of this penalty function is arbitrary. Furthermore, there is the issue of how to choose $\lambda$ (*without* expensive additional computations) which is taken up again in Subsubsection 5.2.2.

Each tree is built on a subsample drawn from one half of the data, while the other half serves for the estimation of the IATE at point $x$. We name this procedure 'two-sample honesty' as data points used to place the splits ("split-construction data") will not be used to estimate the treatment effects ("leaf-populating data") and vice versa.[25] The final estimator based on $B$ trees takes a form:

$$\widehat{IATE^{mcf}}(m,l;x) = \hat{\theta}^{mcf}_{m,l}(x) = \frac{1}{B}\sum_{b=1}^{B}\sum_{i=1}^{N/2} w^{mcf}_{i,b}(d_i,x_i;x,m,l)y_i = \sum_{i=1}^{N/2} w^{mcf}_{i}(d_i,x_i;x,m,l)y_i,$$

where $w^{mcf}_{i,b}(d_i,x_i;x,m,l) = \left(\frac{\underline{1}(d_i=m)}{N^m_{S_b(x)}} - \frac{\underline{1}(d_i=l)}{N^l_{S_b(x)}}\right)\underline{1}(x_i \in S_b(x))$ represent the weights for the $IATE(x)$ estimate in tree $b$. This estimator is differencing the average outcomes of the treatment groups evaluated on the "leaf-populating" subsample in each tree leaf containing point $x$, $S_b(x)$, and subsequently averages these differences across trees. As in the *grf*, the *mcf* forest weight $w^{mcf}_i(d_i,x_i;x,m,l)$ captures how important observation $i$ is for estimation of the parameter of interest at point $x$. While the *mcf* weights weigh the observed outcome directly, *grf* applies the forest weights in a locally weighted moment function. There is also a difference in the number of weights estimated in each method due to different honesty concepts. For a given $B$ and $N$, *grf* estimates $N$ weights on approximately $B/4$ trees, while *mcf* estimates $N/2$ weights on $B$ trees as illustrated in Appendix A.3 in Figures A.1 and A.2. Thus, *grf* has an advantage in smoothing over more observations, while *mcf* has an advantage in the precision of the weights, potentially influencing the finite sample properties of both methods.

The *mcf* estimators of GATE and ATE are averages of corresponding group or sample $\widehat{IATE^{mcf}}(m,l;x_j)$'s. By construction, they take the form of weighted observed outcomes. The *mcf* GATE estimator incorporates into the weighted average not only observed outcomes of the observations in the pre-specified group but potentially also observed outcomes

[25] See Figure A.2 in Appendix A.3 for a graphical representation.

of close neighbouring groups [26], defined by the non-zero forest weights in the IATEs.

Appendix A.1 contains the (so-far missing) theoretical guarantees for the *mcf*. Under similar assumptions as in WA18, the *mcf* estimators of the IATEs, GATEs, and ATEs are consistent and pointwise asymptotically normal. The *mcf* inference procedures exploit the fact that *mcf* predictions, like for many Random Forest based estimators, can be computed as weighted means of $Y$. The details of this weights-based inference are presented in Appendix A.2, while Appendix A.3 explains some of the difference to the *grf* at the implementation level.

## 4.4. Overview

This section summarizes the statistical guarantees of the three CCMLs for estimation of ATE, GATE, and IATE. For the ATE, *dml* and *grf* provide the strongest guarantees, as they are based on the DR efficient score, and are expected to perform well in terms of bias, MSE and inference in the simulations.[27] The *mcf* is also expected to perform well but with a somewhat larger MSE as it does not necessarily reach the efficiency bound.

Regarding GATE, all methods guarantee consistency and asymptotic normality. *dml* and *grf* differ from the *mcf* in the utilization of the observed data points. While *dml* and *grf* compute the average of estimated components of the DR score for observations in the corresponding group of interest, *mcf* builds a weighted average of $Y$'s, potentially incorporating information from groups close to the group of interest as determined by the forest structure, as mentioned in Subsection 4.3. This will affect the efficiency of the GATE estimators in different scenarios as captured in the following proposition.

**PROPOSITION 1.** Let $J$ denote the number of groups for which GATEs are to be estimated and let $\Pi$ be a $J \times 1$ vector containing the non-zero population shares of observations in the corresponding groups. Suppose that the total number of groups increases from $J$ to $J'$ and the population share of group $j$ under $J$ groups, $\pi_j$, decreases to $\pi_{j'}$ under $J'$ groups. Then, the asymptotic variances of the efficient *dml* and *grf* GATE

[26] This in general depends on whether variables $Z$ are among the splitting variables, type of $Z$, and the sample size. E.g., in settings with discrete $Z$ and when some trees do not (yet) split on all groups defined by a single value of $Z$, the observed outcomes from other groups that populate the same final leaves as $x_j$'s from group $z$ are incorporated into the weighted average. When all trees split on all groups, only observed outcomes from group $z$ are incorporated into the weighted average.

[27] Assuming a certain quality of the nuisance parameter estimators as mentioned in Subsection 4.1 and Subsection 4.2

estimators increase for each group $j$ within $J$ proportionally to $\pi_j/\pi_{j'}$, while the variance of *mcf* estimator of the GATE does not depend on the number of groups and remains stable. This implies $Var\left(\widehat{GATE}^{mcf}\left(m,l;z_j\right)\right) \leq Var\left(\widehat{GATE}^{dml/grf}\left(m,l;z_j\right)\right)$ for large $J$.

For the IATE, there are no specific statistical guarantees for *dml*. There is an upper bound on the error of the DR-learner providing a general guarantee. *grf* and *mcf* guarantee pointwise consistency and asymptotic normality and are both expected to yield lower MSEs in the simulations than *dml*-based DR-learner. For *grf*, the convergence rate is known and correspondingly slower due to the non-parametric nature of the estimator. Its speed depends on the subsampling rate which is implicitly restricted by the forest parameters (see Theorem 5 in ATW19). All the above-mentioned conjectures are captured in Table 2.

Table 2: Selected statistical properties of the CCMLs

| **Approach** | *ATE* | *GATE* | *IATE* |
|---|---|---|---|
| ***dml*** | Consistent, asymptotically normal, asymptotically efficient | Consistent, asymptotically normal, uniform convergence | Depends on implementation of the smoothing (regression) step |
| ***grf*** | Consistent, asymptotically normal, asymptotically efficient | Consistent, asymptotically normal, uniform convergence | Consistent, asymptotically normal, pointwise convergence known (slow) convergence rate |
| ***mcf*** | Consistent, asymptotically normal | Consistent, asymptotically normal, pointwise convergence<br><br>More efficient for large number of groups (compared to *dml* and *grf*) | Consistent, asymptotically normal, pointwise convergence |

# 5. Monte Carlo study

## 5.1. Concept

In this section, we compare the finite-sample performance of CCMLs in estimating average effects across three specified aggregation levels in a Monte Carlo simulation. This comparison includes evaluating both the precision of point estimates and the robustness of inference procedures.

In executing this Monte Carlo approach, we meticulously construct artificial datasets, varying numerous elements of the data generating process (DGP). These elements include the degree of selection into treatment (referred to as selectivity henceforth), the type and quantity of covariates, sample size, diverse functional forms, varying degrees of effect heterogeneity, the influence of covariates on outcomes and heterogeneity, the share of treated within the sample, and the number of treatments resulting in 77 different DGPs.[28] Despite its significant computational demands, this approach offers a comprehensive exploration of various scenarios, surpassing the more limited scope of Monte Carlo studies typically employed in papers introducing a new methodology.

The subsequent subsection provides a concise overview of the simulation designs' components. The procedure involves initially generating random data, applying the different estimators to this training dataset, predicting effects on a separate dataset of equivalent size derived from the same DGP, saving the results, and repeating these steps $R$ times. Subsequently, we calculate a range of performance metrics that capture different dimensions of the accuracy and reliability of both estimation and inference processes.

## 5.2. Key features of the simulation study

### *5.2.1. Data Generating Process*

The simulated (i.i.d.) data consist of the covariates, the treatment, and the potential outcomes. The simulation of these components will be discussed in turn (for additional details see Appendix B).

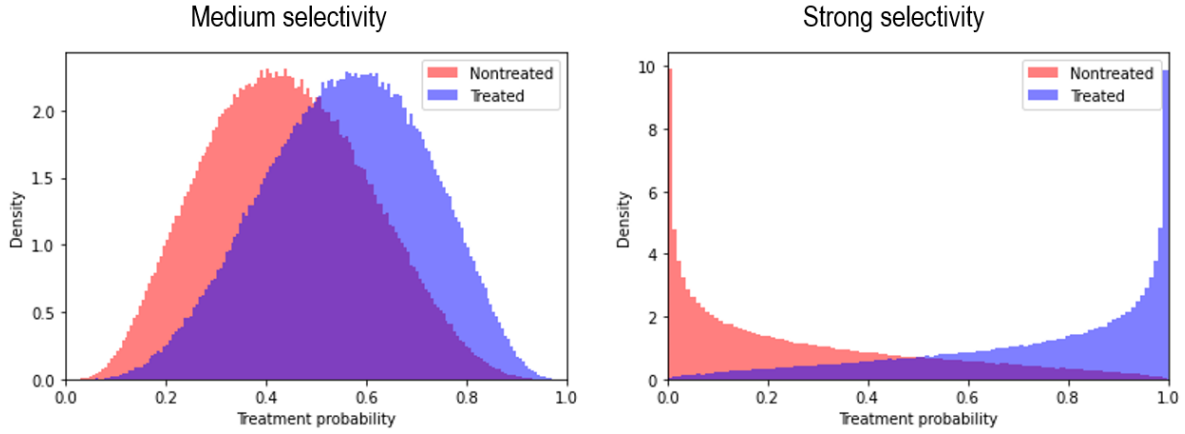
The simulation involves generating a range of 10 to 50 independent covariates ($p$ = 10, 20, 50). These covariates are normally, uniformly, or binomially (dummy variables)

[28] Table B.2 in Appendix B.2.5 gives the list of the scenarios investigated. Appendix C collects the tables that contain the corresponding results.

distributed, or as combinations of the three types. Among these $p$ covariates, the first $k$ covariates influence both the selection into treatment and the potential outcomes. The effect of these $k$ covariates is modelled to decrease linearly, ranging from 1 to $1/k$. The base specification consists of 20 uniformly and normally distributed covariates, half of them relevant.[29]

The selection process is based on a linear index function of the $k$ relevant covariates plus noise. The quantiles of this index function are used to generate the treatments. The base specification considers cases with random selectivity (experiment), medium selectivity (true $R^2$ of about 10%), and strong selectivity (true $R^2$ of about 42%). Cases of two and four treatments with equal as well as asymmetric treatment shares are considered. The base specification consists of 2 treatments with equal treatment shares. Figure 1 shows the distribution of the true treatment probabilities for the strong and medium selectivity case with two treatments and 50% treatment share.

Figure 1: Distribution of treatment probabilities in treated and non-treated population



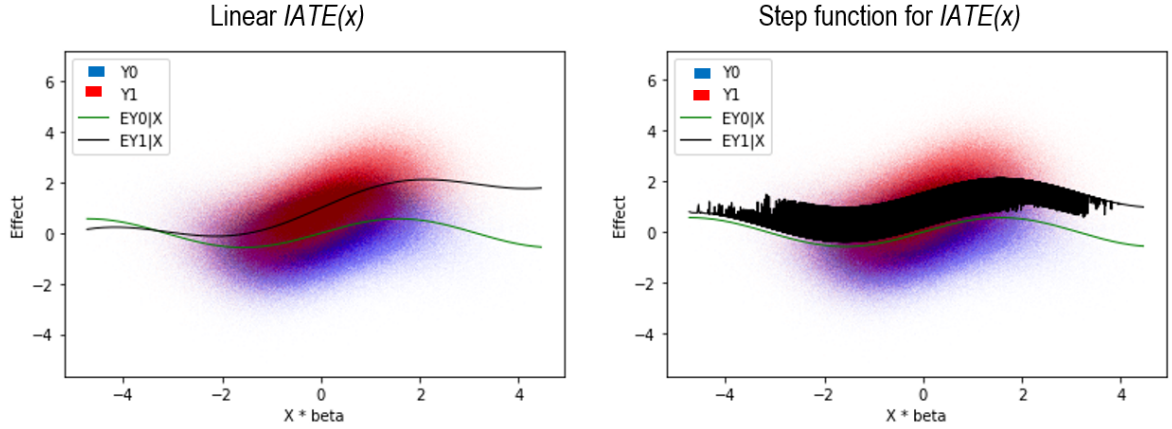

Note: Two treatments and 50% treatment share (base specification). Figures are based on 1'000'000 observations.

The non-treatment potential outcomes are obtained by simulating the expected non-treatment potential outcome as a sine-function of the linear index of the covariates plus noise. The relevance of the sine-function relative to the noise level is varied in the simulations from cases with true $R^2$ of 0 to 45% (base specification: 10%). The potential treatment outcomes are obtained by adding simulated IATEs plus noise to the expected

[29] When covariates are simulated from both, the normal and uniform distributions, the $1^{st}$, $3^{rd}$, $5^{th}$, ... covariates are drawn from the uniform distribution.

non-treatment potential outcomes.[30] The IATEs are generated as functions of the linear index or as a smooth function of the first two, most important, covariates. In the former case, a linear function, a logistic function, and a quadratic function is specified. Finally, the smooth function approach follows closely the specification of WA18. In all these cases the ATE is close to one. The case of zero IATEs, and thus zero ATE, is considered as well. The specification of the IATE as a smooth function is used in the base specification presented in Subsection 5.3.[31] The smooth function, when plotted as a function of the linear index, exhibits a pattern that resembles a step function. Figure 2 shows the expected and realised potential outcomes and their relation to the linear index. Thus, it summarizes main properties of the base specification following WA18, as well as of the linear specification of the IATE. Similar plots for the other specifications can be found in Appendix B.2.4.

Figure 2: Shape of potential outcomes for different shapes of IATEs (base specification)




Note: Figures are based on 1'000'000 observations.

The GATEs depend only on the first covariate, which is uniformly distributed if there are uniformly as well as normally distributed covariates in the DGP. This continuous covariate is split in groups with the same expected sizes and corresponding indicator variables are created. Thus, these dummy variables have no direct effect on the DGP. They are passed to the GATE estimators.

A final aspect of the DGP is the sample size. We mainly consider sample sizes of 2'500

[30] The 2 (2 treatments) or 4 noise terms (4 treatments) used to simulate potential outcomes are independent of each other.

[31] The base specifications with different degrees of selectivity and types of IATEs are captured in rows 1-15 in Table B.2 in Appendix B.2.5.

and 10'000 as compromise between computational costs and practical relevance.[32] To keep the noise from the simulations on the performance measures stable (at least for the estimators of ATE and GATEs, which can be expected to show $\sqrt{N}-$convergence), the number of replications, $R$, declines at the same rate as the sample increases. Thus, the results for $N$=2'500 are based on $R$=1'000, while $R$ declines to 250 for $N$=10'000.

In our simulation study, it's crucial to highlight that we adopt a methodology designed to replicate repeated sampling inference more closely, especially in scenarios involving stochastic covariates. To achieve this, the samples used for computing the out-of-training-sample effects are freshly drawn from the same Data Generating Process (DGP) for each replication. Additionally, these samples are matched in size to the training sample. This approach contrasts with the method used in Knaus et al. (2021), where the sample for calculating effects is drawn once prior to the initial replication and remains constant across subsequent replications, even when similar DGPs are employed.

The methodology chosen for our study, however, comes with certain trade-offs. Notably, for each IATE as a function of covariate values ($IATE(x)$), there exists only a single true and estimated value across all replications. This uniqueness arises because all or at least some covariates are continuous, leading to a scenario where each specific covariate value appears only once in the simulations. Consequently, for the IATEs this aspect of our simulation methodology precludes the possibility of estimating moments other than the expectation. This is an important consideration to bear in mind when interpreting the results and the applicability of our findings.

### *5.2.2. Estimators*

Below, we present the results of the Modified Causal Forest and the Generalized Random Forest, both with the outcome variable centred prior to estimation, *mcf-cent* and *grf-cent*, as well as of Double/debiased Machine Learning with normalized weights, *dml-norm*. For IATE estimation, we also include a more efficient version of the centred *mcf* for which no inference is available (see below), *mcf-cent-eff*. The tables in Appendix C contain additional results for the uncentred *mcf* and *grf*, standard (non-normalized) *dml*, and *OLS*. We describe the implementation of these estimators briefly in this section and refer the interested reader to Appendix B.3 for details.

[32] $N$=40'000 is also considered, but, for computational reasons, only for one specification.

While Appendix B.3.1 details the implementation of the *mcf* further, at least 2 points merit some more discussion. The first point concerns the penalty parameter, $\lambda$. In the simulations, $\lambda$ is set equal to $Var(Y)$. $Var(Y)$ corresponds to the MSE when the effects are estimated by the sample mean without any splits. Thus, it provides some ad-hoc benchmark for plausible values of $\lambda$. Generally, decreasing the penalty increases biases and reduces variances, et vice versa.[33] The simulations below will show that biases are more likely to occur when selectivity is strong. Thus, if a priori knowledge about the importance of selectivity is available, then the researcher might adjust the penalty term accordingly.

Secondly, if inference is not a priority, like when using the IATEs as inputs into the training of an optimal assignment algorithm, the efficiency loss inherent in the *mcf*'s two-sample honesty approach can be avoided by cross-fitting, i.e., by repeating the estimation with exchanged roles of the two samples and averaging the two (or more) estimates. However, in such a case it is unclear how to compute the weights-based inference for the averaged estimator *mcf-cent-eff* as the two components of this average are correlated.[34] A similar efficiency loss can be avoided in the *grf* procedure by switching off the half-sampling when inference is not relevant. However, the current implementation in the *grf* package does not support this option.

As mentioned in Subsection 4.2, *grf* performs local centring inside the Random Forest to remove confounding bias. The default *grf* local centring, labelled here as *uncentred*, refers to local centring that uses out-of-bag predictions of the outcome and treatment assignment from Random Forests trained on the full sample, to calculate the residuals. The simulation study includes an additional local centring strategy, labelled as *centred*. The difference between the centred and the uncentred *mcf* and *grf* is that the former uses a transformed outcome variable. This transformation subtracts a Random Forest prediction of $E(Y|X)$ from the observed $Y$ (obtained with 5-fold-cross-fitting), and thus purges them from much of the influence of $X$.[35] For *grf*, GATEs and ATE are estimated

[33] The sensitivity of the centred version of the *mcf* to the choice of the penalty parameter seems to be generally very low.

[34] For such a cross-fitted estimator (e.g., computed as mean of the single estimators), conservative inference could be obtained by basing inference on normality with a variance taken as average over the variances of the single estimations.

[35] Note that the *grf* package does not provide centring of this kind. We added this estimator as the original version used in the *grf* package performed poorly in many of our DGPs for reasons discussed in Subsection 4.2.

via linear regression in which a variant of a local AIPW estimator is regressed on group indicators or a constant.

Compared to the standard *dml*, the normalized *dml* is more robust to extreme values of the estimated propensity score. It is obtained by normalizing the weights that are implicit in the *dml* scores. While it is straightforward to use *dml* for ATE, it is less straightforward to use them for heterogeneity estimation. Here, GATEs and IATEs are obtained by regression-type approaches in which the estimated components of the *dml* scores serve as dependent variable. The GATEs are obtained as OLS-coefficients of a saturated regression model with the indicators for the groups defined by the discrete variable $Z$ as independent variables. IATEs are computed by using $X$ as independent variables either in a regression Random Forest or in an OLS regression. When OLS is used, inference is based on the heteroscedasticity-robust covariance matrix of the corresponding coefficients. No inference is obtained for the Random Forest based IATEs.

All estimators are not tuned as computational costs would be prohibitive given the already extensive simulations. Instead, the default values provided in the respective software packages are used.

#### *5.2.3. Performance measures*

The main performance measures are the biases of the effects and their standard errors, the standard deviation of the effects, the mean absolute error, and the root mean squared error (RMSE) of the effects. The coverage probability (CovP) for the 95% confidence interval is reported to gauge the quality of the inference.[36]

Note that, as mentioned above, the standard deviation, and thus the bias of the estimated standard error, cannot be computed for the IATE due to the simulation design. Whenever several parameters are involved (as for the GATEs and IATEs), the performance measures are computed for each parameter and then averaged.

## 5.3. Results

In this subsection, we analyse *dml-norm*, *grf-cent*, and *mcf-cent* for the WA18 base specification Data Generating Processes (DGPs) and sample sizes of $N$=2,500 and $N$=10,000

[36] In addition to the CovP for the 80% interval, the tables in the Appendix C also report the skewness and excess kurtosis of the estimators, which are, however, in a large majority of cases in the 'normal', unproblematic ranges.

(when there is no confusion, in this subsection, we will drop the *-norm* and *-cent* ending in the text). Additionally, the degree of selectivity within the data is varied.

Since it turned out that the relative performance of the estimators does not only depend on the strength of selection into treatment, but also on the specific parameter to be estimated, the results for the ATE, the GATEs, and the IATEs are discussed in turn.

### *5.3.1. Average treatment effects*

Table 3 shows the results for the ATE. If there is no selectivity, like in an experiment, we obtain the expected results: the efficient estimators, *dml* and *grf*, are similar and outperform the *mcf*, at least when selectivity is not too strong. All estimators are essentially unbiased, and empirical coverage is close to the nominal level. Furthermore, when the sample size quadruples, the standard deviation and RMSE halve, which is indicative of $\sqrt{N}-$convergence.

The analysis reveals a notable trend with increasing selectivity: the performance of the normalized double/debiased machine learning (*dml-norm*) estimator deteriorates, particularly in terms of bias. This leads to a large increase in the RMSE.[37] One plausible explanation is that (despite the normalisation) the double-robust score becomes more problematic when propensity score values become more extreme. A similar increase in bias is visible for the *grf*, although it is not as extreme as for *dml*. The resulting bias impacts also their coverage rates, which fall to very low levels. These issues appear to a much lesser extent for the *mcf*.

One summary coming from this table is that *mcf* appears to be more robust to stronger selectivity at the cost of some additional RMSE when selectivity does not matter much. The results in Appendix C.1 show that these performance patterns with respect to the degree of selectivity also appear for linear IATEs and non-linear IATEs. However, in the case of quadratic IATEs it turns out that the uncentred (!) *grf* outperforms the centred *grf* in terms of bias even for strong selectivity. One possible explanation for this surprising behaviour is that the Random Forest estimator used to do the centring does not capture the shape of $E(Y|X = x)$ well.

Fixing selectivity to medium levels and varying the other parameters of the DGP (Appendix C.2), suggests that the patterns observed in Table 3 qualitatively appear almost in

[37] As can be seen from Table C.15 in Appendix C.1, this behaviour appears for the not normalized *dml* as well and is driven mainly by the bias of the point estimator.

all other DGPs as well. The most remarkable case is when covariates are more important for the non-treatment potential outcome (Table C.17, Table C.21): all estimators become biased, which increases the RMSE and leads to coverage far below nominal levels.

Table 3: Simulation results for average treatment effect (ATE)

| | | | Estimation of effects | | | | Inference | |
|---|---|---|---|---|---|---|---|---|
| **Estimator** | Selectivity | Sample size | Bias | Mean absolute error | Std. dev | RMSE | Bias (SE) (SE) | CovP (95) in % |
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| **dml-norm** | None | 2'500 | -0.002 | 0.034 | 0.043 | 0.043 | 0.005 | 97 |
| **grf-cent** | | | 0.003 | 0.033 | 0.042 | 0.042 | 0.000 | 95 |
| **mcf-cent** | | | 0.003 | 0.048 | 0.061 | 0.061 | -0.001 | 94 |
| **dml-norm** | Medium | 2'500 | 0.020 | 0.040 | 0.045 | 0.050 | 0.004 | 95 |
| **grf-cent** | | | 0.037 | 0.046 | 0.042 | 0.056 | -0.000 | 85 |
| **mcf-cent** | | | 0.037 | 0.057 | 0.061 | 0.072 | 0.000 | 91 |
| **dml-norm** | Strong | 2'500 | 0.140 | 0.140 | 0.056 | 0.150 | 0.002 | 23 |
| **grf-cent** | | | 0.090 | 0.091 | 0.046 | 0.101 | -0.006 | 41 |
| **mcf-cent** | | | 0.046 | 0.060 | 0.058 | 0.074 | 0.008 | 92 |
| **dml-norm** | None | 10'000 | 0.000 | 0.016 | 0.020 | 0.020 | 0.003 | 98 |
| **grf-cent** | | | 0.002 | 0.017 | 0.021 | 0.021 | 0.000 | 95 |
| **mcf-cent** | | | 0.004 | 0.022 | 0.028 | 0.028 | 0.002 | 97 |
| **dml-norm** | Medium | 10'000 | 0.010 | 0.019 | 0.022 | 0.024 | 0.003 | 94 |
| **grf-cent** | | | 0.015 | 0.022 | 0.022 | 0.027 | -0.001 | 92 |
| **mcf-cent** | | | 0.014 | 0.025 | 0.027 | 0.030 | 0.003 | 96 |
| **dml-norm** | Strong | 10'000 | 0.079 | 0.079 | 0.032 | 0.085 | 0.000 | 28 |
| **grf-cent** | | | 0.042 | 0.044 | 0.027 | 0.050 | -0.006 | 46 |
| **mcf-cent** | | | -0.013 | 0.025 | 0.028 | 0.031 | 0.007 | 96 |

Note: RMSE abbreviates the Root Mean Squared Error. *CovP (95)* denotes the (average) probability that the true value is part of the estimated 95% confidence interval. 1'000 / 250 replications are used for 2'500 / 10'000 observations.

### *5.3.2. Conditional average treatment effects with a small number of groups (GATEs)*

Table 4 shows that the relative performance of the different estimators for the GATEs depends not only on the strength of selectivity, but also on the number of groups for which a GATE is computed for as captured in Proposition 1. Table 4 shows the results for the cases of 5 and 40 groups, while Tables C.28 to C.30 in Appendix C.2 show the intermediate cases of 10 and 20 groups as well.

Concerning the point estimate, *mcf* outperforms *grf* and *dml* once there are 10 and more ($N$=2'500) or 20 and more groups ($N$=10'000), independent of the strength of selectivity.

Table 4: Simulation results for group average treatment effects (GATEs)

| | | | Estimation of effects | | | | Inference | |
|---|---|---|---|---|---|---|---|---|
| **Estimator** | Selec-tivity | Sample size | Bias | Mean absolute error | Std. dev | RMSE | Bias (SE) (SE) | CovP (95) in % |
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| | | | | 5 Groups | | | | |
| **dml-norm** | None | 2'500 | -0.002 | 0.075 | 0.094 | 0.094 | 0.002 | 95 |
| **grf-cent** | | | 0.004 | 0.075 | 0.094 | 0.094 | 0.000 | 95 |
| **mcf-cent** | | | 0.002 | 0.090 | 0.091 | 0.114 | -0.010 | 83 |
| **dml-norm** | Med- | 2'500 | 0.020 | 0.081 | 0.081 | 0.101 | 0.001 | 95 |
| **grf-cent** | ium | | 0.037 | 0.081 | 0.093 | 0.102 | -0.000 | 93 |
| **mcf-cent** | | | 0.037 | 0.094 | 0.090 | 0.121 | -0.006 | 84 |
| **dml-norm** | Strong | 2'500 | 0.140 | 0.157 | 0.123 | 0.187 | -0.005 | 74 |
| **grf-cent** | | | 0.090 | 0.119 | 0.096 | 0.148 | -0.007 | 75 |
| **mcf-cent** | | | 0.046 | 0.112 | 0.085 | 0.146 | 0.009 | 81 |
| **dml-norm** | None | 10'000 | -0.001 | 0.037 | 0.046 | 0.046 | 0.001 | 95 |
| **grf-cent** | | | 0.002 | 0.036 | 0.046 | 0.046 | 0.001 | 95 |
| **mcf-cent** | | | 0.004 | 0.045 | 0.045 | 0.058 | -0.002 | 87 |
| **dml-norm** | Med- | 10'000 | 0.010 | 0.041 | 0.050 | 0.051 | 0.000 | 94 |
| **grf-cent** | ium | | 0.016 | 0.041 | 0.048 | 0.051 | -0.001 | 92 |
| **mcf-cent** | | | 0.013 | 0.049 | 0.045 | 0.063 | 0.004 | 86 |
| **dml-norm** | Strong | 10'000 | 0.078 | 0.090 | 0.073 | 0.108 | -0.007 | 71 |
| **grf-cent** | | | 0.043 | 0.063 | 0.051 | 0.080 | -0.006 | 75 |
| **mcf-cent** | | | -0.014 | 0.085 | 0.045 | 0.100 | 0.007 | 65 |
| | | | | 40 Groups | | | | |
| **dml-norm** | None | 2'500 | -0.002 | 0.214 | 0.268 | 0.269 | -0.003 | 94 |
| **grf-cent** | | | 0.004 | 0.213 | 0.269 | 0.269 | -0.001 | 95 |
| **mcf-cent** | | | 0.002 | 0.096 | 0.096 | 0.120 | -0.013 | 82 |
| **dml-norm** | Med- | 2'500 | 0.020 | 0.226 | 0.283 | 0.284 | -0.005 | 94 |
| **grf-cent** | ium | | 0.037 | 0.213 | 0.265 | 0.268 | -0.001 | 95 |
| **mcf-cent** | | | 0.036 | 0.099 | 0.095 | 0.128 | -0.008 | 83 |
| **dml-norm** | Strong | 2'500 | 0.140 | 0.293 | 0.338 | 0.367 | -0.020 | 89 |
| **grf-cent** | | | 0.090 | 0.223 | 0.255 | 0.280 | -0.003 | 92 |
| **mcf-cent** | | | 0.043 | 0.119 | 0.090 | 0.153 | 0.007 | 80 |
| **dml-norm** | None | 10'000 | -0.001 | 0.106 | 0.132 | 0.132 | 0.000 | 95 |
| **grf-cent** | | | 0.002 | 0.104 | 0.130 | 0.130 | 0.002 | 95 |
| **mcf-cent** | | | 0.004 | 0.050 | 0.050 | 0.063 | -0.005 | 85 |
| **dml-norm** | Med- | 10'000 | 0.010 | 0.113 | 0.140 | 0.141 | -0.001 | 95 |
| **grf-cent** | ium | | 0.016 | 0.106 | 0.131 | 0.132 | 0.000 | 95 |
| **mcf-cent** | | | 0.012 | 0.055 | 0.051 | 0.069 | -0.002 | 84 |
| **dml-norm** | Strong | 10'000 | 0.079 | 0.164 | 0.192 | 0.209 | -0.015 | 89 |
| **grf-cent** | | | 0.043 | 0.115 | 0.130 | 0.145 | -0.002 | 92 |
| **mcf-cent** | | | -0.016 | 0.091 | 0.050 | 0.104 | 0.005 | 63 |

Note: RMSE abbreviates the Root Mean Squared Error. *CovP (95)* denotes the (average) probability that the true value is part of the estimated 95% confidence interval. 1'000 / 250 replications are used for 2'500 / 10'000 observations.

It is not surprising that *mcf* performs better for strong selectivity: the previous section showed that for strong selectivity the *mcf* dominates even for a GATE with only one group, i.e., the ATE.

We observe a differential dependence of the standard deviation on the number of groups: its increase is much slower for the *mcf* than for *dml* and *grf*. The reason is the way the GATE estimators are constructed by the different methods. As mentioned above, *dml* and *grf* are averaging double-robust scores within the cells of the discrete $Z$. Since the estimators are $\sqrt{N}-$convergent, we expect that when the sample is reduced to one quarter of the original sample, the standard deviations of such estimators double. As the cells defined by the group variable in the DGPs are of approximately equal size, the number of groups and observations per group are inverse proportionally related. The resulting doubling of the standard deviation of the *grf* and *dml* estimators is exactly what is observed in Tables C.28 to C.30 when comparing results for 5 vs. 20 groups, or 10 vs. 40 groups, respectively.

The *mcf* aggregates IATEs within these cells. However, as the IATEs also potentially use "leaf-populating data" outside of these cells, the standard deviation of the *mcf* increases much slower than for *grf* and *dml*.[38] In fact, we observe that when the number of groups increases further, the RMSE of the *mcf* approaches the one of the *mcf* IATEs from below, while the RMSE of *dml* and *grf* increases substantially and sometimes exceeds the RMSE of the IATEs which are of much higher dimension.

Concerning inference, the findings are a bit more pronounced: *dml* and *grf* have the correct coverage rates for no and medium selectivity even for 40 groups, while the coverage rates for *mcf* are too low.[39] The figures contained at the end of Tables C.28, C.29, and C.30 in Appendix C.2 indicate that this problem comes mainly from a bias of the GATEs with the smallest true values, while for the other GATEs coverage is close to the nominal level. In fact, the good performance in terms of RMSE and the problems with coverage are two sides of the same coin. Due to their aggregation from IATEs that are averaged

[38] The aggregation of IATEs that potentially weigh information from observations belonging to other groups into a GATE resembles smoothing across (adjacent) categories in the context of categorical regressors. Heiler and Mareckova (2021) show that optimal smoothing parameters in a non-parametric regression involving categorical regressors do not vanish asymptotically. Furthermore, they prove that the variance of the smoothed estimator is a weighted sum of the asymptotic variances from other categories. This finding provides a plausible rationale for the observed steadiness in the variance of the *mcf* estimator across varying group sizes.

[39] Note that when selection into treatment gets stronger, splitting for the *mcf* (and *grf*) stops earlier, as leaves must have both treated and controls after a possible split, thus bias increases.

over trees, *mcf* estimators share the strengths and the weaknesses of many smoothing methods: the variance is reduced at the cost of an increasing bias. To improve on the inference in finite samples, the simulations indicate that it needs some debiasing. This is however beyond the scope of this paper.

### *5.3.3. Conditional average treatment effects at the finest aggregation level (IATEs)*

Table 5 contains the results for the IATEs, averaged over all $N$ IATEs. As inference is usually not the main goal when computing IATEs, for the centred *mcf* the more efficient version is included as well.

For the point estimates the ordering is clear-cut. *grf* and *mcf* are similar and outperform both *dml* versions (*ols*, *rf*). In many cases, the efficient centred *mcf* (*mcf-cent-eff*) performs best in terms of RMSE. The only exception to this rule seems to be, again, the case of a quadratic function for the IATE, in which all methods seem to do (almost) equally bad with large RMSEs and large mean absolute errors. With respect to the other estimators, in a couple of cases OLS outperforms the more sophisticated CML estimators (but substantially underperforms in many other scenarios).

For inference, the results show that all estimators have coverage probabilities that are substantially below their nominal levels. However, given the level of granularity of the IATEs, this finding is of course not surprising. At least for the *mcf*, we conjecture that this problem comes from the biases of the individual IATEs. It is less likely that it comes from a too small estimated standard error, because weights-based standard error estimation is performed in a very similar way as for the ATE and the GATEs, in which it turned out to be almost unbiased.

### *5.3.4. Summary*

*dml* performs particularly well when the target is low dimensional, such as the ATE and GATEs with few groups, and selectivity is not too strong. If selectivity is strong or the number of groups becomes too large, then its performance quickly deteriorates.[40]

*grf* shows a similar good performance for the low dimensional parameters, as well as a

[40] A potential alternative *dml*-based estimator in the strong selectivity setting could be the automatic *dml* that would estimate inverse probability weights instead of propensity scores as nuisance parameters.

Table 5: Simulation results for individualized average treatment effects (IATEs)

| | | | Estimation of effects | | | Inference |
|---|---|---|---|---|---|---|
| **Estimator** | Selectivity | Sample size | Bias | Mean absolute error | RMSE | CovP (95) in % |
| (1) | (2) | (3) | (4) | (5) | (7) | (9) |
| **dml-ols** | None | 2'500 | -0.002 | 0.229 | 0.286 | 23 |
| **dml-rf** | | | -0.002 | 0.270 | 0.342 | - |
| **grf-cent** | | | 0.004 | 0.155 | 0.187 | 77 |
| **mcf-cent** | | | 0.002 | 0.162 | 0.199 | 77 |
| **mcf-cent-eff** | | | 0.004 | 0.151 | 0.184 | - |
| **dml-ols** | Medium | 2'500 | 0.020 | 0.236 | 0.295 | 24 |
| **dml-rf** | | | 0.013 | 0.284 | 0.362 | - |
| **grf-cent** | | | 0.034 | 0.175 | 0.210 | 72 |
| **mcf-cent** | | | 0.037 | 0.174 | 0.212 | 75 |
| **mcf-cent-eff** | | | 0.040 | 0.163 | 0.197 | - |
| **dml-ols** | Strong | 2'500 | 0.140 | 0.285 | 0.357 | 28 |
| **dml-rf** | | | 0.117 | 0.349 | 0.461 | - |
| **grf-cent** | | | 0.092 | 0.254 | 0.297 | 55 |
| **mcf-cent** | | | 0.046 | 0.212 | 0.252 | 68 |
| **mcf-cent-eff** | | | 0.048 | 0.205 | 0.240 | - |
| **dml-ols** | None | 10'000 | -0.001 | 0.179 | 0.219 | 7 |
| **dml-rf** | | | 0.000 | 0.201 | 0.256 | - |
| **grf-cent** | | | 0.002 | 0.082 | 0.103 | 91 |
| **mcf-cent** | | | 0.004 | 0.096 | 0.121 | 83 |
| **mcf-cent-eff** | | | 0.003 | 0.088 | 0.110 | - |
| **dml-ols** | Medium | 10'000 | 0.010 | 0.181 | 0.222 | 8 |
| **dml-rf** | | | 0.005 | 0.214 | 0.274 | - |
| **grf-cent** | | | 0.016 | 0.089 | 0.111 | 89 |
| **mcf-cent** | | | 0.013 | 0.106 | 0.131 | 81 |
| **mcf-cent-eff** | | | 0.013 | 0.100 | 0.122 | - |
| **dml-ols** | Strong | 10'000 | 0.079 | 0.205 | 0.256 | 12 |
| **dml-rf** | | | 0.058 | 0.275 | 0.418 | - |
| **grf-cent** | | | 0.053 | 0.123 | 0.162 | 82 |
| **mcf-cent** | | | -0.014 | 0.156 | 0.185 | 66 |
| **mcf-cent-eff** | | | -0.014 | 0.151 | 0.151 | - |

Note: *dml* is the normalized *dml* (denoted as *dml-norm* in Tables 3 and 4). RMSE abbreviates the Root Mean Squared Error. *CovP (95)* denotes the (average) probability that the true value is part of the estimated 95% confidence interval. 1'000 / 250 replications are used for 2'500 / 10'000 observations.

similarly bad performance for GATEs with many groups.[41] However, it performs well for IATEs. It is also less affected by strong selectivity than *dml*. These results however only hold for the modified version of the *grf* in which the outcome variable in the training data is explicitly centred before it enters the *grf* algorithm. The original uncentred version of *grf* as suggested in Athey et al. (2019) and implemented in their package underperforms

[41] Additional smoothing in a setting with many groups could be considered in empirical settings for *dml*-based and *grf* -based GATE estimators to further improve their RMSE.

in many scenarios due its bias problem (as does the uncentred version of the *mcf*).[42]

Compared to *dml* and *grf*, the *mcf* generally shows a robust and competitive performance in many scenarios. The price to pay for this robustness and competitiveness is the somewhat higher standard deviation for the very low dimensional parameters like ATE and GATEs. Depending on the sample and the resulting precision of the estimators, this price may well be worthwhile paying. A further important advantage of the *mcf* is that its estimates are internally consistent over aggregation levels as ATE and GATEs are computed as averages of IATEs. This advantage becomes particularly apparent when the number of groups for the GATEs increases. While *dml* and *grf* have substantially higher RMSEs than for the more fain-grained IATEs, the uncertainty (and bias) in the *mcf* estimators smoothly increases with the number of groups, until it reaches the level of the most fine-grained heterogeneity parameter, the IATE.

# 6. Conclusion

Estimation of causal effects at different levels of granularity is of great importance for informed decision-making and tailored interventions. Lower levels of granularity capture the effect of a policy or intervention on a large population, guiding decisions on policies that cannot be targeted at individuals but must be deployed universally. Higher levels of granularity capture effects at a group or individual level that can serve for decisions how policies can be tailored more individually or for better understanding of the effects of large-scale policies at more granular levels. The complexity of such interventions requires methods that can estimate fine-grained heterogeneities of causal effects flexibly, such as some Causal Machine Learning (CML) methods.

In this paper, we investigate such CML methods subject to the restrictions that (i) they provide estimators of the causal effects at all aggregations levels, (ii) are essentially non-parametric, (iii) can predict IATE based on covariates only and (iv) that they allow for classical repeated sampling inference. Ideally, they are also internally consistent that aggregation of lower-level effects lead to the higher-level effects. Double/debiased Machine Learning (*dml*), the Generalized Random Forest (*grf*), and the Modified Causal Forest

[42] As described in Subsubsection 5.2.2, labels centred and uncentred refer to whether the outcome variable enters the estimation procedure residualized by $K$-fold cross-fitting or not, respectively. In both cases, the *grf* procedure still performs local centring using out-of-bag predictions as described in Subsection 4.2.

(*mcf*) fulfil these criteria and thus belong to the group of estimation methods which we call 'Comprehensive Causal Machine Learners' (CCMLs).

Here, we describe these estimators and their proven theoretical guarantees. For *dml* and the *grf*, they are already known, but not so for the *mcf*. Therefore, we explicitly provide them. The large-scale simulation study reveals scenarios in which the methods perform well. *dml* with normalized weights performs well in terms of RMSE and coverage probability when the target is low-dimensional, i.e., ATE and GATEs with few groups, and when the selection into treatment is not strong. *grf* shows similar behaviour as *dml*, however its performance for IATEs in terms of RMSE is much better. *mcf* has similar performance as *grf* in case of IATEs and outperforms *dml* and *grf* in scenarios with many GATEs or strong selection.

The results of the simulation study offer several practical recommendations. For low-dimensional targets when selection into treatment is moderate, *dml* is preferred including statistical inference. For IATEs, Causal Forest based methods perform well in terms of point estimation. When inference is not a priority, the more efficient version of *mcf* is recommended to estimate IATEs. For large groups, *mcf* is recommended for point estimation of GATEs. When sample is large enough and slight loss of efficiency is not detrimental, *mcf* can be used to estimate all effects due to its robustness to strong selection into treatment and large GATE groups. When internal consistency of the effects is important, only *mcf* can guarantee it due to its aggregation strategy.

The practical use of these three CCMLs is supported by the availability of well-maintained software packages: *dml* is available as Python and R packages, the g*rf* is available as an R package, and the *mcf* is available as a Python package.[43] On a practical note, the results in this paper indicate that users of the *grf* package are encouraged to perform local centring of the outcome variable via $K$-fold cross-fitting to remove any potential bias (the same holds for *mcf*, where this step is already implemented in the package).

The simulation study also points to topics for further research. For example, the sensitivity of the *dml* estimators to strong selectivity opens a topic of how to make them more robust to extreme propensity scores without impairing its efficiency and statistical inference. Due to the smoothing character of Causal Forests, in finite samples *mcf* es-

[43] As examples can serve: Bach, Chernozhukov, Kurz, and Spindler (2022), Bach, Kurz, et al. (2024), and Knaus (2022) for *dml*; Athey and Wager (2019) for *grf*; and Bodory et al. (2022) for *mcf*. The R packages can be downloaded from CRAN. The Python packages are available on PyPI.

timation of GATEs and IATEs balances the bias-variance trade-off yielding low RMSEs but impairing the coverage probability due to a small bias and low variance. Finding a way how to further de-bias the estimates would improve coverage probability in finite samples.

# References


Abadie, A., and Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, *74*(1), 235–267.

Abrevaya, J., Hsu, Y.-C., and Lieli, R. P. (2015). Estimating conditional average treatment effects. *Journal of Business & Economic Statistics*, *33*(4), 485–505.

Alaa, A. M., Ahmad, Z., and van der Laan, M. (2024). Conformal meta-learners for predictive inference of individual treatment effects. *Advances in Neural Information Processing Systems*, *36*.

Alaa, A. M., and van der Schaar, M. (2017). Bayesian inference of individualized treatment effects using multi-task gaussian processes. *Advances in Neural Information Processing Systems*, *30*.

Athanasopoulos, G., Hyndman, R. J., Kourentzes, N., and Panagiotelis, A. (2024). Forecast reconciliation: A review. *International Journal of Forecasting*, *40*(2), 430–456.

Athey, S. (2017). Beyond prediction: Using big data for policy problems. *Science*, *355*(6324), 483–485.

Athey, S., and Imbens, G. W. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, *113*(27), 7353–7360.

Athey, S., and Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, *31*(2), 3–32.

Athey, S., Imbens, G. W., and Wager, S. (2018). Approximate residual balancing: Debiased inference of average treatment effects in high dimensions. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *80*(4), 597–623.

Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, *47*(2), 1148–1178.

Athey, S., and Wager, S. (2019). Estimating treatment effects with causal forests: An application. *Observational Studies*, *5*(2), 37–51.

Audrino, F., Chassot, J., Huang, C., Knaus, M., Lechner, M., and Ortega, J.-P. (2024). How does post-earnings announcement sentiment affect firms' dynamics? New evidence from causal machine learning. *Journal of Financial Econometrics*, *22*(3), 575–604.

Austin, P. C. (2012). Using ensemble-based methods for directly estimating causal effects: An investigation of tree-based G-computation. *Multivariate Behavioral Research*, *47*(1), 115–135.

Bach, P., Chernozhukov, V., Kurz, M. S., and Spindler, M. (2022). DoubleML-an object-oriented implementation of Double Machine Learning in Python. *Journal of Machine Learning Research*, *23*(53), 1–6.

Bach, P., Kurz, M. S., Chernozhukov, V., Spindler, M., and Klaassen, S. (2024). DoubleML: An object-oriented implementation of Double Machine Learning in R. *Journal of Statistical Software*, *108*(3), 1–56.

Bach, P., Schacht, O., Chernozhukov, V., Klaassen, S., and Spindler, M. (2024). Hyperparameter tuning for causal inference with Double Machine Learning: A simulation study. In *Causal learning and reasoning* (pp. 1065–1117).

Baiardi, A., and Naghi, A. A. (2024). The value added of machine learning to causal inference: Evidence from revisited studies. *The Econometrics Journal*, *27*(2), 213–234.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies*, *81*(2), 608–650.

Bertrand, M., Crépon, B., Marguerie, A., and Premand, P. (2021). *Do workfare programs live up to their promises? Experimental evidence from Cote D'Ivoire.* (NBER working paper)

Billingsley, P. (1968). *Convergence of probability measures* (1st ed.). New York: Wiley.

Bodory, H., Busshoff, H., and Lechner, M. (2022). High resolution treatment effects estimation: Uncovering effect heterogeneities with the modified causal forest. *Entropy*, *24*(8), 1039.

Bodory, H., Camponovo, L., Huber, M., and Lechner, M. (2020). The finite sample performance of inference methods for propensity score matching and weighting estimators. *Journal of Business & Economic Statistics*, *38*(1), 183–200.

Boller, D., Lechner, M., and Okasa, G. (2021). The effect of sport in online dating: Evidence from causal machine learning. *arXiv preprint arXiv:2104.04601*.

Bradley, R. C. (1986). Basic properties of strong mixing conditions. In E. Eberlein and M. S. Taqqu (Eds.), *Dependence in probability and statistics: A survey of recent results* (pp. 165–192). Boston, MA: Birkhäuser Boston.

Burlat, H. (2024). Everybody's got to learn sometime? A causal machine learning evaluation of training programmes for jobseekers in France. *Labour Economics*, 102573.

Cannas, M., and Arpino, B. (2019). A comparison of machine learning algorithms and covariate balance measures for propensity score matching and weighting. *Biometrical Journal*, *61*(4), 1049–1072.

Caron, A., Baio, G., and Manolopoulou, I. (2022). Estimating individual treatment effects using non-parametric regression models: A review. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *185*(3), 1115–1149.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, *21*(1), C1–C68.

Chernozhukov, V., Newey, W. K., and Singh, R. (2022). Automatic debiased machine learning of causal and structural effects. *Econometrica*, *90*(3), 967–1027.

Cockx, B., Lechner, M., and Bollens, J. (2023). Priority to unemployed immigrants? A causal machine learning evaluation of training in Belgium. *Labour Economics*, *80*, 102306.

Davis, J. M., and Heller, S. B. (2017). Using causal forests to predict treatment heterogeneity: An application to summer jobs. *American Economic Review*, *107*(5), 546–550.

Devroye, L., Györfi, L., Krzyżak, A., and Lugosi, G. (1994). On the strong universal consistency of nearest neighbor regression function estimates. *The Annals of Statistics*, *22*(3), 1371–1385.

Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans.* Society for Industrial and Applied Mathematics.

Fan, Q., Hsu, Y.-C., Lieli, R. P., and Zhang, Y. (2022). Estimation of conditional average treatment effects with high-dimensional data. *Journal of Business & Economic Statistics*, *40*(1), 313–327.

Farbmacher, H., Kögel, H., and Spindler, M. (2021). Heterogeneous effects of poverty on attention. *Labour Economics*, *71*, 102028.

Farrell, M. H. (2015). Robust inference on average treatment effects with possibly more covariates than observations. *Journal of Econometrics*, *189*(1), 1–23.

Foster, D. J., and Syrgkanis, V. (2023). Orthogonal statistical learning. *The Annals of Statistics*, *51*(3), 879–908.

Goller, D. (2023). Analysing a built-in advantage in asymmetric darts contests using causal machine learning. *Annals of Operations Research*, *325*(1), 649–679.

Goller, D., Lechner, M., Moczall, A., and Wolff, J. (2020). Does the estimation of the propensity score by machine learning improve matching estimation? The case of Germany's programmes for long term unemployed. *Labour Economics*, *65*, 101855.

Graham, B. S., Pinto, C. C. X., and Egel, D. (2012). Inverse probability tilting for moment condition models with missing data. *The Review of Economic Studies*, *79*(3), 1053–1079.

Györfi, L., Kohler, M., Krzyżak, A., and Walk, H. (2002). *A distribution-free theory of nonparametric regression.* Springer New York.

Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica*, *66*(2), 315–331.

Hansen, B. B. (2008). The prognostic analogue of the propensity score. *Biometrika*, *95*(2), 481–488.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). Springer New York. (10th printing with corrections, 2013)

Heiler, P., and Mareckova, J. (2021). Shrinkage for categorical regressors. *Journal of Econometrics*, *223*(1), 161–189.

Hill, J. L. (2011). Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, *20*(1), 217–240.

Hodler, R., Lechner, M., and Raschky, P. A. (2023). Institutions and the resource curse: New insights from causal machine learning. *Plos One*, *18*(6), e0284968.

Huber, M., Lechner, M., and Wunsch, C. (2013). The performance of estimators based on the propensity score. *Journal of Econometrics*, *175*(1), 1–21.

Hurwicz, L. (1950). Generalization of the concept of identification. In T. C. Koopmans (Ed.), *Statistical inference in dynamic economic models* (pp. 245–57). New York: Wiley.

Imai, K., and Ratkovic, M. (2014). Covariate balancing propensity score. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *76*(1), 243–263.

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika*, *87*(3), 706–710.

Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and statistics*, *86*(1), 4–29.

Imbens, G. W., and Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, *47*(1), 5–86.

Kennedy, E. H. (2023). Towards optimal doubly robust estimation of heterogeneous causal effects. *Electronic Journal of Statistics*, *17*(2), 3008–3049.

Knaus, M. C. (2021). A double machine learning approach to estimate the effects of musical practice on student's skills. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *184*(1), 282–300.

Knaus, M. C. (2022). Double machine learning-based programme evaluation under unconfoundedness. *The Econometrics Journal*, *25*(3), 602–627.

Knaus, M. C., Lechner, M., and Strittmatter, A. (2021). Machine learning estimation of heterogeneous causal effects: Empirical Monte Carlo evidence. *The Econometrics Journal*, *24*(1), 134–161.

Knaus, M. C., Lechner, M., and Strittmatter, A. (2022). Heterogeneous employment effects of job search programs: A machine learning approach. *Journal of Human Resources*, *57*(2), 597–636.

Knittel, C. R., and Stolper, S. (2021). Machine learning about treatment effect heterogeneity: The case of household energy use. In *AEA Papers and Proceedings* (Vol. 111, pp. 440–444).

Künzel, S. R., Sekhon, J. S., Bickel, P. J., and Yu, B. (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the National Academy of Sciences*, *116*(10), 4156–4165.

Langenberger, B., Steinbeck, V., Schöner, L., Busse, R., Pross, C., and Kuklinski, D. (2023). Exploring treatment effect heterogeneity of a proms alert intervention in knee and hip arthroplasty patients: A causal forest application. *Computers in Biology and Medicine*, *163*, 107118.

Lechner, M. (2001). Identification and estimation of causal effects of multiple treatments under the conditional independence assumption. In *Econometric evaluation of labour market policies* (pp. 43–58). Physica-Verlag HD.

Lechner, M. (2018). Modified causal forests for estimating heterogeneous causal effects. *arXiv preprint arXiv:1812.09487*.

Lee, S., Okui, R., and Whang, Y.-J. (2017). Doubly robust uniform confidence band for the conditional average treatment effect function. *Journal of Applied Econometrics*, *32*(7), 1207–1225.

Levy, J., van der Laan, M., Hubbard, A., and Pirracchio, R. (2021). A fundamental measure of treatment effect heterogeneity. *Journal of Causal Inference*, *9*(1), 83–108.

McCaffrey, D. F., Ridgeway, G., and Morral, A. R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological methods*, *9*(4), 403.

Meinshausen, N. (2006). Quantile regression forests. *Journal of Machine Learning Research*, *7*(6), 983–999.

Neumann, M. H. (2013). A central limit theorem for triangular arrays of weakly dependent random variables, with applications in statistics. *ESAIM: Probability and Statistics*, *17*, 120–134.

Neyman, J. (1959). Optimal asymptotic tests of composite hypotheses. *Probability and Statsitics*, 213–234.

Nie, X., and Wager, S. (2021). Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika*, *108*(2), 299–319.

Pytka, K., and Gulyas, A. (2021). Understanding the sources of earnings losses after job displacement: A machine-learning approach. Kiel, Hamburg: ZBW - Leibniz Information Centre for Economics.

Rehill, P. (2024). How do applied researchers use the causal forest? A methodological review of a method. *arXiv preprint arXiv:2404.13356*.

Robins, J. M., and Rotnitzky, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association*, *90*(429), 122–129.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, *89*(427), 846–866.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, *90*(429), 106–121.

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica*, 931–954.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688.

Salditt, M., Eckes, T., and Nestler, S. (2024). A tutorial introduction to heterogeneous treatment effect estimation with meta-learners. *Administration and Policy in Mental Health and Mental Health Services Research*, *51*(5), 650–673.

Semenova, V., and Chernozhukov, V. (2021). Debiased machine learning of conditional average treatment effects and other causal functions. *The Econometrics Journal*, *24*(2), 264–289.

Sexton, J., and Laake, P. (2009). Standard errors for bagged and random forest estimators. *Computational Statistics & Data Analysis*, *53*(3), 801–811.

Strittmatter, A. (2023). What is the value added by using causal machine learning methods in a welfare experiment evaluation? *Labour Economics*, *84*, 102412.

van der Laan, M. J. (2006). Statistical inference for variable importance. *The International Journal of Biostatistics*, *2*(1).

van der Laan, M. J., and Rose, S. (2018). *Targeted learning in data science: Causal inference for complex longitudinal studies*. Springer International Publishing.

van der Laan, M. J., and Rubin, D. (2006). Targeted maximum likelihood learning. *The International Journal of Biostatistics*, *2*(1).

Wager, S., and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, *113*(523), 1228–1242.

Wager, S., Hastie, T., and Efron, B. (2014). Confidence intervals for random forests: The jackknife and the infinitesimal jackknife. *The Journal of Machine Learning Research*, *15*(1), 1625–1651.

Wager, S., and Walther, G. (2015). Adaptive concentration of regression trees, with application to random forests. *arXiv preprint arXiv:1503.06388*.

Wei, W., Petersen, M., van der Laan, M. J., Zheng, Z., Wu, C., and Wang, J. (2023). Efficient targeted learning of heterogeneous treatment effects for multiple subgroups. *Biometrics*, *79*(3), 1934–1946.

Wendling, T., Jung, K., Callahan, A., Schuler, A., Shah, N. H., and Gallego, B. (2018). Comparing methods for estimation of heterogeneous treatment effects using observational data from health care databases. *Statistics in Medicine*, *37*(23), 3309–3324.

Westreich, D., Lessler, J., and Funk, M. J. (2010). Propensity score estimation: Neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. *Journal of Clinical Epidemiology*, *63*(8), 826–833.

Zheng, W., and van der Laan, M. J. (2011). Cross-validated targeted minimum-loss-based estimation. In *Targeted learning: Causal inference for observational and experimental data* (pp. 459–474). New York, NY: Springer New York.

Zimmert, M., and Lechner, M. (2019). Nonparametric estimation of causal heterogeneity under high-dimensional confounding. *arXiv preprint arXiv:1908.08779*.

# Appendix A. Details of the *mcf* and *grf*

In this appendix, we present the proven theoretical guarantees of the *mcf*, its inference procedure, as well as some more details of the Causal Forest algorithms.

## Appendix A.1. Theoretical guarantees for *mcf*

### *Appendix A.1.1. Proof for Proposition 1*

**PROPOSITION 1.** Let $J$ denote the number of groups for which GATEs are to be estimated and let $\Pi$ be a $J \times 1$ vector containing the non-zero population shares of observations in the corresponding groups. Suppose that the total number of groups increases from $J$ to $J'$ and the population share of group $j$ under $J$ groups, $\pi_j$, decreases to $\pi_{j'}$ under $J'$ groups. Then, the asymptotic variances of the efficient *dml* and *grf* GATE estimators increase for each group $j$ within $J$ proportionally to $\pi_j/\pi_{j'}$, while the variance of *mcf* estimator of the GATE does not depend on the number of groups and remains stable. This implies $Var\left(\widehat{GATE^{mcf}}\left(m,l;z_j\right)\right) \leq Var\left(\widehat{GATE^{dml/grf}}\left(m,l;z_j\right)\right)$ for large $J$.

**Proof:** Based on the pointwise asymptotic results from Semenova and Chernozhukov (2021), the change in asymptotic variance for the *dml* and *grf* GATE estimator for each group $j$ within $J$ will be proportional to $\pi_j/\pi_{j'}$ with $J$ increasing to $J'$. The behaviour of the *mcf* estimator of the GATE resembles in this situation a behaviour of a smoothing estimator. Inherently, the weights-based variance of the *mcf* estimator does not depend on $J$. ■

### *Appendix A.1.2. Assumptions and IATE*

Next, we present the asymptotic properties of the *mcf* (abstracting from local centring).[44]

The plain-vanilla *mcf* procedure aggregates individual trees $T$ into a forest in the following way: The data are split into the training set ("split-construction data"), $\Im_{tr}$, and estimation set ("leaf-populating data"), $\Im_{est}$, to form two fixed non-overlapping halves of

[44] More details of the assumptions needed as well as the formal proofs are contained in Appendix A.1.4 below.

the full data set of sizes $N_1 = N_2 = N/2$. The subsampling rates for each tree are $\beta_1$ and $\beta_2$, $1/2 < \beta_1 < \beta_2 \leq 1$, for the training and estimation set respectively, leading to a subsample size of $s_1 = c_1 N_1^{\beta_1}$ for the training set and of $s_2 = c_2 N_2^{\beta_2}$ for the estimation set, where the constant $c_1$ represents how the subsample is further split for the tree building and $c_2$ represents further potential split for the estimation set.[45] We name this procedure 'two-sample honesty' as estimation data will not become training data and vice versa. This honesty concept underpins the weights-based inference introduced in Lechner (2018), which is covered in Appendix A.2 below.

Each tree of the form $T(m, l, x; \xi, \Im_{tr}, \Im_{est})$, where $\xi \sim \Xi$ is a source of auxiliary randomness in the tree building process (such as random choice of splitting variables), can be used to estimate $IATE(m, l; x)$. When not important for proofs, $\xi$ and/or $\Im_{tr}$ and $\Im_{est}$ will be suppressed in the notation for better readability. *mcf* is an average over $B$ trees:

$$F(m, l, x; \Im_{tr}, \Im_{est}) = \frac{1}{B} \sum_{b=1}^{B} T_b(m, l, x; \xi_b, \Im_{tr,b}, \Im_{est,b}),$$

where $\Im_{tr,b}$ and $\Im_{est,b}$ are drawn without replacement from $\Im_{tr}$ and $\Im_{est}$ at subsampling rates $\beta_1$ and $\beta_2$, respectively. $\xi_b$ is a random draw from $\Xi$. The $b$ subscript will be supressed when it will not lead to any confusion. Like in Wager and Athey (2018, WA18), the trees $T$ need to be *honest*, as defined in Definition 1:

**DEFINITION 1.** A tree grown on a training sample $\Im_{tr,b}$ is *honest* if the tree does not use the responses $Y$ from the estimation sample $\Im_{est,b}$ to place its splits.

Honesty is crucial for inference and to bound the bias of the *mcf*. The splitting and subsampling procedure described above guarantees that the trees are honest. The main difference to the definition of honesty of WA18 and Athey et al. (2019) is the reverse order of steps. WA18 first subsample and then split the data for their double-sample trees. Here, the data set is split first and then the training and estimation sets are subsampled from the given split. This allows to better control bias and variance rates as the size of the training set will codetermine the size of the final leaves translating into bias and the size

[45] If $N$ is an odd number, one of the two halves will contain one observation more than the other one without any consequences. Note also that the dependence of the subsample sizes $s_1$ and $s_2$ on $N$ is suppressed in most of the following notation. Setting $c_1$ to lower values helps to build decorrelated trees, setting $c_1$ to higher values helps to reduce bias. Practically, for high values of $\beta_1$, $c_1$ should be set to lower values such as 1/2. Further results indicate that larger estimation samples help to reduce variance, therefore $c_2$ and $\beta_2$ are recommended to be set to 1.

of the estimation set will influence the variance rate.

To guarantee consistency, the final leaves must asymptotically shrink in all dimensions similarly as in Meinshausen (2006) and WA18, invoking the following definition of a random-split tree:

**DEFINITION 2.** A tree is considered a *random-split* tree if at every splitting step, marginalizing over $\xi$, there is a guaranteed minimum probability $\pi/p$ that the next split occurs along the $u$-th covariate for some $0 < \pi \leq 1$, for all $u = 1, ..., p$.

There are several options how to obtain a random-split tree. For the strategy implemented in *mcf*, see Appendix B.3.1.

The following Definition 3 controls the shape of the leaves. Definition 4 imposes *symmetry*. Both are required to derive the asymptotic results.

**DEFINITION 3.** A tree predictor grown by recursive partitioning is $(\alpha, \nu)$-*regular* for some $\alpha > 0$ if (1) each split leaves at least a fraction $\alpha$ of the available training observations of each treatment on each side of the split, (2) the final leaf containing $x$ has at least $\nu$ observations from each of the $M$ treatment groups for some $\nu \in \mathbb{N}$, and (3) the final leaf containing $x$ has at least one treatment with less than $2\nu - 1$ observations.

Regarding the role of the splitting rule on $(\alpha, \nu)$-regularity, the algorithm first determines splits that do not violate the regularity condition. For these, the splitting criterion is calculated and the split that achieves the minimum value of the objective function is chosen as the best one. By following this procedure, any influence of the splitting criterion (including the penalty) on the regularity of the final leaves can be ruled out.

**DEFINITION 4.** A predictor is *symmetric* if the output of the predictor does not depend on the order in which the observations are indexed in the training and estimation samples.

Consider $B$ trees satisfying Definitions 1-4, that training and estimation data are obtained from a two-sample honesty procedure, and let $\hat{\theta}^{mcf}_{m,l}(x)$ denote an *mcf* estimator of $IATE(m, l; x)$ obtained as

$$\widehat{IATE^{mcf}}(m,l;x) = \hat{\theta}^{mcf}_{m,l}(x) = \frac{1}{B}\sum_{b=1}^{B}\sum_{i=1}^{N_2} w^{mcf}_{i,b}(d_i, x_i; x, m, l) y_i$$
$$= \sum_{i=1}^{N_2} w^{mcf}_{i}(d_i, x_i; x, m, l) y_i,$$

where $w^{mcf}_{i,b}(d_i, x_i; x, m, l) = \left(\frac{\underline{1}(d_i=m)}{N^m_{S_b(x)}} - \frac{\underline{1}(d_i=l)}{N^l_{S_b(x)}}\right)\underline{1}(x_i \in S_b(x))$ represent the weights for the IATE estimate in tree $b$. This estimator is differencing the average outcomes of the treatment groups evaluated on the estimation subsample in the tree leaf containing point $x$, $S_b(x)$. Averaging across trees allows for a weighted representation of the forest estimator of the IATE with forest weights $w^{mcf}_{i}(d_i, x_i; x, m, l) = \frac{1}{B}\sum_{b=1}^{B} w^{mcf}_{i,b}(d_i, x_i; x, m, l)$.

Furthermore, in order to derive the bias bound and asymptotic distribution of the estimator, the common support assumption must hold in its stricter version as in WA18, i.e. $0 < \varepsilon < p_d(x) < 1 - \varepsilon$ for all $d$ and $x$ and some small $\varepsilon$. Given these concepts, we state the main theorems guaranteeing consistency and asymptotic Gaussianity of the *mcf* estimator of the IATE. Further assumptions necessary for achieving this asymptotic distribution in the case of i.i.d. sampling are (i) Lipschitz continuity of first and second order moments of the outcome variable conditional on the covariates, (ii) using subsampling to obtain the training data for tree building (subsamples should increase with $N$, slower than $N$, but not too slow), and (iii) conditions on the covariates (independent, continuous with bounded support, $p$ is fixed, i.e. low-dimensional). All proofs are collected in the Appendix A.1.4. Each section in Appendix A.1.4 contains all the necessary proofs and intermediate results for the corresponding main theorem. One of the main differences to the proofs in WA18 is caused by the different splitting and subsampling approach to achieve honesty. The Causal Forest estimator in WA18 approximates a U-statistic. Thus, their proofs are based on the corresponding Gaussian theory. Since the *mcf* allows for $\beta_2 = 1$, the estimator cannot be generally interpreted as a U-statistic and so the results in WA18 do not apply. Instead, the weighted representation of the forest estimate is leveraged to obtain asymptotic properties that also cover the case of $\beta_2 = 1$.

The first result is the bound on the bias of the forest estimator. The proof is like the proof in WA18. In the first step, we show that the leaves get small in volume as $s_1$ gets large. In the second step, we show that the estimation observations in the final leaf can be seen as a subset of nearest neighbours around the point $x$. Thus, their expected distance and Lipschitz continuity help to bound the bias. Lemma 1 below in Appendix A.1.4.1

gives rates at which the Lebesgue measure of the final leaves in a regular tree shrinks under the assumption of the covariates being independent from each other and uniformly distributed.

**THEOREM 1.** In addition to the conditions of Lemma 1 (regular trees and independent uniformly distributed covariates), suppose that trees $T$ are random-split and honest and all $E\left[Y^d \middle| X = x\right]$ are Lipschitz continuous. Then, the absolute bias of the *mcf* IATE estimator at a given value of $x$ is bounded by

$$\left|E\left[\hat{\theta}^{mcf}_{m,l}(x)\right] - \theta^0_{m,l}(x)\right| = O\left(s_1^{-\log(1-\alpha)/p\log(\alpha)}\right).$$

Note that the bias rate of the forest is the same as the bias rate of a single tree, as a forest prediction is the average of the tree predictions. The bias rate is mainly driven by the shrinkage of the Lebesgue measure of the leaf that is influenced by the choice of parameter $\alpha$. The upper bound in Theorem 1 resembles the bias rate of nearest-neighbours regression estimators under Lipschitz continuity in a $p$-dimensional space for which the rate of the expected number of nearest-neighbours is influenced by the parameter $\alpha$. This stems from the fact that the final leaves can be bounded by balls whose volume shrink at the same rate as the volume of the final leaves, as shown in the proof in the Appendix A.1.4.1. Note that this rate is rather conservative as it stems from bounding the shallowest final leaf with a leaf that would always end up with a $(1-\alpha)$ share of observations at the same level of depth. The rate is faster than in WA18 because we bound by the expected distance between the nearest neighbours and the point $x$ instead of the longest expected diameter of the leaf.

The second result is the asymptotic distribution of the IATE estimator at point $x$.

**THEOREM 2.** Assume that there is a sample of size $N$ containing i.i.d. data $(X_i, Y_i, D_i) \in [0,1]^p \times \mathbb{R} \times \{0, 1, ..., M-1\}$ for a given value of $x$. Moreover, the covariates are independently and uniformly distributed $X \sim U([0,1]^p)$. Let $T$ be an honest, regular, and symmetric random-split tree. Further assume that $E\left[Y^d \middle| X = x\right]$ and $E\left[(Y^d)^2 \middle| X = x\right]$ are Lipschitz continuous and $Var\left(Y^d \middle| X = x\right) > 0$. Then for $1/2 < \beta_1 < \beta_2 < \frac{p+2}{p}\frac{\log(1-\alpha)}{\log(\alpha)}\beta_1$,

$$\frac{\hat{\theta}^{mcf}_{m,l}(x) - \theta^0_{m,l}(x)}{\sqrt{Var\left(\hat{\theta}^{mcf}_{m,l}(x)\right)}} \xrightarrow{d} N(0,1).$$

The restrictions on the sampling rates require that $(2+p)\log(1-\alpha)/(p\log(\alpha)) > 1$. This suggests that $\alpha$ needs to be set closer and closer to 0.5 for larger $p$. This is a consequence of the conservative bias bound and the curse of dimensionality as values of $\alpha$ closer to 0.5 make sure that the shallowest final leaves get tighter upper bounds ensuring that the bias vanishes fast enough. Additionally, the relationship between the subsampling rates further ensures that the final leaf does not end up with too many honest observations and the squared bias–variance ratio converges to 0. The results further indicate that variance is converging to zero the fastest when $\beta_2 = 1$. Similarly, higher values of $\beta_1$ make the bias converge to zero faster. Due to the conservativeness of the bias bound, it is not straightforward to suggest a value for parameter $\alpha$. Following the conservative bound, $\alpha$ should be set closer and closer to 1/2 for higher dimensions (increasing $p$). Practically, lower values of $\alpha$ increase the flexibility of the splitting procedure allowing daughter leaves to contain at least an $\alpha$ share of the observations in the parent node, i.e. still allowing for splits that would be enforced by higher $\alpha$. The convergence of the bias-variance ratio to 0 together with a side result of the variance converging to 0 guarantees the consistency of the *mcf* estimator of the IATE as captured in the following corollary.

**COROLLARY 2.** Let all assumptions from Theorem 2 hold. Then, $\hat{\theta}^{mcf}_{m,l}(x) \xrightarrow{p} \theta^{0}_{m,l}(x)$.

*Appendix A.1.3. GATEs and ATE*

Estimators for GATEs and ATEs are obtained by averaging the IATEs in the respective subsamples defined by $z$ (assuming discrete $Z$). Although estimating ATEs and GATEs directly instead of aggregating IATEs could lead to more efficient estimators (see Subsection 4.1 and Subsection 4.2), the computational burden would also be higher, in particular if the number of GATEs of interest is large, as is common in many empirical studies. Furthermore, there is no guarantee that the results are internally consistent, i.e., the respective averages of the IATEs are indeed close to their ATE and GATE counterparts. Therefore, letting $\hat{\theta}^{mcf}_{m,l}(x)$ be an estimator of $IATE(m,l,x)$, the *mcf* estimator computes GATEs and ATEs as appropriate averages of $\hat{\theta}^{mcf}_{m,l}(x)$'s:

$$\begin{aligned}\widehat{GATE}^{mcf}(m,l;z) = \hat{\theta}^{mcf}_{m,l}(z) &= \frac{1}{N_2^z}\sum_{i=1}^{N_2} \underline{1}(z_i = z)\hat{\theta}^{mcf}_{m,l}(x_i) \\ &= \sum_{i=1}^{N_2} w_i^{mcf}(z,m,l)y_i;\end{aligned}$$

$$w_i^{mcf}(z,m,l) = \frac{1}{N_2^z}\sum_{j=1}^{N_2} \underline{1}(z_j = z) w_i^{mcf}(d_i, x_i; x_j, m, l); \qquad N_2^z = \sum_{i=1}^{N_2} \underline{1}(z_i = z).$$

$$\begin{aligned}
\widehat{ATE}^{mcf}(m,l) = \hat{\theta}_{m,l}^{mcf} &= \frac{1}{N_2}\sum_{i=1}^{N_2} \hat{\theta}_{m,l}^{mcf}(x_i) \\
&= \sum_{i=1}^{N_2} w_i^{mcf}(m,l) y_i; \\
w_i^{mcf}(m,l) &= \frac{1}{N_2}\sum_{j=1}^{N_2} w_i^{mcf}(d_i, x_i; x_j, m, l).
\end{aligned}$$

These expressions show that ATEs and GATEs have the same weights-based representation as the IATEs. Hence, asymptotic Gaussianity can be established in a similar way.

**THEOREM 3.** Let all assumptions of Theorem 2 hold. Then,

$$\frac{\hat{\theta}_{m,l}^{mcf} - \theta_{m,l}^0}{\sqrt{Var\left(\hat{\theta}_{m,l}^{mcf}\right)}} \xrightarrow{d} N(0,1).$$

Utilizing the same Central Limit Theorem for triangular arrays as in Theorem 3 and assuming that $Z$ is among the splitting variables yields the following corollary:

**COROLLARY 3.** Let all assumptions from Theorem 2 hold and assume that a discrete $Z$ is among the splitting variables. Then for $Z = z$,

$$\frac{\hat{\theta}_{m,l}^{mcf}(z) - \theta_{m,l}^0(z)}{\sqrt{Var\left(\hat{\theta}_{m,l}^{mcf}(z)\right)}} \xrightarrow{d} N(0,1).$$

*Appendix A.1.4. Proofs*

The following notation as in WA18 is used for the asymptotic scaling: $f(s) \gtrsim g(s)$ means that $\liminf_{s\to\infty} f(s)/g(s) \geq 1$, and $f(s) \lesssim g(s)$ means that $\liminf_{s\to\infty} f(s)/g(s) \leq 1$. Further, $f(s) = \Omega(g(s))$ means that $\liminf_{s\to\infty} |f(s)|/g(s) > 0$, i.e., that $|f(s)|$ is bounded below by $g(s)$ asymptotically.

*Appendix A.1.4.1. Proofs for Theorem 1*

In this section, we collect all necessary results for the bias bound. In the first step, we show that the volume of the final tree leaf containing point $x$ shrinks with larger

subsample size $s_1$. We focus on the volume instead of the diameter in comparison to WA18 as the volume is important to determine the expected value of estimation samples in the leaf as the subsampling rates in the training and estimation set differ.

**LEMMA 1.** (based on Lemma 1 in WA18) Let $S(x)$ be a final leaf containing the point $x$ in a regular tree according to the definitions above and let $\lambda(S(x))$ be its Lebesgue measure. Suppose that $X \sim U([0,1]^p)$ independently. Then for $\alpha \leq 0.5$, the expected value of the Lebesgue measure of the final leaf has the following bounds

$$E[\lambda(S(x))] = O\left(s_1^{-\log(1-\alpha)/\log(\alpha)}\right),$$
$$E[\lambda(S(x))] = \Omega\left(s_1^{-1}\right).$$

**Proof:** Let $c(x)$ denote number of splits leading to the leaf $S(x)$ and $s_1^d$ be the number of observations treated with treatment $d$ in the training subsample. By the results in Wager and Walther (2015), in particular using their Lemma 12, Lemma 13 and Corollary 14, with high probability and simultaneously for all but last $O(\log(\log s_1))$ parent nodes above $S(x)$, the number of training observations in the node divided by $s_1$ is within a factor $1 + o(1)$ of the Lebesgue measure of the node. Therefore, for large enough $s_1$ with probability greater than $1 - 1/s_1$ it holds that

$$\lambda(S(x)) \leq (1 - \alpha + o(1))^{c(x)}.$$

To further evaluate the upper bound for $\lambda(S(x))$, the smallest number of splits that could lead to a leaf $S(x)$ needs to be determined. Let $s_{min} = \min_d s_{1,d}$ denote the smallest treatment in the training subsample. By regularity, the following holds $s_{min}\alpha^{c(x)} \leq 2\nu - 1$. Since $s_{min} \gtrsim s_1\varepsilon$, then $c(x) \geq \log((2\nu - 1)/(s_1\varepsilon))/\log(\alpha)$ for large $s_1$ and $\lambda(S(x))$ is bounded by

$$\lambda(S(x)) = O\left(s_1^{-\log(1-\alpha)/\log(\alpha)}\right).$$

This also translates into the upper bound for $E[\lambda(S(x))]$.

Let $N_{S(x)}$ and $N_{S(x)}^d$ denote the number of observations in leaf $S(x)$ and the number of observations treated with $d$ in leaf $S(x)$, respectively. Regarding the lower bound, by regularity, $N_{S(x)}^d \geq \nu$ holds for all treatments. It can be shown that the probability of $N_{S(x)} \geq \nu M$ when $\lambda(S(x)) \leq \nu M/s_1$ decays at least at a rate of $1/s_1$ for uniformly

distributed covariates. Therefore, $\nu M/s_1 \leq \lambda(S(x))$ with probability $1-O(1/s_1)$ yielding $E\left[\lambda(S(x))\right] = \Omega\left(s_1^{-1}\right)$. The upper bound is above the lower bound for $\alpha \leq 1/2$. ■

**THEOREM 1.** Under the conditions of Lemma 1, suppose that trees $T$ are random-split and honest and $E\left[Y^d \middle| X = x\right]$ are Lipschitz continuous. Then, the bias of the *mcf* estimator of the IATE for a given value of $x$ is bounded by

$$\left|E\left[\hat{\theta}^{mcf}_{m,l}(x)\right] - \theta^0_{m,l}(x)\right| = O\left(s_1{}^{-\log(1-\alpha)/p\log(\alpha)}\right).$$

**Proof:** As a forest prediction is an average of the tree predictions, the rate of the bias of a tree prediction is also the rate of the bias of the forest prediction. Let $T_b(x,d)$ denote a tree prediction of a potential outcome of treatment $d$ at point $x$ of a form

$$T_b(x,d) = \sum_{j=1}^{s_2} w^{mcf}_{bj,b}(d_{bj}, x_{bj}; x, d) y_{bj},$$

where $b_j$ denotes the $j$-th observation in the $b$-th estimation subsample. The respective weights are given by:

$$w^{mcf}_{bj,b}(d_{bj}, x_{bj}; x, d) = \begin{cases} \left(N^d_{S_b(x)}\right)^{-1} & \text{if} \quad x_{bj} \in S_b(x) \quad \text{and} \quad d_{bj} = d \\ 0 & \text{else} \end{cases}.$$

The tree estimator of the treatment effect is $T(m,l,x) = T(x,m) - T(x,l)$. As the treatment effect estimator is a difference of two potential outcome estimators, the absolute bias can be bounded by the rate of the bias of the potential outcome estimators as

$$\left|E\left[T(m,l,x)\right] - IATE(m,l;x)\right| \leq \sum_{d\in\{m,l\}} \left|E\left[T(x,d)\right] - E\left[Y^d \middle| X = x\right]\right|.$$

Therefore, in the following, the bias of the tree estimator for the potential outcome will be analysed.

The next observation is that if $E\left[Y^d \middle| X = x\right]$ is Lipschitz, then $E\left[Y \middle| X = x, D = d\right]$ is also Lipschitz as the two expectations coincide under the CIA, CS and observation rule. By using Jensen's inequality and Lipschitz continuity, the absolute bias can be bounded by

$$
\begin{aligned}
\left|E\left[T_b(x,d)\right] - E\left[Y^d \,|X=x\right]\right| &= \left|E\left[E\left[\frac{1}{N^d_{S_b(x)}} \sum_{i\in S_b(x), D_i=d} Y_i \middle| N^d_{S_b(x)}\right]\right] - E\left[Y^d \,|X=x\right]\right| \\
&= \left|E\left[\frac{1}{N^d_{S_b(x)}} \sum_{i\in S_b(x), D_i=d} E\left[Y_i \middle| N^d_{S_b(x)}\right]\right] - E\left[Y^d \,|X=x\right]\right| \\
&\leq E\left[\frac{1}{N^d_{S_b(x)}} \sum_{i\in S_b(x), D_i=d} \left|E\left[Y_i \middle| N^d_{S_b(x)}\right] - E\left[Y^d \,|X=x\right]\right|\right] \\
&= E\left[\frac{1}{N^d_{S_b(x)}} \sum_{i\in S_b(x), D_i=d} \left|E\left[Y_i \middle| N^d_{S_b(x)}\right] - E\left[Y_i \,|X_i=x, D_i=d\right]\right|\right] \\
&\leq E\left[\frac{1}{N^d_{S_b(x)}} \sum_{i\in S_b(x), D_i=d} C_d \left\|X_i - x\right\|\right].
\end{aligned}
$$

Using Lipschitz continuity moves the analysis of the bound from the $Y|X$ space to the covariate space. This allows bounding the bias using the random-split assumption. In the rest of the proof, we can assume that $\pi = 1$ since the rate on a distance between conditional expectation of $Y$ on the leaf and conditional expectation of $Y$ at $x$ is the slowest for $\pi = 1$. The rationale for the bias speeding up for $0 < \pi < 1$ is that less relevant dimensions are split less frequently and partitioning focuses on dimensions more relevant for approximating $E[Y|X = x]$. As the $\alpha$ regularity would yield a very loose bound on the expected distance, we take a different approach here based on the nearest neighbours (NN) of the point $x$ that also contain all the observations in the final leaf. Each final leaf can be bounded by a ball with the centre at $x$ and radius equal to the longest segment of the leaf which we denote as $diam(S_b(x))$. Then the number of treated observations in the ball, denoted as $N^d_{B(S_b(x))}$, follows a binomial distribution for $s_2$ observations and the success probability being the Lebesgue measure of the ball that can be seen as a function of the $diam(S_b(x))$ since the covariates are independent and uniformly distributed.[46] At the same time the number of observations in the final leaf $N^d_{S_b(x)}$ follows also a binomial distribution for $s_2$ observations and the success probability being the Lebesgue measure of the final leaf that can be seen as a function of the $diam(S_b(x))$ and the angles to the vertices from the "origin" vertex using a high-dimensional polar system. As both random variables depend on the same power of $diam(S_b(x))$ that shrinks to 0 due to random-split property such that we can conclude that $E[N^d_{B(S_b(x))}] \leq O\left(E[N^d_{S_b(x)}]\right)$ with a constant larger than 1. As the ball contains the final leaf, $\sum_{i\in S_b(x), D_i=d} \|X_i - x\| \leq \sum_{i\in B(S_b(x)), D_i=d} \|X_i - x\|$.

[46] Without loss of generality, we assume constant treatment probability for all $x$. The proof can be generalized to treatment probabilities varying with $x$, leading to a different constant in the following bounds.

When we randomly split all data points in the estimation subsample with treatment $d$ into $N^d_{B(S_b(x))}+1$ segments, the first $N^d_{B(S_b(x))}$ segments will have a length $s_{2,d}/N^d_{B(S_b(x))}$. Denote $\tilde{X}^x_j$ as the first nearest neighbour in the $j^{\text{th}}$ segment. Then,

$$\sum_{i\in B(S_b(x)),D_i=d} \|X_i - x\| \leq \sum_{\substack{j=1\\ D_j=d}}^{N^d_{B(S_b(x))}} \left\|\tilde{X}^x_j - x\right\|.$$

Therefore, the absolute bias can be further bounded by

$$\begin{aligned}
\left|E\left[T_b(x,d)\right] - E\left[Y^d \,|X=x\right]\right| &\leq C_d E\left[\frac{1}{N^d_{S_b(x)}} \sum_{\substack{j=1\\ D_j=d}}^{N^d_{B(S_b(x))}} \left\|\tilde{X}^x_j - x\right\|\right] \\
&= C_d E\left[E\left[\frac{1}{N^d_{S_b(x)}} \sum_{\substack{j=1\\ D_j=d}}^{N^d_{B(S_b(x))}} \left\|\tilde{X}^x_j - x\right\| \middle| N^d_{S_b(x)}, N^d_{B(S_b(x))}\right]\right] \\
&= C_d E\left[\frac{N^d_{B(S_b(x))}}{N^d_{S_b(x)}} E\left[\left\|\tilde{X}^x_1 - x\right\| \middle| N^d_{B(S_b(x))}\right]\right] \\
&\leq C_{d,B} E\left[E\left[\left\|\tilde{X}^x_1 - x\right\| \middle| N^d_{B(S_b(x))}\right]\right] + O\left(1/s_2\right) \\
&= C_{d,B} E\left[E\left[\left\|X_{\left(1,\left\lfloor s_{2,d}/N^d_{B(S_b(x))}\right\rfloor\right)} - x\right\| \middle| N^d_{B(S_b(x))}\right]\right] + O\left(1/s_2\right),
\end{aligned}$$

where $X_{(1,N)}$ denotes the first nearest neighbour among $N$ observations and $C_{d,B}$ collects the Lipschitz constant and the constant from the ratio of the expected number of observations in the ball and the final leaf. The result is obtained by applying a second order Taylor approximation of $N^d_{B(S_b(x))}/N^d_{S_b(x)}$ around $(E[N^d_{B(S_b(x))}], E[N^d_{S_b(x)}])$, the bound on the ratio of expected number of observations in the ball and the final leaf, the fact that $\left\|\tilde{X}^x_1 - x\right\|$ is bounded and Cauchy-Schwarz inequality. For a fixed $N^d_{B(S_b(x))}$ in a ball containing a leaf from a regular, random-split tree, we can use results in Györfi, Kohler, Krzyżak, and Walk (2002) for nearest neighbour (NN) estimators that also use the expected distance of the first neighbours in their Theorem 6.2 and Lemma 6.4 yielding the bound

$$C_{d,B} E\left[E\left[\left\|X_{\left(1,\left\lfloor s_{2,d}/N^d_{B(S_b(x))}\right\rfloor\right)} - x\right\| \middle| N^d_{B(S_b(x))}\right]\right] \leq \frac{\tilde{c}_d}{s_{2,d}^{1/p}} E\left[{N^d_{S_b(x)}}^{\frac{1}{p}}\right],$$

where $\tilde{c}_d$ collects the constant $C_{d,B}$ and all constants that emerge in the NN proof. The upper bound for the last expectation can be then found by applying Jensen's inequality,

$$E\left[E\left[\left(\frac{N^d_{S_b(x)}}{s_{2,d}}\right)^{\frac{1}{p}}\middle|\lambda(S_b(x))\right]\right] \leq E\left[\left(\frac{E\left[N^d_{S_b(x)}\middle|\lambda(S_b(x))\right]}{s_{2,d}}\right)^{\frac{1}{p}}\right]$$
$$\leq \tilde{c}_\varepsilon E\left[(\lambda(S_b(x)))^{\frac{1}{p}}\right] \leq \tilde{c}_\varepsilon (E\left[\lambda(S_b(x))\right])^{\frac{1}{p}},$$

where $\tilde{c}_\varepsilon$ collects constants stemming from the common support assumption. Using the results from Lemma 1, the result at the tree level is

$$\left|E\left[T_b(x,d)\right] - E\left[Y^d\,|X=x\right]\right| = O\left(s_1{}^{-\log(1-\alpha)/p\log(\alpha)}\right).$$

The final constant is a function of the Lipschitz constant, the constant from the ratio of the observations in the ball and the final leaf, the common support parameter $\varepsilon$, and $\nu$, the regularity parameter controlling the number of the observations in the final leaf. As the forest estimator is the average of the tree estimators, the result above holds also for the forest estimators. ■

*Appendix A.1.4.2. Proofs for Theorem 2*

The proofs of asymptotic Gaussianity build on a central limit theorem for weakly dependent random variables of a form $A_{i,N_2} = W_iY_i - E\left[W_iY_i\right]$, introduced in Neumann (2013). In this section, we prove that $A_{i,N_2}$ satisfy the conditions of the CLT yielding the first necessary result. As not all $Y_i$ are unbiased estimators of $\mu_d(x)$, the weighted average is not an unbiased estimator. Therefore, in the second step it is necessary to show that the ratio of bias and variance converges to zero as the sample gets larger and the final leaf shrinks asymptotically.

For the analysis of the variance of the forest estimator of the IATE, we observe the following:

**COROLLARY 1.** For any forest estimator $F$ that averages $B$ tree estimators $T$, the rate of the forest variance $Var(F)$ is bounded from above and below by the rate of the individual tree variance $Var(T)$.

**Proof:** The variance of a forest estimator in a simplified notation is

$$Var(F) = \frac{1}{B^2}\left[\sum_{b=1}^{B} Var(T_b) + \sum_{b=1}^{B}\sum_{b'\neq b} Cov(T_b, T_{b'})\right].$$

As a forest is an average of trees, the worst upper bound of the variance of the forest estimator is the variance of an individual tree estimator. Lemma 3 shows the upper bound for the tree estimator that converges to zero for $\beta_2 > \beta_1$.

Since any covariance has to go to zero at least as fast as the variance, the lower bound of the rate of the variance is also determined by the variance of an individual tree. ■

In the following, we therefore focus on deriving the properties of the tree weights.

**LEMMA 2.** Let $W_{i,b} := W^{mcf}_{bi,b}(D, X; x, d)$ where the subscript $bi$ represents $i$-th observation in the estimation sample of size $N_2$ that was subsampled to the subsample $b$ of size $s_2$. Suppose that the assumptions from Lemma 1 hold and the tree is honest and symmetric. Moreover $\beta_2 > \beta_1/2 > 1/4$. Then, the moments of the tree weights have the following rates and bounds:

a) $E\left[W_{i,b}\right] \sim \frac{1}{N^{\beta_2}}$ ,

b) $E\left[W_{i,b}^2\right] = \Omega\left(N^{\frac{\log(1-\alpha)}{\log(\alpha)}\beta_1 - 2\beta_2}\right)$ and $E\left[W_{i,b}^2\right] = O\left(N^{\beta_1 - 2\beta_2}\right)$,

c) $Var(W_{i,b})$ has the same bounds as $E\left[W_{i,b}^2\right]$,

d) $\left|Cov\left(W_{i,b}, W_{j,b}\right)\right| = \Omega\left(N^{\frac{\log(1-\alpha)}{\log(\alpha)}\beta_1 - 3\beta_2}\right)$ and $\left|Cov\left(W_{i,b}, W_{j,b}\right)\right| = O\left(N^{\beta_1 - 3\beta_2}\right)$.

**Proof:**

a) Due to symmetry, the expected value of the first moment of the tree weights can be expressed as:

$$\begin{aligned} E\left[W_{i,b}\right] &= E\left[E\left[W_{i,b}\middle|N^d_{S_b(x)}\right]\right] \\ &= E\left[\frac{s_2 - N^d_{S_b(x)}}{s_2}\cdot 0 + \frac{N^d_{S_b(x)}}{s_2}\frac{1}{N^d_{S_b(x)}}\right] \\ &= \frac{1}{s_2} \sim \frac{1}{N^{\beta_2}}. \end{aligned}$$

b) Let $p_d(S(x))$ be the propensity score of getting treatment $d$ when in leaf $S(x)$. Due to the common support assumption $p_d(S(x))$ must lie in an interval $(\varepsilon, 1-\varepsilon)$ for large $N$. Thus, under symmetry, the expected value of the second moment of the tree weights can be expressed as:

$$
\begin{aligned}
E\left[W_{i,b}^2\right] &= E\left[E\left[W_{i,b}^2 \middle| N_{S_b(x)}^d\right]\right] \\
&= E\left[\frac{s_2 - N_{S_b(x)}^d}{s_2} \cdot 0 + \frac{N_{S_b(x)}^d}{s_2} \frac{1}{{N_{S_b(x)}^d}^2}\right] \\
&= \frac{1}{s_2} E\left[\frac{1}{N_{S_b(x)}^d}\right] = \frac{1}{s_2} E\left[\frac{1-(1-\lambda(S(x))p_d(S(x)))^{s_2}}{s_2\lambda(S(x))p_d(S(x))}\right],
\end{aligned}
$$

where the last equality uses the law of iterated expectations and the fact that $N_{S_b(x)}^d$ is a positive binomial random variable. Since $E\left[1/N_{S_b(x)}^d\right] > 0$, the lower bound is determined by $E\left[1/s_2\lambda(S(x))p_d(S(x))\right]$. The lower bound can then be derived as follows using Jensen's inequality and results in Lemma 1:

$$
E\left[W_{i,b}^2\right] = \Omega\left(\frac{1}{s_2 s_2 E\left[\lambda(S(x))\right]}\right) = \Omega\left(N^{-2\beta_2 + \frac{\log(1-\alpha)}{\log(\alpha)}\beta_1}\right).
$$

The upper bound can be similarly derived for $\beta_2 > \beta_1/2 > 1/4$ as

$$
\begin{aligned}
E\left[\frac{1}{N_{S_b(x)}^d}\right] &= E\left[\frac{1-(1-\lambda(S(x))p_d(S(x)))^{s_2}}{s_2\lambda(S(x))p_d(S(x))}\right] \\
&= E\left[\frac{1-(1-\lambda(S(x))p_d(S(x)))^{s_2}}{s_2\lambda(S(x))p_d(S(x))} \middle| \lambda(S(x)) > \nu M/s_1\right] \Pr\left(\lambda(S(x)) < \nu M/s_1\right) \\
&+ E\left[\frac{1-(1-\lambda(S(x))p_d(S(x)))^{s_2}}{s_2\lambda(S(x))p_d(S(x))} \middle| \lambda(S(x)) \geq \nu M/s_1\right] \Pr\left(\lambda(S(x)) \geq \nu M/s_1\right) \\
&\leq 1 \cdot O(1/s_1) + O(s_1/s_2) \cdot 1 = O\left(1/N^{\beta_2 - \beta_1}\right).
\end{aligned}
$$

The two results yield

$$
\begin{aligned}
E\left[W_{i,b}^2\right] &= \Omega\left(N^{-2\beta_2 + \frac{\log(1-\alpha)}{\log(\alpha)}\beta_1}\right), \\
E\left[W_{i,b}^2\right] &= O\left(N^{-2\beta_2 + \beta_1}\right).
\end{aligned}
$$

Both rates have constants that depend on $\varepsilon$, the common support parameter and regularity parameter $\nu$. The constant for the upper bound also depends on the number of treatments as the number of treatments influences the smallest expected Lebesgue measure of the leaf.

c) Since $E\left[W_{i,b}^2\right] \to 0$ for $\beta_2 > \beta_1/2 > 1/4$, the bounds for $Var(W_{i,b})$ are the same as for $E\left[W_{i,b}^2\right]$.

d) The covariance can be expressed as

$$Cov(W_{i,b}, W_{j,b}) = Cov\left(W_{i,b}, 1 - \sum_{k \neq j} W_{k,b}\right) = -Var(W_{i,b}) - \sum_{k \neq i,j} Cov(W_{i,b}, W_{k,b})$$

yielding

$$Cov(W_{i,b}, W_{j,b}) = -\frac{Var(W_{i,b})}{s_2 - 1}.$$

The bounds for the covariance are

$$|Cov\left(W_{i,b}, W_{j,b}\right)| = \Omega\left(N^{-3\beta_2 + \frac{\log(1-\alpha)}{\log(\alpha)}\beta_1}\right),$$
$$|Cov\left(W_{i,b}, W_{j,b}\right)| = O\left(N^{-3\beta_2+\beta_1}\right).$$

Note that

$$\sum_{k \neq j} Cov(W_{i,b}, W_{k,b}) = \frac{Var(W_{i,b})}{s_2 - 1}.$$

■

**LEMMA 3.** Suppose that the tree conditions from Lemma 2 hold i.e., we build regular, honest, symmetric trees. Additionally, $E\left[Y^d \middle| X = x\right]$ is Lipschitz continuous. Moreover, assume that $E\left[\left(Y^d\right)^2 \middle| X = x\right]$ is also Lipschitz continuous and $Var\left(Y^d | X = x\right) > 0$. Further, the sampling rates satisfy $\beta_2 > \beta_1 > 1/2$. Then the tree variance has the following rates

$$Var\left(\sum_{i=1}^{s_2} W_{i,b} Y_i\right) = O\left(N^{\beta_1 - \beta_2}\right),$$
$$Var\left(\sum_{i=1}^{s_2} W_{i,b} Y_i\right) = \Omega\left(N^{\frac{\log(1-\alpha)}{\log(\alpha)}\beta_1 - \beta_2}\right).$$

**Proof:** The variance of a tree estimator of $\mu_d(x)$ can be decomposed as

$$Var\left(\sum_{i=1}^{s_2} W_{i,b} Y_i\right) = \sum_{i=1}^{s_2} Var\left(W_{i,b} Y_i\right) + \sum_{i=1}^{s_2} \sum_{j \neq i} Cov\left(W_{i,b} Y_i, W_{j,b} Y_j\right).$$

The upper bound for $Var\left(W_{i,b} Y_i\right)$ is

$$\begin{aligned} Var\left(W_{i,b} Y_i\right) &= E\left[W_{i,b}^2 Y_i^2\right] - E^2\left[W_{i,b} Y_i\right] \\ &\leq E\left[W_{i,b}^2 Y_i^2\right] = E\left[E\left[W_{i,b}^2 \middle| X_i\right] E\left[Y_i^2 \middle| X_i\right]\right] \end{aligned}$$

$$
\begin{aligned}
&\leq E\left[E\left[W_{i,b}^2\middle| X_i\right] E\left[\sup_{d,x} E\left[\left(Y_i\right)^2\middle| X_i = x, D_i = d\right]\right]\right] \\
&= E\left[E\left[W_{i,b}^2\middle| X_i\right] E\left[\sup_{d,x} E\left[\left(Y_i^d\right)^2\middle| X_i = x\right]\right]\right] \\
&\leq E\left[W_{i,b}^2\right] C_2 = O\left(N^{-2\beta_2+\beta_1}\right),
\end{aligned}
$$

using the identifying assumptions, Lipschitz continuity of $E\left[\left(Y_i^d\right)^2\middle| X_i\right]$ on a bounded covariate space and noting that $W_{i,b}$ and $Y_i$ are independent conditional on $X_i$. This yields that

$$
\sum_{i=1}^{s_2} Var\left(W_{i,b}Y_i\right) \leq s_2 E\left[W_{i,b}^2\right] C_2 = O\left(N^{\beta_1-\beta_2}\right).
$$

Note that this upper bound converges to zero for $\beta_2 > \beta_1 > 1/2$. In order to derive the upper bound for the covariance part, we rewrite the covariance as

$$
\begin{aligned}
\left|Cov(W_{i,b}Y_i, W_{j,b}Y_j)\right| &= \left|Cov\left(W_{i,b}Y_i, \left(1 - \sum_{k\neq j} W_{k,b}\right) Y_j\right)\right| \\
&= \left|-\sum_{k\neq j} Cov\left(W_{i,b}Y_i, W_{k,b}Y_j\right)\right| \\
&= \left|\sum_{k\neq j} Cov\left(W_{i,b}Y_i, W_{k,b}Y_j\right)\right|.
\end{aligned}
$$

Next, we derive the upper bound of $\sum_{k\neq j} Cov\left(W_{i,b}Y_i, W_{k,b}Y_j\right)$

$$
\begin{aligned}
\sum_{k\neq j} Cov\left(W_{i,b}Y_i, W_{k,b}Y_j\right) &= \sum_{k\neq j} E\left[W_{i,b}Y_iW_{k,b}Y_j\right] - E\left[W_{i,b}Y_i\right] E\left[W_{k,b}Y_j\right] \\
&= \sum_{k\neq j} E\left[W_{i,b}Y_iW_{k,b}\right] E\left[Y_j\right] - E\left[W_{i,b}Y_i\right] E\left[W_{k,b}\right] E\left[Y_j\right] \\
&= \sum_{k\neq j} \left(E\left[W_{i,b}Y_iW_{k,b}\right] - E\left[W_{i,b}Y_i\right] E\left[W_{k,b}\right]\right) E\left[Y_j\right] \\
&= \sum_{k\neq j} \left(E\left[E\left[W_{i,b}W_{k,b}\middle| X_i, X_k\right] E\left[Y_i\middle| X_i\right]\right]\right. \\
&\quad \left. - E\left[E\left[W_{i,b}\middle| X_i\right] E\left[Y_i\middle| X_i\right]\right] E\left[W_{k,b}\right]\right) E\left[Y_j\right] \\
&= \sum_{k\neq j} \left(E\left[\left(E\left[W_{i,b}W_{k,b}\middle| X_i, X_k\right] - E\left[W_{i,b}\middle| X_i\right] E\left[W_{k,b}\right]\right) E\left[Y_i\middle| X_i\right]\right]\right) E\left[Y_j\right] \\
&\leq \sum_{k\neq j} \left(E\left[W_{i,b}W_{k,b}\right] - E\left[W_{i,b}\right] E\left[W_{k,b}\right]\right) C_1^2 \\
&= \sum_{k\neq j} Cov(W_{i,b}, W_{k,b}) C_1^2 = \frac{Var(W_{i,b})}{s_2 - 1} C_1^2,
\end{aligned}
$$

where the inequality uses the same supremum argument as the proof for the variance, the fact that the expected potential outcomes are Lipschitz continuous on a bounded covariate space, i.e., $E\left[Y_i^d \,|X_i\right] \in [-C_1, C_1]$ and $E\left[Y_i^d\right] \in [-C_1, C_1]$ for some positive constant $C_1$, and the fact that the final sum of covariances is positive and therefore the product certainly bounds the original sum from above. This yields an upper bound for

$$\begin{aligned}\sum_{i=1}^{s_2}\sum_{j\neq i}\left|Cov\left(W_{i,b}Y_i, W_{j,b}Y_j\right)\right| &\leq s_2(s_2-1)\frac{Var(W_{i,b})}{s_2-1}C_1^2 \\ &= O\left(N^{\beta_1-\beta_2}\right).\end{aligned}$$

The overall upper bound for the tree variance is

$$Var\left(\sum_{i=1}^{s_2} W_{i,b}Y_i\right) = O\left(N^{\beta_1-\beta_2}\right).$$

The constant depends on $\varepsilon$, the common support parameter, regularity parameter $\nu$, number of treatments $M$ and a constant related to Lipschitz continuity of the outcome variable.

Due to non-negativity of the variance, it is enough to find the lower bound for $E\left[W_{i,b}^2 Y_i^2\right]$ to analyse the lower bound of $Var\left(W_{i,b}Y_i\right)$.

$$\begin{aligned}E\left[W_{i,b}^2 Y_i^2\right] &= E\left[E\left[W_{i,b}^2\,|X_i\right] E\left[Y_i^2\,|X_i\right]\right] \\ &\geq E\left[E\left[W_{i,b}^2\,|X_i\right] E\left[\inf_{d,x} E\left[Y_i^2\,|X_i = x, D_i = d\right]\right]\right] \\ &= E\left[E\left[W_{i,b}^2\,|X_i\right]\right] E\left[\inf_{d,x} E\left[\left(Y_i^d\right)^2\,|X_i = x\right]\right] \\ &= \Omega\left(N^{\frac{\log(1-\alpha)}{\log(\alpha)}\beta_1 - 2\beta_2}\right).\end{aligned}$$

Therefore, $\sum_{i=1}^{s_2} Var\left(W_{i,b}Y_i\right) = \Omega\left(N^{\frac{\log(1-\alpha)}{\log(\alpha)}\beta_1-\beta_2}\right)$.

The lower bound for the covariance can be derived using the triangular inequality,

$$\begin{aligned}\left|Cov(W_{i,b}Y_i, W_{j,b}Y_j)\right| &= \left|E\left[W_{i,b}Y_i W_{j,b}Y_j\right] - E\left[W_{i,b}Y_i\right]E\left[W_{j,b}Y_j\right]\right| \\ &\geq \Big|\left|E\left[W_{i,b}Y_i W_{j,b}Y_j\right]\right| - \left|E\left[W_{i,b}Y_i\right]E\left[W_{j,b}Y_j\right]\right|\Big|\end{aligned}$$

and finding lower bounds for the individual elements:

$$
\begin{aligned}
|E\left[W_{i,b}Y_iW_{j,b}Y_j\right]| &= |E\left[E\left[W_{i,b}Y_iW_{j,b}Y_j\,|\lambda(S(x))\right]\right]| \\
&= \left|E\left[E\left[1/\left(N^d_{S(x)}\right)^2 Y_iY_j\,|\lambda(S(x)), X_i, X_j \in S(x)\right]\Pr\left(X_i, X_j \in S(x)\,|\lambda(S(x))\right)\right]\right| \\
&\geq \frac{1}{\left(s_{2,d}\right)^2}\inf_{d,x} E\left[|Y_iY_j|\,|\,D_i = d, D_j = d, X_i = x, X_j = x\right] E\left[\left(\lambda(S(x))\right)^2\right] \\
&= \Omega\left(N^{-2\beta_2-2\beta_1}\right) \\
|E\left[W_{i,b}Y_i\right]| &= |E\left[E\left[W_{i,b}Y_i\,|\lambda(S(x))\right]\right]| \\
&= |E\left[E\left[W_{i,b}Y_i\,|X_i \in S(x), \lambda(S(x))\right]\Pr\left(X_i \in S(x)\,|\lambda(S(x))\right)\right]| \\
&\geq \frac{1}{s_{2,d}}\inf_{d,x} E\left[|Y_i|\,|\,D_i = d, X_i = x\right] E\left[\lambda(S(x))\right] \\
&= \Omega\left(N^{-\beta_2-\beta_1}\right)
\end{aligned}
$$

Since $|E\left[W_{i,b}Y_iW_{j,b}Y_j\right]| = \Omega\left(N^{-2\beta_2-2\beta_1}\right)$ and $|E\left[W_{i,b}Y_i\right]|\,|E\left[W_{j,b}Y_j\right]| = \Omega\left(N^{-2\beta_2-2\beta_1}\right)$, the final lower bound is

$$
|Cov(W_{i,b}Y_i, W_{j,b}Y_j)| = \Omega\left(N^{-2\beta_2-2\beta_1}\right).
$$

Putting these results together yields a lower bound for the covariance:

$$
\sum_{i=1}^{s_2}\sum_{j\neq i}|Cov\left(W_{i,b}Y_i, W_{j,b}Y_j\right)| = \Omega\left(N^{-2\beta_1}\right).
$$

As the lower bound for the covariance exceeds the lower bound for the variance, the lower bound for the tree variance is determined by the one converging slower to zero i.e.

$$
Var\left(\sum_{i=1}^{s_2} W_{i,b}Y_i\right) = \Omega\left(N^{\frac{\log(1-\alpha)}{\log(\alpha)}\beta_1-\beta_2}\right).
$$

The constant depends on $\varepsilon$, the common support parameter, regularity parameter $\nu$ and a constant related to Lipschitz continuity of the outcome variable. ■

**LEMMA 4.** Let the assumptions from Lemma 3 hold and $A_{i,N_2} := W_iY_i - E\left[W_iY_i\right]$, where $W_i$ are the forest weights of the *mcf*. Then $(A_{i,N_2})_{i=1,\ldots,N_2}$ is a triangular array satisfying:

a) $E[A_{i,N_2}] = 0$,

b) $\sum_{i=1}^{N_2} E[A^2_{i,N_2}] < \infty$ for all $N$ and $i$,

c) $\sigma^2_N := Var(A_{1,N_2} + \ldots + A_{N_2,N_2}) \xrightarrow[N\to\infty]{} 0$,

d) $\sum_{i=1}^{N_2} E\left[A_{i,N_2}^2 \underline{1}\left(|A_{i,N_2}| > \tilde{\varepsilon}\right)\right] \xrightarrow[N\to\infty]{} 0$ for all $\tilde{\varepsilon} > 0$,

e) There is a summable sequence $(\tilde{\pi}_r)_{r\in\mathbb{N}}$ such that for all $\tilde{u} \in \mathbb{N}$ and all indices $1 \leq q_1 < q_2 < ... < q_{\tilde{u}} < q_{\tilde{u}} + r = t_1 \leq t_2 \leq N_2$, the following upper bounds for covariances hold true for all measurable functions $g : \mathbb{R}^{\tilde{u}} \to \mathbb{R}$ with $\|g\|_\infty = \sup_{x\in\mathbb{R}^{\tilde{u}}} |g(x)| \leq 1$:

$$|Cov(g(A_{q_1,N_2}, ..., A_{q_{\tilde{u}},N_2})A_{q_{\tilde{u}},N_2}, A_{t_1,N_2})| \leq \left(E[A_{q_{\tilde{u}},N_2}^2] + E[A_{t_1,N_2}^2] + N_2^{-1}\right)\tilde{\pi}_r$$

and

$$|Cov(g(A_{q_1,N_2}, ..., A_{q_{\tilde{u}},N_2}), A_{t_1,N_2}, A_{t_2,N_2})| \leq \left(E[A_{t_1,N_2}^2] + E[A_{t_2,N_2}^2] + N_2^{-1}\right)\tilde{\pi}_r.$$

**Proof:**

a) $E\left[W_i Y_i - E\left[W_i Y_i\right]\right] = 0.$

b) $E\left[A_{i,N_2}^2\right] = Var\left(W_i Y_i\right) = E\left[W_i^2 Y_i^2\right] - E^2\left[W_i Y_i\right]$

To derive upper bounds for $E[W_i Y_i]$ and $E[W_i^2 Y_i^2]$, we use that $W_i$ and $Y_i$ are independent conditional on $X_i$ and $E\left[Y^d \middle| X = x\right]$ and $E\left[\left(Y^d\right)^2 \middle| X = x\right]$ are Lipschitz continuous and therefore can be bounded from above by $C_1 < \infty$ and $C_2 < \infty$ respectively on a bounded covariate space. The first and second moment of the forest weights are

$$E\left[W_i\right] = \frac{N_2 - s_2}{N_2} \cdot 0 + \frac{s_2}{N_2} E\left[\frac{1}{B}\sum_{b=1}^{B} W_{i,b}\right] = \frac{s_2}{N_2}\frac{1}{B} B \frac{1}{s_2} = \frac{1}{N_2} \sim N^{-1}$$

and

$$\begin{aligned}
E\left[W_i^2\right] &= \frac{N_2 - s_2}{N_2} \cdot 0 + \frac{s_2}{N_2} E\left[\left(\frac{1}{B}\sum_{b=1}^{B} W_{i,b}\right)^2\right] \\
&= \frac{s_2}{N_2}\frac{1}{B^2}\left[\sum_{b=1}^{B} E\left[W_{i,b}^2\right] + \sum_{b=1}^{B}\sum_{b'\neq b} E\left[W_{i,b} W_{i,b'}\right]\right] \\
&= \frac{s_2}{N_2}\frac{1}{B^2}\left[BE\left[W_{i,b}^2\right] + B(B-1)\left(Cov(W_{i,b}, W_{i,b'}) + \left(E\left[W_{i,b}\right]\right)^2\right)\right] \\
&\leq \frac{s_2}{N_2}\frac{1}{B^2}\left[BE\left[W_{i,b}^2\right] + B(B-1)\left(Var(W_{i,b}) + \left(E\left[W_{i,b}\right]\right)^2\right)\right] \\
&= \frac{s_2}{N_2}\frac{1}{B^2}\left[B^2 E\left[W_{i,b}^2\right]\right] = O\left(N^{-1-\beta_2+\beta_1}\right).
\end{aligned}$$

These results together with identifying assumptions and Lipschitz continuity can be used to further bound the following expectations:

$$E\left[W_i Y_i\right] = E\left[E\left[W_i \left| X_i\right.\right] E\left[Y_i \left| X_i\right.\right]\right] = O\left(N^{-1}\right),$$
$$E[W_i^2 Y_i^2] = E\left[E\left[W_i^2 \left| X_i\right.\right] E\left[Y_i^2 \left| X_i\right.\right]\right] = O\left(N^{-1-\beta_2+\beta_1}\right)$$

It follows that

$$\sum_{i=1}^{N_2} E[A_{i,N_2}^2] = O\left(N^{-\beta_2+\beta_1}\right) < \infty.$$

c)

$$\begin{aligned} Var(A_{1,N_2} + ... + A_{N_2,N_2}) &= Var\left(\sum_{i=1}^{N_2} W_i Y_i - \sum_{i=1}^{N_2} E\left[W_i Y_i\right]\right) = Var\left(\sum_{i=1}^{N_2} W_i Y_i\right) \\ &= \sum_{i=1}^{N_2} Var\left(W_i Y_i\right) + \sum_{i=1}^{N_2} \sum_{j \neq i} Cov\left(W_i Y_i, W_j Y_j\right) \\ &\leq \sum_{i=1}^{N_2} Var\left(W_i Y_i\right) + \sum_{i=1}^{N_2} \sum_{j \neq i} \left|Cov\left(W_i Y_i, W_j Y_j\right)\right| \end{aligned}$$

Using the results from b), the first element is bounded at rate $O\left(N^{-\beta_2+\beta_1}\right)$ and by a similar logic as in the tree case the sum of all covariances has to decay to zero as fast as the sum of variances. These results yield that $Var(A_{1,N_2}+...+A_{N_2,N_2}) \xrightarrow[N\to\infty]{} 0$.

d) Due to monotonicity and the result in b):

$$\sum_{i=1}^{N_2} E\left[A_{i,N_2}^2 \underline{1}\left(\left|A_{i,N_2}\right| > \tilde{\varepsilon}\right)\right] \leq \sum_{i=1}^{N_2} E\left[A_{i,N_2}^2\right] \to 0 \text{ for all } \tilde{\varepsilon} > 0.$$

e) Due to exchangeability which implies strict stationarity and the result in c), it is possible to interpret $\left(A_{i,N_2}\right)_{i=1,...,N_2}$ as a $\rho$-mixing process. Since every $\phi$-mixing process is also $\rho$-mixing, then along the Lemma 20.1 in Billingsley (1968) for $\phi$-mixing processes, we can also bound the covariances of a $\rho$-mixing process as follows:[47]

$$\begin{aligned} &\left|Cov\left(g\left(A_{q_1,N_2}, ..., A_{q_{\tilde{u}},N_2}\right) A_{q_{\tilde{u}},N_2}, A_{t_1,N_2}\right)\right| \\ &\qquad \leq 2\sqrt{\phi_{t_1-q_{\tilde{u}}}} \sqrt{E\left[g^2\left(A_{q_1,N_2}, ..., A_{q_{\tilde{u}},N_2}\right) A_{q_{\tilde{u}},N_2}^2\right]} \sqrt{E\left[A_{t_1,N_2}^2\right]} \end{aligned}$$

[47] Bradley (1986) showed that the coefficients of the two processes fulfil the following inequality: $\rho_r \leq 2\sqrt{\phi_r}$.

and

$$|Cov(g(A_{q_1,N_2},...,A_{q_{\tilde{u}},N_2}),A_{t_1,N_2}A_{t_2,N_2}| \leq 2\phi_{t_1-q_{\tilde{u}}}E\left[|A_{t_1,N_2}A_{t_2,N_2}|\right]$$
$$\leq 2\phi_{t_1-q_{\tilde{u}}}E\left[A^2_{t_1,N_2}\right]$$

where the last inequality follows from the stationarity of the process. The two results can be further bounded

$$\begin{aligned}|Cov(g(A_{q_1,N_2},...,A_{q_{\tilde{u}},N_2})A_{q_{\tilde{u}},N_2},A_{t_1,N_2}| &\leq 2\sqrt{\phi_{t_1-q_{\tilde{u}}}}\sqrt{E\left[A^2_{q_{\tilde{u}},N_2}\right]}\sqrt{E\left[A^2_{t_1,N_2}\right]}\\ &\leq \sqrt{\phi_{t_1-q_{\tilde{u}}}}\left(E\left[A^2_{q_{\tilde{u}},N_2}\right]+E\left[A^2_{t_1,N_2}\right]\right)\\ &\leq \sqrt{\phi_{t_1-q_{\tilde{u}}}}\left(E\left[A^2_{q_{\tilde{u}},N_2}\right]+E\left[A^2_{t_1,N_2}\right]+N_2^{-1}\right)\end{aligned}$$

by property of the function $g()$ and inequality of arithmetic and geometric mean and

$$|Cov(g(A_{q_1,N_2},...,A_{q_{\tilde{u}},N_2}),A_{t_1,N_2}A_{t_2,N_2}| \leq \sqrt{\phi_{t_1-q_{\tilde{u}}}}\left(E\left[A^2_{t_1,N_2}\right]+E\left[A^2_{t_2,N_2}\right]+N_2^{-1}\right)$$

by stationarity and the fact that the mixing coefficients are smaller or equal to 1. Then the weak dependence conditions are fulfilled for $\tilde{\pi}_r = \sqrt{\phi_r}$.

■

**LEMMA 5.** Define $A_{i,N_2}$ as in Lemma 4, then $\hat{\theta}^{mcf}_{m,l}(x) - E\left[\hat{\theta}^{mcf}_{m,l}(x)\right] = \sum_{i=1}^{N_2} A_{i,N_2}$ and

$$\frac{\hat{\theta}^{mcf}_{m,l}(x) - E\left[\hat{\theta}^{mcf}_{m,l}(x)\right]}{\sqrt{Var\left(\hat{\theta}^{mcf}_{m,l}(x)\right)}} \to N(0,1).$$

**Proof:** The normality proof for triangular arrays satisfying the conditions in Lemma 4 is proven in Neumann (2013) and can be applied on estimation of potential outcomes $\mu_m(x)$ and $\mu_l(x)$. As those are estimated on the estimation data set, the difference of the two quantities will also follow a normal distribution. ■

**THEOREM 2.** Assume that there is a sample of size $N$ containing i.i.d. data $(X_i, Y_i, D_i) \in [0,1]^p \times \mathbb{R} \times \{0,1,...,M-1\}$ for a given value of $x$. Moreover, the covari-

ates are independently and uniformly distributed $X \sim U([0,1]^p)$. Let $T$ be an honest, regular, and symmetric random-split tree. Further assume that $E\left[Y^d \middle| X = x\right]$ and $E\left[(Y^d)^2 \middle| X = x\right]$ are Lipschitz continuous and $Var\left(Y^d \middle| X = x\right) > 0$. Then for $1/2 < \beta_1 < \beta_2 < \frac{p+2}{p}\frac{\log(1-\alpha)}{\log(\alpha)}\beta_1$,

$$\frac{\hat{\theta}^{mcf}_{m,l}(x) - \theta^0_{m,l}(x)}{\sqrt{Var\left(\hat{\theta}^{mcf}_{m,l}(x)\right)}} \xrightarrow{d} N(0,1).$$

**Proof:** Given the result in Lemma 5, it remains to show that

$$\frac{E\left[\hat{\theta}^{mcf}_{m,l}(x)\right] - \theta^0_{m,l}(x)}{\sqrt{Var\left(\hat{\theta}^{mcf}_{m,l}(x)\right)}} \to 0.$$

The final result will follow then from Slutsky's lemma. By Theorem 1, we have

$$\left|E\left[\hat{\theta}^{mcf}_{m,l}(x)\right] - \theta^0_{m,l}(x)\right| = O\left(s_1{}^{-\log(1-\alpha)/p\log(\alpha)}\right).$$

From Corollary 1 and Lemma 3, we get

$$Var\left(\hat{\theta}^{mcf}_{m,l}(x)\right) = \Omega\left(N^{\frac{\log(1-\alpha)}{\log(\alpha)}\beta_1 - \beta_2}\right).$$

It follows that

$$\frac{\left(E\left[\hat{\theta}^{mcf}_{m,l}(x)\right] - \theta^0_{m,l}(x)\right)}{\sqrt{Var\left(\hat{\theta}^{mcf}_{m,l}(x)\right)}} = O\left(N^{-\left(\frac{1}{p}+\frac{1}{2}\right)\frac{\log(1-\alpha)}{\log(\alpha)}\beta_1 + \frac{\beta_2}{2}}\right).$$

The ratio converges to zero when

$$\frac{p}{2+p}\beta_2 < \frac{\log(1-\alpha)}{\log(\alpha)}\beta_1.$$

■

*Appendix A.1.4.3. Proofs for Theorem 3*

**THEOREM 3.** Let all assumptions from Theorem 2 hold and define $\hat{\theta}^{mcf}_{m,l}$ as an average of all corresponding $\hat{\theta}^{mcf}_{m,l}(x)$. Then,

$$\frac{\hat{\theta}_{m,l}^{mcf} - \theta_{m,l}^{0}}{\sqrt{Var\left(\hat{\theta}_{m,l}^{mcf}\right)}} \rightarrow N(0,1).$$

**Proof:** Using the CLT for triangular arrays of weakly dependent random variables requires to check that all requirements in Lemma 4 hold for $A_{i,N_2} := W_i^{m,l} Y_i - E\left[W_i^{m,l} Y_i\right]$, where $W_i^{m,l} := W_i^{mcf}(m,l)$ represents the weights of the ATE. The proof uses the observation that the rates for the $ATE$ cannot be worse than for the $IATE$ and, therefore, the conditions in Lemma 4 will be satisfied for the weights of the ATE. The convergence rate may be affected due to the pointwise convergence at different points $x$.[48]

a) $E\left[A_{i,N_2}\right] = 0$ holds trivially.

b) $\sum_{i=1}^{N_2} E\left[A_{i,N_2}^2\right] = \sum_{i=1}^{N_2} Var\left(W_i^{m,l} Y_i\right)$. The upper bound on the individual variances is the upper bound of $E\left[\left(W_i^{m,l}\right)^2 Y_i^2\right]$:

$$\begin{aligned} E\left[\left(W_i^{m,l}\right)^2 Y_i^2\right] &= E\left[\left(\frac{1}{N_2}\sum_{j=1}^{N_2} W_i^{mcf}(D_i, X_i; X_j, m, l)\right)^2 Y_i^2\right] \\ &\leq \frac{1}{(N_2)^2} E\left[(N_2)^2 E\left[\left(W_i^{mcf}(D_i, X_i; X_j, m, l)\right)^2 \middle| X_i\right] E\left[Y_i^2 \,|X_i\right]\right] \\ &= O\left(N^{-1-\beta_2+\beta_1}\right). \end{aligned}$$

The second requirement is also satisfied as $\sum_{i=1}^{N_2} Var\left(W_i^{m,l} Y_i\right) = O\left(N^{-\beta_2+\beta_1}\right) < \infty$.

c)

$$\begin{aligned} Var(A_{1,N_2} + ... + A_{N_2,N_2}) &= Var\left(\sum_{i=1}^{N_2} W_i^{m,l} Y_i - \sum_{i=1}^{N_2} E\left[W_i^{m,l} Y_i\right]\right) \\ &= Var\left(\sum_{i=1}^{N_2} W_i^{m,l} Y_i\right) \\ &\leq \sum_{i=1}^{N_2} Var\left(W_i^{m,l} Y_i\right) + \sum_{i=1}^{N_2}\sum_{j\neq i} \left|Cov\left(W_i^{m,l} Y_i, W_j^{m,l} Y_j\right)\right| \end{aligned}$$

Using the results from b), the first element is bounded at rate $O\left(N^{-\beta_2+\beta_1}\right)$ and by a similar logic as in the tree case, the sum of all covariances must decay to

[48] Nevertheless, the simulation results show that the RMSE for the *mcf* estimator of the ATE decreases at a $\sqrt{N}-$rate for many DGPs, suggesting that the estimator exhibits $\sqrt{N}-$convergence.

zero as fast as the sum of the variances. These results yield that $Var(A_{1,N_2} + ... + A_{N_2,N_2}) \xrightarrow[N\to\infty]{} 0$.

d) Due to monotonicity and result in b):

$$\sum_{i=1}^{N_2} E\left[A_{i,N_2}^2 \underline{1}\left(|A_{i,N_2}| > \tilde{\varepsilon}\right)\right] \leq \sum_{i=1}^{N_2} E\left[A_{i,N_2}^2\right] \to 0 \text{ for all } \tilde{\varepsilon} > 0.$$

e) The proof follows the same logic as in Lemma 4.

With this, all requirements for the CLT for triangular arrays of weakly dependent random variables hold, so that we get

$$\frac{\hat{\theta}_{m,l}^{mcf} - \theta_{m,l}^{0}}{\sqrt{Var\left(\hat{\theta}_{m,l}^{mcf}\right)}} \to N(0,1).$$

Note that $E\left[\sum_{i=1}^{N_2} W_i^d Y_i\right] = E\left[Y^d\right]$ holds due to exchangeability and the fact that the weights sum to 1. Therefore, we applied the CLT directly to the quantity of interest. ■

**COROLLARY 3.** Let all assumptions from Theorem 2 hold and assume that a discrete $Z$ is among the splitting variables. Then for $Z = z$,

$$\frac{\hat{\theta}_{m,l}^{mcf}(z) - \theta_{m,l}^{0}(z)}{\sqrt{Var\left(\hat{\theta}_{m,l}^{mcf}(z)\right)}} \xrightarrow{d} N(0,1).$$

**Proof:** In a similar fashion as in Theorem 2, by setting $A_{i,N_2} := W_i^{z,m,l} Y_i - E\left[W_i^{z,m,l} Y_i\right]$, where $W_i^{z,m,l} := W_i^{mcf}(z,m,l)$, and checking that all requirements in Lemma 4 hold for $A_{i,N_2}$, the CLT for triangular arrays of weakly dependent random variables can be applied to show

$$\frac{\hat{\theta}_{m,l}^{mcf}(z) - E\left[\hat{\theta}_{m,l}^{mcf}(z)\right]}{\sqrt{Var\left(\hat{\theta}_{m,l}^{mcf}(z)\right)}} \xrightarrow{d} N(0,1).$$

Similarly as for $IATE$, when some $Y_i$ with a $z_i$ that does not belong to group $z$ has a non-zero weight, the final estimator is not an unbiased estimator of $\mu_d(z)$. Assuming that $Z$ is among the splitting variables guarantees that there exists some sample size $\tilde{N}$ such

that for all $N > \tilde{N}$ all trees split on $Z$. This implies that only $Y_i$ which belong to group $z$ have non-zero weights and $E\left[\hat{\theta}^{mcf}_{m,l}(z)\right] = \theta^0_{m,l}(z)$. ■

## Appendix A.2. Weights-based inference

There are several suggestions in the literature on how to conduct inference and how to compute standard errors of Random Forest based predictions (e.g., Wager, Hastie, and Efron, 2014; Wager and Athey, 2018, and the references therein). Although these methods can be used to conduct inference on the IATEs, it is yet unexplored how these methods could be readily generalized to take account of the aggregation steps needed for the GATE and ATE parameters.

Therefore, the *mcf* uses an alternative inference method useful for estimators that have a representation as weighted averages of the observed outcome. This perspective is attractive for Random Forest based estimators as they consist of trees that first stratify the data (when building a tree), and subsequently average over these strata (when building the forest). Thus, the *mcf* exploits the weights-based representation explicitly for inference (see also Abadie and Imbens, 2006, for a related approach).

Considering a weights-based estimator with random weights $W_i$ (that are normalized such that all weights add up to $N$) for $\hat{\theta}$ (i.e., IATEs, GATEs, or ATEs):

$$\hat{\theta} = \frac{1}{N}\sum_{i=1}^{N} W_i Y_i; \qquad Var(\hat{\theta}) = Var\left(\frac{1}{N}\sum_{i=1}^{N} W_i Y_i\right).$$

The variance of the estimator can be rewritten as:

$$Var\left(\frac{1}{N}\sum_{i=1}^{N} W_i Y_i\right) = E_W\left(\frac{1}{N^2}\sum_{i=1}^{N} {W_i}^2 \sigma^2_{Y|W}(W_i)\right) + Var_W\left(\frac{1}{N}\sum_{i=1}^{N} W_i \mu_{Y|W}(W_i)\right),$$

where $\mu_{Y|W}(W_i) = E(Y_i|W_i)$ and $\sigma^2_{Y|W}(W_i) = Var(Y_i|W_i)$. The derivation exploits the combination of two-sample honesty with an i.i.d sample. Remember that under two-sample honesty, observations can be either used for training or for estimation, without switching the roles.[49] Further recall that the forest weights of the *mcf* are computed as a

[49] In contrast, the 'one-sample honesty' in Wager and Athey (2018) is based on continuously switching the role of observations used for tree building and effect estimation in their Causal Forest. Under this splitting procedure, each weight may still depend on many observations, and the conditioning set cannot be reduced as in the case of two-sample honesty.

function of training data determining the final leaves and value $x_i$ picks the corresponding weight, leading to the following representation: $W_i = w(x_i, \Im_{tr})$. Therefore, under i.i.d. sampling, $Y_i$ and $W_j$ are independent. Thus, the conditioning set $W_1, ..., W_N$ can be reduced to $W_i$ for each conditional mean and conditional variance.

This leads to the following expression of the variance of the proposed estimators:[50]

$$Var\left(\frac{1}{N_2}\sum_{i=1}^{N_2} W_i Y_i\right) = E_W\left(\frac{1}{N_2^2}\sum_{i=1}^{N_2} {W_i}^2 \sigma^2_{Y|W}(W_i)\right) + Var_W\left(\frac{1}{N_2}\sum_{i=1}^{N_2} W_i \mu_{Y|W}(W_i)\right).$$

The above expression suggests using the following estimator:

$$\widehat{Var(\hat{\theta})} = \frac{1}{N_2^2}\sum_{i=1}^{N_2} w_i^2 \hat{\sigma}^2_{Y|W}(w_i) + \frac{1}{N_2(N_2-1)}\sum_{i=1}^{N_2}\left(w_i \hat{\mu}_{Y|\hat{W}}(w_i) - \frac{1}{N_2}\sum_{i=1}^{N_2} w_i \hat{\mu}_{Y|W}(w_i)\right)^2.$$

The conditional expectations and variances may be computed by standard non-parametric or ML methods, as this is a one-dimensional problem for which many well-established estimators exist. Bodory, Camponovo, Huber, and Lechner (2020) investigate $k$-nearest neighbour estimators to obtain estimates for these quantities. They found good results in a binary treatment setting for the ATET. The same method is used here.[51] As both, $\mu_{Y|W}$ and $\sigma^2_{Y|W}$, are bounded and the number of nearest neighbours in their estimation is chosen such that it grows slower than $N_2$, consistency of the conditional expectations is guaranteed, see, e.g., Devroye, Györfi, Krzyżak, and Lugosi (1994). Additionally, for larger $N$ the non-zero weights concentrate at a close neighbourhood of the given point $x$ as the leaves are shrinking towards a point. It is, however, beyond the scope of this paper to analyse rigorously the exact statistical conditions needed for this estimator to lead to valid inference.

## Appendix A.3. Comparison of *mcf* and *grf*

Comparing forest weights of *grf*, $w_i^{grf}(x)$, and the *mcf*, $w_i^{mcf}(d_i, x_i; x, m, l)$, both represent how important observation $i$ is for estimation of any quantity at point $x$. Meanwhile, the *mcf* weights weigh the observed outcome directly, while *grf* applies the forest weights

[50] Note that the weighting estimator uses only the honest data, therefore the weights here sum up to $N_2$.

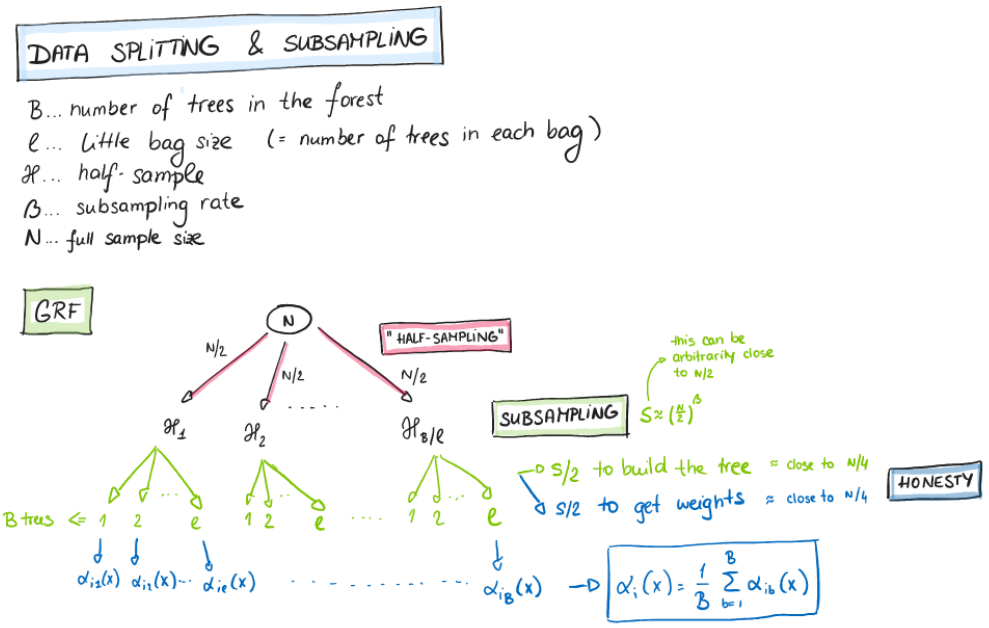
[51] They also found a considerable robustness on how exactly to compute the conditional means and variances. Note that since their results relate to aggregate treatment effect parameters, their generalisability to the level of IATE's is unclear.

in a locally weighted moment function. Thus, a weighted representation of the final *grf* estimator of the IATE at point $x$ weighs the observed outcome $y_i$ by weights that combine $w_i^{grf}(x)$ with treatment assignments in a non-linear fashion (see the result in Subsection 4.2).

A further difference between the *grf* and the *mcf* is related to the number of forest weights used in the IATE estimator, and how many trees contribute to the calculation of each weight. Both are determined by the corresponding honesty procedures. Subsection 4.2 summarizes that the *grf* forest weights represent by how much each observation is relevant for the $IATE(m, l; x)$. This means that there will be $N$ estimated forest weights because each half-sample is drawn from the full sample. For large enough $B$, each observation has a chance to end up in the estimation sample on which the weights are estimated. However, the half-sampling and further honesty split lead every forest weight $w_i^{grf}(x)$ to be based on approximately $B/4$ trees. On the other hand, *mcf* computes only $N/2$ of forest weights due to the primary half-split of data into training and estimation data sets. However, since the same $N/2$ observations are used in each tree to estimate the weights for $\beta_2 = 1$, each forest weight is averaged across all $B$ trees. For given $B$ and $N$, the *grf* has an advantage in smoothing over more observations and the *mcf* in the precision of the weights, potentially influencing finite sample properties of both methods.

Figures A.1 and A.2 capture graphically the main differences between the *mcf* and *grf* procedures regarding the one-sample and two-sample honesty and estimation of the weights.

Figure A.1: Diagram of *grf*




→ GRF will estimate N weights using cca B/4 values for each weight:

• ... represents that observation i is used for weight calculation

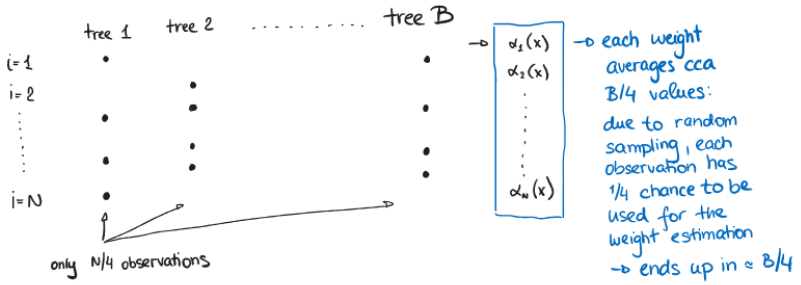
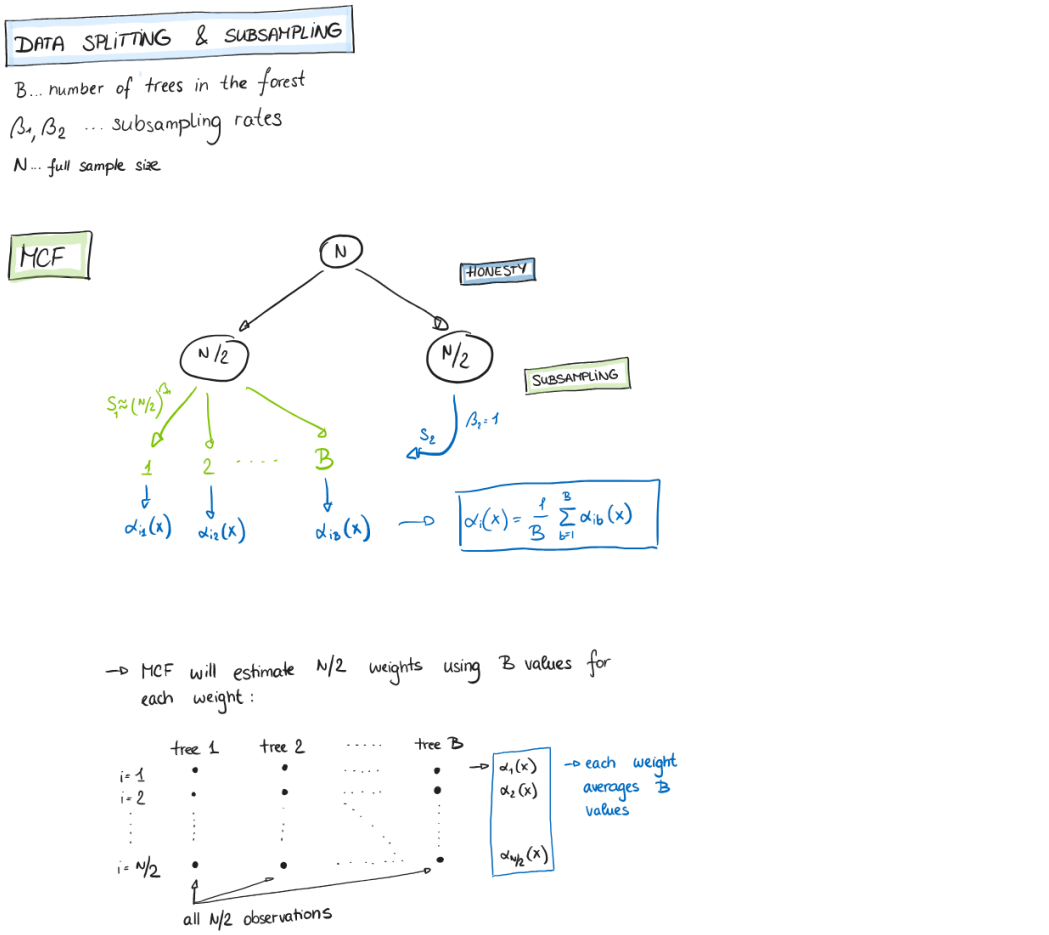


Figure A.2: Diagram of *mcf*




## Appendix A.4. Results for *grf* estimation of ATE

The ATE estimator for treatment pair $m$ and 0 implemented in *grf* is based on the following score function:

$$\psi^{grf}_{m,0}(O;\theta_{m,0},\eta^{grf}_{ATE(m,0)}(X)) = \Gamma^{grf}_{m,0}(O;\eta^{grf}_{ATE(m,0)}(X)) - \theta_{m,0},$$

$$\Gamma^{grf}_{m,0}(O;\eta^{grf}_{ATE(m,0)}(X)) = \theta_{m,0}(X) + \frac{\underline{1}(D=m)(Y-\mu(X)+\sum_{d>0} p_d(X)\theta_{d,0}(X)-\theta_{m,0}(X))}{p_m(X)} - \frac{\underline{1}(D=0)(Y-\mu(X)+\sum_{d>0} p_d(X)\theta_{d,0}(X))}{1-\sum_{d>0} p_d(X)},$$

where $\mu(x) = E[Y|X=x]$ and $\eta^{grf}_{ATE(m,0)}(X) = (\theta_{1,0}(X),\ldots,\theta_{M-1,0}(X),\mu(X),p_1(X),\ldots,p_{M-1}(X))$. Since

$$\begin{aligned}
\mu_d(x) &= E\left[Y\,|X=x\,, D=d\right], \\
\mu(x) &= \sum_{d=0}^{M-1} p_d(x)\mu_d(x), \\
\theta_{d,0}(x) &= \mu_d(x) - \mu_0(x), \\
\mu_0(x) &= \mu(x) - \sum_{d>0} p_d(x)\theta_{d,0}(x), \\
\mu_m(x) &= \mu(x) - \sum_{d>0} p_d(x)\theta_{d,0}(x) + \theta_{m,0}(x),
\end{aligned}$$

it directly follows that $\psi^{grf}_{m,0}(O;\theta_{m,0},\eta^{grf}_{ATE(m,0)}(X)) = \psi^{dml}_{m,0}(O;\theta_{m,0},\eta^{dml}_{m,0}(X))$ which is the DR efficient score of Robins and Rotnitzky (1995) for the ATE. The identification result $E\left[\psi^{grf}_{m,0}(O;\theta^0_{m,0},\eta^{grf,0}_{ATE(m,0)}(X))\right] = 0$ follows directly as well. Since the *dml* score is asymptotically efficient, the *grf* score can be expected to be asymptotically efficient when all nuisance parameters are consistently estimated.

The following proofs investigate consistent estimation of the ATE and its robustness to misspecification of nuisance parameters. Assume that $\hat{\eta}^{grf}_{ATE(m,0)}$ converges in probability to some values $\bar{\eta}_{ATE(m,0)}$. If $\bar{\eta}_{ATE(m,0)} = \eta^0_{ATE(m,0)}$, all nuisance parameters are consistently estimated. Since the ATE estimator is a sample average,

$$\begin{aligned}
\widehat{ATE^{grf}}(m,0) &= \frac{1}{N}\sum_{i=1}^{N} \Gamma^{grf}_{m,0}(o_i;\hat{\eta}^{grf}_{ATE(m,0)}(x_i)) \\
&= \frac{1}{N}\sum_{i=1}^{N} \widehat{\theta}_{m,0}(x_i) + \frac{\underline{1}(d_i=m)}{\widehat{p}_m(x_i)}\left(y_i - \widehat{\mu}(x_i) + \sum_{d>0}\widehat{p}_d(x_i)\widehat{\theta}_{d,0}(x_i) - \widehat{\theta}_{m,0}(x_i)\right) \\
&\quad - \frac{\underline{1}(d_i=0)}{1-\sum_{d>0}\widehat{p}_d(X)}\left(y_i - \widehat{\mu}(x_i) + \sum_{d>0}\widehat{p}_d(x_i)\widehat{\theta}_{d,0}(x_i)\right),
\end{aligned}$$

it converges in probability to

$$E\left[\bar{\theta}_{m,0}(X) + \frac{\underline{1}(D=m)}{\bar{p}_m(X)}\left(Y - \bar{\mu}(X) + \sum_{d>0}\bar{p}_d(X)\bar{\theta}_{d,0}(X) - \bar{\theta}_{m,0}(X)\right)\right.$$
$$\left. - \frac{\underline{1}(D=0)}{1 - \sum\limits_{d>0}\bar{p}_d(X)}\left(Y - \bar{\mu}(X) + \sum_{d>0}\bar{p}_d(X)\bar{\theta}_{d,0}(X)\right)\right].$$

If the propensity scores are consistently estimated, i.e., $\bar{p}_d(X) = p_d^0(X)$ for all $d > 0$, then by using the Law of Iterated Expectations and the identification assumptions, the probability limit of the ATE estimator becomes

$$E\left[\bar{\theta}_{m,0}(X) + \mu_m^0(X) - \bar{\mu}(X) + \sum_{d>0}p_d^0(X)\bar{\theta}_{d,0}(X) - \bar{\theta}_{m,0}(X) - \mu_0^0(X) + \bar{\mu}(X) - \sum_{d>0}p_d^0(X)\bar{\theta}_{d,0}(X)\right]$$
$$= E\left[\mu_m^0(X) - \mu_0^0(X)\right] = ATE_{m,0}^0.$$

This implies that the ATE estimator remains consistent when propensity scores are estimated consistently, even if the response function $\mu(x)$ and the treatment effects $\theta_{1,0}(X)$, $\ldots, \theta_{M-1,0}(X)$ are not consistently estimated. The presence of the propensity score in the estimation formula for the potential outcomes yields an inconsistent estimator of the ATE when the propensity scores are not consistently estimated, even if $\mu(x)$ and $\theta_{1,0}(X), \ldots, \theta_{M-1,0}(X)$ are consistently estimated. Therefore, consistent estimation of the ATE based on the *grf* score is not robust to misspecification in propensity scores.

Proving Neyman-orthogonality of $\psi_{m,0}^{grf}(O; \theta_{m,0}, \eta_{ATE(m,0)}^{grf}(X))$, it needs to be shown that at $r = 0$:

$$\frac{\partial E\left[\psi_{m,0}^{grf}\left(O; \theta_{m,0}^0, \eta_{ATE(m,0)}^{grf,0}(X) + r\left(\eta_{ATE(m,0)}^{grf}(X) - \eta_{ATE(m,0)}^{grf,0}(X)\right)\right)\right]}{\partial r} = 0.$$

For better readability, denote $\delta := \eta_{ATE(m,0)}^{grf}(X) - \eta_{ATE(m,0)}^{grf,0}(X)$ as the vector containing the deviations of all nuisance parameters from their true values. The individual elements, denoted as $\delta_{\theta(d,0)}$, $\delta_\mu$, and $\delta_{p(d)}$, capture the deviations of the treatment effects, response function $\mu(x)$, and propensity scores, respectively, for all treatments $d > 0$. The *grf* score evaluated at $(O; \theta_{m,0}^0, \eta_{ATE(m,0)}^{grf,0}(X) + r\delta)$ takes the form:

$$E\left[\psi_{m,0}^{grf}(O; \theta_{m,0}^0, \eta_{ATE(m,0)}^{grf,0}(X) + r\delta)\right] = E\left[\theta_{m,0}^0(X) + r\delta_{\theta(m,0)} + \right.$$
$$\left.\frac{\underline{1}(D=m)\left(Y - \mu^0(X) - r\delta_\mu + \sum\limits_{d>0}(p_d^0(X) + r\delta_{p(d)})(\theta_{d,0}^0(X) + r\delta_{(d,0)}) - \theta_{m,0}^0(X) - r\delta_{\theta(m,0)}\right)}{p_m^0(X) + r\delta_{p(m)}}\right.$$

$$- \frac{\underline{1}(D=0)\left(Y - \mu^0(X) - r\delta_\mu + \sum_{d>0}(p_d^0(X) + r\delta_{p(d)})(\theta_{d,0}^0(X) + r\delta_{\theta(d,0)})\right)}{1 - \sum_{d>0}\left(p_d^0(X) + r\delta_{p(d)}\right)} - \theta_{m,0}^0\Bigg].$$

Noting that

$$\begin{aligned}
\delta_{p(0)} &= -\sum_{d>0}\delta_{p(d)}, \\
\delta_{\mu(0)} &= \delta_\mu - \sum_{d>0} p_d^0(X)\delta_{\theta(d,0)} - \sum_{d>0}\delta_{p(d)}\theta_{d,0}^0(X) - r\sum_{d>0}\delta_{p(d)}\delta_{\theta(d,0)}, \\
\delta_{\mu(m)} &= \delta_\mu - \sum_{d>0} p_d^0(X)\delta_{\theta(d,0)} - \sum_{d>0}\delta_{p(d)}\theta_{d,0}^0(X) - r\sum_{d>0}\delta_{p(d)}\delta_{\theta(d,0)} + \delta_{\theta(m,0)},
\end{aligned}$$

differentiating under the expectation sign at $r = 0$ leads to:[52]

$$\begin{aligned}
\frac{\partial E\left[\psi_{m,0}^{grf}(O;\theta_{m,0}^0, \eta_{ATE(m,0)}^{grf,0}(X) + r\delta)\right]}{\partial r} &= E\Bigg[\delta_{\theta(m,0)} \\
&+ \frac{\underline{1}(D=m)\delta_{\mu(m)}p_m^0(X) - \underline{1}(D=m)(Y - \mu_m^0(X))\delta_{p(m)}}{\left(p_m^0(X)\right)^2} \\
&- \frac{\underline{1}(D=0)\delta_{\mu(0)}p_0^0(X) - \underline{1}(D=0)(Y - \mu_0^0(X))\delta_{p(0)}}{\left(p_0^0(X)\right)^2}\Bigg].
\end{aligned}$$

Using the law of iterated expectations and the identifying assumptions yields at $r = 0$:

$$\frac{\partial E\left[\psi_{m,0}^{grf}(O;\theta_{m,0}^0, \eta_{ATE(m,0)}^{grf,0}(X) + r\delta)\right]}{\partial r} = E\left[\delta_{\theta(m,0)} + \delta_{\mu(m)} - \delta_{\mu(0)}\right] = 0,$$

proving Neyman-orthogonality.

# Appendix B. Additional details of the Monte Carlo study

Appendix B.1 explains the general simulation protocol used, while Appendix B.2 explains the data generating processes in detail. Appendix B.3 gives the details of the implementation of the various versions of the different estimators used.

[52] The identifying assumptions and regularity conditions, such as Lipschitz continuity, guarantee that sufficient conditions for interchanging the expectation and derivation signs are fulfilled.

## Appendix B.1. Simulation protocol

Table B.1 shows the protocol employed in the Monte Carlo study.

Table B.1: Protocol of the Monte Carlo Study

| | |
|---|---|
| 1 | Specify the data generating process with respect to (i) sample size, (ii) strength of selectivity into treatments, (iii) type of covariates, (iv) influence of covariates on non-treatment potential outcomes (including the degree of sparsity), (v) size of treatment effect and its heterogeneity, (vi) treatment share, and (vii) number of treatments. |
| 2 | Draw training data of size $N$. |
| 3 | Draw prediction data of (same) size $N$. |
| 4 | Compute the true values of ATE and GATE on a large sample (1,000,000 observations) and get the true IATE from the DGP. |
| 5 | Train the different estimators on the training data. |
| 6 | Predict ATE, GATEs, and IATEs on the prediction data. |
| 7 | Repeat steps 2 to 6 $R$ times ($R = 1'000 \times (2'500/N)$). |
| 8 | Compute the performance measures. |
| 9 | Repeat steps 1 to 8 for different specifications. |

Note: The number of replications ($R$) declines such that the simulation noise remains approximately constant if the estimator is $\sqrt{N}-$convergent.

## Appendix B.2. Data generating processes

The description of the data generating process covers the covariates, the selection (treatment) process, the outcome process, and the effects and their heterogeneity.

### *Appendix B.2.1. Covariates*

The covariates are independent from each other and either normally or uniformly distributed,[53] both with mean zero and variance one.

$$
\begin{aligned}
&x_i^U \sim uniform(-\sqrt{12}/2, \sqrt{12}/2), && \dim(x_i^U) = p^U \\
&x_i^N \sim N(0,1), && \dim(x_i^N) = p^N \\
&p = p^U + p^N, \quad \forall i = 1, ..., N.
\end{aligned}
$$

We also consider a scenario where the first 5 variables of $X^N$ and $X^U$ are generated as dummy variables, with $X^D = 2 \cdot \underline{1}(X > 0) - 1$, also with mean zero.

[53] The reason for the uniform distribution is that the results for the *grf* and the *mcf* assume uniformly distributed covariates.

In the simulations, we consider values of the following triples ($p^N$, $p^U$, $p^D$): (10, 10, 0), (5, 5, 0), (25, 25, 0), (20, 0, 0), (0, 20, 0), and (5, 5, 10). Thus, we capture cases of 10 to 50 covariates that may be of different types. (10, 10, 0) is the base specification.

*Appendix B.2.2. Selection (treatment) process*

We consider cases of two and four treatments where all treatments have equal shares. As an extension, in the binary treatment case there are also treatment shares of 25%. The treatment indicators are obtained from the quantiles of the following index function:

$$\breve{d}_i = \frac{\breve{\lambda}\, x_i'\beta}{\sqrt{\sum_{j=1}^{p} \beta_j^2/1.25}} + u_i, \quad \beta = (1, 1-1/k, \ldots, 1/k, 0, \ldots, 0), \quad \dim(\beta) = p, \quad \breve{\lambda} \in \{0, 0.42, 1.25\}$$

$$u_i \sim N(0,1), \qquad \forall i = 1, \ldots, N.$$

The impact of the covariates declines with their order. The parameter $\breve{\lambda}$ determines the strength of the selection process, from a randomized experiment ($\breve{\lambda}$=0) to the case of strong selection ($\breve{\lambda}$=1.25 for uniform and mixed covariates; in case of normally distributed covariates grid for $\breve{\lambda}$ is 0, 0.45, and 1.5). In the intermediate setting, a machine learning estimation (using 1'000'000 observations) of the selection equation will obtain an out-of-sample $R^2$ of about 10%, while in the extreme case this value rises to about 42%. As it turns out that the strength of selectivity is an important parameter when considering the performance of estimators, the base scenario considers all three selectivity levels. The baseline scenario in Appendix C.2 is based on the medium selectivity when estimating effects at all granularity levels.

*Appendix B.2.3. IATEs*

The IATEs are deterministic functions of the covariates. All specifications of the IATEs are based on the following linear index, which is like the one used for the selection process:

$$\breve{\Delta}_i = \frac{\zeta\, x_i'\beta}{\sqrt{\sum_{j=1}^{p} \beta_j^2/1.25}}, \quad \beta = (1, 1-1/k, \ldots, 1/k, 0, \ldots, 0), \quad \dim(\beta) = p, \quad \zeta \in \{0, \zeta^{med}\}$$

$$\forall i = 1, \ldots, N.$$

The value of $\zeta^{med}$ captures the strength of heterogeneity. Its exact value depends on

the type of non-linearity and the $R^2$ of the potential non-treatment outcome process to ensure that the implications of the different effect strengths on the outcome remain stable.

The following variations of heterogeneity are considered for the IATE:

| IATE Formula | As a Function of $X\beta$ |
|---|---|
| $IATE_i = \breve{\Delta}_i + 1$ | Linear |
| $IATE_i = F^{\text{logistic}}(\breve{\Delta}_i) + 0.5$ | Nonlinear |
| $IATE_i = \frac{\breve{\Delta}_i^2 - 1.25}{\sqrt{3}} + 1$ | Quadratic |
| $IATE_i = f^{WA}(\tilde{x}_{1,i}) f^{WA}(\tilde{x}_{2,i}) - 2.8$ | Step function (like Wager and Athey, 2018) |

$F^{\text{logistic}}$ : c.d.f. of logistic distribution, $\quad f^{WA}(x) = 1 + \frac{1}{1+e^{-20(x-1/3)}}, \quad \forall i = 1, \ldots, N.$

They represent a linear function and three non-linear functions. Note that the last function is very similar to the DGP used in WA18. For the latter, the heterogeneity depends only on two variables (which are the ones that are most important for the selection and the outcome processes). $\tilde{x}_{1,i}$ and $\tilde{x}_{2,i}$ are transformed versions of $x_{1,i}$ and $x_{2,i}$, where the exact transformation used depends on the distribution of each variable. The baseline scenario in Appendix C.2 is based on the step function only.

The resulting ATE in these cases is always one, independent of the heterogeneity specification. In addition to these cases, we also consider a case of the IATEs all being equal to zero.

*Appendix B.2.4. Outcome processes*

The outcome process consists first of specifying a process for the non-treatment potential outcome. The potential outcome with treatment is obtained by adding the IATE plus noise to the non-treatment outcome.

The following non-linear outcome process is specified for the potential non-treatment outcome:

$$\breve{y}_i = \frac{x_i'\beta}{\sqrt{\sum_{j=1}^{p} \beta_j^2 / 1.25}}, \qquad \beta = (1, 1 - 1/k, \ldots, 1/k, 0, \ldots, 0), \qquad \dim(\beta) = p,$$

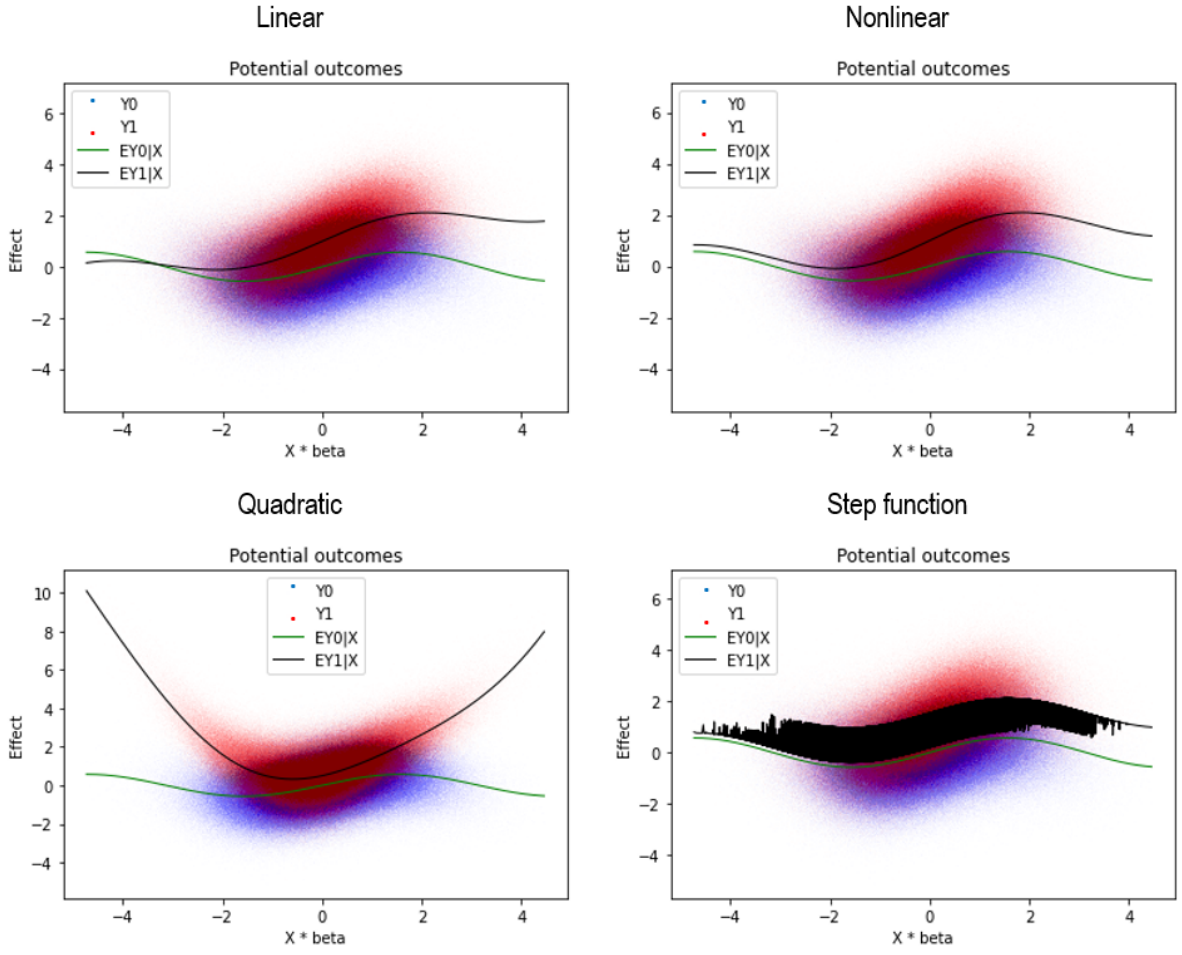
$$y_i^0 = \delta \sin(\breve{y}_i) + \varepsilon_i^0, \qquad \delta \in \{0, \delta^{med}, \delta^{strong}\}, \quad \varepsilon_i^0 \sim N(0,1)$$

$$y_i^1 = \delta \sin(\breve{y}_i) + IATE_i + \varepsilon_i^1, \qquad \varepsilon_i^1 \sim N(0,1) \qquad \forall i = 1, \ldots, N.$$

$\delta^{med}$ and $\delta^{strong}$ are chosen such that the $R^2$ in the outcome process of the potential non-treatment outcome is about 10% (base specification) and 45%.

Figure B.1 shows the relation of the potential outcomes and their expectations with the index $x_i'\beta$ for the case of $p^N = p^U = 10$, $p = 20$, $k^N = k^U = 5$, $k = 10$, and $\delta^{med}$ with respect to the different heterogeneities. Figure B.2 shows the same relationship for the stronger effect size $\delta^{strong}$.

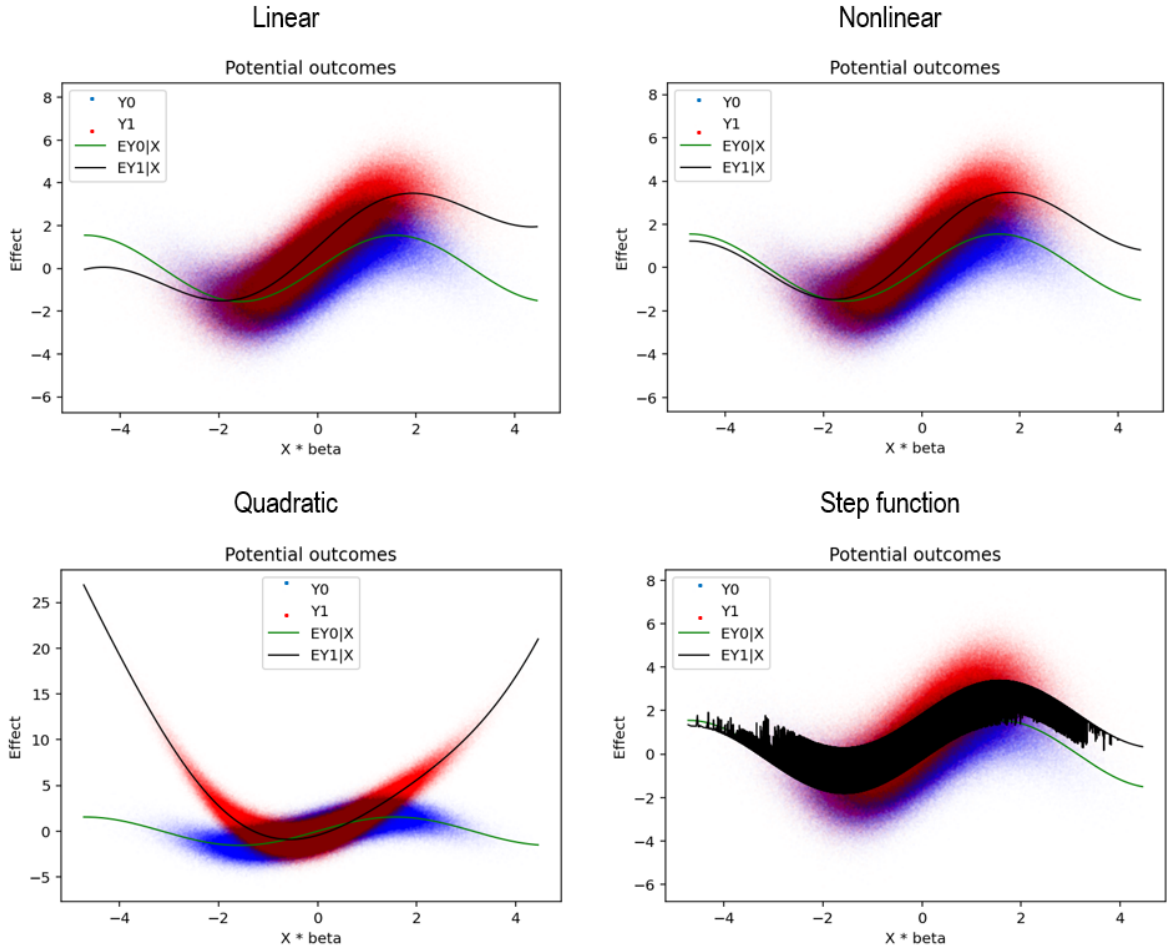
Figure B.1: Shape of potential outcomes for different shapes of IATEs ($\delta^{med}$)




Note: Figures are based on 1'000'000 observations.

Figures B.1 show that the differences between the linear and the non-linear case are rather small, at least compared to the quadratic IATEs and those based on the step function. In particular the latter, for which the IATEs depend only on the first two most important covariates, show behaviour that is not being well approximated by any

Figure B.2: Shape of potential outcomes for different shapes of IATEs ($\delta^{strong}$)




Note: Figures are based on 1'000'000 observations.

parsimonious parametric function of the linear index. Thus, we conjecture that this type of effect heterogeneity will be most difficult to estimate well.

Increasing the predictiveness of the covariates for the non-treatment potential outcome (Figure B.2), denoted by the blue dots, and its conditional expected values, denoted by the green line, leads to substantially more pronounced non-linearity with respect to the linear index. This is true for all specifications of the IATEs, but particularly so for the quadratic one.

*Appendix B.2.5. Overview of simulations and outcome tables*

Table B.2 gives an overview where to find the simulations results for the different data generating processes in Appendix C.

Table B.2: Overview of specifications and their locations

| Table | Hetero-geneity | Selec-tivity | $k$ | $p$ | $R^2$ $(y^0)$ | $X$ | $N$ | No of treat-ments | Treat-ment share | GATE groups |
|---|---|---|---|---|---|---|---|---|---|---|
| C.1 | **none** | none | 10 | 20 | 10% | U, N | stand | 2 | 50% | 5 |
| C.2 | **none** | middle | 10 | 20 | 10% | U, N | stand | 2 | 50% | 5 |
| C.3 | **none** | strong | 10 | 20 | 10% | U, N | stand | 2 | 50% | 5 |
| C.4 | **linear** | none | 10 | 20 | 10% | U, N | stand | 2 | 50% | 5 |
| C.5 | **linear** | middle | 10 | 20 | 10% | U, N | stand | 2 | 50% | 5 |
| C.6 | **linear** | strong | 10 | 20 | 10% | U, N | stand | 2 | 50% | 5 |
| C.7 | **nonlin** | none | 10 | 20 | 10% | U, N | stand | 2 | 50% | 5 |
| C.8 | **nonlin** | middle | 10 | 20 | 10% | U, N | stand | 2 | 50% | 5 |
| C.9 | **nonlin** | strong | 10 | 20 | 10% | U, N | stand | 2 | 50% | 5 |
| C.10 | **quadrat** | none | 10 | 20 | 10% | U, N | stand | 2 | 50% | 5 |
| C.11 | **quadrat** | middle | 10 | 20 | 10% | U, N | stand | 2 | 50% | 5 |
| C.12 | **quadrat** | strong | 10 | 20 | 10% | U, N | stand | 2 | 50% | 5 |
| C.13 | step | none | 10 | 20 | 10% | U, N | stand | 2 | 50% | 5 |
| C.14 | step | middle | 10 | 20 | 10% | U, N | stand | 2 | 50% | 5 |
| C.15 | step | strong | 10 | 20 | 10% | U, N | stand | 2 | 50% | 5 |
| C.16 | step | middle | **5** | **10** | 10% | U, N | stand | 2 | 50% | 5 |
| C.17 | step | middle | **25** | **50** | 10% | U, N | stand | 2 | 50% | 5 |
| C.18 | step | middle | **4** | 20 | 10% | U, N | stand | 2 | 50% | 5 |
| C.19 | step | middle | **16** | 20 | 10% | U, N | stand | 2 | 50% | 5 |
| C.20 | step | middle | 10 | 20 | **0%** | U, N | stand | 2 | 50% | 5 |
| C.21 | step | middle | 10 | 20 | **45%** | U, N | stand | 2 | 50% | 5 |
| C.22 | step | middle | 10 | 20 | 10% | **U** | stand | 2 | 50% | 5 |
| C.23 | step | middle | 10 | 20 | 10% | **N** | stand | 2 | 50% | 5 |
| C.24 | step | middle | 10 | 20 | 10% | **U,N,D** | stand | 2 | 50% | 5 |
| C.25 | step | middle | 10 | 20 | 10% | U, N | stand | **4** | equal | 5 |
| C.26 | step | middle | 10 | 20 | 10% | U, N | **40'000** | 2 | 50% | 5 |
| C.27 | step | middle | 10 | 20 | 10% | U, N | stand | 2 | **25%** | 5 |
| C.28 | step | none | 10 | 20 | 10% | U, N | stand | 2 | 50% | **5-40** |
| C.29 | step | middle | 10 | 20 | 10% | U, N | stand | 2 | 50% | **5-40** |
| C.30 | step | strong | 10 | 20 | 10% | U, N | stand | 2 | 50% | **5-40** |

**Note:** Deviations from the base benchmark specifications are in bold font. $N$: *stand* implies that $N = 2'500$ and $N = 10'000$ are in the same table. $X$: *U, N, D* denotes uniformly and normally distributed covariates, as well as dummy variables.

## Appendix B.3. Implementation of the estimators

To avoid additional computational costs in an already computationally very expensive simulation study, none of the estimators has been explicitly tuned. Either default values are used for the tuning parameters, or tuning parameters have been set to a fixed value beforehand. In an empirical application, the quality of estimation could be improved by

hyperparameter tuning. For more details on hyperparameter tuning for *dml*, see Bach, Schacht, Chernozhukov, Klaassen, and Spindler (2024). Forest estimators often minimize out-of-bag error to find optimal hyperparameters. For details on the tuning procedures, please refer to the documentation of *grf* and *mcf*.

*Appendix B.3.1. MCF*

The results of the *mcf* are computed with the Python versions of the package which are available on PyPI. Most of the simulations have been performed with version 0.3.3., but some of the simulations used also the newer and faster versions up to 0.4.3. Although some default values changed a bit between the versions, results were very similar. In the following, sample A denotes one half of the data set used to build the forest ("split-construction" data) and sample B denotes the other half of the sample used to estimate the effects ("leaf-populating" data).

Contrary to the default values, the minimum leaf size in each tree equals 5. The number of trees contained in any forest equals 1000. Trees are formed on random subsamples drawn without replacement (subsampling) with a sample size of 50% of the size of sample A. Regarding estimation of MCE and finding close neighbours, closeness is based on a Random Forest based prognostic score (see Hansen, 2008), $[\hat{\mu}_0(x), ...., \hat{\mu}_{M-1}(x)]$ weighted by the Mahalanobis distance, as forming the neighbours by simplified Mahalanobis matching suffers in large-dimensional problems.

To construct random-split trees, the number of variables tested for a split is determined through the following process: Initially, a random number is drawn from a Poisson distribution with parameter $(0.35 * p - 1)$. To this result, 1 is added, ensuring that at least 1 variable is considered for a split. If the final result exceeds $p$, all variables are tested for a split. Otherwise, the specified number of variables is randomly selected from the pool of all $p$ variables.

Two variants of this estimator are used, with and without local centring. Local centring is implemented in the *mcf* package by running a regression random forest (using the Python package scikit-learn) to predict $E[Y|X]$ ($X$ does not include the treatment). These (out-of-training-sample) predictions are obtained in a 5-fold cross-fitting scheme. The predictions are subtracted from the observed $Y$ to obtain a centred outcome, $Y^{cent}$. $Y^{cent}$ (instead of $Y$) is subsequently used as outcome in the *mcf* algorithm. The simulations

show that this centred version outperforms the uncentred version when there is non-random selection. Recentring is implemented in the following way:

1) Estimate the trees that define the Random Forest for $E[Y|X = x]$ in sample A.

2) Recentring of outcomes in sample A: Split sample A randomly into $K$ equally sized parts, $A$–1 to $A$–$K$. Use the outcomes in the union of the $K - 1$ folds $A$–1 to $A$–$(K - 1)$ to obtain the Random Forest predictions given the forest estimated in step 1). Use these predictions to predict $E[Y|X = x]$ in fold $A$–$K$. Do this for all-possible combinations of folds (cross-fitting as used in $k$-fold cross-fitting). Subtract the predictions from the actual values of $Y$.

3) Redo step 2 in sample B using the estimated forests of sample A.

Concerning the specifics of the local centring algorithm, there are a couple of points worth mentioning. First, to avoid overfitting, the outcomes of observation '$i$' are not used to predict itself. Therefore, the chosen implementation is based on cross-fitting.

Second, weights-based inference requires avoiding a dependence of the weights in sample B on outcomes of sample A. However, since recentring uses outcome variables independent of the treatment state, this could induce a correlation between the recentred outcomes in different treatment states. This finite sample correlation will be ignored here (as in Athey et al., 2019).

Third, the number of folds is a tuning parameter that influences the noise added to the recentred outcome by subtracting an estimated quantity. The simulation results indicate that the computationally most attractive choice of $K = 2$ may be too small in medium sized samples and that a somewhat larger number of folds may be needed to avoid much additional noise to the estimators.

The nonparametric regressions that enter the estimation of the standard errors are based on $k$-NN estimation with number of neighbours equal to *2 sqrt*$(N)$.

If inference is not of interest, a more efficient *mcf* estimator (denoted as IATE eff in Appendix C) can be obtained by switching the roles of the samples used for building the forest and populating the leaves with outcome values, and subsequently averaging the two estimates.

*Appendix B.3.2. GRF*

The reported results are based on the R package *grf* (version 2.3.1). The default forest parameters are 2000 trees (twice more than in *mcf)*, minimum node size is 5 and number of variables tested for a split is a random draw from a Poisson distribution with parameter *min(p, round_up(sqrt(p+20)))*.

The default implementation of the *grf* uses out-of-bag predictions of nuisance parameters to remove confounding bias in order to find the best splits on which the IATEs are estimated. As this implementation is not in line with the theory in Athey et al. (2019), a centred version of the algorithm, using $K$-fold cross-fitting to calculate the residualized outcome that is later fed as an input to the *grf*, is added to the simulation.

For later estimation of GATEs and ATEs, so-called DR scores need to be constructed. GATE and ATE can be estimated directly via functions *average_treatment_effect()* and *best_linear_predictor()*. Both functions need the estimated forest as an input to estimate the effects on the data used for building the forest and estimating the forest weights. Here, the simulation deviates from the protocol, as currently *grf* does not provide an option to estimate the DR scores on a new (prediction) data set. The package provides function *get_scores()* which computes the estimated component of a DR score on the data set that was used for the forest if needed. The current implementation of GATE estimation omits the *best_linear_predictor()* and directly regresses the scores from *get_scores()* on group dummies exploiting the homoscedastic error terms for estimation of standard errors.

*Appendix B.3.3. DML*

The key element of all *dml* estimators is the DR component of the *dml* score. For treatments $m$ and $l$, this DR component is defined as:

$$\Gamma^{dml}_{m,l}(X,Y,D;\eta^{dml}_{m,l}(X)) = \mu_m(X) - \mu_l(X) + \frac{\underline{1}(D=m)(Y-\mu_m(X))}{p_m(X)} - \frac{\underline{1}(D=l)(Y-\mu_l(X))}{p_l(X)},$$
$$\mu_d(x) = E\left[Y\,|D=d, X=x\right], \quad p_d(x) = P(D=d\,|X=x).$$

The nuisance parameters, $\eta^{dml}_{m,l}(X) = (\mu_m(X), \mu_l(X), p_m(X), p_l(X))$, are estimated with regression forests. As usual, the necessary cross-fitting is implemented via 5-fold cross-fitting. To obtain effects, the estimated DR components of the score for the desired treatment contrasts are formed. The ATE is obtained by averaging these differences.

GATEs and IATEs are obtained by regression-type approaches in which the estimated DR component of the score serves as dependent variable. The GATEs are computed as OLS-coefficients of a (saturated) regression model with the indicators for the GATEs groups acting as independent variables. IATEs are estimated by using $X$ as independent variables either in a regression random forest or in an OLS estimation. When OLS regressions are used, inference is based on heteroscedasticity-robust covariance matrix of the corresponding coefficients. No inference is obtained for the random forest based IATEs.

As the weights may lead to small sample issues, in particular when selection probabilities get close to zero, two versions of the *dml* estimator are considered. The first one is taking the weights as they are, while the second version normalises the sum of the weights to one ($N$)[54] and truncates 'too large' weights. Too large here means that a single weight is larger than 5% of the sum of all weights (if so, it is truncated at 5%). Since the normalised versions appears to outperform the non-normalized one in many simulations, the latter is presented in the main body of the text.

*Appendix B.3.4. OLS*

As a benchmark estimator, a two-sample (for binary treatments) OLS estimator is implemented in a standard way. However, since there are substantial non-linearities in the DGP, it is not surprising that there are many scenarios in which this estimator performs badly. Therefore, OLS results are not reported in the main part of the paper.

# Appendix C. Detailed results of the Monte Carlo study

## Appendix C.1. Base specifications

In this section, we vary the selectivity and the type of the IATE jointly and keep the other parameters fixed at their base values (i.e., $k = 10$, $p = 20$, $R^2(y^0) = 10\%$, $X^N$ and $X^U$, binary treatment, treatment share 50%, $N = 2'500$ and $N = 10'000$).

[54] Similarly, as in the normalized DR learner in Knaus (2022).

Table C.1: No IATE, no selectivity

| | | | Estimation of effects | | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | N = 2'500 | | | | | | |
| **ATE** | *1* | *mcf* | -0.003 | 0.048 | 0.060 | 0.060 | 0.01 | -0.05 | 0.000 | 96 | 79 |
| **GATE** | *5* | | -0.003 | 0.052 | 0.065 | 0.065 | 0.01 | 0.05 | 0.002 | 96 | 81 |
| **IATE** | *N* | | -0.003 | 0.079 | - | 0.100 | - | - | - | 97 | 85 |
| **IATE eff** | *N* | | -0.001 | 0.056 | - | 0.070 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.001 | 0.048 | 0.061 | 0.061 | 0.03 | 0.07 | -0.002 | 94 | 77 |
| **GATE** | *5* | *cent* | 0.001 | 0.051 | 0.065 | 0.065 | 0.03 | 0.13 | -0.001 | 95 | 78 |
| **IATE** | *N* | | 0.001 | 0.080 | - | 0.100 | - | - | - | 96 | 82 |
| **IATE eff** | *N* | | 0.001 | 0.056 | - | 0.070 | - | - | - | - | - |
| **ATE** | *1* | *grf* | -0.001 | 0.034 | 0.042 | 0.042 | -0.14 | -0.01 | -0.000 | 95 | 82 |
| **GATE** | *5* | | -0.001 | 0.075 | 0.094 | 0.094 | 0.04 | 0.07 | -0.001 | 95 | 80 |
| **IATE** | *N* | | -0.001 | 0.048 | - | 0.060 | - | - | - | 99 | 92 |
| **ATE** | *1* | *grf* | 0.003 | 0.032 | 0.041 | 0.041 | 0.12 | 0.01 | 0.001 | 95 | 82 |
| **GATE** | *5* | *cent* | 0.003 | 0.074 | 0.093 | 0.093 | 0.03 | -0.00 | -0.000 | 95 | 80 |
| **IATE** | *N* | | 0.003 | 0.047 | - | 0.059 | - | - | - | 99 | 92 |
| **ATE** | *1* | *dml* | -0.002 | 0.033 | 0.042 | 0.042 | 0.07 | -0.07 | 0.005 | 98 | 84 |
| **GATE** | *5* | | -0.002 | 0.074 | 0.093 | 0.093 | -0.02 | 0.04 | 0.001 | 95 | 81 |
| **IATE** | *N* | *ols* | -0.002 | 0.167 | - | 0.210 | - | - | - | 32 | 21 |
| **IATE** | *N* | *rf* | -0.002 | 0.221 | - | 0.281 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | -0.002 | 0.033 | 0.042 | 0.042 | 0.08 | -0.07 | 0.004 | 97 | 84 |
| **GATE** | *5* | *norm* | -0.002 | 0.073 | 0.093 | 0.093 | -0.02 | 0.06 | 0.000 | 95 | 80 |
| **IATE** | *N* | *ols* | -0.002 | 0.166 | - | 0.209 | - | - | - | 31 | 21 |
| **IATE** | *N* | *rf* | -0.002 | 0.218 | - | 0.278 | - | - | - | - | - |
| **ATE** | *1* | *ols* | -0.001 | 0.033 | 0.041 | 0.041 | 0.08 | -0.05 | -0.012 | 83 | 62 |
| **GATE** | *5* | | -0.001 | 0.072 | 0.091 | 0.091 | -0.03 | 0.10 | -0.026 | 84 | 64 |
| **IATE** | *N* | | -0.001 | 0.163 | - | 0.205 | - | - | - | 83 | 64 |

**Note:** Table to be continued.

Table C.1 - continued: No IATE, no selectivity

| | | | | Estimation of effects | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 10'000 | | | | | |
| **ATE** | *1* | *mcf* | 0.001 | 0.022 | 0.028 | 0.028 | 0.12 | 0.03 | 0.003 | 96 | 82 |
| **GATE** | *5* | | 0.001 | 0.025 | 0.031 | 0.031 | 0.09 | -0.30 | 0.002 | 97 | 84 |
| **IATE** | *N* | | 0.001 | 0.049 | - | 0.062 | - | - | - | 98 | 87 |
| **IATE eff** | *N* | | 0.000 | 0.035 | - | 0.043 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.003 | 0.022 | 0.027 | 0.028 | 0.14 | -0.12 | 0.002 | 97 | 80 |
| **GATE** | *5* | *cent* | 0.003 | 0.020 | 0.030 | 0.030 | 0.11 | -0.21 | 0.002 | 96 | 81 |
| **IATE** | *N* | | 0.003 | 0.048 | - | 0.060 | - | - | - | 97 | 84 |
| **IATE eff** | *N* | | 0.002 | 0.034 | - | 0.042 | - | - | - | - | - |
| **ATE** | *1* | *grf* | -0.001 | 0.015 | 0.020 | 0.020 | -0.11 | 0.57 | 0.001 | 96 | 81 |
| **GATE** | *5* | | -0.001 | 0.037 | 0.047 | 0.047 | -0.13 | 0.15 | -0.001 | 94 | 80 |
| **IATE** | *N* | | -0.001 | 0.030 | - | 0.038 | - | - | - | 99.8 | 98 |
| **ATE** | *1* | *grf* | 0.002 | 0.016 | 0.020 | 0.020 | 0.02 | 0.13 | 0.000 | 96 | 79 |
| **GATE** | *5* | *cent* | 0.002 | 0.036 | 0.045 | 0.045 | -0.10 | -0.12 | 0.000 | 95 | 79 |
| **IATE** | *N* | | 0.002 | 0.030 | - | 0.038 | - | - | - | 99.9 | 98 |
| **ATE** | *1* | *dml* | -0.001 | 0.015 | 0.019 | 0.019 | -0.11 | -0.09 | 0.004 | 98 | 87 |
| **GATE** | *5* | | -0.001 | 0.036 | 0.045 | 0.045 | 0.04 | 0.03 | 0.001 | 95 | 81 |
| **IATE** | *N* | *ols* | -0.001 | 0.080 | - | 0.100 | - | - | - | 16 | 11 |
| **IATE** | *N* | *rf* | -0.001 | 0.153 | - | 0.195 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | -0.001 | 0.015 | 0.019 | 0.019 | -0.11 | -0.08 | 0.004 | 98 | 87 |
| **GATE** | *5* | *norm* | -0.001 | 0.036 | 0.045 | 0.045 | 0.04 | 0.02 | 0.001 | 95 | 81 |
| **IATE** | *N* | *ols* | -0.001 | 0.079 | - | 0.100 | - | - | - | 16 | 11 |
| **IATE** | *N* | *rf* | -0.001 | 0.152 | - | 0.194 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.000 | 0.015 | 0.019 | 0.019 | -0.17 | -0.04 | -0.005 | 86 | 66 |
| **GATE** | *5* | | 0.000 | 0.036 | 0.045 | 0.045 | 0.01 | -0.03 | -0.013 | 85 | 63 |
| **IATE** | *N* | | 0.035 | 0.079 | - | 0.100 | - | - | - | 84 | 64 |

**Note:** For GATE and IATE the results are averaged over all effects. *CovP(95, 80)* denotes the (average) probability that the true value is part of the 95%/80% confidence interval. 1'000 / 500 replications used for 2'500 / 10'000 obs.

Table C.2: No IATE, medium selectivity

| | | | Estimation of effects | | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | N = 2'500 | | | | | | |
| **ATE** | *1* | *mcf* | 0.160 | 0.160 | 0.061 | 0.171 | 0.12 | 0.10 | 0.000 | 25 | 9 |
| **GATE** | *5* | | 0.160 | 0.161 | 0.073 | 0.176 | 0.09 | 0.10 | 0.003 | 42 | 19 |
| **IATE** | *N* | | 0.160 | 0.167 | - | 0.194 | - | - | - | 77 | 48 |
| **IATE eff** | *N* | | 0.163 | 0.164 | - | 0.164 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.022 | 0.051 | 0.060 | 0.064 | 0.14 | 0.31 | -0.001 | 94 | 76 |
| **GATE** | *5* | *cent* | 0.022 | 0.060 | 0.073 | 0.076 | 0.13 | 0.27 | 0.001 | 95 | 79 |
| **IATE** | *N* | | 0.022 | 0.086 | - | 0.109 | - | - | - | 96 | 82 |
| **IATE eff** | *N* | | 0.025 | 0.063 | - | 0.079 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.091 | 0.092 | 0.043 | 0.101 | -0.03 | -0.23 | -0.002 | 41 | 19 |
| **GATE** | *5* | | 0.091 | 0.108 | 0.093 | 0.131 | 0.02 | -0.00 | -0.001 | 83 | 60 |
| **IATE** | *N* | | 0.090 | 0.094 | - | 0.109 | - | - | - | 88 | 66 |
| **ATE** | *1* | *grf* | 0.027 | 0.040 | 0.041 | 0.049 | 0.02 | -0.25 | -0.000 | 90 | 71 |
| **GATE** | *5* | *cent* | 0.027 | 0.076 | 0.091 | 0.095 | 0.01 | 0.11 | 0.001 | 94 | 78 |
| **IATE** | *N* | | 0.028 | 0.053 | - | 0.066 | - | - | - | 98 | 90 |
| **ATE** | *1* | *dml* | 0.016 | 0.038 | 0.044 | 0.048 | 0.00 | -0.23 | 0.004 | 96 | 81 |
| **GATE** | *5* | | 0.016 | 0.079 | 0.097 | 0.099 | -0.05 | -0.05 | 0.001 | 95 | 80 |
| **IATE** | *N* | *ols* | 0.017 | 0.176 | - | 0.221 | - | - | - | 33 | 22 |
| **IATE** | *N* | *rf* | 0.009 | 0.232 | - | 0.297 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.017 | 0.038 | 0.044 | 0.047 | -0.01 | -0.26 | 0.004 | 96 | 79 |
| **GATE** | *5* | *norm* | 0.017 | 0.079 | 0.097 | 0.099 | -0.05 | -0.05 | 0.001 | 95 | 79 |
| **IATE** | *N* | *ols* | 0.017 | 0.175 | - | 0.222 | - | - | - | 32 | 22 |
| **IATE** | *N* | *rf* | 0.010 | 0.231 | - | 0.299 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.007 | 0.035 | 0.043 | 0.044 | -0.01 | -0.19 | -0.012 | 84 | 62 |
| **GATE** | *5* | | 0.007 | 0.077 | 0.093 | 0.096 | -0.06 | -0.01 | -0.026 | 83 | 61 |
| **IATE** | *N* | | 0.007 | 0.165 | - | 0.208 | - | - | - | 83 | 63 |

**Note:** Table to be continued.

Table C.2 - continued: No IATE, medium selectivity

| | | | | **Estimation of effects** | | | | | | **Estimation of std. errors** | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 10'000 | | | | | |
| **ATE** | *1* | *mcf* | 0.138 | 0.138 | 0.028 | 0.141 | 0.15 | -0.02 | 0.002 | 0 | 0 |
| **GATE** | *5* | | 0.138 | 0.138 | 0.037 | 0.143 | 0.12 | -0.14 | 0.001 | 11 | 4 |
| **IATE** | *N* | | 0.138 | 0.140 | - | 0.156 | - | - | - | 62 | 33 |
| **IATE eff** | *N* | | 0.138 | 0.138 | - | 0.148 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.001 | 0.021 | 0.027 | 0.027 | 0.27 | 0.03 | 0.003 | 97 | 84 |
| **GATE** | *5* | *cent* | 0.001 | 0.031 | 0.038 | 0.039 | 0.21 | 0.04 | 0.003 | 97 | 84 |
| **IATE** | *N* | | 0.001 | 0.056 | - | 0.071 | - | - | - | 97 | 84 |
| **IATE eff** | *N* | | 0.001 | 0.042 | - | 0.053 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.059 | 0.059 | 0.020 | 0.062 | -0.23 | -0.20 | 0.000 | 16 | 7 |
| **GATE** | *5* | | 0.059 | 0.064 | 0.046 | 0.075 | -0.07 | -0.14 | -0.000 | 74 | 48 |
| **IATE** | *N* | | 0.058 | 0.061 | - | 0.070 | - | - | - | 98 | 85 |
| **ATE** | *1* | *grf* | 0.011 | 0.019 | 0.021 | 0.024 | -0.13 | 0.21 | -0.001 | 90 | 74 |
| **GATE** | *5* | *cent* | 0.011 | 0.038 | 0.046 | 0.048 | -0.05 | -0.10 | -0.001 | 94 | 76 |
| **IATE** | *N* | | 0.013 | 0.033 | - | 0.041 | - | - | - | 99.8 | 97 |
| **ATE** | *1* | *dml* | 0.008 | 0.014 | 0.023 | 0.023 | -0.02 | 0.31 | 0.003 | 96 | 82 |
| **GATE** | *5* | | 0.009 | 0.040 | 0.049 | 0.050 | 0.12 | 0.05 | 0.000 | 94 | 80 |
| **IATE** | *N* | *ols* | 0.009 | 0.086 | - | 0.108 | - | - | - | 17 | 11 |
| **IATE** | *N* | *rf* | 0.004 | 0.161 | - | 0.207 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.009 | 0.018 | 0.021 | 0.023 | -0.03 | 0.32 | 0.003 | 96 | 82 |
| **GATE** | *5* | *norm* | 0.009 | 0.040 | 0.049 | 0.050 | 0.12 | 0.04 | 0.000 | 94 | 80 |
| **IATE** | *N* | *ols* | 0.009 | 0.086 | - | 0.108 | - | - | - | 17 | 11 |
| **IATE** | *N* | *rf* | 0.004 | 0.160 | - | 0.207 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.007 | 0.017 | 0.020 | 0.022 | 0.02 | 0.19 | -0.005 | 84 | 64 |
| **GATE** | *5* | | 0.007 | 0.038 | 0.046 | 0.053 | 0.09 | 0.13 | -0.013 | 78 | 55 |
| **IATE** | *N* | | 0.007 | 0.083 | - | 0.104 | - | - | - | 83 | 63 |

**Note:** For GATE and IATE the results are averaged over all effects. *CovP(95, 80)* denotes the (average) probability that the true value is part of the 95%/80% confidence interval. 1'000 / 500 replications used for 2'500 / 10'000 obs.

Table C.3: No IATE, strong selectivity

| | | | Estimation of effects | | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 2'500 | | | | | |
| **ATE** | *1* | *mcf* | 0.375 | 0.375 | 0.063 | 0.380 | 0.03 | -0.03 | 0.000 | 0 | 0 |
| **GATE** | *5* | | 0.375 | 0.375 | 0.082 | 0.384 | 0.02 | 0.17 | 0.010 | 3 | 1 |
| **IATE** | *N* | | 0.375 | 0.375 | - | 0.394 | - | - | - | 21 | 6 |
| **IATE eff** | *N* | | 0.378 | 0.378 | - | 0.388 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.011 | 0.045 | 0.056 | 0.057 | 0.03 | -0.13 | 0.008 | 97 | 85 |
| **GATE** | *5* | *cent* | 0.011 | 0.060 | 0.076 | 0.077 | 0.04 | 0.11 | 0.013 | 98 | 86 |
| **IATE** | *N* | | 0.011 | 0.091 | - | 0.115 | - | - | - | 97 | 86 |
| **IATE eff** | *N* | | 0.013 | 0.066 | - | 0.084 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.264 | 0.264 | 0.049 | 0.268 | -0.07 | -0.03 | -0.010 | 0 | 0 |
| **GATE** | *5* | | 0.264 | 0.264 | 0.093 | 0.280 | -0.06 | -0.00 | -0.005 | 17 | 6 |
| **IATE** | *N* | | 0.256 | 0.256 | - | 0.265 | - | - | - | 32 | 8 |
| **ATE** | *1* | *grf* | 0.068 | 0.070 | 0.045 | 0.081 | 0.03 | -0.06 | -0.006 | 58 | 35 |
| **GATE** | *5* | *cent* | 0.068 | 0.090 | 0.089 | 0.112 | 0.02 | 0.14 | -0.001 | 87 | 67 |
| **IATE** | *N* | | 0.076 | 0.084 | - | 0.101 | - | - | - | 94 | 78 |
| **ATE** | *1* | *dml* | 0.136 | 0.137 | 0.058 | 0.148 | -0.32 | 1.36 | -0.003 | 32 | 14 |
| **GATE** | *5* | | 0.136 | 0.155 | 0.124 | 0.185 | -0.67 | 5.35 | -0.014 | 69 | 48 |
| **IATE** | *N* | *ols* | 0.136 | 0.235 | - | 0.297 | - | - | - | 30 | 20 |
| **IATE** | *N* | *rf* | 0.114 | 0.290 | - | 0.480 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.129 | 0.129 | 0.055 | 0.140 | -0.03 | 0.01 | 0.002 | 38 | 16 |
| **GATE** | *5* | *norm* | 0.129 | 0.148 | 0.121 | 0.177 | -0.13 | 0.02 | -0.005 | 76 | 54 |
| **IATE** | *N* | *ols* | 0.129 | 0.177 | - | 0.292 | - | - | - | 33 | 22 |
| **IATE** | *N* | *rf* | 0.129 | 0.296 | - | 0.401 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.109 | 0.110 | 0.053 | 0.121 | 0.00 | 0.04 | -0.014 | 25 | 12 |
| **GATE** | *5* | | 0.109 | 0.117 | 0.102 | 0.152 | -0.03 | 0.01 | -0.029 | 60 | 41 |
| **IATE** | *N* | | 0.109 | 0.192 | - | 0.240 | - | - | - | 77 | 56 |

**Note:** Table to be continued.

Table C.3 - continued: No IATE, strong selectivity

| | | | | Estimation of effects | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | N = 10'000 | | | | | | |
| **ATE** | *1* | *mcf* | 0.334 | 0.334 | 0.030 | 0.336 | -0.03 | 0.60 | 0.001 | 0 | 0 |
| **GATE** | *5* | | 0.334 | 0.334 | 0.045 | 0.337 | -0.07 | 0.08 | 0.006 | 0 | 0 |
| **IATE** | *N* | | 0.334 | 0.334 | - | 0.345 | - | - | - | 5 | 1 |
| **IATE eff** | *N* | | 0.335 | 0.335 | - | 0.340 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | -0.041 | 0.043 | 0.027 | 0.049 | 0.17 | 0.38 | 0.006 | 80 | 48 |
| **GATE** | *5* | *cent* | -0.041 | 0.048 | 0.041 | 0.058 | -0.07 | 0.06 | 0.008 | 90 | 70 |
| **IATE** | *N* | | -0.041 | 0.073 | - | 0.091 | - | - | - | 94 | 69 |
| **IATE eff** | *N* | | -0.041 | 0.060 | - | 0.074 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.192 | 0.192 | 0.024 | 0.194 | -0.37 | 0.18 | -0.004 | 0 | 0 |
| **GATE** | *5* | | 0.192 | 0.192 | 0.047 | 0.198 | -0.00 | -0.01 | -0.002 | 2 | 0 |
| **IATE** | *N* | | 0.187 | 0.187 | - | 0.193 | - | - | - | 53 | 18 |
| **ATE** | *1* | *grf* | 0.027 | 0.031 | 0.025 | 0.037 | -0.08 | -0.07 | -0.005 | 68 | 45 |
| **GATE** | *5* | *cent* | 0.027 | 0.044 | 0.047 | 0.055 | -0.09 | 0.06 | -0.003 | 90 | 69 |
| **IATE** | *N* | | 0.039 | 0.049 | - | 0.061 | - | - | - | 99 | 94 |
| **ATE** | *1* | *dml* | 0.079 | 0.079 | 0.034 | 0.086 | 0.034 | 0.54 | 3.01 | 27 | 11 |
| **GATE** | *5* | | 0.079 | 0.091 | 0.076 | 0.111 | 0.076 | 0.20 | 6.61 | 66 | 46 |
| **IATE** | *N* | *ols* | 0.079 | 0.133 | - | 0.171 | - | - | - | 16 | 11 |
| **IATE** | *N* | *rf* | 0.061 | 0.208 | - | 0.411 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.070 | 0.070 | 0.031 | 0.077 | 0.000 | 0.15 | 0.33 | 37 | 18 |
| **GATE** | *5* | *norm* | 0.070 | 0.084 | 0.071 | 0.101 | -0.007 | -0.07 | 0.13 | 75 | 54 |
| **IATE** | *N* | *ols* | 0.070 | 0.129 | - | 0.162 | - | - | - | 18 | 12 |
| **IATE** | *N* | *rf* | 0.051 | 0.213 | - | 0.336 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.111 | 0.111 | 0.026 | 0.111 | 0.10 | -0.32 | -0.007 | 0 | 0 |
| **GATE** | *5* | | 0.111 | 0.113 | 0.051 | 0.126 | 0.04 | -0.10 | -0.015 | 26 | 14 |
| **IATE** | *N* | | 0.111 | 0.129 | - | 0.156 | - | - | - | 61 | 41 |

**Note:** For GATE and IATE the results are averaged over all effects. *CovP(95, 80)* denotes the (average) probability that the true value is part of the 95%/80% confidence interval. 1'000 / 500 replications used for 2'500 / 10'000 obs.

Table C.4: Linear IATE, no selectivity

| | | | Estimation of effects | | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | N = 2'500 | | | | | | |
| **ATE** | *1* | *mcf* | -0.004 | 0.049 | 0.062 | 0.062 | -0.02 | -0.06 | -0.021 | 96 | 80 |
| **GATE** | *5* | | -0.004 | 0.091 | 0.078 | 0.114 | 0.00 | 0.19 | 0.001 | 83 | 61 |
| **IATE** | *N* | | -0.004 | 0.200 | - | 0.252 | - | - | - | 69 | 49 |
| **IATE eff** | *N* | | -0.004 | 0.193 | - | 0.242 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.002 | 0.049 | 0.062 | 0.062 | 0.01 | 0.08 | -0.001 | 95 | 78 |
| **GATE** | *5* | *cent* | 0.001 | 0.095 | 0.080 | 0.118 | 0.01 | 0.22 | -0.008 | 75 | 55 |
| **IATE** | *N* | | 0.002 | 0.208 | - | 0.261 | - | - | - | 62 | 43 |
| **IATE eff** | *N* | | 0.004 | 0.201 | - | 0.252 | - | - | - | - | - |
| **ATE** | *1* | *grf* | -0.001 | 0.034 | 0.043 | 0.043 | -0.15 | 0.01 | 0.000 | 95 | 82 |
| **GATE** | *5* | | -0.001 | 0.077 | 0.097 | 0.097 | 0.03 | 0.07 | -0.001 | 95 | 80 |
| **IATE** | *N* | | -0.001 | 0.214 | - | 0.268 | - | - | - | 57 | 40 |
| **ATE** | *1* | *grf* | 0.004 | 0.033 | 0.042 | 0.042 | 0.09 | 0.02 | 0.001 | 94 | 81 |
| **GATE** | *5* | *cent* | 0.004 | 0.075 | 0.094 | 0.094 | 0.04 | 0.04 | 0.000 | 95 | 80 |
| **IATE** | *N* | | 0.004 | 0.212 | - | 0.266 | - | - | - | 56 | 40 |
| **ATE** | *1* | *dml* | -0.002 | 0.034 | 0.042 | 0.042 | 0.06 | -0.07 | 0.005 | 98 | 84 |
| **GATE** | *5* | | -0.002 | 0.075 | 0.095 | 0.095 | -0.02 | 0.04 | 0.001 | 95 | 81 |
| **IATE** | *N* | *ols* | -0.002 | 0.168 | - | 0.212 | - | - | - | 34 | 22 |
| **IATE** | *N* | *rf* | -0.002 | 0.245 | - | 0.310 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | -0.002 | 0.034 | 0.042 | 0.042 | 0.06 | -0.07 | 0.005 | 97 | 83 |
| **GATE** | *5* | *norm* | -0.002 | 0.075 | 0.095 | 0.095 | -0.02 | 0.06 | 0.000 | 95 | 80 |
| **IATE** | *N* | *ols* | -0.002 | 0.167 | - | 0.210 | - | - | - | 32 | 21 |
| **IATE** | *N* | *rf* | -0.002 | 0.244 | - | 0.308 | - | - | - | - | - |
| **ATE** | *1* | *ols* | -0.001 | 0.033 | 0.042 | 0.042 | 0.01 | -0.08 | -0.013 | 81 | 62 |
| **GATE** | *5* | | -0.002 | 0.073 | 0.092 | 0.092 | -0.04 | 0.09 | -0.027 | 84 | 64 |
| **IATE** | *N* | | -0.001 | 0.163 | - | 0.205 | - | - | - | 83 | 63 |

**Note:** Table to be continued.

Table C.4 - continued: Linear IATE, no selectivity

| | | | | Estimation of effects | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 10'000 | | | | | |
| **ATE** | *1* | *mcf* | 0.001 | 0.023 | 0.029 | 0.029 | 0.108 | -0.20 | 0.004 | 98 | 82 |
| **GATE** | *5* | | 0.000 | 0.066 | 0.042 | 0.080 | -0.04 | -0.27 | 0.000 | 67 | 46 |
| **IATE** | *N* | | 0.001 | 0.162 | - | 0.205 | - | - | - | 61 | 43 |
| **IATE eff** | *N* | | 0.000 | 0.158 | - | 0.201 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.003 | 0.022 | 0.028 | 0.028 | 0.13 | -0.15 | 0.002 | 97 | 79 |
| **GATE** | *5* | *cent* | 0.003 | 0.062 | 0.047 | 0.077 | 0.01 | -0.25 | -0.008 | 65 | 46 |
| **IATE** | *N* | | 0.003 | 0.171 | - | 0.216 | - | - | - | 52 | 36 |
| **IATE eff** | *N* | | 0.002 | 0.168 | - | 0.212 | - | - | - | - | - |
| **ATE** | *1* | *grf* | -0.000 | 0.015 | 0.020 | 0.020 | -0.11 | 0.51 | 0.001 | 96 | 82 |
| **GATE** | *5* | | -0.001 | 0.037 | 0.047 | 0.047 | -0.15 | 0.18 | -0.000 | 94 | 81 |
| **IATE** | *N* | | -0.001 | 0.175 | - | 0.221 | - | - | - | 61 | 44 |
| **ATE** | *1* | *grf* | 0.003 | 0.016 | 0.020 | 0.021 | 0.02 | 0.15 | -0.000 | 96 | 82 |
| **GATE** | *5* | *cent* | 0.003 | 0.036 | 0.045 | 0.045 | -0.07 | -0.10 | 0.001 | 95 | 80 |
| **IATE** | *N* | | 0.003 | 0.174 | - | 0.219 | - | - | - | 61 | 44 |
| **ATE** | *1* | *dml* | -0.001 | 0.016 | 0.020 | 0.020 | -0.10 | 0.00 | 0.004 | 98 | 86 |
| **GATE** | *5* | | -0.001 | 0.037 | 0.046 | 0.046 | 0.02 | 0.01 | 0.001 | 95 | 82 |
| **IATE** | *N* | *ols* | -0.001 | 0.080 | - | 0.101 | - | - | - | 16 | 11 |
| **IATE** | *N* | *rf* | -0.001 | 0.187 | - | 0.236 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | -0.001 | 0.016 | 0.020 | 0.020 | -0.11 | -0.01 | 0.003 | 98 | 86 |
| **GATE** | *5* | *norm* | -0.001 | 0.037 | 0.046 | 0.046 | 0.02 | 0.01 | 0.001 | 95 | 82 |
| **IATE** | *N* | *ols* | -0.001 | 0.080 | - | 0.101 | - | - | - | 17 | 11 |
| **IATE** | *N* | *rf* | -0.001 | 0.186 | - | 0.235 | - | - | - | - | - |
| **ATE** | *1* | *ols* | -0.000 | 0.016 | 0.020 | 0.020 | -0.16 | 0.01 | -0.005 | 84 | 65 |
| **GATE** | *5* | | -0.001 | 0.037 | 0.045 | 0.046 | -0.01 | 0.01 | -0.013 | 84 | 64 |
| **IATE** | *N* | | 0.000 | 0.079 | - | 0.100 | - | - | - | 84 | 64 |

**Note:** For GATE and IATE the results are averaged over all effects. *CovP(95, 80)* denotes the (average) probability that the true value is part of the 95%/80% confidence interval. 1'000 / 500 replications used for 2'500 / 10'000 obs.

Table C.5: Linear IATE, medium selectivity

| | | | | Estimation of effects | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 2'500 | | | | | |
| **ATE** | *1* | *mcf* | 0.225 | 0.225 | 0.063 | 0.234 | 0.11 | 0.11 | 0.001 | 6 | 1 |
| **GATE** | *5* | | 0.225 | 0.227 | 0.081 | 0.252 | 0.05 | 0.09 | 0.001 | 29 | 16 |
| **IATE** | *N* | | 0.225 | 0.280 | - | 0.340 | - | - | - | 53 | 36 |
| **IATE eff** | *N* | | 0.228 | 0.276 | - | 0.333 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.056 | 0.067 | 0.061 | 0.082 | 0.09 | 0.28 | -0.001 | 84 | 63 |
| **GATE** | *5* | *cent* | 0.055 | 0.102 | 0.080 | 0.126 | 0.06 | 0.17 | -0.003 | 76 | 54 |
| **IATE** | *N* | | 0.056 | 0.215 | - | 0.269 | - | - | - | 62 | 43 |
| **IATE eff** | *N* | | 0.059 | 0.208 | - | 0.260 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.128 | 0.128 | 0.044 | 0.135 | -0.04 | -0.23 | -0.002 | 16 | 5 |
| **GATE** | *5* | | 0.128 | 0.137 | 0.096 | 0.161 | 0.01 | -0.02 | -0.002 | 72 | 47 |
| **IATE** | *N* | | 0.122 | 0.248 | - | 0.309 | - | - | - | 49 | 33 |
| **ATE** | *1* | *grf* | 0.050 | 0.055 | 0.042 | 0.065 | 0.02 | -0.18 | -0.000 | 77 | 53 |
| **GATE** | *5* | *cent* | 0.050 | 0.086 | 0.093 | 0.108 | 0.01 | 0.08 | 0.000 | 91 | 72 |
| **IATE** | *N* | | 0.049 | 0.230 | - | 0.287 | - | - | - | 51 | 36 |
| **ATE** | *1* | *dml* | 0.030 | 0.044 | 0.045 | 0.054 | -0.02 | -0.23 | 0.004 | 93 | 72 |
| **GATE** | *5* | | 0.030 | 0.083 | 0.099 | 0.104 | -0.06 | -0.05 | 0.001 | 94 | 78 |
| **IATE** | *N* | *ols* | 0.031 | 0.179 | - | 0.225 | - | - | - | 33 | 22 |
| **IATE** | *N* | *rf* | 0.020 | 0.258 | - | 0.328 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.032 | 0.044 | 0.045 | 0.055 | -0.02 | -0.21 | 0.004 | 93 | 73 |
| **GATE** | *5* | *norm* | 0.032 | 0.083 | 0.099 | 0.104 | -0.05 | -0.05 | 0.001 | 94 | 78 |
| **IATE** | *N* | *ols* | 0.032 | 0.178 | - | 0.225 | - | - | - | 33 | 22 |
| **IATE** | *N* | *rf* | 0.021 | 0.257 | - | 0.327 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.007 | 0.036 | 0.031 | 0.044 | -0.04 | -0.16 | -0.013 | 83 | 61 |
| **GATE** | *5* | | 0.006 | 0.078 | 0.094 | 0.097 | -0.06 | -0.01 | -0.027 | 82 | 61 |
| **IATE** | *N* | | 0.007 | 0.165 | - | 0.208 | - | - | - | 83 | 61 |

**Note:** Table to be continued.

Table C.5 - continued: Linear IATE, medium selectivity

| | | | | Estimation of effects | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 10'000 | | | | | |
| **ATE** | *1* | *mcf* | 0.191 | 0.191 | 0.029 | 0.192 | 0.14 | -0.10 | 0.003 | 0 | 0 |
| **GATE** | *5* | | 0.191 | 0.191 | 0.044 | 0.206 | 0.12 | -0.16 | 0.000 | 14 | 7 |
| **IATE** | *N* | | 0.191 | 0.236 | - | 0.284 | - | - | - | 43 | 28 |
| **IATE eff** | *N* | | 0.191 | 0.233 | - | 0.279 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.026 | 0.031 | 0.027 | 0.037 | 0.21 | 0.20 | 0.003 | 90 | 66 |
| **GATE** | *5* | *cent* | 0.025 | 0.064 | 0.044 | 0.078 | 0.19 | 0.06 | -0.002 | 69 | 48 |
| **IATE** | *N* | | 0.026 | 0.177 | - | 0.223 | - | - | - | 53 | 36 |
| **IATE eff** | *N* | | 0.026 | 0.173 | - | 0.219 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.084 | 0.084 | 0.020 | 0.087 | -0.22 | -0.32 | 0.001 | 2 | 0 |
| **GATE** | *5* | | 0.084 | 0.086 | 0.047 | 0.098 | -0.07 | -0.11 | -0.000 | 55 | 30 |
| **IATE** | *N* | | 0.078 | 0.204 | - | 0.254 | - | - | - | 53 | 37 |
| **ATE** | *1* | *grf* | 0.024 | 0.027 | 0.022 | 0.033 | -0.10 | 0.20 | -0.001 | 79 | 50 |
| **GATE** | *5* | *cent* | 0.025 | 0.046 | 0.047 | 0.057 | -0.02 | -0.08 | -0.001 | 88 | 68 |
| **IATE** | *N* | | 0.024 | 0.195 | - | 0.244 | - | - | - | 55 | 38 |
| **ATE** | *1* | *dml* | 0.018 | 0.023 | 0.022 | 0.028 | -0.07 | 0.51 | 0.003 | 90 | 74 |
| **GATE** | *5* | | 0.018 | 0.042 | 0.049 | 0.053 | 0.12 | 0.00 | 0.000 | 93 | 78 |
| **IATE** | *N* | *ols* | 0.018 | 0.087 | - | 0.110 | - | - | - | 17 | 11 |
| **IATE** | *N* | *rf* | 0.011 | 0.195 | - | 0.248 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.018 | 0.023 | 0.022 | 0.028 | -0.08 | 0.53 | 0.003 | 90 | 74 |
| **GATE** | *5* | *norm* | 0.018 | 0.042 | 0.049 | 0.053 | 0.12 | -0.01 | 0.000 | 93 | 78 |
| **IATE** | *N* | *ols* | 0.019 | 0.087 | - | 0.110 | - | - | - | 17 | 11 |
| **IATE** | *N* | *rf* | 0.011 | 0.195 | - | 0.248 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.007 | 0.017 | 0.021 | 0.022 | 0.05 | 0.17 | -0.015 | 81 | 66 |
| **GATE** | *5* | | 0.007 | 0.043 | 0.042 | 0.053 | 0.08 | 0.12 | -0.014 | 77 | 56 |
| **IATE** | *N* | | 0.007 | 0.083 | - | 0.104 | - | - | - | 83 | 66 |

**Note:** For GATE and IATE the results are averaged over all effects. *CovP(95, 80)* denotes the (average) probability that the true value is part of the 95%/80% confidence interval. 1'000 / 500 replications used for 2'500 / 10'000 obs.

Table C.6: Linear IATE, strong selectivity

| | | | Estimation of effects | | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 2'500 | | | | | |
| **ATE** | *1* | *mcf* | 0.528 | 0.528 | 0.065 | 0.532 | 0.03 | 0.06 | -0.001 | 0 | 0 |
| **GATE** | *5* | | 0.528 | 0.528 | 0.086 | 0.532 | 0.03 | 0.09 | 0.009 | 1 | 0 |
| **IATE** | *N* | | 0.528 | 0.535 | - | 0.597 | - | - | - | 20 | 11 |
| **IATE eff** | *N* | | 0.531 | 0.536 | - | 0.593 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.085 | 0.088 | 0.057 | 0.102 | 0.03 | -0.06 | 0.007 | 76 | 49 |
| **GATE** | *5* | *cent* | 0.084 | 0.120 | 0.078 | 0.157 | 0.01 | 0.11 | 0.012 | 73 | 52 |
| **IATE** | *N* | | 0.085 | 0.241 | - | 0.301 | - | - | - | 61 | 42 |
| **IATE eff** | *N* | | 0.087 | 0.234 | - | 0.292 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.348 | 0.348 | 0.050 | 0.352 | -0.05 | -0.07 | -0.010 | 0 | 0 |
| **GATE** | *5* | | 0.348 | 0.348 | 0.096 | 0.372 | -0.06 | -0.03 | -0.005 | 10 | 3 |
| **IATE** | *N* | | 0.331 | 0.382 | - | 0.461 | - | - | - | 32 | 21 |
| **ATE** | *1* | *grf* | 0.117 | 0.118 | 0.046 | 0.126 | 0.04 | -0.07 | -0.006 | 19 | 7 |
| **GATE** | *5* | *cent* | 0.118 | 0.142 | 0.090 | 0.175 | 0.02 | 0.15 | -0.002 | 65 | 47 |
| **IATE** | *N* | | 0.120 | 0.274 | - | 0.342 | - | - | - | 46 | 31 |
| **ATE** | *1* | *dml* | 0.197 | 0.198 | 0.060 | 0.207 | -0.57 | 3.72 | -0.003 | 9 | 2 |
| **GATE** | *5* | | 0.197 | 0.208 | 0.129 | 0.208 | -0.95 | 10.36 | -0.016 | 54 | 33 |
| **IATE** | *N* | *ols* | 0.198 | 0.278 | - | 0.347 | - | - | - | 26 | 17 |
| **IATE** | *N* | *rf* | 0.170 | 0.344 | - | 0.568 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.188 | 0.188 | 0.056 | 0.196 | -0.02 | -0.02 | 0.002 | 9 | 2 |
| **GATE** | *5* | *norm* | 0.188 | 0.197 | 0.123 | 0.229 | -0.13 | 0.02 | -0.016 | 61 | 39 |
| **IATE** | *N* | *ols* | 0.189 | 0.271 | - | 0.336 | - | - | - | 29 | 20 |
| **IATE** | *N* | *rf* | 0.159 | 0.344 | - | 0.450 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.109 | 0.110 | 0.053 | 0.121 | -0.03 | -0.09 | -0.015 | 25 | 12 |
| **GATE** | *5* | | 0.109 | 0.127 | 0.103 | 0.153 | -0.04 | 0.01 | -0.030 | 60 | 41 |
| **IATE** | *N* | | 0.109 | 0.193 | - | 0.240 | - | - | - | 77 | 57 |

**Note:** Table to be continued.

Table C.6 - continued: Linear IATE, strong selectivity

| | | | | **Estimation of effects** | | | | | **Estimation of std. errors** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 10'000 | | | | | |
| **ATE** | *1* | *mcf* | 0.469 | 0.469 | 0.031 | 0.470 | -0.08 | 0.68 | 0.001 | 0 | 0 |
| **GATE** | *5* | | 0.468 | 0.468 | 0.048 | 0.477 | -0.06 | 0.08 | 0.005 | 0 | 0 |
| **IATE** | *N* | | 0.469 | 0.475 | - | 0.526 | - | - | - | 13 | 7 |
| **IATE eff** | *N* | | 0.469 | 0.475 | - | 0.523 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.012 | 0.024 | 0.028 | 0.030 | 0.26 | 0.68 | 0.005 | 98 | 85 |
| **GATE** | *5* | *cent* | 0.012 | 0.093 | 0.042 | 0.113 | -0.03 | 0.15 | 0.007 | 57 | 38 |
| **IATE** | *N* | | 0.012 | 0.208 | - | 0.263 | - | - | - | 51 | 35 |
| **IATE eff** | *N* | | 0.012 | 0.204 | - | 0.258 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.256 | 0.256 | 0.024 | 0.257 | -0.31 | 0.07 | -0.004 | 0 | 0 |
| **GATE** | *5* | | 0.255 | 0.255 | 0.049 | 0.274 | 0.00 | -0.05 | -0.002 | 4 | 2 |
| **IATE** | *N* | | 0.240 | 0.317 | - | 0.388 | - | - | - | 38 | 25 |
| **ATE** | *1* | *grf* | 0.058 | 0.058 | 0.026 | 0.064 | -0.08 | -0.07 | -0.006 | 24 | 10 |
| **GATE** | *5* | *cent* | 0.058 | 0.096 | 0.048 | 0.119 | -0.10 | 0.04 | -0.003 | 54 | 38 |
| **IATE** | *N* | | 0.065 | 0.251 | - | 0.314 | - | - | - | 47 | 32 |
| **ATE** | *1* | *dml* | 0.127 | 0.127 | 0.035 | 0.132 | 0.47 | 3.25 | -0.003 | 6 | 2 |
| **GATE** | *5* | | 0.127 | 0.133 | 0.078 | 0.153 | -0.02 | 7.79 | -0.024 | 44 | 26 |
| **IATE** | *N* | *ols* | 0.128 | 0.169 | - | 0.211 | - | - | - | 13 | 9 |
| **IATE** | *N* | *rf* | 0.104 | 0.259 | - | 0.462 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.117 | 0.117 | 0.032 | 0.121 | 0.47 | 3.25 | -0.003 | 6 | 2 |
| **GATE** | *5* | *norm* | 0.117 | 0.122 | 0.072 | 0.141 | -0.02 | 7.79 | -0.024 | 44 | 26 |
| **IATE** | *N* | *ols* | 0.117 | 0.161 | - | 0.199 | - | - | - | 13 | 9 |
| **IATE** | *N* | *rf* | 0.092 | 0.259 | - | 0.374 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.111 | 0.111 | 0.027 | 0.114 | -0.08 | -0.28 | -0.008 | 0 | 0 |
| **GATE** | *5* | | 0.111 | 0.112 | 0.051 | 0.126 | 0.02 | -0.10 | -0.015 | 25 | 14 |
| **IATE** | *N* | | 0.111 | 0.129 | - | 0.156 | - | - | - | 60 | 40 |

**Note:** For GATE and IATE the results are averaged over all effects. *CovP(95, 80)* denotes the (average) probability that the true value is part of the 95%/80% confidence interval. 1'000 / 500 replications used for 2'500 / 10'000 obs.

Table C.7: Nonlinear IATE, no selectivity

| | | | Estimation of effects | | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | N = 2'500 | | | | | | |
| **ATE** | *1* | *mcf* | 0.000 | 0.048 | 0.060 | 0.060 | 0.03 | 0.00 | 0.006 | 97 | 85 |
| **GATE** | *5* | | -0.001 | 0.080 | 0.077 | 0.110 | 0.06 | 0.25 | 0.004 | 86 | 65 |
| **IATE** | *N* | | -0.001 | 0.208 | - | 0.252 | - | - | - | 70 | 48 |
| **IATE eff** | *N* | | -0.001 | 0.201 | - | 0.240 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.002 | 0.049 | 0.062 | 0.062 | 0.00 | 0.06 | -0.001 | 94 | 78 |
| **GATE** | *5* | *cent* | 0.002 | 0.095 | 0.081 | 0.118 | 0.02 | 0.20 | -0.008 | 76 | 55 |
| **IATE** | *N* | | 0.002 | 0.220 | - | 0.264 | - | - | - | 56 | 37 |
| **IATE eff** | *N* | | 0.004 | 0.214 | - | 0.255 | - | - | - | - | - |
| **ATE** | *1* | *grf* | -0.001 | 0.034 | 0.043 | 0.043 | -0.15 | 0.03 | 0.000 | 95 | 82 |
| **GATE** | *5* | | -0.001 | 0.077 | 0.097 | 0.097 | 0.03 | 0.07 | -0.001 | 95 | 80 |
| **IATE** | *N* | | -0.001 | 0.229 | - | 0.271 | - | - | - | 50 | 33 |
| **ATE** | *1* | *grf* | 0.004 | 0.033 | 0.042 | 0.042 | 0.10 | 0.04 | 0.000 | 94 | 81 |
| **GATE** | *5* | *cent* | 0.004 | 0.075 | 0.094 | 0.095 | 0.04 | 0.06 | 0.000 | 95 | 80 |
| **IATE** | *N* | | 0.004 | 0.228 | - | 0.269 | - | - | - | 49 | 33 |
| **ATE** | *1* | *dml* | -0.002 | 0.034 | 0.042 | 0.043 | 0.05 | -0.07 | 0.005 | 98 | 84 |
| **GATE** | *5* | | -0.002 | 0.076 | 0.095 | 0.095 | -0.01 | 0.04 | 0.001 | 95 | 81 |
| **IATE** | *N* | *ols* | -0.002 | 0.172 | - | 0.218 | - | - | - | 32 | 21 |
| **IATE** | *N* | *rf* | -0.002 | 0.248 | - | 0.313 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | -0.002 | 0.034 | 0.043 | 0.043 | 0.06 | -0.06 | 0.005 | 97 | 83 |
| **GATE** | *5* | *norm* | -0.002 | 0.075 | 0.095 | 0.095 | -0.02 | 0.06 | 0.000 | 95 | 80 |
| **IATE** | *N* | *ols* | -0.002 | 0.171 | - | 0.218 | - | - | - | 31 | 21 |
| **IATE** | *N* | *rf* | -0.002 | 0.246 | - | 0.313 | - | - | - | - | - |
| **ATE** | *1* | *ols* | -0.001 | 0.034 | 0.042 | 0.042 | 0.05 | -0.09 | -0.013 | 82 | 74 |
| **GATE** | *5* | | -0.002 | 0.073 | 0.092 | 0.092 | -0.03 | 0.09 | -0.027 | 84 | 64 |
| **IATE** | *N* | | -0.001 | 0.168 | - | 0.212 | - | - | - | 84 | 64 |

**Note:** Table to be continued.

Table C.7 - continued: Nonlinear IATE, no selectivity

| | | | | Estimation of effects | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 10'000 | | | | | |
| **ATE** | *1* | *mcf* | 0.000 | 0.022 | 0.030 | 0.030 | 0.25 | 0.03 | 0.003 | 98 | 86 |
| **GATE** | *5* | | -0.001 | 0.059 | 0.042 | 0.073 | 0.08 | -0.04 | 0.001 | 75 | 53 |
| **IATE** | *N* | | 0.000 | 0.169 | - | 0.203 | - | - | - | 61 | 41 |
| **IATE eff** | *N* | | 0.000 | 0.165 | - | 0.197 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.003 | 0.022 | 0.028 | 0.028 | 0.11 | -0.15 | 0.002 | 97 | 78 |
| **GATE** | *5* | *cent* | 0.003 | 0.062 | 0.048 | 0.077 | 0.01 | -0.23 | -0.009 | 66 | 47 |
| **IATE** | *N* | | 0.003 | 0.183 | - | 0.218 | - | - | - | 45 | 30 |
| **IATE eff** | *N* | | 0.002 | 0.180 | - | 0.214 | - | - | - | - | - |
| **ATE** | *1* | *grf* | -0.000 | 0.016 | 0.020 | 0.020 | -0.10 | 0.50 | 0.001 | 96 | 83 |
| **GATE** | *5* | | -0.001 | 0.037 | 0.048 | 0.048 | -0.16 | 0.19 | -0.000 | 94 | 81 |
| **IATE** | *N* | | -0.001 | 0.189 | - | 0.223 | - | - | - | 54 | 36 |
| **ATE** | *1* | *grf* | 0.003 | 0.016 | 0.020 | 0.021 | 0.02 | 0.15 | 0.000 | 96 | 79 |
| **GATE** | *5* | *cent* | 0.003 | 0.036 | 0.045 | 0.045 | -0.07 | -0.09 | 0.001 | 95 | 80 |
| **IATE** | *N* | | 0.003 | 0.187 | - | 0.221 | - | - | - | 54 | 36 |
| **ATE** | *1* | *dml* | -0.001 | 0.016 | 0.020 | 0.020 | -0.09 | -0.01 | 0.004 | 98 | 86 |
| **GATE** | *5* | | -0.001 | 0.037 | 0.046 | 0.046 | 0.02 | 0.01 | 0.001 | 95 | 82 |
| **IATE** | *N* | *ols* | -0.001 | 0.088 | - | 0.111 | - | - | - | 15 | 10 |
| **IATE** | *N* | *rf* | -0.001 | 0.189 | - | 0.239 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | -0.001 | 0.016 | 0.020 | 0.020 | -0.09 | -0.02 | 0.003 | 98 | 86 |
| **GATE** | *5* | *norm* | -0.001 | 0.037 | 0.046 | 0.046 | 0.02 | 0.00 | 0.001 | 95 | 82 |
| **IATE** | *N* | *ols* | -0.001 | 0.087 | - | 0.111 | - | - | - | 15 | 10 |
| **IATE** | *N* | *rf* | -0.001 | 0.189 | - | 0.239 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.000 | 0.016 | 0.020 | 0.020 | 0.01 | -0.01 | -0.005 | 84 | 66 |
| **GATE** | *5* | | -0.001 | 0.037 | 0.045 | 0.046 | 0.01 | 0.01 | -0.013 | 84 | 64 |
| **IATE** | *N* | | 0.000 | 0.087 | - | 0.111 | - | - | - | 81 | 64 |

**Note:** For GATE and IATE the results are averaged over all effects. *CovP(95, 80)* denotes the (average) probability that the true value is part of the 95%/80% confidence interval. 1'000 / 500 replications used for 2'500 / 10'000 obs.

Table C.8: Nonlinear IATE, medium selectivity

| | | | | Estimation of effects | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 2'500 | | | | | |
| **ATE** | *1* | *mcf* | 0.219 | 0.219 | 0.060 | 0.227 | -0.03 | -0.14 | 0.004 | 6 | 1 |
| **GATE** | *5* | | 0.219 | 0.221 | 0.080 | 0.244 | 0.07 | 0.07 | 0.005 | 31 | 18 |
| **IATE** | *N* | | 0.219 | 0.276 | - | 0.336 | - | - | - | 58 | 41 |
| **IATE eff** | *N* | | 0.219 | 0.268 | - | 0.336 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.056 | 0.067 | 0.061 | 0.083 | 0.09 | 0.29 | -0.001 | 84 | 64 |
| **GATE** | *5* | *cent* | 0.056 | 0.102 | 0.081 | 0.126 | 0.05 | 0.15 | -0.003 | 77 | 54 |
| **IATE** | *N* | | 0.056 | 0.225 | - | 0.271 | - | - | - | 57 | 38 |
| **IATE eff** | *N* | | 0.059 | | - | 0.261 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.128 | 0.128 | 0.044 | 0.135 | -0.04 | -0.24 | -0.002 | 16 | 5 |
| **GATE** | *5* | | 0.128 | 0.137 | 0.097 | 0.162 | 0.01 | -0.04 | -0.002 | 71 | 47 |
| **IATE** | *N* | | 0.123 | 0.259 | - | 0.310 | - | - | - | 45 | 30 |
| **ATE** | *1* | *grf* | 0.050 | 0.055 | 0.042 | 0.065 | 0.03 | -0.16 | -0.000 | 77 | 52 |
| **GATE** | *5* | *cent* | 0.050 | 0.087 | 0.093 | 0.108 | 0.01 | 0.08 | -0.000 | 91 | 72 |
| **IATE** | *N* | | 0.050 | 0.244 | - | 0.289 | - | - | - | 45 | 30 |
| **ATE** | *1* | *dml* | 0.030 | 0.043 | 0.045 | 0.054 | -0.02 | -0.21 | 0.004 | 93 | 73 |
| **GATE** | *5* | | 0.029 | 0.083 | 0.099 | 0.104 | -0.05 | -0.05 | 0.001 | 94 | 79 |
| **IATE** | *N* | *ols* | 0.030 | 0.183 | - | 0.230 | - | - | - | 32 | 22 |
| **IATE** | *N* | *rf* | 0.019 | 0.260 | - | 0.331 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.031 | 0.044 | 0.045 | 0.055 | -0.03 | -0.24 | 0.004 | 93 | 72 |
| **GATE** | *5* | *norm* | 0.031 | 0.083 | 0.099 | 0.104 | -0.05 | -0.05 | 0.001 | 94 | 78 |
| **IATE** | *N* | *ols* | 0.031 | 0.182 | - | 0.230 | - | - | - | 32 | 21 |
| **IATE** | *N* | *rf* | 0.020 | 0.259 | - | 0.330 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.008 | 0.036 | 0.044 | 0.045 | -0.04 | -0.16 | -0.013 | 83 | 60 |
| **GATE** | *5* | | 0.008 | 0.079 | 0.094 | 0.098 | -0.06 | -0.06 | -0.027 | 82 | 61 |
| **IATE** | *N* | | 0.008 | 0.170 | - | 0.215 | - | - | - | 82 | 62 |

**Note:** Table to be continued.

Table C.8 - continued: Nonlinear IATE, medium selectivity

| | | | | Estimation of effects | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 10'000 | | | | | |
| **ATE** | *1* | *mcf* | 0.181 | 0.181 | 0.030 | 0.184 | 0.22 | -0.47 | 0.002 | 0 | 0 |
| **GATE** | *5* | | 0.181 | 0.181 | 0.044 | 0.194 | 0.14 | 0.20 | 0.002 | 13 | 7 |
| **IATE** | *N* | | 0.181 | 0.228 | - | 0.275 | - | - | - | 50 | 35 |
| **IATE eff** | *N* | | 0.181 | 0.223 | - | 0.269 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.026 | 0.031 | 0.027 | 0.038 | 0.21 | 0.26 | 0.003 | 90 | 68 |
| **GATE** | *5* | *cent* | 0.025 | 0.063 | 0.045 | 0.077 | 0.18 | 0.05 | -0.002 | 70 | 50 |
| **IATE** | *N* | | 0.026 | 0.187 | - | 0.223 | - | - | - | 47 | 31 |
| **IATE eff** | *N* | | 0.026 | 0.184 | - | 0.219 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.084 | 0.084 | 0.020 | 0.086 | -0.22 | -0.35 | 0.001 | 2 | 0 |
| **GATE** | *5* | | 0.084 | 0.086 | 0.047 | 0.097 | -0.08 | -0.14 | -0.000 | 55 | 30 |
| **IATE** | *N* | | 0.078 | 0.214 | - | 0.254 | - | - | - | 48 | 32 |
| **ATE** | *1* | *grf* | 0.024 | 0.027 | 0.022 | 0.033 | -0.08 | 0.18 | -0.001 | 78 | 51 |
| **GATE** | *5* | *cent* | 0.024 | 0.046 | 0.047 | 0.056 | -0.03 | -0.10 | -0.001 | 89 | 69 |
| **IATE** | *N* | | 0.024 | 0.207 | - | 0.244 | - | - | - | 48 | 32 |
| **ATE** | *1* | *dml* | 0.017 | 0.022 | 0.022 | 0.028 | -0.05 | 0.49 | 0.003 | 91 | 75 |
| **GATE** | *5* | | 0.017 | 0.042 | 0.049 | 0.053 | 0.12 | -0.01 | 0.000 | 94 | 78 |
| **IATE** | *N* | *ols* | 0.018 | 0.094 | - | 0.120 | - | - | - | 16 | 10 |
| **IATE** | *N* | *rf* | 0.010 | 0.198 | - | 0.251 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.017 | 0.022 | 0.022 | 0.028 | -0.07 | 0.50 | 0.003 | 90 | 74 |
| **GATE** | *5* | *norm* | 0.017 | 0.042 | 0.049 | 0.053 | 0.12 | -0.01 | 0.000 | 93 | 78 |
| **IATE** | *N* | *ols* | 0.018 | 0.094 | - | 0.120 | - | - | - | 16 | 10 |
| **IATE** | *N* | *rf* | 0.011 | 0.198 | - | 0.251 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.009 | 0.018 | 0.021 | 0.023 | 0.07 | 0.16 | -0.005 | 81 | 65 |
| **GATE** | *5* | | 0.008 | 0.039 | 0.047 | 0.055 | 0.08 | 0.12 | -0.013 | 75 | 53 |
| **IATE** | *N* | | 0.009 | 0.091 | - | 0.116 | - | - | - | 79 | 59 |

**Note:** For GATE and IATE the results are averaged over all effects. *CovP(95, 80)* denotes the (average) probability that the true value is part of the 95%/80% confidence interval. 1'000 / 500 replications used for 2'500 / 10'000 obs.

Table C.9: Nonlinear IATE, strong selectivity

| | | | | Estimation of effects | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 2'500 | | | | | |
| **ATE** | *1* | *mcf* | 0.517 | 0.517 | 0.062 | 0.521 | -0.07 | -0.05 | 0.002 | 0 | 0 |
| **GATE** | *5* | | 0.517 | 0.517 | 0.087 | 0.532 | 0.01 | 0.05 | 0.012 | 1 | 0 |
| **IATE** | *N* | | 0.518 | 0.520 | - | 0.589 | - | - | - | 28 | 16 |
| **IATE eff** | *N* | | 0.518 | 0.520 | - | 0.581 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.091 | 0.094 | 0.057 | 0.108 | 0.03 | -0.04 | 0.007 | 73 | 44 |
| **GATE** | *5* | *cent* | 0.091 | 0.129 | 0.078 | 0.160 | 0.04 | 0.12 | 0.012 | 72 | 51 |
| **IATE** | *N* | | 0.091 | 0.252 | - | 0.303 | - | - | - | 56 | 38 |
| **IATE eff** | *N* | | 0.093 | 0.246 | - | 0.294 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.356 | 0.356 | 0.051 | 0.359 | -0.04 | -0.09 | -0.010 | 0 | 0 |
| **GATE** | *5* | | 0.356 | 0.356 | 0.096 | 0.379 | -0.06 | -0.05 | -0.006 | 9 | 3 |
| **IATE** | *N* | | 0.341 | 0.386 | - | 0.469 | - | - | - | 35 | 24 |
| **ATE** | *1* | *grf* | 0.124 | 0.125 | 0.046 | 0.133 | 0.04 | -0.08 | -0.006 | 15 | 5 |
| **GATE** | *5* | *cent* | 0.125 | 0.146 | 0.091 | 0.178 | 0.02 | 0.15 | -0.002 | 64 | 45 |
| **IATE** | *N* | | 0.128 | 0.289 | - | 0.346 | - | - | - | 41 | 27 |
| **ATE** | *1* | *dml* | 0.198 | 0.199 | 0.061 | 0.207 | -0.53 | 3.20 | -0.003 | 9 | 2 |
| **GATE** | *5* | | 0.198 | 0.208 | 0.130 | 0.240 | -0.89 | 9.06 | -0.016 | 53 | 33 |
| **IATE** | *N* | *ols* | 0.199 | 0.280 | - | 0.350 | - | - | - | 26 | 17 |
| **IATE** | *N* | *rf* | 0.170 | 0.345 | - | 0.563 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.189 | 0.189 | 0.057 | 0.197 | -0.02 | -0.01 | 0.002 | 11 | 3 |
| **GATE** | *5* | *norm* | 0.188 | 0.198 | 0.124 | 0.228 | -0.13 | 0.03 | -0.005 | 62 | 39 |
| **IATE** | *N* | *ols* | 0.189 | 0.273 | - | 0.339 | - | - | - | 30 | 20 |
| **IATE** | *N* | *rf* | 0.159 | 0.346 | - | 0.452 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.124 | 0.125 | 0.053 | 0.135 | -0.03 | 0.10 | -0.015 | 17 | 8 |
| **GATE** | *5* | | 0.124 | 0.139 | 0.103 | 0.165 | -0.04 | 0.01 | -0.030 | 55 | 36 |
| **IATE** | *N* | | 0.124 | 0.202 | - | 0.253 | - | - | - | 75 | 54 |

**Note:** Table to be continued.

Table C.9 - continued: Nonlinear IATE, strong selectivity

| | | | Estimation of effects | | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | N = 10'000 | | | | | | |
| **ATE** | *1* | *mcf* | 0.455 | 0.455 | 0.032 | 0.456 | -0.02 | -0.16 | 0.001 | 0 | 0 |
| **GATE** | *5* | | 0.455 | 0.455 | 0.048 | 0.463 | -0.02 | -0.14 | 0.007 | 0 | 0 |
| **IATE** | *N* | | 0.455 | 0.458 | - | 0.513 | - | - | - | 21 | 12 |
| **IATE eff** | *N* | | 0.453 | 0.454 | - | 0.506 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.017 | 0.026 | 0.028 | 0.033 | 0.26 | 0.69 | 0.005 | 96 | 83 |
| **GATE** | *5* | *cent* | 0.017 | 0.091 | 0.043 | 0.110 | -0.03 | 0.12 | 0.007 | 59 | 40 |
| **IATE** | *N* | | 0.017 | 0.218 | - | 0.260 | - | - | - | 45 | 30 |
| **IATE eff** | *N* | | 0.018 | 0.215 | - | 0.255 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.261 | 0.261 | 0.024 | 0.262 | -0.27 | -0.01 | -0.004 | 0 | 0 |
| **GATE** | *5* | | 0.261 | 0.261 | 0.049 | 0.278 | 0.03 | -0.02 | -0.002 | 3 | 1 |
| **IATE** | *N* | | 0.248 | 0.322 | - | 0.392 | - | - | - | 38 | 25 |
| **ATE** | *1* | *grf* | 0.063 | 0.063 | 0.026 | 0.068 | -0.06 | -0.11 | -0.006 | 20 | 8 |
| **GATE** | *5* | *cent* | 0.064 | 0.095 | 0.049 | 0.118 | -0.10 | 0.01 | -0.003 | 54 | 37 |
| **IATE** | *N* | | 0.072 | 0.265 | - | 0.314 | - | - | - | 41 | 27 |
| **ATE** | *1* | *dml* | 0.126 | 0.126 | 0.035 | 0.130 | 0.51 | 3.24 | -0.003 | 6 | 2 |
| **GATE** | *5* | | 0.125 | 0.131 | 0.078 | 0.150 | 0.02 | 7.37 | -0.014 | 45 | 26 |
| **IATE** | *N* | *ols* | 0.126 | 0.170 | - | 0.212 | - | - | - | 13 | 9 |
| **IATE** | *N* | *rf* | 0.102 | 0.259 | - | 0.459 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.114 | 0.114 | 0.032 | 0.119 | 0.16 | 0.35 | 0.000 | 6 | 2 |
| **GATE** | *5* | *norm* | 0.114 | 0.120 | 0.072 | 0.138 | -0.10 | 0.17 | -0.006 | 54 | 35 |
| **IATE** | *N* | *ols* | 0.115 | 0.162 | - | 0.200 | - | - | - | 15 | 10 |
| **IATE** | *N* | *rf* | 0.089 | 0.260 | - | 0.375 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.126 | 0.126 | 0.027 | 0.129 | 0.08 | -0.12 | -0.008 | 0 | 0 |
| **GATE** | *5* | | 0.126 | 0.127 | 0.052 | 0.140 | 0.02 | -0.11 | -0.015 | 20 | 11 |
| **IATE** | *N* | | 0.126 | 0.145 | - | 0.175 | - | - | - | 55 | 36 |

**Note:** For GATE and IATE the results are averaged over all effects. *CovP(95, 80)* denotes the (average) probability that the true value is part of the 95%/80% confidence interval. 1'000 / 500 replications used for 2'500 / 10'000 obs.

Table C.10: Quadratic IATE, no selectivity

| | | | Estimation of effects | | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | N = 2'500 | | | | | | |
| **ATE** | *1* | *mcf* | -0.003 | 0.053 | 0.065 | 0.065 | 0.01 | -0.22 | 0.001 | 96 | 80 |
| **GATE** | *5* | | -0.003 | 0.090 | 0.074 | 0.110 | 0.06 | 0.02 | 0.003 | 81 | 59 |
| **IATE** | *N* | | -0.003 | 0.450 | - | 0.648 | - | - | - | 30 | 18 |
| **IATE eff** | *N* | | -0.003 | 0.448 | - | 0.644 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.002 | 0.053 | 0.065 | 0.065 | 0.00 | -0.07 | -0.001 | 95 | 80 |
| **GATE** | *5* | *cent* | 0.002 | 0.091 | 0.074 | 0.110 | 0.07 | 0.20 | -0.002 | 79 | 55 |
| **IATE** | *N* | | 0.002 | 0.455 | - | 0.653 | - | - | - | 27 | 55 |
| **IATE eff** | *N* | | 0.004 | 0.453 | - | 0.649 | - | - | - | - | - |
| **ATE** | *1* | *grf* | -0.002 | 0.037 | 0.046 | 0.046 | -0.13 | 0.06 | -0.001 | 95 | 80 |
| **GATE** | *5* | | -0.002 | 0.082 | 0.103 | 0.103 | 0.01 | -0.03 | -0.001 | 95 | 80 |
| **IATE** | *N* | | -0.001 | 0.465 | - | 0.668 | - | - | - | 24 | 15 |
| **ATE** | *1* | *grf* | 0.003 | 0.036 | 0.045 | 0.045 | 0.04 | -0.03 | 0.000 | 95 | 81 |
| **GATE** | *5* | *cent* | 0.003 | 0.081 | 0.101 | 0.101 | -0.02 | -0.02 | -0.000 | 95 | 79 |
| **IATE** | *N* | | 0.004 | 0.465 | - | 0.665 | - | - | - | 23 | 15 |
| **ATE** | *1* | *dml* | -0.001 | 0.036 | 0.045 | 0.046 | -0.01 | 0.01 | 0.005 | 97 | 84 |
| **GATE** | *5* | | -0.002 | 0.080 | 0.101 | 0.101 | 0.00 | 0.06 | 0.002 | 95 | 81 |
| **IATE** | *N* | *ols* | -0.002 | 0.505 | - | 0.720 | - | - | - | 13 | 8 |
| **IATE** | *N* | *rf* | -0.004 | 0.446 | - | 0.626 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | -0.002 | 0.036 | 0.046 | 0.046 | -0.01 | 0.01 | 0.004 | 97 | 83 |
| **GATE** | *5* | *norm* | -0.002 | 0.080 | 0.101 | 0.101 | 0.00 | 0.07 | 0.000 | 95 | 80 |
| **IATE** | *N* | *ols* | -0.002 | 0.505 | - | 0.720 | - | - | - | 12 | 8 |
| **IATE** | *N* | *rf* | -0.004 | 0.444 | - | 0.625 | - | - | - | - | - |
| **ATE** | *1* | *ols* | -0.002 | 0.037 | 0.046 | 0.046 | 0.04 | 0.03 | -0.014 | 85 | 61 |
| **GATE** | *5* | | -0.002 | 0.080 | 0.101 | 0.101 | 0.01 | 0.09 | -0.029 | 83 | 65 |
| **IATE** | *N* | | -0.002 | 0.505 | - | 0.720 | - | - | - | 39 | 25 |

**Note:** Table to be continued.

Table C.10 - continued: Quadratic IATE, no selectivity

| | | | | Estimation of effects | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 10'000 | | | | | |
| **ATE** | *1* | *mcf* | 0.001 | 0.024 | 0.030 | 0.030 | 0.22 | -0.02 | 0.004 | 97 | 84 |
| **GATE** | *5* | | 0.001 | 0.062 | 0.040 | 0.073 | 0.15 | -0.15 | 0.001 | 68 | 45 |
| **IATE** | *N* | | 0.001 | 0.405 | - | 0.584 | - | - | - | 21 | 13 |
| **IATE eff** | *N* | | 0.000 | 0.404 | - | 0.583 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.002 | 0.024 | 0.029 | 0.029 | 0.21 | 0.01 | 0.002 | 96 | 82 |
| **GATE** | *5* | *cent* | 0.002 | 0.059 | 0.042 | 0.070 | 0.16 | -0.09 | -0.002 | 69 | 45 |
| **IATE** | *N* | | 0.002 | 0.409 | - | 0.588 | - | - | - | 19 | 12 |
| **IATE eff** | *N* | | 0.002 | 0.408 | - | 0.586 | - | - | - | - | - |
| **ATE** | *1* | *grf* | -0.001 | 0.017 | 0.021 | 0.021 | -0.02 | 0.21 | 0.001 | 95 | 83 |
| **GATE** | *5* | | -0.001 | 0.041 | 0.051 | 0.051 | -0.06 | -0.115 | -0.001 | 94 | 79 |
| **IATE** | *N* | | -0.001 | 0.423 | - | 0.607 | - | - | - | 25 | 16 |
| **ATE** | *1* | *grf* | 0.002 | 0.018 | 0.022 | 0.022 | -0.01 | -0.13 | 0.000 | 96 | 79 |
| **GATE** | *5* | *cent* | 0.002 | 0.039 | 0.048 | 0.048 | -0.08 | -0.08 | 0.001 | 96 | 81 |
| **IATE** | *N* | | 0.002 | 0.422 | - | 0.605 | - | - | - | 25 | 16 |
| **ATE** | *1* | *dml* | 0.000 | 0.016 | 0.020 | 0.020 | 0.00 | -0.12 | 0.004 | 99 | 86 |
| **GATE** | *5* | | 0.000 | 0.039 | 0.049 | 0.049 | 0.06 | -0.07 | 0.001 | 95 | 81 |
| **IATE** | *N* | *ols* | 0.000 | 0.483 | - | 0.693 | - | - | - | 3 | 2 |
| **IATE** | *N* | *rf* | -0.004 | 0.339 | - | 0.479 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.000 | 0.016 | 0.020 | 0.020 | 0.00 | -0.13 | 0.004 | 99 | 86 |
| **GATE** | *5* | *norm* | 0.000 | 0.039 | 0.049 | 0.049 | 0.06 | -0.06 | 0.001 | 96 | 81 |
| **IATE** | *N* | *ols* | 0.000 | 0.483 | - | 0.693 | - | - | - | 3 | 2 |
| **IATE** | *N* | *rf* | -0.004 | 0.339 | - | 0.479 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.001 | 0.016 | 0.021 | 0.021 | 0.02 | 0.09 | -0.005 | 84 | 68 |
| **GATE** | *5* | | 0.000 | 0.040 | 0.050 | 0.050 | 0.00 | -0.04 | -0.014 | 83 | 64 |
| **IATE** | *N* | | 0.000 | 0.483 | - | 0.693 | - | - | - | 17 | 11 |

**Note:** For GATE and IATE the results are averaged over all effects. *CovP(95, 80)* denotes the (average) probability that the true value is part of the 95%/80% confidence interval. 1'000 / 500 replications used for 2'500 / 10'000 obs.

Table C.11: Quadratic IATE, medium selectivity

| | | | Estimation of effects | | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | N = 2'500 | | | | | | |
| **ATE** | *1* | *mcf* | 0.123 | 0.124 | 0.064 | 0.139 | 0.10 | 0.08 | 0.001 | 52 | 27 |
| **GATE** | *5* | | 0.123 | 0.155 | 0.082 | 0.179 | 0.11 | 0.02 | 0.003 | 54 | 35 |
| **IATE** | *N* | | 0.123 | 0.504 | - | 0.667 | - | - | - | 25 | 16 |
| **IATE eff** | *N* | | 0.123 | 0.503 | - | 0.663 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | -0.036 | 0.059 | 0.064 | 0.074 | 0.05 | 0.25 | -0.002 | 90 | 73 |
| **GATE** | *5* | *cent* | -0.036 | 0.099 | 0.080 | 0.127 | 0.25 | 0.18 | -0.001 | 79 | 60 |
| **IATE** | *N* | | -0.036 | 0.446 | - | 0.660 | - | - | - | 31 | 19 |
| **IATE eff** | *N* | | -0.034 | 0.444 | - | 0.656 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.030 | 0.044 | 0.045 | 0.054 | -0.05 | -0.14 | -0.001 | 90 | 68 |
| **GATE** | *5* | | 0.030 | 0.100 | 0.099 | 0.125 | 0.05 | 0.01 | -0.001 | 88 | 68 |
| **IATE** | *N* | | 0.014 | 0.480 | - | 0.682 | - | - | - | 22 | 14 |
| **ATE** | *1* | *grf* | -0.046 | 0.053 | 0.044 | 0.064 | 0.05 | -0.30 | -0.001 | 79 | 57 |
| **GATE** | *5* | *cent* | -0.046 | 0.100 | 0.096 | 0.128 | -0.02 | 0.10 | 0.001 | 86 | 69 |
| **IATE** | *N* | | -0.057 | 0.456 | - | 0.682 | - | - | - | 25 | 16 |
| **ATE** | *1* | *dml* | -0.014 | 0.040 | 0.048 | 0.050 | -0.04 | -0.22 | 0.004 | 96 | 82 |
| **GATE** | *5* | | -0.014 | 0.090 | 0.104 | 0.113 | -0.04 | 0.02 | 0.002 | 93 | 77 |
| **IATE** | *N* | *ols* | -0.014 | 0.509 | - | 0.730 | - | - | - | 14 | 9 |
| **IATE** | *N* | *rf* | -0.019 | 0.467 | - | 0.667 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | -0.013 | 0.040 | 0.048 | 0.050 | -0.04 | -0.25 | 0.004 | 96 | 82 |
| **GATE** | *5* | *norm* | -0.013 | 0.090 | 0.104 | 0.113 | -0.02 | 0.02 | 0.002 | 93 | 77 |
| **IATE** | *N* | *ols* | -0.014 | 0.508 | - | 0.730 | - | - | - | 14 | 9 |
| **IATE** | *N* | *rf* | -0.019 | 0.467 | - | 0.666 | - | - | - | - | - |
| **ATE** | *1* | *ols* | -0.104 | 0.105 | 0.049 | 0.115 | 0.01 | -0.10 | -0.016 | 20 | 10 |
| **GATE** | *5* | | -0.104 | 0.156 | 0.101 | 0.156 | -0.05 | 0.02 | -0.031 | 56 | 41 |
| **IATE** | *N* | | -0.104 | 0.526 | - | 0.798 | - | - | - | 56 | 41 |

**Note:** Table to be continued.

Table C.11 - continued: Quadratic IATE, medium selectivity

| | | | **Estimation of effects** | | | | | | **Estimation of std. errors** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 10'000 | | | | | |
| **ATE** | *1* | *mcf* | 0.092 | 0.092 | 0.029 | 0.097 | 0.25 | 0.05 | 0.003 | 18 | 3 |
| **GATE** | *5* | | 0.091 | 0.124 | 0.045 | 0.137 | 0.26 | 0.14 | 0.002 | 33 | 17 |
| **IATE** | *N* | | 0.091 | 0.452 | - | 0.612 | - | - | - | 19 | 12 |
| **IATE eff** | *N* | | 0.091 | 0.451 | - | 0.609 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | -0.059 | 0.060 | 0.029 | 0.066 | 0.29 | 0.42 | 0.002 | 53 | 24 |
| **GATE** | *5* | *cent* | -0.060 | 0.079 | 0.044 | 0.106 | 0.28 | 0.21 | 0.000 | 68 | 51 |
| **IATE** | *N* | | -0.060 | 0.401 | - | 0.611 | - | - | - | 23 | 15 |
| **IATE eff** | *N* | | -0.059 | 0.399 | - | 0.608 | - | - | - | - | - |
| **ATE** | *1* | *grf* | -0.002 | 0.017 | 0.021 | 0.021 | -0.22 | 0.03 | 0.001 | 95 | 82 |
| **GATE** | *5* | | -0.002 | 0.060 | 0.049 | 0.076 | -0.04 | -0.13 | -0.001 | 82 | 60 |
| **IATE** | *N* | | -0.030 | 0.446 | - | 0.655 | - | - | - | 22 | 14 |
| **ATE** | *1* | *grf* | -0.061 | 0.061 | 0.023 | 0.065 | -0.06 | 0.41 | -0.001 | 20 | 8 |
| **GATE** | *5* | *cent* | -0.061 | 0.073 | 0.050 | 0.095 | -0.02 | -0.21 | -0.001 | 71 | 53 |
| **IATE** | *N* | | -0.083 | 0.432 | - | 0.662 | - | - | - | 25 | 16 |
| **ATE** | *1* | *dml* | -0.012 | 0.021 | 0.023 | 0.026 | 0.02 | 0.03 | 0.003 | 94 | 80 |
| **GATE** | *5* | | -0.013 | 0.048 | 0.053 | 0.062 | 0.07 | 0.02 | 0.000 | 91 | 74 |
| **IATE** | *N* | *ols* | -0.013 | 0.482 | - | 0.698 | - | - | - | 3 | 2 |
| **IATE** | *N* | *rf* | -0.016 | 0.370 | - | 0.537 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | -0.012 | 0.021 | 0.023 | 0.026 | 0.02 | 0.03 | 0.003 | 94 | 80 |
| **GATE** | *5* | *norm* | -0.013 | 0.048 | 0.053 | 0.062 | 0.07 | 0.01 | 0.000 | 91 | 74 |
| **IATE** | *N* | *ols* | -0.013 | 0.482 | - | 0.698 | - | - | - | 3 | 2 |
| **IATE** | *N* | *rf* | -0.016 | 0.370 | - | 0.537 | - | - | - | - | - |
| **ATE** | *1* | *ols* | -0.104 | 0.104 | 0.023 | 0.107 | -0.01 | -0.11 | -0.007 | 0 | 0 |
| **GATE** | *5* | | -0.104 | 0.135 | 0.052 | 0.186 | 0.03 | 0.15 | -0.017 | 44 | 31 |
| **IATE** | *N* | | -0.105 | 0.505 | - | 0.776 | - | - | - | 18 | 11 |

**Note:** For GATE and IATE the results are averaged over all effects. *CovP(95, 80)* denotes the (average) probability that the true value is part of the 95%/80% confidence interval. 1'000 / 500 replications used for 2'500 / 10'000 obs.

Table C.12: Quadratic IATE, strong selectivity

| | | | Estimation of effects | | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | N = 2'500 | | | | | | |
| **ATE** | *1* | *mcf* | 0.261 | 0.261 | 0.066 | 0.269 | -0.03 | 0.06 | -0.001 | 2 | 0 |
| **GATE** | *5* | | 0.261 | 0.280 | 0.091 | 0.316 | 0.03 | 0.09 | 0.002 | 27 | 19 |
| **IATE** | *N* | | 0.261 | 0.601 | - | 0.743 | - | - | - | 20 | 12 |
| **IATE eff** | *N* | | 0.264 | 0.600 | - | 0.739 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | -0.182 | 0.182 | 0.057 | 0.191 | -0.02 | -0.06 | 0.007 | 16 | 4 |
| **GATE** | *5* | *cent* | -0.182 | 0.187 | 0.078 | 0.231 | 0.03 | 0.03 | 0.008 | 52 | 33 |
| **IATE** | *N* | | -0.182 | 0.442 | - | 0.717 | - | - | - | 46 | 29 |
| **IATE eff** | *N* | | -0.180 | 0.437 | - | 0.712 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.028 | 0.046 | 0.049 | 0.056 | -0.10 | -0.08 | -0.008 | 85 | 62 |
| **GATE** | *5* | | 0.029 | 0.120 | 0.094 | 0.146 | -0.06 | 0.01 | -0.003 | 76 | 54 |
| **IATE** | *N* | | -0.020 | 0.479 | - | 0.697 | - | - | - | 25 | 16 |
| **ATE** | *1* | *grf* | -0.208 | 0.208 | 0.046 | 0.213 | 0.02 | -0.04 | -0.006 | 0 | 0 |
| **GATE** | *5* | *cent* | -0.208 | 0.212 | 0.090 | 0.253 | 0.02 | 0.14 | -0.001 | 46 | 28 |
| **IATE** | *N* | | -0.235 | 0.435 | - | 0.732 | - | - | - | 44 | 27 |
| **ATE** | *1* | *dml* | -0.037 | 0.056 | 0.059 | 0.069 | -0.02 | -0.09 | -0.003 | 87 | 69 |
| **GATE** | *5* | | -0.037 | 0.154 | 0.123 | 0.194 | -0.10 | 0.48 | -0.010 | 73 | 55 |
| **IATE** | *N* | *ols* | -0.037 | 0.550 | - | 0.798 | - | - | - | 14 | 9 |
| **IATE** | *N* | *rf* | -0.058 | 0.513 | - | 0.759 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | -0.048 | 0.061 | 0.057 | 0.074 | 0.01 | -0.07 | 0.000 | 86 | 66 |
| **GATE** | *5* | *norm* | -0.048 | 0.156 | 0.121 | 0.197 | -0.08 | 0.04 | -0.003 | 75 | 57 |
| **IATE** | *N* | *ols* | -0.047 | 0.550 | - | 0.803 | - | - | - | 16 | 10 |
| **IATE** | *N* | *rf* | -0.070 | 0.515 | - | 0.750 | - | - | - | - | - |
| **ATE** | *1* | *ols* | -0.369 | 0.369 | 0.057 | 0.373 | -0.01 | 0.03 | -0.018 | 0 | 0 |
| **GATE** | *5* | | -0.369 | 0.392 | 0.108 | 0.521 | -0.07 | -0.02 | -0.034 | 33 | 24 |
| **IATE** | *N* | | -0.369 | 0.675 | - | 1.091 | - | - | - | 38 | 24 |

**Note:** Table to be continued.

Table C.12 - continued: Quadratic IATE, strong selectivity

| | | | | Estimation of effects | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 10'000 | | | | | |
| **ATE** | *1* | *mcf* | 0.198 | 0.198 | 0.032 | 0.201 | -0.04 | 0.57 | 0.001 | 0 | 0 |
| **GATE** | *5* | | 0.198 | 0.231 | 0.052 | 0.256 | 0.09 | 0.09 | -0.001 | 15 | 8 |
| **IATE** | *N* | | 0.198 | 0.552 | - | 0.702 | - | - | - | 15 | 9 |
| **IATE eff** | *N* | | 0.198 | 0.551 | - | 0.700 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | -0.254 | 0.254 | 0.029 | 0.255 | 0.34 | 0.51 | 0.004 | 0 | 0 |
| **GATE** | *5* | *cent* | -0.254 | 0.254 | 0.044 | 0.280 | 0.11 | 0.13 | 0.003 | 3 | 1 |
| **IATE** | *N* | | -0.254 | 0.420 | - | 0.721 | - | - | - | 40 | 25 |
| **IATE eff** | *N* | | -0.254 | 0.418 | - | 0.719 | - | - | - | - | - |
| **ATE** | *1* | *grf* | -0.065 | 0.065 | 0.025 | 0.069 | -0.21 | -0.02 | -0.004 | 16 | 6 |
| **GATE** | *5* | | -0.065 | 0.094 | 0.049 | 0.129 | -0.03 | -0.02 | -0.002 | 66 | 49 |
| **IATE** | *N* | | -0.124 | 0.448 | - | 0.703 | - | - | - | 30 | 18 |
| **ATE** | *1* | *grf* | -0.264 | 0.264 | 0.026 | 0.265 | -0.04 | -0.21 | -0.006 | 0 | 0 |
| **GATE** | *5* | *cent* | -0.264 | 0.264 | 0.049 | 0.287 | -0.05 | 0.06 | -0.003 | 2 | 0 |
| **IATE** | *N* | | -0.299 | 0.433 | - | 0.754 | - | - | - | 49 | 31 |
| **ATE** | *1* | *dml* | -0.069 | 0.071 | 0.036 | 0.078 | 0.58 | 1.44 | -0.004 | 38 | 19 |
| **GATE** | *5* | | -0.069 | 0.123 | 0.079 | 0.167 | 0.73 | 5.73 | -0.014 | 65 | 49 |
| **IATE** | *N* | *ols* | -0.069 | 0.507 | - | 0.761 | - | - | - | 5 | 3 |
| **IATE** | *N* | *rf* | -0.083 | 0.446 | - | 0.772 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | -0.078 | 0.079 | 0.033 | 0.085 | 0.20 | 0.18 | -0.001 | 31 | 14 |
| **GATE** | *5* | *norm* | -0.078 | 0.124 | 0.074 | 0.169 | 0.07 | 0.28 | -0.007 | 67 | 50 |
| **IATE** | *N* | *ols* | -0.078 | 0.505 | - | 0.762 | - | - | - | 5 | 3 |
| **IATE** | *N* | *rf* | -0.094 | 0.448 | - | 0.702 | - | - | - | - | - |
| **ATE** | *1* | *ols* | -0.366 | 0.366 | 0.030 | 0.368 | 0.14 | 0.02 | -0.011 | 0 | 0 |
| **GATE** | *5* | | -0.367 | 0.375 | 0.055 | 0.510 | 0.06 | -0.08 | -0.019 | 25 | 17 |
| **IATE** | *N* | | -0.367 | 0.656 | - | 1.078 | - | - | - | 17 | 11 |

**Note:** For GATE and IATE the results are averaged over all effects. *CovP(95, 80)* denotes the (average) probability that the true value is part of the 95%/80% confidence interval. 1'000 / 500 replications used for 2'500 / 10'000 obs.

Table C.13: Step-function IATE, no selectivity

| | | | | Estimation of effects | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | N = 2'500 | | | | | | |
| **ATE** | *1* | *mcf* | -0.003 | 0.048 | 0.061 | 0.061 | 0.005 | 0.00 | -0.10 | 97 | 83 |
| **GATE** | *5* | | -0.004 | 0.092 | 0.087 | 0.115 | 0.000 | 0.01 | 0.02 | 86 | 67 |
| **IATE** | *N* | | -0.004 | 0.158 | - | 0.193 | - | - | - | 83 | 62 |
| **IATE eff** | *N* | | -0.002 | 0.147 | - | 0.178 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.003 | 0.048 | 0.061 | 0.061 | 0.03 | 0.06 | -0.001 | 94 | 78 |
| **GATE** | *5* | *cent* | 0.002 | 0.090 | 0.091 | 0.114 | 0.02 | 0.09 | -0.010 | 83 | 64 |
| **IATE** | *N* | | 0.002 | 0.162 | - | 0.199 | - | - | - | 77 | 56 |
| **IATE eff** | *N* | | 0.004 | 0.151 | - | 0.184 | - | - | - | - | - |
| **ATE** | *1* | *grf* | -0.001 | 0.035 | 0.043 | 0.043 | -0.13 | -0.04 | -0.000 | 95 | 80 |
| **GATE** | *5* | | -0.001 | 0.076 | 0.096 | 0.096 | 0.06 | 0.09 | -0.001 | 95 | 80 |
| **IATE** | *N* | | -0.001 | 0.158 | - | 0.190 | - | - | - | 76 | 55 |
| **ATE** | *1* | *grf* | 0.003 | 0.033 | 0.042 | 0.042 | 0.07 | -0.02 | 0.000 | 95 | 80 |
| **GATE** | *5* | *cent* | 0.004 | 0.075 | 0.094 | 0.094 | 0.01 | 0.01 | 0.000 | 95 | 80 |
| **IATE** | *N* | | 0.004 | 0.155 | - | 0.187 | - | - | - | 77 | 57 |
| **ATE** | *1* | *dml* | -0.002 | 0.034 | 0.043 | 0.043 | 0.04 | -0.12 | 0.005 | 97 | 85 |
| **GATE** | *5* | | -0.002 | 0.075 | 0.094 | 0.095 | -0.02 | 0.04 | 0.002 | 95 | 81 |
| **IATE** | *N* | *ols* | -0.002 | 0.230 | - | 0.287 | - | - | - | 24 | 16 |
| **IATE** | *N* | *rf* | -0.002 | 0.273 | - | 0.345 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | -0.002 | 0.034 | 0.043 | 0.043 | 0.04 | -0.11 | 0.005 | 97 | 84 |
| **GATE** | *5* | *norm* | -0.002 | 0.075 | 0.094 | 0.094 | -0.02 | 0.05 | 0.002 | 95 | 80 |
| **IATE** | *N* | *ols* | -0.002 | 0.229 | - | 0.286 | - | - | - | 23 | 15 |
| **IATE** | *N* | *rf* | -0.002 | 0.270 | - | 0.342 | - | - | - | - | - |
| **ATE** | *1* | *ols* | -0.001 | 0.034 | 0.042 | 0.042 | 0.04 | -0.11 | -0.013 | 82 | 61 |
| **GATE** | *5* | | -0.002 | 0.073 | 0.092 | 0.093 | -0.04 | 0.09 | -0.027 | 84 | 64 |
| **IATE** | *N* | | -0.001 | 0.228 | - | 0.284 | - | - | - | 68 | 48 |

**Note:** Table to be continued.

Table C.13 - continued: Step-function IATE, no selectivity

| | | | | Estimation of effects | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 10'000 | | | | | |
| **ATE** | *1* | *mcf* | 0.001 | 0.022 | 0.028 | 0.028 | 0.13 | -0.15 | 0.005 | 98 | 84 |
| **GATE** | *5* | | 0.001 | 0.049 | 0.045 | 0.063 | 0.07 | -0.21 | 0.002 | 87 | 66 |
| **IATE** | *N* | | 0.001 | 0.097 | - | 0.121 | - | - | - | 86 | 68 |
| **IATE eff** | *N* | | 0.000 | 0.089 | - | 0.111 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.004 | 0.022 | 0.028 | 0.028 | 0.028 | 0.15 | 0.002 | 97 | 81 |
| **GATE** | *5* | *cent* | 0.004 | 0.045 | 0.045 | 0.058 | 0.045 | 0.17 | -0.002 | 87 | 68 |
| **IATE** | *N* | | 0.004 | 0.096 | - | 0.121 | - | - | - | 83 | 63 |
| **IATE eff** | *N* | | 0.003 | 0.088 | - | 0.110 | - | - | - | - | - |
| **ATE** | *1* | *grf* | -0.000 | 0.016 | 0.020 | 0.020 | -0.17 | 0.39 | 0.001 | 96 | 83 |
| **GATE** | *5* | | -0.001 | 0.037 | 0.047 | 0.047 | -0.09 | 0.13 | -0.001 | 94 | 80 |
| **IATE** | *N* | | -0.001 | 0.083 | - | 0.105 | - | - | - | 91 | 76 |
| **ATE** | *1* | *grf* | 0.002 | 0.017 | 0.021 | 0.021 | 0.01 | 0.02 | 0.000 | 96 | 79 |
| **GATE** | *5* | *cent* | 0.002 | 0.036 | 0.046 | 0.046 | -0.07 | -0.05 | 0.001 | 95 | 80 |
| **IATE** | *N* | | 0.002 | 0.082 | - | 0.103 | - | - | - | 91 | 76 |
| **ATE** | *1* | *dml* | 0.000 | 0.016 | 0.020 | 0.020 | -0.13 | 0.25 | 0.003 | 97 | 86 |
| **GATE** | *5* | | -0.001 | 0.037 | 0.046 | 0.046 | 0.06 | -0.02 | 0.001 | 95 | 82 |
| **IATE** | *N* | *ols* | -0.001 | 0.179 | - | 0.219 | - | - | - | 7 | 5 |
| **IATE** | *N* | *rf* | 0.000 | 0.202 | - | 0.257 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.000 | 0.016 | 0.020 | 0.020 | -0.14 | 0.25 | 0.003 | 98 | 85 |
| **GATE** | *5* | *norm* | -0.001 | 0.037 | 0.046 | 0.046 | 0.05 | -0.03 | 0.001 | 95 | 82 |
| **IATE** | *N* | *ols* | -0.001 | 0.179 | - | 0.219 | - | - | - | 7 | 4 |
| **IATE** | *N* | *rf* | 0.000 | 0.201 | - | 0.256 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.000 | 0.016 | 0.020 | 0.020 | -0.09 | -0.02 | -0.005 | 84 | 66 |
| **GATE** | *5* | | -0.001 | 0.037 | 0.046 | 0.046 | 0.01 | -0.03 | -0.013 | 85 | 62 |
| **IATE** | *N* | | -0.001 | 0.179 | - | 0.219 | - | - | - | 45 | 30 |

**Note:** For GATE and IATE the results are averaged over all effects. *CovP(95, 80)* denotes the (average) probability that the true value is part of the 95%/80% confidence interval. 1'000 / 500 replications used for 2'500 / 10'000 obs.

Table C.14: Step-function IATE, medium selectivity

| | | | | Estimation of effects | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 2'500 | | | | | |
| **ATE** | *1* | *mcf* | 0.175 | 0.175 | 0.062 | 0.186 | 0.11 | 0.13 | 0.002 | 21 | 7 |
| **GATE** | *5* | | 0.175 | 0.178 | 0.090 | 0.211 | 0.06 | -0.04 | 0.008 | 53 | 32 |
| **IATE** | *N* | | 0.175 | 0.214 | - | 0.268 | - | - | - | 67 | 50 |
| **IATE eff** | *N* | | 0.178 | 0.205 | - | 0.258 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.037 | 0.057 | 0.061 | 0.072 | 0.09 | 0.26 | 0.000 | 91 | 72 |
| **GATE** | *5* | *cent* | 0.037 | 0.094 | 0.090 | 0.121 | 0.05 | 0.07 | -0.006 | 84 | 65 |
| **IATE** | *N* | | 0.037 | 0.174 | - | 0.212 | - | - | - | 75 | 53 |
| **IATE eff** | *N* | | 0.040 | 0.163 | - | 0.197 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.097 | 0.097 | 0.044 | 0.107 | -0.03 | -0.24 | -0.002 | 38 | 17 |
| **GATE** | *5* | | 0.097 | 0.113 | 0.096 | 0.137 | 0.03 | -0.00 | -0.003 | 81 | 57 |
| **IATE** | *N* | | 0.092 | 0.188 | - | 0.231 | - | - | - | 70 | 50 |
| **ATE** | *1* | *grf* | 0.037 | 0.046 | 0.042 | 0.056 | -0.00 | -0.20 | -0.000 | 85 | 64 |
| **GATE** | *5* | *cent* | 0.037 | 0.081 | 0.093 | 0.102 | -0.01 | 0.07 | -0.000 | 93 | 76 |
| **IATE** | *N* | | 0.034 | 0.175 | - | 0.210 | - | - | - | 72 | 51 |
| **ATE** | *1* | *dml* | 0.019 | 0.040 | 0.045 | 0.045 | -0.03 | -0.24 | 0.004 | 96 | 79 |
| **GATE** | *5* | | 0.019 | 0.081 | 0.099 | 0.101 | -0.04 | -0.05 | 0.001 | 95 | 80 |
| **IATE** | *N* | *ols* | 0.019 | 0.237 | - | 0.296 | - | - | - | 25 | 16 |
| **IATE** | *N* | *rf* | 0.012 | 0.285 | - | 0.364 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.020 | 0.040 | 0.045 | 0.050 | -0.04 | -0.27 | 0.004 | 95 | 78 |
| **GATE** | *5* | *norm* | 0.020 | 0.081 | 0.081 | 0.101 | -0.04 | -0.05 | 0.001 | 95 | 79 |
| **IATE** | *N* | *ols* | 0.020 | 0.236 | - | 0.295 | - | - | - | 24 | 16 |
| **IATE** | *N* | *rf* | 0.013 | 0.284 | - | 0.362 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.007 | 0.037 | 0.044 | 0.045 | -0.07 | -0.20 | -0.013 | 83 | 63 |
| **GATE** | *5* | | 0.007 | 0.080 | 0.094 | 0.100 | -0.06 | 0.01 | -0.027 | 81 | 61 |
| **IATE** | *N* | | 0.007 | 0.231 | - | 0.287 | - | - | - | 67 | 47 |

**Note:** Table to be continued.

Table C.14 - continued: Step-function IATE, medium selectivity

| | | | | Estimation of effects | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 10'000 | | | | | |
| **ATE** | *1* | *mcf* | 0.144 | 0.144 | 0.029 | 0.147 | 0.16 | -0.02 | 0.003 | 0 | 0 |
| **GATE** | *5* | | 0.144 | 0.144 | 0.048 | 0.158 | 0.16 | -0.03 | -0.001 | 21 | 7 |
| **IATE** | *N* | | 0.144 | 0.158 | - | 0.193 | - | - | - | 45 | 63 |
| **IATE eff** | *N* | | 0.144 | 0.153 | - | 0.186 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.014 | 0.025 | 0.027 | 0.030 | 0.21 | 0.24 | 0.003 | 96 | 82 |
| **GATE** | *5* | *cent* | 0.013 | 0.049 | 0.045 | 0.063 | 0.22 | 0.17 | 0.004 | 86 | 69 |
| **IATE** | *N* | | 0.013 | 0.106 | - | 0.131 | - | - | - | 81 | 59 |
| **IATE eff** | *N* | | 0.013 | 0.100 | - | 0.122 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.064 | 0.064 | 0.020 | 0.067 | -0.18 | -0.41 | 0.001 | 13 | 4 |
| **GATE** | *5* | | 0.064 | 0.068 | 0.047 | 0.080 | -0.03 | -0.15 | -0.001 | 71 | 46 |
| **IATE** | *N* | | 0.063 | 0.096 | - | 0.126 | - | - | - | 86 | 71 |
| **ATE** | *1* | *grf* | 0.015 | 0.022 | 0.022 | 0.027 | -0.08 | 0.05 | -0.001 | 87 | 69 |
| **GATE** | *5* | *cent* | 0.016 | 0.041 | 0.048 | 0.051 | -0.04 | -0.03 | -0.001 | 92 | 75 |
| **IATE** | *N* | | 0.016 | 0.089 | - | 0.111 | - | - | - | 89 | 73 |
| **ATE** | *1* | *dml* | 0.010 | 0.019 | 0.022 | 0.024 | -0.16 | 0.94 | 0.003 | 94 | 82 |
| **GATE** | *5* | | 0.010 | 0.041 | 0.050 | 0.051 | 0.14 | 0.04 | 0.000 | 94 | 80 |
| **IATE** | *N* | *ols* | 0.010 | 0.181 | - | 0.222 | - | - | - | 8 | 5 |
| **IATE** | *N* | *rf* | 0.005 | 0.214 | - | 0.275 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.010 | 0.019 | 0.022 | 0.024 | -0.17 | 0.97 | 0.003 | 94 | 82 |
| **GATE** | *5* | *norm* | 0.010 | 0.041 | 0.050 | 0.051 | 0.13 | 0.04 | 0.000 | 94 | 80 |
| **IATE** | *N* | *ols* | 0.010 | 0.181 | - | 0.222 | - | - | - | 8 | 5 |
| **IATE** | *N* | *rf* | 0.005 | 0.214 | - | 0.274 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.008 | 0.017 | 0.021 | 0.022 | 0.08 | 0.21 | -0.005 | 81 | 65 |
| **GATE** | *5* | | 0.007 | 0.046 | 0.047 | 0.058 | 0.09 | 0.22 | -0.014 | 74 | 53 |
| **IATE** | *N* | | 0.007 | 0.182 | - | 0.222 | - | - | - | 44 | 30 |

**Note:** For GATE and IATE the results are averaged over all effects. *CovP(95, 80)* denotes the (average) probability that the true value is part of the 95%/80% confidence interval. 1'000 / 500 replications used for 2'500 / 10'000 obs.

Table C.15: Step-function IATE, strong selectivity

| | | | | Estimation of effects | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 2'500 | | | | | |
| **ATE** | *1* | *mcf* | 0.437 | 0.437 | -0.001 | 0.441 | 0.03 | -0.02 | -0.001 | 0 | 0 |
| **GATE** | *5* | | 0.436 | 0.436 | 0.001 | 0.455 | 0.00 | 0.06 | 0.001 | 1 | 0 |
| **IATE** | *N* | | 0.436 | 0.438 | - | 0.494 | - | - | - | 32 | 16 |
| **IATE eff** | *N* | | 0.439 | 0.439 | - | 0.489 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.046 | 0.060 | 0.058 | 0.074 | 0.01 | 0.05 | 0.008 | 92 | 72 |
| **GATE** | *5* | *cent* | 0.046 | 0.112 | 0.085 | 0.146 | -0.03 | 0.11 | 0.009 | 81 | 62 |
| **IATE** | *N* | | 0.046 | 0.212 | - | 0.252 | - | - | - | 68 | 45 |
| **IATE eff** | *N* | | 0.048 | 0.205 | - | 0.240 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.284 | 0.284 | 0.051 | 0.288 | -0.06 | -0.06 | -0.010 | 0 | 0 |
| **GATE** | *5* | | 0.284 | 0.284 | 0.101 | 0.309 | -0.06 | 0.06 | -0.011 | 19 | 8 |
| **IATE** | *N* | | 0.272 | 0.315 | - | 0.394 | - | - | - | 49 | 37 |
| **ATE** | *1* | *grf* | 0.090 | 0.091 | 0.046 | 0.101 | 0.04 | -0.04 | -0.006 | 41 | 20 |
| **GATE** | *5* | *cent* | 0.090 | 0.119 | 0.096 | 0.148 | 0.00 | 0.09 | -0.007 | 75 | 56 |
| **IATE** | *N* | | 0.092 | 0.254 | - | 0.297 | - | - | - | 55 | 33 |
| **ATE** | *1* | *dml* | 0.147 | 0.148 | 0.059 | 0.159 | -0.30 | 1.42 | -0.003 | 27 | 11 |
| **GATE** | *5* | | 0.147 | 0.165 | 0.127 | 0.195 | -0.61 | 4.66 | -0.014 | 66 | 46 |
| **IATE** | *N* | *ols* | 0.148 | 0.287 | - | 0.362 | - | - | - | 25 | 17 |
| **IATE** | *N* | *rf* | 0.125 | 0.344 | - | 0.533 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.140 | 0.140 | 0.056 | 0.150 | 0.00 | -0.07 | 0.002 | 23 | 13 |
| **GATE** | *5* | *norm* | 0.140 | 0.157 | 0.123 | 0.187 | -0.13 | 0.01 | -0.005 | 74 | 52 |
| **IATE** | *N* | *ols* | 0.140 | 0.285 | - | 0.357 | - | - | - | 28 | 18 |
| **IATE** | *N* | *rf* | 0.117 | 0.349 | - | 0.461 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.111 | 0.112 | 0.054 | 0.123 | -0.02 | 0.01 | -0.015 | 25 | 13 |
| **GATE** | *5* | | 0.111 | 0.133 | 0.103 | 0.160 | -0.04 | 0.03 | -0.029 | 58 | 39 |
| **IATE** | *N* | | 0.111 | 0.990 | - | 0.314 | - | - | - | 65 | 47 |

**Note:** Table to be continued.

Table C.15 - continued: Step-function IATE, strong selectivity

| | | | | Estimation of effects | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 10'000 | | | | | |
| **ATE** | *1* | *mcf* | 0.375 | 0.375 | 0.031 | 0.377 | -0.06 | 0.70 | 0.001 | 0 | 0 |
| **GATE** | *5* | | 0.375 | 0.375 | 0.053 | 0.384 | -0.10 | -0.10 | -0.002 | 0 | 0 |
| **IATE** | *N* | | 0.375 | 0.375 | - | 0.409 | - | - | - | 18 | 6 |
| **IATE eff** | *N* | | 0.375 | 0.375 | - | 0.407 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | -0.013 | 0.025 | 0.028 | 0.031 | 0.18 | 0.26 | 0.007 | 96 | 86 |
| **GATE** | *5* | *cent* | -0.014 | 0.085 | 0.045 | 0.100 | 0.01 | 0.03 | 0.007 | 65 | 41 |
| **IATE** | *N* | | -0.014 | 0.156 | - | 0.185 | - | - | - | 66 | 43 |
| **IATE eff** | *N* | | -0.014 | 0.151 | - | 0.151 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.211 | 0.211 | 0.025 | 0.213 | -0.31 | -0.09 | -0.004 | 0 | 0 |
| **GATE** | *5* | | 0.211 | 0.211 | 0.052 | 0.220 | 0.00 | -0.04 | -0.005 | 3 | 0 |
| **IATE** | *N* | | 0.205 | 0.208 | - | 0.253 | - | - | - | 59 | 39 |
| **ATE** | *1* | *grf* | 0.042 | 0.044 | 0.027 | 0.050 | -0.00 | -0.18 | -0.006 | 46 | 27 |
| **GATE** | *5* | *cent* | 0.043 | 0.063 | 0.051 | 0.080 | -0.09 | 0.03 | -0.006 | 75 | 56 |
| **IATE** | *N* | | 0.053 | 0.123 | - | 0.162 | - | - | - | 82 | 66 |
| **ATE** | *1* | *dml* | 0.088 | 0.088 | 0.035 | 0.095 | 0.45 | 2.84 | -0.004 | 22 | 8 |
| **GATE** | *5* | | 0.088 | 0.099 | 0.091 | 0.119 | 0.10 | 6.78 | -0.015 | 62 | 42 |
| **IATE** | *N* | *ols* | 0.088 | 0.208 | - | 0.262 | - | - | - | 11 | 7 |
| **IATE** | *N* | *rf* | 0.069 | 0.268 | - | 0.479 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.079 | 0.079 | 0.032 | 0.085 | 0.12 | 0.21 | 0.000 | 28 | 12 |
| **GATE** | *5* | *norm* | 0.078 | 0.090 | 0.073 | 0.108 | -0.08 | 0.04 | -0.007 | 71 | 49 |
| **IATE** | *N* | *ols* | 0.079 | 0.205 | - | 0.256 | - | - | - | 12 | 8 |
| **IATE** | *N* | *rf* | 0.058 | 0.275 | - | 0.418 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.113 | 0.113 | 0.027 | 0.116 | 0.07 | -0.30 | -0.008 | 0 | 0 |
| **GATE** | *5* | | 0.113 | 0.118 | 0.052 | 0.134 | 0.05 | -0.07 | -0.015 | 27 | 17 |
| **IATE** | *N* | | 0.113 | 0.200 | - | 0.253 | - | - | - | 46 | 31 |

**Note:** For GATE and IATE the results are averaged over all effects. *CovP(95, 80)* denotes the (average) probability that the true value is part of the 95%/80% confidence interval. 1'000 / 500 replications used for 2'500 / 10'000 obs.

## Appendix C.2. Extensions to the base specification

So far, we investigate the performance of the estimations when we vary the IATE and the degree of selectivity, as these dimensions are key parameters in any unconfoundedness setting. However, there are many other dimensions for which it is interesting to see if the conclusions concerning estimator behaviour change. Since it is infeasible to vary all of them simultaneously, we chose the setting with medium selectivity, the step function specification of IATE heterogeneity, as well as the other parameters from the previous specification ($k = 10$, $p = 20$, $R^2(y^0) = 10\%$, $X^U$ and $X^N$, 2 treatment groups, a treatment share of 50%, 5 GATEs, $N = 2'500$ and $N = 10'000$). Then, we vary these parameters only individually, instead of varying them jointly. Therefore, the following tables will only show the dimension that deviates from this baseline specification.

Table C.16: Fewer covariates ($p = 10$, $k = p/2$)

| | | | Estimation of effects | | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | N = 2'500 | | | | | | |
| **ATE** | *1* | *mcf* | 0.134 | 0.135 | 0.060 | 0.147 | -0.09 | -0.04 | 0.005 | 44 | 21 |
| **GATE** | *5* | | 0.133 | 0.144 | 0.091 | 0.177 | -0.01 | 0.09 | 0.005 | 70 | 50 |
| **IATE** | *N* | | 0.134 | 0.196 | - | 0.246 | - | - | - | 80 | 62 |
| **IATE eff** | *N* | | 0.136 | 0.181 | - | 0.228 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.006 | 0.049 | 0.060 | 0.061 | -0.09 | 0.01 | 0.002 | 96 | 83 |
| **GATE** | *5* | *cent* | 0.005 | 0.088 | 0.093 | 0.112 | -0.02 | -0.01 | 0.002 | 90 | 73 |
| **IATE** | *N* | | 0.005 | 0.171 | - | 0.211 | - | - | - | 86 | 65 |
| **IATE eff** | *N* | | 0.005 | 0.154 | - | 0.188 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.043 | 0.050 | 0.044 | 0.061 | 0.05 | -0.02 | -0.001 | 82 | 60 |
| **GATE** | *5* | | 0.043 | 0.085 | 0.097 | 0.107 | -0.05 | -0.01 | -0.003 | 92 | 73 |
| **IATE** | *N* | | 0.040 | 0.158 | - | 0.194 | - | - | - | 78 | 58 |
| **ATE** | *1* | *grf* | 0.011 | 0.036 | 0.045 | 0.046 | 0.16 | 0.04 | -0.002 | 93 | 76 |
| **GATE** | *5* | *cent* | 0.012 | 0.079 | 0.098 | 0.099 | 0.03 | 0.12 | -0.003 | 94 | 78 |
| **IATE** | *N* | | 0.013 | 0.157 | - | 0.190 | - | - | - | 78 | 58 |
| **ATE** | *1* | *dml* | 0.000 | 0.039 | 0.049 | 0.049 | -0.08 | 0.14 | 0.001 | 97 | 83 |
| **GATE** | *5* | | 0.000 | 0.084 | 0.105 | 0.105 | 0.00 | -0.02 | 0.003 | 95 | 81 |
| **IATE** | *N* | *ols* | 0.000 | 0.214 | - | 0.266 | - | - | - | 19 | 12 |
| **IATE** | *N* | *rf* | -0.002 | 0.336 | - | 0.433 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.002 | 0.039 | 0.049 | 0.049 | -0.08 | 0.13 | 0.002 | 96 | 81 |
| **GATE** | *5* | *norm* | 0.002 | 0.083 | 0.103 | 0.104 | -0.01 | -0.04 | 0.000 | 95 | 80 |
| **IATE** | *N* | *ols* | 0.002 | 0.212 | - | 0.264 | - | - | - | 17 | 11 |
| **IATE** | *N* | *rf* | 0.000 | 0.329 | - | 0.423 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.009 | 0.037 | 0.046 | 0.047 | -0.02 | 0.06 | -0.015 | 81 | 63 |
| **GATE** | *5* | | 0.008 | 0.086 | 0.096 | 0.107 | 0.01 | 0.14 | -0.029 | 79 | 57 |
| **IATE** | *N* | | 0.008 | 0.207 | - | 0.257 | - | - | - | 60 | 41 |

**Note:** Table to be continued.

Table C.16 - continued: Fewer covariates ($p = 10$, $k = p/2$)

| | | | | Estimation of effects | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 10'000 | | | | | |
| **ATE** | *1* | *mcf* | 0.092 | 0.092 | 0.030 | 0.097 | 0.09 | -0.35 | 0.002 | 18 | 5 |
| **GATE** | *5* | | 0.091 | 0.094 | 0.051 | 0.113 | 0.12 | 0.07 | 0.003 | 62 | 40 |
| **IATE** | *N* | | 0.091 | 0.132 | - | 0.166 | - | - | - | 83 | 66 |
| **IATE eff** | *N* | | 0.093 | 0.120 | - | 0.152 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | -0.009 | 0.026 | 0.030 | 0.032 | 0.10 | -0.29 | 0.001 | 95 | 80 |
| **GATE** | *5* | *cent* | -0.010 | 0.049 | 0.049 | 0.061 | 0.13 | 0.10 | 0.003 | 92 | 72 |
| **IATE** | *N* | | -0.009 | 0.113 | - | 0.140 | - | - | - | 90 | 71 |
| **IATE eff** | *N* | | -0.008 | 0.098 | - | 0.121 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.020 | 0.023 | 0.021 | 0.029 | -0.13 | -0.04 | 0.000 | 82 | 62 |
| **GATE** | *5* | | 0.019 | 0.042 | 0.049 | 0.053 | 0.08 | 0.03 | -0.002 | 92 | 76 |
| **IATE** | *N* | | 0.020 | 0.086 | - | 0.111 | - | - | - | 90 | 75 |
| **ATE** | *1* | *grf* | 0.003 | 0.017 | 0.022 | 0.022 | 0.13 | 0.58 | -0.001 | 93 | 78 |
| **GATE** | *5* | *cent* | 0.003 | 0.040 | 0.049 | 0.050 | 0.02 | 0.12 | -0.002 | 95 | 80 |
| **IATE** | *N* | | 0.007 | 0.086 | - | 0.109 | - | - | - | 90 | 75 |
| **ATE** | *1* | *dml* | -0.003 | 0.019 | 0.023 | 0.023 | 0.00 | -0.45 | 0.003 | 98 | 82 |
| **GATE** | *5* | | -0.003 | 0.041 | 0.051 | 0.051 | 0.10 | 0.04 | 0.001 | 96 | 81 |
| **IATE** | *N* | *ols* | -0.003 | 0.176 | - | 0.214 | - | - | - | 5 | 3 |
| **IATE** | *N* | *rf* | -0.003 | 0.214 | - | 0.340 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | -0.003 | 0.019 | 0.023 | 0.023 | 0.00 | -0.45 | 0.001 | 98 | 82 |
| **GATE** | *5* | *norm* | -0.003 | 0.041 | 0.051 | 0.051 | 0.01 | 0.04 | 0.000 | 95 | 80 |
| **IATE** | *N* | *ols* | -0.003 | 0.176 | - | 0.214 | - | - | - | 5 | 3 |
| **IATE** | *N* | *rf* | -0.003 | 0.263 | - | 0.334 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.004 | 0.017 | 0.021 | 0.021 | -0.01 | -0.23 | -0.006 | 82 | 65 |
| **GATE** | *5* | | 0.003 | 0.053 | 0.048 | 0.065 | 0.09 | 0.13 | -0.014 | 67 | 50 |
| **IATE** | *N* | | 0.003 | 0.178 | - | 0.215 | - | - | - | 35 | 23 |

**Note:** For GATE and IATE the results are averaged over all effects. *CovP(95, 80)* denotes the (average) probability that the true value is part of the 95%/80% confidence interval. 1'000 / 500 replications used for 2'500 / 10'000 obs.

Table C.17: More covariates ($p = 50$, $k = p/2$)

| | | | Estimation of effects | | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | N = 2'500 | | | | | | |
| **ATE** | *1* | *mcf* | 0.223 | 0.223 | 0.061 | 0.231 | -0.02 | 0.05 | 0.002 | 5 | 1 |
| **GATE** | *5* | | 0.223 | 0.224 | 0.085 | 0.253 | -0.04 | 0.22 | -0.001 | 33 | 14 |
| **IATE** | *N* | | 0.223 | 0.244 | - | 0.304 | - | - | - | 53 | 40 |
| **IATE eff** | *N* | | 0.223 | 0.238 | - | 0.297 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.124 | 0.125 | 0.060 | 0.138 | 0.02 | -0.03 | 0.001 | 48 | 22 |
| **GATE** | *5* | *cent* | 0.124 | 0.136 | 0.085 | 0.174 | -0.01 | 0.22 | -0.006 | 66 | 46 |
| **IATE** | *N* | | 0.124 | 0.196 | - | 0.247 | - | - | - | 60 | 43 |
| **IATE eff** | *N* | | 0.125 | 0.189 | - | 0.239 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.178 | 0.178 | 0.043 | 0.183 | -0.01 | 0.06 | 0.000 | 2 | 0 |
| **GATE** | *5* | | 0.178 | 0.180 | 0.095 | 0.202 | -0.08 | 0.10 | 0.000 | 53 | 27 |
| **IATE** | *N* | | 0.172 | 0.252 | - | 0.308 | - | - | - | 52 | 35 |
| **ATE** | *1* | *grf* | 0.107 | 0.107 | 0.042 | 0.115 | 0.01 | -0.03 | 0.001 | 29 | 11 |
| **GATE** | *5* | *cent* | 0.107 | 0.119 | 0.093 | 0.142 | 0.03 | -0.06 | 0.000 | 79 | 55 |
| **IATE** | *N* | | 0.102 | 0.231 | - | 0.271 | - | - | - | 53 | 32 |
| **ATE** | *1* | *dml* | 0.092 | 0.093 | 0.045 | 0.103 | -0.04 | -0.07 | 0.003 | 52 | 25 |
| **GATE** | *5* | | 0.093 | 0.111 | 0.098 | 0.135 | -0.03 | 0.09 | -0.003 | 83 | 61 |
| **IATE** | *N* | *ols* | 0.093 | 0.308 | - | 0.386 | - | - | - | 38 | 26 |
| **IATE** | *N* | *rf* | 0.084 | 0.250 | - | 0.314 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.092 | 0.092 | 0.046 | 0.102 | -0.06 | -0.06 | 0.003 | 53 | 25 |
| **GATE** | *5* | *norm* | 0.092 | 0.111 | 0.099 | 0.135 | -0.02 | 0.09 | -0.002 | 84 | 62 |
| **IATE** | *N* | *ols* | 0.092 | 0.310 | - | 0.389 | - | - | - | 39 | 26 |
| **IATE** | *N* | *rf* | 0.083 | 0.251 | - | 0.316 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.012 | 0.039 | 0.048 | 0.049 | -0.09 | -0.02 | -0.016 | 77 | 58 |
| **GATE** | *5* | | 0.012 | 0.080 | 0.099 | 0.101 | 0.00 | 0.09 | -0.032 | 81 | 62 |
| **IATE** | *N* | | 0.012 | 0.295 | - | 0.369 | - | - | - | 75 | 55 |

**Note:** Table to be continued.

Table C.17 - continued: More covariates ($p = 50$, $k = p/2$)

| | | | | Estimation of effects | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 10'000 | | | | | |
| **ATE** | *1* | *mcf* | 0.201 | 0.201 | 0.031 | 0.204 | 0.01 | -0.07 | 0.001 | 0 | 0 |
| **GATE** | *5* | | 0.202 | 0.202 | 0.047 | 0.213 | 0.05 | -0.08 | -0.002 | 1 | 0 |
| **IATE** | *N* | | 0.202 | 0.205 | - | 0.240 | - | - | - | 38 | 23 |
| **IATE eff** | *N* | | 0.202 | 0.204 | - | 0.235 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.093 | 0.093 | 0.029 | 0.097 | 0.14 | 0.12 | 0.001 | 11 | 3 |
| **GATE** | *5* | *cent* | 0.093 | 0.094 | 0.045 | 0.116 | 0.11 | -0.03 | -0.003 | 50 | 31 |
| **IATE** | *N* | | 0.093 | 0.130 | - | 0.164 | - | - | - | 59 | 43 |
| **IATE eff** | *N* | | 0.093 | 0.164 | - | 0.159 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.144 | 0.144 | 0.020 | 0.145 | -0.17 | 0.29 | 0.001 | 0 | 0 |
| **GATE** | *5* | | 0.144 | 0.144 | 0.046 | 0.152 | 0.05 | -0.09 | 0.001 | 13 | 3 |
| **IATE** | *N* | | 0.141 | 0.148 | - | 0.188 | - | - | - | 66 | 50 |
| **ATE** | *1* | *grf* | 0.073 | 0.073 | 0.020 | 0.076 | -0.10 | 0.06 | 0.000 | 6 | 2 |
| **GATE** | *5* | *cent* | 0.072 | 0.074 | 0.045 | 0.085 | 0.07 | -0.24 | 0.001 | 66 | 39 |
| **IATE** | *N* | | 0.068 | 0.111 | - | 0.142 | - | - | - | 80 | 63 |
| **ATE** | *1* | *dml* | 0.070 | 0.070 | 0.021 | 0.073 | 0.10 | 0.03 | 0.003 | 14 | 4 |
| **GATE** | *5* | | 0.070 | 0.073 | 0.049 | 0.085 | -0.03 | 0.00 | -0.001 | 66 | 42 |
| **IATE** | *N* | *ols* | 0.070 | 0.208 | - | 0.259 | - | - | - | 14 | 9 |
| **IATE** | *N* | *rf* | 0.062 | 0.183 | - | 0.232 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.068 | 0.068 | 0.021 | 0.071 | 0.11 | 0.02 | 0.003 | 16 | 4 |
| **GATE** | *5* | *norm* | 0.069 | 0.072 | 0.049 | 0.084 | -0.03 | 0.01 | -0.001 | 69 | 44 |
| **IATE** | *N* | *ols* | 0.068 | 0.208 | - | 0.260 | - | - | - | 15 | 10 |
| **IATE** | *N* | *rf* | 0.055 | 0.185 | - | 0.234 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.011 | 0.020 | 0.022 | 0.024 | 0.03 | -0.19 | -0.006 | 79 | 56 |
| **GATE** | *5* | | 0.011 | 0.041 | 0.041 | 0.051 | -0.02 | 0.09 | -0.014 | 79 | 60 |
| **IATE** | *N* | | 0.011 | 0.200 | - | 0.248 | - | - | - | 59 | 41 |

**Note:** For GATE and IATE the results are averaged over all effects. *CovP(95, 80)* denotes the (average) probability that the true value is part of the 95%/80% confidence interval. 1'000 / 500 replications used for 2'500 / 10'000 obs.

Table C.18: Sparser model ($p = 20$, $k = 4$)

| | | | Estimation of effects | | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | N = 2'500 | | | | | | |
| **ATE** | *1* | *mcf* | 0.132 | 0.133 | 0.061 | 0.146 | 0.04 | 0.01 | 0.003 | 47 | 22 |
| **GATE** | *5* | | 0.132 | 0.142 | 0.091 | 0.175 | 0.01 | -0.07 | 0.001 | 69 | 49 |
| **IATE** | *N* | | 0.132 | 0.189 | - | 0.237 | - | - | - | 74 | 57 |
| **IATE eff** | *N* | | 0.135 | 0.189 | - | 0.237 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.011 | 0.050 | 0.063 | 0.063 | 0.07 | 0.08 | 0.000 | 95 | 79 |
| **GATE** | *5* | *cent* | 0.010 | 0.091 | 0.093 | 0.115 | 0.05 | -0.02 | -0.004 | 87 | 68 |
| **IATE** | *N* | | 0.010 | 0.172 | - | 0.209 | - | - | - | 77 | 55 |
| **IATE eff** | *N* | | 0.013 | 0.160 | - | 0.192 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.043 | 0.051 | 0.044 | 0.062 | -0.09 | -0.03 | -0.001 | 80 | 58 |
| **GATE** | *5* | | 0.043 | 0.085 | 0.096 | 0.107 | 0.07 | 0.06 | -0.003 | 91 | 74 |
| **IATE** | *N* | | 0.040 | 0.184 | - | 0.220 | - | - | - | 69 | 48 |
| **ATE** | *1* | *grf* | 0.022 | 0.040 | 0.044 | 0.049 | 0.03 | 0.04 | -0.002 | 91 | 73 |
| **GATE** | *5* | *cent* | 0.022 | 0.079 | 0.095 | 0.099 | 0.00 | -0.01 | -0.001 | 94 | 77 |
| **IATE** | *N* | | 0.021 | 0.180 | - | 0.214 | - | - | - | 70 | 49 |
| **ATE** | *1* | *dml* | 0.000 | 0.036 | 0.045 | 0.045 | -0.01 | -0.14 | 0.005 | 97 | 85 |
| **GATE** | *5* | | 0.000 | 0.082 | 0.102 | 0.102 | -0.10 | -0.04 | 0.001 | 95 | 80 |
| **IATE** | *N* | *ols* | 0.000 | 0.240 | - | 0.299 | - | - | - | 28 | 17 |
| **IATE** | *N* | *rf* | -0.004 | 0.292 | - | 0.375 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.002 | 0.036 | 0.045 | 0.045 | -0.01 | -0.14 | 0.004 | 97 | 84 |
| **GATE** | *5* | *norm* | 0.002 | 0.081 | 0.101 | 0.102 | -0.10 | -0.02 | 0.000 | 95 | 80 |
| **IATE** | *N* | *ols* | 0.002 | 0.238 | - | 0.297 | - | - | - | 25 | 16 |
| **IATE** | *N* | *rf* | -0.003 | 0.288 | - | 0.370 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.002 | 0.030 | 0.044 | 0.044 | -0.06 | -0.14 | -0.012 | 84 | 63 |
| **GATE** | *5* | | 0.002 | 0.086 | 0.096 | 0.107 | -0.03 | -0.01 | -0.028 | 78 | 58 |
| **IATE** | *N* | | 0.002 | 0.232 | - | 0.289 | - | - | - | 67 | 47 |

**Note:** Table to be continued.

Table C.18 - continued: Sparser model ($p = 20$, $k = 4$)

| | | | | Estimation of effects | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 10'000 | | | | | |
| **ATE** | *1* | *mcf* | 0.097 | 0.097 | 0.029 | 0.102 | 0.16 | -0.31 | 0.003 | 14 | 2 |
| **GATE** | *5* | | 0.097 | 0.099 | 0.050 | 0.119 | 0.08 | -0.27 | 0.001 | 56 | 36 |
| **IATE** | *N* | | 0.097 | 0.127 | - | 0.161 | - | - | - | 75 | 57 |
| **IATE eff** | *N* | | 0.097 | 0.119 | - | 0.151 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | -0.002 | 0.023 | 0.029 | 0.029 | 0.07 | -0.19 | 0.003 | 98 | 84 |
| **GATE** | *5* | *cent* | -0.002 | 0.049 | 0.049 | 0.062 | 0.12 | -0.13 | 0.002 | 89 | 71 |
| **IATE** | *N* | | -0.003 | 0.106 | - | 0.131 | - | - | - | 83 | 61 |
| **IATE eff** | *N* | | -0.002 | 0.098 | - | 0.119 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.022 | 0.026 | 0.021 | 0.031 | -0.15 | -0.25 | -0.000 | 81 | 57 |
| **GATE** | *5* | | 0.022 | 0.043 | 0.049 | 0.054 | 0.05 | -0.07 | -0.002 | 92 | 73 |
| **IATE** | *N* | | 0.023 | 0.091 | - | 0.118 | - | - | - | 87 | 72 |
| **ATE** | *1* | *grf* | 0.007 | 0.019 | 0.023 | 0.024 | 0.19 | 0.07 | -0.002 | 91 | 74 |
| **GATE** | *5* | *cent* | 0.008 | 0.040 | 0.048 | 0.050 | 0.01 | -0.05 | -0.001 | 93 | 77 |
| **IATE** | *N* | | 0.010 | 0.091 | - | 0.114 | - | - | - | 88 | 71 |
| **ATE** | *1* | *dml* | 0.001 | 0.017 | 0.022 | 0.022 | -0.14 | 0.63 | 0.003 | 96 | 86 |
| **GATE** | *5* | | 0.000 | 0.040 | 0.050 | 0.051 | 0.09 | 0.06 | 0.001 | 95 | 80 |
| **IATE** | *N* | *ols* | 0.000 | 0.182 | - | 0.223 | - | - | - | 8 | 5 |
| **IATE** | *N* | *rf* | -0.002 | 0.218 | - | 0.282 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.001 | 0.017 | 0.022 | 0.022 | -0.15 | 0.62 | 0.003 | 96 | 87 |
| **GATE** | *5* | *norm* | 0.001 | 0.040 | 0.050 | 0.050 | 0.09 | 0.06 | 0.000 | 95 | 79 |
| **IATE** | *N* | *ols* | 0.001 | 0.182 | - | 0.223 | - | - | - | 8 | 5 |
| **IATE** | *N* | *rf* | -0.002 | 0.216 | - | 0.279 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.004 | 0.017 | 0.021 | 0.022 | -0.09 | 0.30 | -0.006 | 84 | 66 |
| **GATE** | *5* | | 0.004 | 0.056 | 0.052 | 0.068 | -0.01 | 0.08 | -0.014 | 64 | 44 |
| **IATE** | *N* | | 0.004 | 0.184 | - | 0.225 | - | - | - | 44 | 29 |

**Note:** For GATE and IATE the results are averaged over all effects. *CovP(95, 80)* denotes the (average) probability that the true value is part of the 95%/80% confidence interval. 1'000 / 500 replications used for 2'500 / 10'000 obs.

Table C.19: Less sparse model ($p = 20$, $k = 16$)

| | | | Estimation of effects | | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 2'500 | | | | | |
| **ATE** | *1* | *mcf* | 0.192 | 0.192 | 0.061 | 0.201 | 0.08 | 0.23 | 0.002 | 14 | 4 |
| **GATE** | *5* | | 0.191 | 0.193 | 0.089 | 0.224 | 0.07 | 0.17 | -0.003 | 47 | 25 |
| **IATE** | *N* | | 0.191 | 0.225 | - | 0.279 | - | - | - | 64 | 48 |
| **IATE eff** | *N* | | 0.194 | 0.216 | - | 0.270 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.053 | 0.065 | 0.060 | 0.080 | 0.05 | 0.43 | 0.001 | 87 | 65 |
| **GATE** | *5* | *cent* | 0.052 | 0.096 | 0.088 | 0.126 | 0.05 | 0.29 | -0.005 | 82 | 65 |
| **IATE** | *N* | | 0.052 | 0.174 | - | 0.214 | - | - | - | 75 | 53 |
| **IATE eff** | *N* | | 0.055 | 0.164 | - | 0.200 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.125 | 0.125 | 0.044 | 0.133 | -0.01 | -0.09 | -0.002 | 17 | 5 |
| **GATE** | *5* | | 0.125 | 0.134 | 0.097 | 0.159 | 0.00 | -0.03 | -0.003 | 72 | 48 |
| **IATE** | *N* | | 0.119 | 0.193 | - | 0.240 | - | - | - | 70 | 51 |
| **ATE** | *1* | *grf* | 0.048 | 0.054 | 0.042 | 0.064 | 0.04 | -0.18 | -0.000 | 79 | 55 |
| **GATE** | *5* | *cent* | 0.047 | 0.085 | 0.093 | 0.106 | -0.01 | 0.09 | -0.001 | 92 | 74 |
| **IATE** | *N* | | 0.044 | 0.174 | - | 0.209 | - | - | - | 73 | 52 |
| **ATE** | *1* | *dml* | 0.032 | 0.044 | 0.044 | 0.055 | -0.02 | -0.15 | 0.005 | 93 | 73 |
| **GATE** | *5* | | 0.032 | 0.083 | 0.098 | 0.103 | -0.02 | -0.05 | 0.000 | 93 | 79 |
| **IATE** | *N* | *ols* | 0.032 | 0.236 | - | 0.295 | - | - | - | 24 | 16 |
| **IATE** | *N* | *rf* | 0.024 | 0.282 | - | 0.359 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.032 | 0.044 | 0.044 | 0.055 | -0.02 | -0.17 | 0.004 | 92 | 73 |
| **GATE** | *5* | *norm* | 0.032 | 0.083 | 0.099 | 0.104 | -0.01 | -0.06 | 0.000 | 93 | 78 |
| **IATE** | *N* | *ols* | 0.032 | 0.236 | - | 0.295 | - | - | - | 24 | 16 |
| **IATE** | *N* | *rf* | 0.024 | 0.283 | - | 0.359 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.008 | 0.036 | 0.044 | 0.045 | -0.07 | -0.19 | -0.013 | 84 | 61 |
| **GATE** | *5* | | 0.008 | 0.078 | 0.094 | 0.097 | -0.05 | 0.03 | -0.027 | 82 | 62 |
| **IATE** | *N* | | 0.008 | 0.230 | - | 0.286 | - | - | - | 67 | 48 |

**Note:** Table to be continued.

Table C.19 - continued: Less sparse model ($p = 20$, $k = 16$)

| | | | | **Estimation of effects** | | | | | **Estimation of std. errors** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 10'000 | | | | | |
| **ATE** | *1* | *mcf* | 0.168 | 0.168 | 0.029 | 0.170 | 0.06 | 0.07 | 0.003 | 0 | 0 |
| **GATE** | *5* | | 0.167 | 0.167 | 0.047 | 0.179 | 0.15 | 0.04 | -0.001 | 7 | 2 |
| **IATE** | *N* | | 0.167 | 0.176 | - | 0.210 | - | - | - | 57 | 39 |
| **IATE eff** | *N* | | 0.167 | 0.172 | - | 0.203 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.029 | 0.032 | 0.026 | 0.039 | -0.05 | 0.10 | 0.004 | 90 | 65 |
| **GATE** | *5* | *cent* | 0.028 | 0.051 | 0.044 | 0.069 | 0.15 | 0.12 | 0.001 | 84 | 67 |
| **IATE** | *N* | | 0.028 | 0.107 | - | 0.133 | - | - | - | 80 | 59 |
| **IATE eff** | *N* | | 0.028 | 0.100 | - | 0.124 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.093 | 0.093 | 0.020 | 0.095 | -0.15 | -0.44 | 0.001 | 0 | 0 |
| **GATE** | *5* | | 0.092 | 0.093 | 0.047 | 0.104 | 0.04 | -0.11 | -0.001 | 49 | 25 |
| **IATE** | *N* | | 0.089 | 0.109 | - | 0.141 | - | - | - | 82 | 66 |
| **ATE** | *1* | *grf* | 0.024 | 0.027 | 0.021 | 0.032 | -0.05 | 0.06 | 0.000 | 80 | 51 |
| **GATE** | *5* | *cent* | 0.025 | 0.043 | 0.046 | 0.054 | -0.04 | -0.07 | -0.000 | 90 | 72 |
| **IATE** | *N* | | 0.024 | 0.089 | - | 0.112 | - | - | - | 89 | 73 |
| **ATE** | *1* | *dml* | 0.020 | 0.025 | 0.22 | 0.030 | -0.31 | 1.16 | 0.002 | 89 | 68 |
| **GATE** | *5* | | 0.020 | 0.042 | 0.49 | 0.053 | 0.07 | -0.03 | 0.000 | 93 | 77 |
| **IATE** | *N* | *ols* | 0.020 | 0.181 | - | 0.222 | - | - | - | 7 | 5 |
| **IATE** | *N* | *rf* | 0.014 | 0.212 | - | 0.271 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.020 | 0.025 | 0.22 | 0.030 | -0.33 | 1.20 | 0.002 | 89 | 70 |
| **GATE** | *5* | *norm* | 0.020 | 0.042 | 0.49 | 0.053 | 0.07 | -0.03 | 0.000 | 93 | 77 |
| **IATE** | *N* | *ols* | 0.020 | 0.181 | - | 0.222 | - | - | - | 7 | 5 |
| **IATE** | *N* | *rf* | 0.014 | 0.212 | - | 0.272 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.009 | 0.018 | 0.021 | 0.023 | -0.11 | 0.33 | -0.006 | 79 | 64 |
| **GATE** | *5* | | 0.009 | 0.043 | 0.047 | 0.053 | 0.05 | 0.07 | -0.014 | 77 | 57 |
| **IATE** | *N* | | 0.009 | 0.181 | - | 0.221 | - | - | - | 45 | 30 |

**Note:** For GATE and IATE the results are averaged over all effects. *CovP(95, 80)* denotes the (average) probability that the true value is part of the 95%/80% confidence interval. 1'000 / 500 replications used for 2'500 / 10'000 obs.

Table C.20: Covariates irrelevant for $Y^0$ ($R^2(y^0) = 0\%$)

| | | | Estimation of effects | | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | N = 2'500 | | | | | | |
| **ATE** | *1* | *mcf* | 0.024 | 0.050 | 0.059 | 0.064 | 0.10 | 0.07 | 0.00 | 93 | 76 |
| **GATE** | *5* | | 0.023 | 0.089 | 0.087 | 0.115 | 0.05 | 0.04 | -0.003 | 86 | 68 |
| **IATE** | *N* | | 0.023 | 0.168 | - | 0.206 | - | - | - | 79 | 58 |
| **IATE eff** | *N* | | 0.026 | 0.157 | - | 0.192 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | -0.034 | 0.056 | 0.060 | 0.069 | 0.059 | 0.184 | 0.000 | 92 | 73 |
| **GATE** | *5* | *cent* | -0.034 | 0.098 | 0.089 | 0.120 | 0.044 | 0.124 | -0.005 | 81 | 61 |
| **IATE** | *N* | | -0.034 | 0.176 | - | 0.213 | - | - | - | 75 | 53 |
| **IATE eff** | *N* | | -0.031 | 0.165 | - | 0.197 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.013 | 0.036 | 0.043 | 0.045 | -0.02 | -0.27 | -0.002 | 93 | 73 |
| **GATE** | *5* | | 0.013 | 0.077 | 0.094 | 0.096 | 0.06 | 0.03 | -0.003 | 93 | 78 |
| **IATE** | *N* | | 0.009 | 0.161 | - | 0.197 | - | - | - | 75 | 56 |
| **ATE** | *1* | *grf* | -0.011 | 0.035 | 0.042 | 0.044 | 0.03 | -0.22 | -0.001 | 94 | 78 |
| **GATE** | *5* | *cent* | -0.010 | 0.075 | 0.093 | 0.094 | -0.01 | 0.06 | -0.001 | 94 | 78 |
| **IATE** | *N* | | -0.012 | 0.164 | - | 0.198 | - | - | - | 74 | 55 |
| **ATE** | *1* | *dml* | -0.002 | 0.037 | 0.045 | 0.045 | -0.04 | -0.23 | 0.004 | 97 | 83 |
| **GATE** | *5* | | -0.002 | 0.078 | 0.098 | 0.098 | -0.04 | -0.06 | 0.000 | 95 | 80 |
| **IATE** | *N* | *ols* | -0.001 | 0.247 | - | 0.307 | - | - | - | 22 | 15 |
| **IATE** | *N* | *rf* | -0.002 | 0.283 | - | 0.362 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | -0.001 | 0.037 | 0.045 | 0.045 | -0.04 | -0.26 | 0.003 | 97 | 83 |
| **GATE** | *5* | *norm* | -0.001 | 0.078 | 0.098 | 0.098 | -0.04 | -0.06 | 0.000 | 95 | 80 |
| **IATE** | *N* | *ols* | -0.001 | 0.246 | - | 0.307 | - | - | - | 22 | 15 |
| **IATE** | *N* | *rf* | -0.001 | 0.282 | - | 0.360 | - | - | - | - | - |
| **ATE** | *1* | *ols* | -0.002 | 0.036 | 0.044 | 0.044 | -0.06 | -0.22 | -0.013 | 83 | 62 |
| **GATE** | *5* | | -0.002 | 0.076 | 0.093 | 0.094 | -0.07 | 0.01 | -0.027 | 84 | 63 |
| **IATE** | *N* | | -0.002 | 0.241 | - | 0.299 | - | - | - | 64 | 45 |

**Note:** Table to be continued.

Table C.20 - continued: Covariates irrelevant for $Y^0$ ($R^2(y^0) = 0\%$)

| | | | | **Estimation of effects** | | | | | | **Estimation of std. errors** | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 10'000 | | | | | |
| **ATE** | *1* | *mcf* | 0.016 | 0.026 | 0.027 | 0.032 | 0.27 | 0.04 | 0.003 | 94 | 79 |
| **GATE** | *5* | | 0.015 | 0.046 | 0.044 | 0.061 | 0.25 | 0.20 | 0.002 | 89 | 72 |
| **IATE** | *N* | | 0.015 | 0.101 | - | 0.128 | - | - | - | 84 | 72 |
| **IATE eff** | *N* | | 0.015 | 0.093 | - | 0.117 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | -0.035 | 0.038 | 0.03 | 0.044 | 0.24 | 0.24 | 0.004 | 81 | 57 |
| **GATE** | *5* | *cent* | -0.036 | 0.062 | 0.04 | 0.073 | 0.21 | 0.19 | 0.002 | 76 | 50 |
| **IATE** | *N* | | -0.036 | 0.111 | - | 0.137 | - | - | - | 79 | 57 |
| **IATE eff** | *N* | | -0.036 | 0.104 | - | 0.127 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.006 | 0.017 | 0.020 | 0.021 | -0.13 | -0.35 | 0.001 | 95 | 77 |
| **GATE** | *5* | | 0.006 | 0.038 | 0.047 | 0.047 | -0.01 | -0.04 | -0.001 | 95 | 78 |
| **IATE** | *N* | | 0.005 | 0.089 | - | 0.114 | - | - | - | 89 | 73 |
| **ATE** | *1* | *grf* | -0.013 | 0.020 | 0.022 | 0.026 | -0.16 | 0.31 | -0.001 | 90 | 67 |
| **GATE** | *5* | *cent* | -0.013 | 0.040 | 0.047 | 0.050 | -0.06 | 0.01 | -0.001 | 92 | 77 |
| **IATE** | *N* | | -0.012 | 0.095 | - | 0.117 | - | - | - | 87 | 69 |
| **ATE** | *1* | *dml* | -0.001 | 0.016 | 0.022 | 0.022 | -0.25 | 1.08 | 0.003 | 96 | 84 |
| **GATE** | *5* | | -0.001 | 0.039 | 0.049 | 0.050 | 0.09 | 0.01 | -0.001 | 94 | 80 |
| **IATE** | *N* | *ols* | -0.001 | 0.197 | - | 0.241 | - | - | - | 7 | 4 |
| **IATE** | *N* | *rf* | -0.001 | 0.213 | - | 0.273 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | -0.001 | 0.017 | 0.022 | 0.022 | -0.28 | 1.13 | 0.002 | 96 | 85 |
| **GATE** | *5* | *norm* | -0.001 | 0.039 | 0.049 | 0.050 | 0.09 | 0.01 | -0.001 | 95 | 80 |
| **IATE** | *N* | *ols* | -0.001 | 0.197 | - | 0.240 | - | - | - | 7 | 4 |
| **IATE** | *N* | *rf* | -0.001 | 0.212 | - | 0.273 | - | - | - | - | - |
| **ATE** | *1* | *ols* | -0.001 | 0.016 | 0.021 | 0.021 | 0.05 | 0.13 | -0.006 | 84 | 66 |
| **GATE** | *5* | | -0.001 | 0.039 | 0.047 | 0.049 | 0.07 | 0.20 | -0.014 | 82 | 62 |
| **IATE** | *N* | | -0.001 | 0.049 | - | 0.196 | - | - | - | 40 | 27 |

**Note:** For GATE and IATE the results are averaged over all effects. *CovP(95, 80)* denotes the (average) probability that the true value is part of the 95%/80% confidence interval. 1'000 / 500 replications used for 2'500 / 10'000 obs.

Table C.21: Covariates very important for $Y^0$ ($R^2(y^0) = 45\%$)

| | | | | Estimation of effects | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 2'500 | | | | | |
| **ATE** | *1* | *mcf* | 0.421 | 0.421 | 0.074 | 0.427 | 0.09 | 0.17 | 0.016 | 0 | 0 |
| **GATE** | *5* | | 0.420 | 0.420 | 0.111 | 0.447 | 0.06 | -0.07 | 0.007 | 9 | 2 |
| **IATE** | *N* | | 0.420 | 0.427 | - | 0.497 | - | - | - | 45 | 31 |
| **IATE eff** | *N* | | 0.422 | 0.427 | - | 0.488 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.165 | 0.165 | 0.068 | 0.178 | 0.13 | 0.31 | 0.002 | 33 | 12 |
| **GATE** | *5* | *cent* | 0.164 | 0.173 | 0.104 | 0.215 | 0.09 | 0.06 | -0.007 | 63 | 42 |
| **IATE** | *N* | | 0.164 | 0.248 | - | 0.308 | - | - | - | 65 | 47 |
| **IATE eff** | *N* | | 0.164 | 0.237 | - | 0.295 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.228 | 0.228 | 0.050 | 0.234 | -0.02 | -0.20 | -0.000 | 1 | 0 |
| **GATE** | *5* | | 0.228 | 0.230 | 0.110 | 0.254 | -0.01 | -0.03 | -0.000 | 45 | 22 |
| **IATE** | *N* | | 0.218 | 0.253 | - | 0.323 | - | - | - | 69 | 50 |
| **ATE** | *1* | *grf* | 0.117 | 0.117 | 0.047 | 0.126 | 0.01 | -0.06 | -0.000 | 29 | 12 |
| **GATE** | *5* | *cent* | 0.117 | 0.131 | 0.102 | 0.158 | 0.02 | 0.02 | -0.000 | 78 | 54 |
| **IATE** | *N* | | 0.109 | 0.193 | - | 0.246 | - | - | - | 77 | 58 |
| **ATE** | *1* | *dml* | 0.061 | 0.067 | 0.050 | 0.079 | -0.04 | -0.20 | 0.006 | 84 | 57 |
| **GATE** | *5* | | 0.061 | 0.101 | 0.110 | 0.126 | -0.02 | 0.00 | 0.004 | 92 | 74 |
| **IATE** | *N* | *ols* | 0.062 | 0.318 | - | 0.395 | - | - | - | 23 | 15 |
| **IATE** | *N* | *rf* | 0.041 | 0.340 | - | 0.434 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.064 | 0.068 | 0.050 | 0.081 | -0.05 | -0.20 | 0.005 | 82 | 54 |
| **GATE** | *5* | *norm* | 0.063 | 0.102 | 0.110 | 0.127 | -0.03 | 0.00 | 0.003 | 92 | 74 |
| **IATE** | *N* | *ols* | 0.064 | 0.317 | - | 0.395 | - | - | - | 23 | 15 |
| **IATE** | *N* | *rf* | 0.044 | 0.339 | - | 0.432 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.022 | 0.044 | 0.049 | 0.054 | -0.09 | -0.14 | -0.015 | 78 | 56 |
| **GATE** | *5* | | 0.022 | 0.105 | 0.103 | 0.129 | -0.04 | -0.02 | -0.029 | 73 | 51 |
| **IATE** | *N* | | 0.022 | 0.314 | - | 0.388 | - | - | - | 56 | 38 |

**Note:** Table to be continued.

Table C.21 - continued: Covariates very important for $Y^0$ ($R^2(y^0) = 45\%$)

| | | | | Estimation of effects | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | N = 10'000 | | | | | | |
| **ATE** | *1* | *mcf* | 0.346 | 0.346 | 0.034 | 0.347 | -0.09 | -0.12 | 0.011 | 0 | 0 |
| **GATE** | *5* | | 0.345 | 0.345 | 0.059 | 0.356 | 0.04 | -0.12 | 0.004 | 0 | 0 |
| **IATE** | *N* | | 0.344 | 0.347 | - | 0.387 | - | - | - | 35 | 20 |
| **IATE eff** | *N* | | 0.345 | 0.346 | - | 0.381 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.104 | 0.104 | 0.029 | 0.108 | 0.08 | -0.05 | 0.005 | 10 | 2 |
| **GATE** | *5* | *cent* | 0.103 | 0.105 | 0.050 | 0.126 | 0.17 | 0.01 | 0.001 | 53 | 32 |
| **IATE** | *N* | | 0.103 | 0.158 | - | 0.194 | - | - | - | 67 | 48 |
| **IATE eff** | *N* | | 0.103 | 0.153 | - | 0.187 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.152 | 0.152 | 0.022 | 0.153 | -0.20 | -0.51 | 0.002 | 0 | 0 |
| **GATE** | *5* | | 0.151 | 0.151 | 0.052 | 0.161 | -0.05 | -0.18 | 0.001 | 20 | 6 |
| **IATE** | *N* | | 0.147 | 0.162 | - | 0.200 | - | - | - | 72 | 53 |
| **ATE** | *1* | *grf* | 0.060 | 0.060 | 0.024 | 0.064 | -0.00 | -0.20 | -0.001 | 29 | 9 |
| **GATE** | *5* | *cent* | 0.061 | 0.068 | 0.051 | 0.081 | -0.04 | -0.02 | -0.001 | 74 | 50 |
| **IATE** | *N* | | 0.061 | 0.118 | - | 0.146 | - | - | - | 83 | 64 |
| **ATE** | *1* | *dml* | 0.036 | 0.037 | 0.023 | 0.043 | -0.14 | 0.80 | 0.004 | 77 | 46 |
| **GATE** | *5* | | 0.036 | 0.052 | 0.053 | 0.065 | 0.20 | 0.10 | 0.001 | 90 | 73 |
| **IATE** | *N* | *ols* | 0.036 | 0.267 | - | 0.325 | - | - | - | 6 | 4 |
| **IATE** | *N* | *rf* | 0.021 | 0.325 | - | 0.320 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.037 | 0.038 | 0.023 | 0.043 | -0.14 | 0.81 | 0.004 | 77 | 44 |
| **GATE** | *5* | *norm* | 0.036 | 0.052 | 0.053 | 0.065 | 0.19 | 0.09 | 0.001 | 89 | 72 |
| **IATE** | *N* | *ols* | 0.036 | 0.267 | - | 0.325 | - | - | - | 6 | 4 |
| **IATE** | *N* | *rf* | 0.022 | 0.319 | - | 0.319 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.023 | 0.027 | 0.022 | 0.032 | 0.21 | 0.34 | -0.005 | 69 | 49 |
| **GATE** | *5* | | 0.022 | 0.080 | 0.051 | 0.095 | 0.11 | 0.28 | 0.109 | 48 | 30 |
| **IATE** | *N* | | 0.022 | 0.273 | - | 0.332 | - | - | - | 32 | 21 |

**Note:** For GATE and IATE the results are averaged over all effects. *CovP(95, 80)* denotes the (average) probability that the true value is part of the 95%/80% confidence interval. 1'000 / 500 replications used for 2'500 / 10'000 obs.

Table C.22: Uniformly distributed covariates ($X^U$)

| | | | | Estimation of effects | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 2'500 | | | | | |
| **ATE** | *1* | *mcf* | 0.179 | 0.179 | 0.058 | 0.188 | 0.115 | -0.24 | 0.006 | 19 | 5 |
| **GATE** | *5* | | 0.178 | 0.180 | 0.088 | 0.212 | 0.030 | -0.12 | 0.000 | 53 | 31 |
| **IATE** | *N* | | 0.179 | 0.214 | - | 0.267 | - | - | - | 67 | 50 |
| **IATE eff** | *N* | | 0.180 | 0.205 | - | 0.256 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.038 | 0.046 | 0.056 | 0.068 | 0.12 | -0.24 | 0.005 | 92 | 75 |
| **GATE** | *5* | *cent* | 0.038 | 0.092 | 0.087 | 0.118 | -0.01 | -0.10 | -0.003 | 85 | 67 |
| **IATE** | *N* | | 0.039 | 0.171 | - | 0.209 | - | - | - | 76 | 55 |
| **IATE eff** | *N* | | 0.039 | 0.161 | - | 0.195 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.095 | 0.095 | 0.042 | 0.103 | 0.03 | -0.30 | 0.000 | 39 | 17 |
| **GATE** | *5* | | 0.094 | 0.110 | 0.093 | 0.133 | -0.02 | 0.09 | 0.000 | 83 | 59 |
| **IATE** | *N* | | 0.089 | 0.188 | - | 0.230 | - | - | - | 71 | 50 |
| **ATE** | *1* | *grf* | 0.034 | 0.044 | 0.043 | 0.055 | 0.02 | 0.22 | -0.001 | 86 | 67 |
| **GATE** | *5* | *cent* | 0.035 | 0.081 | 0.094 | 0.102 | -0.03 | -0.01 | -0.002 | 93 | 76 |
| **IATE** | *N* | | 0.032 | 0.174 | - | 0.209 | - | - | - | 72 | 51 |
| **ATE** | *1* | *dml* | 0.019 | 0.039 | 0.044 | 0.048 | -0.07 | -0.09 | 0.005 | 96 | 81 |
| **GATE** | *5* | | 0.019 | 0.082 | 0.101 | 0.103 | -0.01 | -0.13 | -0.001 | 94 | 78 |
| **IATE** | *N* | *ols* | 0.019 | 0.227 | - | 0.283 | - | - | - | 26 | 17 |
| **IATE** | *N* | *rf* | 0.012 | 0.285 | - | 0.363 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.020 | 0.040 | 0.044 | 0.049 | -0.06 | -0.12 | 0.005 | 96 | 80 |
| **GATE** | *5* | *norm* | 0.020 | 0.082 | 0.101 | 0.103 | -0.01 | -0.13 | -0.001 | 94 | 78 |
| **IATE** | *N* | *ols* | 0.020 | 0.227 | - | 0.283 | - | - | - | 26 | 17 |
| **IATE** | *N* | *rf* | 0.013 | 0.284 | - | 0.362 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.007 | 0.035 | 0.043 | 0.044 | -0.03 | -0.10 | -0.012 | 83 | 64 |
| **GATE** | *5* | | 0.007 | 0.082 | 0.096 | 0.102 | -0.01 | -0.13 | -0.011 | 80 | 60 |
| **IATE** | *N* | | 0.008 | 0.220 | - | 0.273 | - | - | - | 70 | 50 |

**Note:** Table to be continued.

Table C.22 - continued: Uniformly distributed covariates ($X^U$)

| | | | | Estimation of effects | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 10'000 | | | | | |
| **ATE** | *1* | *mcf* | 0.145 | 0.145 | 0.028 | 0.148 | 0.01 | -0.58 | 0.004 | 0 | 0 |
| **GATE** | *5* | | 0.145 | 0.145 | 0.046 | 0.158 | 0.01 | -0.31 | 0.001 | 22 | 5 |
| **IATE** | *N* | | 0.145 | 0.158 | - | 0.192 | - | - | - | 63 | 45 |
| **IATE eff** | *N* | | 0.144 | 0.152 | - | 0.184 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.014 | 0.025 | 0.028 | 0.031 | 0.02 | -0.40 | 0.003 | 95 | 78 |
| **GATE** | *5* | *cent* | 0.013 | 0.048 | 0.044 | 0.063 | 0.00 | -0.23 | 0.002 | 87 | 68 |
| **IATE** | *N* | | 0.013 | 0.107 | - | 0.131 | - | - | - | 81 | 59 |
| **IATE eff** | *N* | | 0.011 | 0.099 | - | 0.120 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.062 | 0.062 | 0.022 | 0.066 | 0.02 | -0.32 | -0.001 | 18 | 6 |
| **GATE** | *5* | | 0.062 | 0.065 | 0.046 | 0.077 | 0.00 | -0.17 | 0.001 | 73 | 49 |
| **IATE** | *N* | | 0.060 | 0.095 | - | 0.124 | - | - | - | 87 | 72 |
| **ATE** | *1* | *grf* | 0.014 | 0.022 | 0.023 | 0.027 | 0.03 | 0.12 | -0.002 | 88 | 66 |
| **GATE** | *5* | *cent* | 0.015 | 0.041 | 0.048 | 0.051 | 0.01 | -0.06 | -0.002 | 92 | 76 |
| **IATE** | *N* | | 0.015 | 0.090 | - | 0.113 | - | - | - | 88 | 72 |
| **ATE** | *1* | *dml* | 0.009 | 0.020 | 0.023 | 0.025 | -0.02 | -0.21 | 0.002 | 96 | 79 |
| **GATE** | *5* | | 0.009 | 0.040 | 0.049 | 0.050 | -0.01 | -0.06 | 0.000 | 95 | 79 |
| **IATE** | *N* | *ols* | 0.009 | 0.168 | - | 0.205 | - | - | - | 8 | 5 |
| **IATE** | *N* | *rf* | 0.004 | 0.215 | - | 0.276 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.009 | 0.020 | 0.023 | 0.025 | -0.03 | -0.17 | 0.002 | 96 | 79 |
| **GATE** | *5* | *norm* | 0.009 | 0.040 | 0.049 | 0.050 | -0.01 | -0.05 | 0.000 | 94 | 79 |
| **IATE** | *N* | *ols* | 0.010 | 0.168 | - | 0.205 | - | - | - | 8 | 5 |
| **IATE** | *N* | *rf* | 0.004 | 0.214 | - | 0.275 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.008 | 0.019 | 0.022 | 0.023 | -0.06 | 0.08 | -0.007 | 82 | 62 |
| **GATE** | *5* | | 0.007 | 0.047 | 0.048 | 0.058 | -0.03 | -0.17 | -0.015 | 73 | 52 |
| **IATE** | *N* | | 0.007 | 0.168 | - | 0.204 | - | - | - | 47 | 31 |

**Note:** For GATE and IATE the results are averaged over all effects. *CovP(95, 80)* denotes the (average) probability that the true value is part of the 95%/80% confidence interval. 1'000 / 500 replications used for 2'500 / 10'000 obs.

Table C.23: Normally distributed covariates ($X^N$)

| | | | | Estimation of effects | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 2'500 | | | | | |
| **ATE** | *1* | *mcf* | 0.191 | 0.191 | 0.062 | 0.201 | -0.06 | 0.14 | 0.002 | 15 | 4 |
| **GATE** | *5* | | 0.191 | 0.194 | 0.089 | 0.226 | -0.12 | 0.17 | -0.001 | 49 | 26 |
| **IATE** | *N* | | 0.190 | 0.226 | - | 0.281 | - | - | - | 64 | 48 |
| **IATE eff** | *N* | | 0.191 | 0.226 | - | 0.270 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.044 | 0.061 | 0.060 | 0.075 | -0.10 | -0.01 | 0.001 | 89 | 69 |
| **GATE** | *5* | *cent* | 0.044 | 0.097 | 0.088 | 0.125 | -0.12 | 0.08 | -0.004 | 82 | 64 |
| **IATE** | *N* | | 0.044 | 0.176 | - | 0.214 | - | - | - | 75 | 53 |
| **IATE eff** | *N* | | 0.045 | 0.165 | - | 0.200 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.107 | 0.108 | 0.044 | 0.116 | -0.12 | 0.01 | -0.001 | 27 | 11 |
| **GATE** | *5* | | 0.107 | 0.121 | 0.096 | 0.145 | -0.05 | -0.08 | -0.002 | 78 | 54 |
| **IATE** | *N* | | 0.102 | 0.193 | - | 0.238 | - | - | - | 70 | 50 |
| **ATE** | *1* | *grf* | 0.041 | 0.048 | 0.041 | 0.058 | 0.08 | -0.03 | 0.000 | 84 | 62 |
| **GATE** | *5* | *cent* | 0.041 | 0.084 | 0.095 | 0.105 | -0.03 | -0.12 | -0.002 | 92 | 74 |
| **IATE** | *N* | | 0.039 | 0.179 | - | 0.214 | - | - | - | 71 | 50 |
| **ATE** | *1* | *dml* | 0.026 | 0.042 | 0.046 | 0.053 | -0.06 | 0.01 | 0.003 | 93 | 77 |
| **GATE** | *5* | | 0.026 | 0.083 | 0.101 | 0.104 | -0.03 | -0.03 | -0.001 | 94 | 78 |
| **IATE** | *N* | *ols* | 0.026 | 0.240 | - | 0.300 | - | - | - | 24 | 16 |
| **IATE** | *N* | *rf* | 0.017 | 0.286 | - | 0.365 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.027 | 0.043 | 0.046 | 0.053 | -0.06 | 0.02 | 0.003 | 93 | 77 |
| **GATE** | *5* | *norm* | 0.027 | 0.083 | 0.101 | 0.104 | -0.03 | -0.03 | -0.001 | 94 | 78 |
| **IATE** | *N* | *ols* | 0.027 | 0.239 | - | 0.299 | - | - | - | 24 | 16 |
| **IATE** | *N* | *rf* | 0.017 | 0.285 | - | 0.363 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.012 | 0.038 | 0.046 | 0.047 | -0.02 | 0.25 | -0.014 | 81 | 61 |
| **GATE** | *5* | | 0.011 | 0.083 | 0.097 | 0.103 | -0.04 | -0.04 | -0.029 | 80 | 58 |
| **IATE** | *N* | | 0.011 | 0.233 | - | 0.291 | - | - | - | 66 | 47 |

**Note:** Table to be continued.

Table C.23 - continued: Normally distributed covariates ($X^N$)

| | | | | Estimation of effects | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 10'000 | | | | | |
| **ATE** | *1* | *mcf* | 0.156 | 0.156 | 0.030 | 0.159 | -0.15 | -0.03 | 0.002 | 0 | 0 |
| **GATE** | *5* | | 0.156 | 0.156 | 0.049 | 0.170 | -0.08 | -0.16 | -0.001 | 16 | 5 |
| **IATE** | *N* | | 0.155 | 0.168 | - | 0.205 | - | - | - | 60 | 43 |
| **IATE eff** | *N* | | 0.155 | 0.163 | - | 0.198 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.018 | 0.028 | 0.029 | 0.034 | -0.08 | -0.29 | 0.001 | 93 | 73 |
| **GATE** | *5* | *cent* | 0.018 | 0.051 | 0.046 | 0.067 | 0.03 | -0.09 | 0.000 | 84 | 67 |
| **IATE** | *N* | | 0.018 | 0.110 | - | 0.136 | - | - | - | 80 | 67 |
| **IATE eff** | *N* | | 0.017 | 0.102 | - | 0.125 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.073 | 0.073 | 0.021 | 0.076 | -0.04 | 0.24 | -0.000 | 8 | 1 |
| **GATE** | *5* | | 0.073 | 0.076 | 0.047 | 0.088 | 0.04 | 0.07 | -0.001 | 66 | 39 |
| **IATE** | *N* | | 0.071 | 0.103 | - | 0.133 | - | - | - | 84 | 68 |
| **ATE** | *1* | *grf* | 0.019 | 0.026 | 0.024 | 0.031 | -0.30 | 0.01 | -0.003 | 82 | 56 |
| **GATE** | *5* | *cent* | 0.019 | 0.042 | 0.048 | 0.053 | 0.08 | -0.11 | -0.002 | 92 | 73 |
| **IATE** | *N* | | 0.021 | 0.089 | - | 0.113 | - | - | - | 89 | 73 |
| **ATE** | *1* | *dml* | 0.016 | 0.022 | 0.023 | 0.028 | 0.07 | -0.07 | 0.002 | 92 | 71 |
| **GATE** | *5* | | 0.015 | 0.042 | 0.050 | 0.053 | 0.05 | -0.05 | -0.001 | 93 | 78 |
| **IATE** | *N* | *ols* | 0.015 | 0.185 | - | 0.227 | - | - | - | 7 | 5 |
| **IATE** | *N* | *rf* | 0.009 | 0.216 | - | 0.276 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.016 | 0.022 | 0.023 | 0.028 | 0.07 | -0.05 | 0.002 | 93 | 70 |
| **GATE** | *5* | *norm* | 0.016 | 0.042 | 0.050 | 0.053 | 0.06 | -0.06 | -0.001 | 93 | 77 |
| **IATE** | *N* | *ols* | 0.015 | 0.185 | - | 0.227 | - | - | - | 7 | 5 |
| **IATE** | *N* | *rf* | 0.009 | 0.216 | - | 0.276 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.013 | 0.021 | 0.022 | 0.025 | -0.05 | -0.23 | -0.006 | 75 | 55 |
| **GATE** | *5* | | 0.012 | 0.050 | 0.048 | 0.062 | 0.02 | -0.16 | -0.014 | 70 | 51 |
| **IATE** | *N* | | 0.012 | 0.186 | - | 0.227 | - | - | - | 43 | 29 |

**Note:** For GATE and IATE the results are averaged over all effects. *CovP(95, 80)* denotes the (average) probability that the true value is part of the 95%/80% confidence interval. 1'000 / 500 replications used for 2'500 / 10'000 obs.

Table C.24: Normally, uniformly distributed, and indicator covariates ($X^D, X^N, X^U$)

| | | | | Estimation of effects | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 2'500 | | | | | |
| **ATE** | *1* | *mcf* | 0.160 | 0.161 | 0.062 | 0.172 | 0.21 | 0.30 | 0.003 | 30 | 10 |
| **GATE** | *5* | | 0.151 | 0.157 | 0.086 | 0.192 | 0.08 | -0.01 | -0.001 | 60 | 39 |
| **IATE** | *N* | | 0.160 | 0.200 | - | 0.247 | - | - | - | 77 | 58 |
| **IATE eff** | *N* | | 0.163 | 0.189 | - | 0.231 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.054 | 0.066 | 0.062 | 0.082 | 0.18 | 0.46 | 0.001 | 87 | 65 |
| **GATE** | *5* | *cent* | 0.045 | 0.094 | 0.158 | 0.122 | 0.04 | -0.01 | -0.005 | 83 | 65 |
| **IATE** | *N* | | 0.054 | 0.158 | - | 0.197 | - | - | - | 86 | 65 |
| **IATE eff** | *N* | | 0.057 | 0.143 | - | 0.175 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.092 | 0.093 | 0.044 | 0.102 | -0.03 | -0.20 | -0.001 | 41 | 21 |
| **GATE** | *5* | | 0.093 | 0.112 | 0.099 | 0.137 | 0.03 | -0.09 | -0.003 | 82 | 60 |
| **IATE** | *N* | | 0.093 | 0.180 | - | 0.219 | - | - | - | 74 | 55 |
| **ATE** | *1* | *grf* | 0.052 | 0.057 | 0.044 | 0.068 | 0.11 | -0.11 | -0.001 | 75 | 51 |
| **GATE** | *5* | *cent* | 0.054 | 0.089 | 0.097 | 0.113 | 0.05 | -0.01 | -0.002 | 90 | 73 |
| **IATE** | *N* | | 0.049 | 0.162 | - | 0.198 | - | - | - | 78 | 59 |
| **ATE** | *1* | *dml* | 0.047 | 0.054 | 0.045 | 0.065 | 0.08 | -0.21 | 0.004 | 86 | 64 |
| **GATE** | *5* | | 0.049 | 0.092 | 0.103 | 0.115 | -0.02 | -0.01 | 0.001 | 93 | 75 |
| **IATE** | *N* | *ols* | 0.047 | 0.203 | - | 0.255 | - | - | - | 30 | 20 |
| **IATE** | *N* | *rf* | 0.047 | 0.320 | - | 0.408 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.048 | 0.055 | 0.045 | 0.066 | 0.08 | -0.22 | 0.004 | 84 | 63 |
| **GATE** | *5* | *norm* | 0.050 | 0.092 | 0.103 | 0.115 | -0.02 | -0.02 | 0.000 | 92 | 75 |
| **IATE** | *N* | *ols* | 0.049 | 0.202 | - | 0.254 | - | - | - | 29 | 20 |
| **IATE** | *N* | *rf* | 0.040 | 0.318 | - | 0.406 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.001 | 0.036 | 0.045 | 0.045 | 0.06 | -0.21 | 0.045 | 82 | 62 |
| **GATE** | *5* | | 0.009 | 0.083 | 0.096 | 0.104 | 0.01 | -0.05 | 0.096 | 80 | 60 |
| **IATE** | *N* | | 0.001 | 0.188 | - | 0.236 | - | - | - | 77 | 57 |

**Note:** Table to be continued.

Table C.24 - continued: Normally, uniformly distributed, and indicator covariates ($X^D$, $X^N, X^U$)

| | | | | | Estimation of effects | | | | | Estimation of std. errors | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 10'000 | | | | | |
| **ATE** | *1* | *mcf* | 0.120 | 0.120 | 0.028 | 0.124 | 0.32 | 0.41 | 0.004 | 2 | 0 |
| **GATE** | *5* | | 0.115 | 0.116 | 0.047 | 0.136 | 0.16 | -0.07 | 0.000 | 41 | 24 |
| **IATE** | *N* | | 0.120 | 0.145 | - | 0.178 | - | - | - | 78 | 59 |
| **IATE eff** | *N* | | 0.120 | 0.135 | - | 0.165 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.022 | 0.028 | 0.027 | 0.035 | 0.13 | 0.29 | 0.004 | 93 | 75 |
| **GATE** | *5* | *cent* | 0.017 | 0.052 | 0.047 | 0.066 | 0.17 | -0.08 | -0.001 | 85 | 67 |
| **IATE** | *N* | | 0.022 | 0.100 | - | 0.125 | - | - | - | 91 | 73 |
| **IATE eff** | *N* | | 0.022 | 0.086 | - | 0.107 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.053 | 0.054 | 0.021 | 0.057 | 0.15 | 0.16 | 0.000 | 30 | 10 |
| **GATE** | *5* | | 0.054 | 0.061 | 0.050 | 0.074 | 0.01 | -0.28 | -0.002 | 78 | 53 |
| **IATE** | *N* | | 0.056 | 0.083 | - | 0.107 | - | - | - | 90 | 76 |
| **ATE** | *1* | *grf* | 0.021 | 0.025 | 0.023 | 0.031 | -0.14 | 0.17 | -0.002 | 83 | 58 |
| **GATE** | *5* | *cent* | 0.022 | 0.043 | 0.048 | 0.054 | -0.06 | 0.03 | -0.000 | 92 | 74 |
| **IATE** | *N* | | 0.022 | 0.073 | - | 0.094 | - | - | - | 93 | 81 |
| **ATE** | *1* | *dml* | 0.023 | 0.026 | 0.022 | 0.032 | 0.03 | 0.35 | 0.003 | 90 | 67 |
| **GATE** | *5* | | 0.024 | 0.045 | 0.052 | 0.058 | 0.12 | 0.11 | 0.001 | 92 | 76 |
| **IATE** | *N* | *ols* | 0.023 | 0.126 | - | 0.156 | - | - | - | 12 | 8 |
| **IATE** | *N* | *rf* | 0.016 | 0.248 | - | 0.318 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.024 | 0.027 | 0.022 | 0.032 | 0.03 | 0.36 | 0.003 | 88 | 64 |
| **GATE** | *5* | *norm* | 0.025 | 0.046 | 0.052 | 0.058 | 0.12 | 0.11 | 0.000 | 92 | 75 |
| **IATE** | *N* | *ols* | 0.024 | 0.126 | - | 0.156 | - | - | - | 12 | 8 |
| **IATE** | *N* | *rf* | 0.017 | 0.245 | - | 0.315 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.003 | 0.016 | 0.021 | 0.021 | 0.10 | 0.03 | -0.005 | 85 | 67 |
| **GATE** | *5* | | 0.010 | 0.050 | 0.049 | 0.061 | 0.06 | 0.11 | -0.015 | 71 | 50 |
| **IATE** | *N* | | 0.002 | 0.122 | - | 0.152 | - | - | - | 64 | 50 |

**Note:** For GATE and IATE the results are averaged over all effects. *CovP(95, 80)* denotes the (average) probability that the true value is part of the 95%/80% confidence interval. 1'000 / 500 replications used for 2'500 / 10'000 obs.

Table C.25: Four treatments

| | | | Estimation of effects | | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | N = 2'500 | | | | | | |
| **ATE** | *1* | *mcf* | 0.086 | 0.099 | 0.087 | 0.122 | 0.205 | -0.06 | 0.008 | 87 | 67 |
| **GATE** | *5* | | 0.086 | 0.125 | 0.112 | 0.160 | 0.09 | -0.08 | 0.009 | 85 | 68 |
| **IATE** | *N* | | 0.086 | 0.183 | - | 0.228 | - | - | - | 83 | 65 |
| **IATE eff** | *N* | | 0.084 | 0.167 | - | 0.206 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | -0.019 | 0.069 | 0.084 | 0.086 | 0.21 | -0.04 | 0.004 | 95 | 80 |
| **GATE** | *5* | *cent* | -0.019 | 0.138 | 0.112 | 0.167 | 0.06 | 0.04 | -0.005 | 77 | 55 |
| **IATE** | *N* | | -0.019 | 0.222 | - | 0.264 | - | - | - | 69 | 46 |
| **IATE eff** | *N* | | -0.022 | 0.214 | - | 0.250 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.027 | 0.052 | 0.060 | 0.066 | -0.01 | 0.41 | -0.001 | 92 | 77 |
| **GATE** | *5* | | 0.027 | 0.107 | 0.132 | 0.136 | 0.05 | 0.05 | 0.001 | 94 | 80 |
| **IATE** | *N* | | -0.025 | 0.200 | - | 0.236 | - | - | - | 69 | 48 |
| **ATE** | *1* | *grf* | 0.004 | 0.049 | 0.062 | 0.062 | -0.06 | 0.23 | -0.003 | 95 | 78 |
| **GATE** | *5* | *cent* | 0.004 | 0.106 | 0.133 | 0.134 | 0.03 | -0.03 | -0.001 | 95 | 79 |
| **IATE** | *N* | | -0.046 | 0.194 | - | 0.232 | - | - | - | 70 | 50 |
| **ATE** | *1* | *dml* | 0.012 | 0.057 | 0.070 | 0.071 | -0.12 | -0.14 | 0.003 | 96 | 80 |
| **GATE** | *5* | | 0.012 | 0.122 | 0.153 | 0.154 | 0.03 | 0.11 | 0.000 | 95 | 81 |
| **IATE** | *N* | *ols* | 0.012 | 0.310 | - | 0.390 | - | - | - | 43 | 29 |
| **IATE** | *N* | *rf* | 0.004 | 0.405 | - | 0.545 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.014 | 0.056 | 0.068 | 0.069 | -0.09 | -0.12 | 0.002 | 96 | 78 |
| **GATE** | *5* | *norm* | 0.014 | 0.119 | 0.149 | 0.150 | 0.04 | 0.03 | -0.002 | 94 | 79 |
| **IATE** | *N* | *ols* | 0.014 | 0.303 | - | 0.381 | - | - | - | 41 | 27 |
| **IATE** | *N* | *rf* | 0.006 | 0.402 | - | 0.523 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.019 | 0.054 | 0.064 | 0.067 | -0.13 | -0.13 | -0.033 | 64 | 44 |
| **GATE** | *5* | | 0.018 | 0.113 | 0.137 | 0.142 | -0.01 | -0.02 | -0.070 | 65 | 46 |
| **IATE** | *N* | | 0.019 | 0.287 | - | 0.361 | - | - | - | 62 | 44 |

**Note:** Table to be continued.

Table C.25 - continued: Four treatments

| | | | Estimation of effects | | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| N = 10'000 | | | | | | | | | | | |
| **ATE** | *1* | *mcf* | 0.075 | 0.077 | 0.044 | 0.087 | 0.00 | 0.65 | 0.005 | 69 | 38 |
| **GATE** | *5* | | 0.075 | 0.087 | 0.043 | 0.110 | 0.00 | 0.41 | 0.004 | 74 | 55 |
| **IATE** | *N* | | 0.075 | 0.124 | - | 0.155 | - | - | - | 66 | 47 |
| **IATE eff** | *N* | | 0.077 | 0.114 | - | 0.293 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | -0.014 | 0.035 | 0.043 | 0.045 | -0.01 | 0.97 | 0.002 | 94 | 81 |
| **GATE** | *5* | *cent* | -0.014 | 0.069 | 0.061 | 0.084 | -0.01 | 0.07 | -0.002 | 84 | 67 |
| **IATE** | *N* | | -0.015 | 0.129 | - | 0.157 | - | - | - | 77 | 54 |
| **IATE eff** | *N* | | -0.012 | 0.121 | - | 0.144 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.035 | 0.038 | 0.030 | 0.045 | 0.03 | -0.21 | 0.000 | 81 | 56 |
| **GATE** | *5* | | 0.035 | 0.060 | 0.067 | 0.076 | 0.08 | -0.01 | -0.001 | 91 | 74 |
| **IATE** | *N* | | 0.005 | 0.089 | - | 0.111 | - | - | - | 90 | 74 |
| **ATE** | *1* | *grf* | 0.005 | 0.024 | 0.029 | 0.030 | -0.12 | 0.20 | 0.000 | 96 | 82 |
| **GATE** | *5* | *cent* | 0.005 | 0.053 | 0.066 | 0.066 | -0.04 | -0.13 | -0.000 | 94 | 78 |
| **IATE** | *N* | | -0.022 | 0.092 | - | 0.113 | - | - | - | 88 | 72 |
| **ATE** | *1* | *dml* | 0.008 | 0.026 | 0.031 | 0.032 | 0.08 | -0.32 | 0.004 | 96 | 83 |
| **GATE** | *5* | | 0.008 | 0.056 | 0.070 | 0.072 | 0.12 | 0.10 | 0.004 | 95 | 82 |
| **IATE** | *N* | *ols* | 0.008 | 0.204 | - | 0.254 | - | - | - | 16 | 10 |
| **IATE** | *N* | *rf* | 0.003 | 0.308 | - | 0.419 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.009 | 0.026 | 0.031 | 0.032 | 0.08 | -0.39 | 0.004 | 97 | 84 |
| **GATE** | *5* | *norm* | 0.008 | 0.056 | 0.070 | 0.071 | 0.14 | 0.09 | 0.003 | 95 | 81 |
| **IATE** | *N* | *ols* | 0.008 | 0.204 | - | 0.253 | - | - | - | 15 | 10 |
| **IATE** | *N* | *rf* | 0.004 | 0.305 | - | 0.412 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.022 | 0.028 | 0.029 | 0.036 | 0.06 | -0.04 | -0.013 | 61 | 44 |
| **GATE** | *5* | | 0.021 | 0.057 | 0.064 | 0.072 | 0.06 | 0.12 | -0.030 | 65 | 46 |
| **IATE** | *N* | | 0.021 | 0.202 | - | 0.253 | - | - | - | 42 | 28 |

**Note:** For GATE and IATE the results are averaged over all effects. *CovP(95, 80)* denotes the (average) probability that the true value is part of the 95%/80% confidence interval. 1'000 / 500 replications used for 2'500 / 10'000 obs. Results are shown for the comparison of treatments 1 to 0.
For the *grf*, the results differ substantially for the different treatment comparisons (which all have the same effect size), with an RMSE of the ATE/GATE/IATE in the range of 0.066-0.154 / 0.136-0.196 / 0.236-0.266 for $N = 2'500$ and 0.045-0.110 / 0.076-0.126 / 0.111-0.147 for $N = 10'000$.
For the centred *grf*, the RMSE of the ATE/GATE/IATE is in the range of 0.062-0.083 / 0.134-0.145 / 0.228-0.239 for $N = 2'500$ and 0.030-0.042 / 0.066-0.073 / 0.113-0.122 for $N = 10'000$.

Table C.26: Larger sample

| | | | Estimation of effects | | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 40'000 | | | | | |
| **ATE** | *1* | *mcf* | 0.111 | 0.111 | 0.012 | 0.112 | 0.12 | -0.92 | 0.004 | 0 | 0 |
| **GATE** | *5* | | 0.111 | 0.111 | 0.022 | 0.115 | 0.30 | 0.33 | 0.004 | 0 | 0 |
| **IATE** | *N* | | 0.111 | 0.120 | - | 0.142 | - | - | - | 45 | 67 |
| **IATE eff** | *N* | | 0.112 | 0.117 | - | 0.137 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.004 | 0.010 | 0.013 | 0.014 | -0.01 | 2.40 | 0.002 | 97 | 87 |
| **GATE** | *5* | *cent* | 0.003 | 0.028 | 0.021 | 0.035 | -0.07 | 1.01 | 0.003 | 84 | 62 |
| **IATE** | *N* | | 0.003 | 0.073 | - | 0.090 | - | - | - | 83 | 62 |
| **IATE eff** | *N* | | 0.001 | 0.067 | - | 0.082 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.044 | 0.044 | 0.010 | 0.045 | -0.23 | 0.73 | 0.000 | 3 | 0 |
| **GATE** | *5* | | 0.043 | 0.045 | 0.024 | 0.050 | -0.23 | 0.14 | -0.001 | 51 | 26 |
| **IATE** | *N* | | 0.042 | 0.066 | - | 0.084 | - | - | - | 93 | 80 |
| **ATE** | *1* | *grf* | 0.005 | 0.010 | 0.011 | 0.012 | 0.12 | -0.44 | -0.001 | 87 | 71 |
| **GATE** | *5* | *cent* | 0.006 | 0.020 | 0.023 | 0.025 | -0.08 | -0.08 | -0.000 | 92 | 76 |
| **IATE** | *N* | | 0.007 | 0.059 | - | 0.073 | - | - | - | 96 | 84 |
| **ATE** | *1* | *dml* | 0.008 | 0.011 | 0.009 | 0.012 | -0.28 | 0.14 | 0.003 | 95 | 79 |
| **GATE** | *5* | | 0.007 | 0.020 | 0.023 | 0.025 | 0.23 | 0.40 | 0.001 | 93 | 82 |
| **IATE** | *N* | *ols* | 0.008 | 0.167 | - | 0.201 | - | - | - | 2 | 1 |
| **IATE** | *N* | *rf* | 0.008 | 0.164 | - | 0.212 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.008 | 0.011 | 0.009 | 0.012 | -0.28 | 0.12 | 0.003 | 95 | 79 |
| **GATE** | *5* | *norm* | 0.008 | 0.020 | 0.023 | 0.025 | 0.23 | 0.40 | 0.001 | 93 | 82 |
| **IATE** | *N* | *ols* | 0.008 | 0.167 | - | 0.201 | - | - | - | 2 | 1 |
| **IATE** | *N* | *rf* | 0.008 | 0.164 | - | 0.212 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.010 | 0.012 | 0.009 | 0.014 | -0.11 | -0.37 | -0.002 | 66 | 42 |
| **GATE** | *5* | | 0.010 | 0.034 | 0.022 | 0.039 | 0.09 | 0.26 | -0.005 | 49 | 33 |
| **IATE** | *N* | | 0.010 | 0.169 | - | 0.203 | - | - | - | 24 | 16 |

**Note:** For GATE and IATE the results are averaged over all effects. *CovP(95, 80)* denotes the (average) probability that the true value is part of the 95%/80% confidence interval. 62 replications.

Table C.27: Only 25% treated

| | | | Estimation of effects | | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | N = 2'500 | | | | | | |
| **ATE** | *1* | *mcf* | 0.247 | 0.247 | 0.073 | 0.257 | -0.04 | -0.12 | 0.001 | 8 | 2 |
| **GATE** | *5* | | 0.246 | 0.247 | 0.094 | 0.300 | -0.05 | -0.09 | 0.001 | 45 | 26 |
| **IATE** | *N* | | 0.247 | 0.269 | - | 0.369 | - | - | - | 53 | 42 |
| **IATE eff** | *N* | | 0.247 | 0.287 | - | 0.360 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.087 | 0.094 | 0.071 | 0.112 | -0.04 | 0.07 | 0.000 | 76 | 50 |
| **GATE** | *5* | *cent* | 0.086 | 0.133 | 0.099 | 0.178 | -0.02 | -0.02 | -0.007 | 75 | 57 |
| **IATE** | *N* | | 0.086 | 0.234 | - | 0.280 | - | - | - | 61 | 40 |
| **IATE eff** | *N* | | 0.089 | 0.227 | - | 0.269 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.144 | 0.144 | 0.051 | 0.152 | -0.02 | 0.11 | -0.002 | 17 | 6 |
| **GATE** | *5* | | 0.143 | 0.156 | 0.111 | 0.187 | -0.00 | 0.00 | -0.003 | 71 | 49 |
| **IATE** | *N* | | 0.172 | 0.293 | - | 0.351 | - | - | - | 46 | 28 |
| **ATE** | *1* | *grf* | 0.075 | 0.078 | 0.050 | 0.090 | 0.05 | -0.13 | -0.001 | 65 | 39 |
| **GATE** | *5* | *cent* | 0.075 | 0.110 | 0.108 | 0.138 | -0.09 | 0.07 | -0.001 | 88 | 67 |
| **IATE** | *N* | | 0.108 | 0.274 | - | 0.317 | - | - | - | 47 | 27 |
| **ATE** | *1* | *dml* | 0.025 | 0.051 | 0.059 | 0.064 | -0.15 | 0.10 | 0.005 | 94 | 79 |
| **GATE** | *5* | | 0.025 | 0.108 | 0.134 | 0.138 | 0.01 | 0.28 | -0.003 | 93 | 78 |
| **IATE** | *N* | *ols* | 0.025 | 0.279 | - | 0.351 | - | - | - | 34 | 23 |
| **IATE** | *N* | *rf* | 0.010 | 0.359 | - | 0.525 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.029 | 0.051 | 0.056 | 0.063 | -0.09 | 0.07 | 0.004 | 93 | 78 |
| **GATE** | *5* | *norm* | 0.029 | 0.105 | 0.129 | 0.133 | 0.04 | 0.04 | -0.003 | 93 | 77 |
| **IATE** | *N* | *ols* | 0.030 | 0.272 | - | 0.341 | - | - | - | 32 | 22 |
| **IATE** | *N* | *rf* | 0.016 | 0.351 | - | 0.472 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.022 | 0.046 | 0.053 | 0.058 | -0.06 | 0.07 | -0.021 | 71 | 54 |
| **GATE** | *5* | | 0.022 | 0.097 | 0.114 | 0.122 | 0.03 | -0.04 | -0.046 | 72 | 52 |
| **IATE** | *N* | | 0.022 | 0.257 | - | 0.321 | - | - | - | 62 | 43 |

**Note:** Table to be continued.

Table C.27 - continued: Only 25% treated

| | | | | Estimation of effects | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | | N = 10'000 | | | | | |
| **ATE** | *1* | *mcf* | 0.189 | 0.189 | 0.035 | 0.192 | 0.06 | -0.10 | 0.002 | 0 | 0 |
| **GATE** | *5* | | 0.188 | 0.188 | 0.053 | 0.214 | 0.03 | -0.03 | 0.001 | 20 | 7 |
| **IATE** | *N* | | 0.188 | 0.208 | - | 0.257 | - | - | - | 54 | 42 |
| **IATE eff** | *N* | | 0.186 | 0.201 | - | 0.251 | - | - | - | - | - |
| **ATE** | *1* | *mcf* | 0.043 | 0.046 | 0.034 | 0.055 | 0.16 | 0.02 | 0.002 | 79 | 54 |
| **GATE** | *5* | *cent* | 0.042 | 0.066 | 0.052 | 0.091 | 0.02 | 0.01 | 0.001 | 80 | 63 |
| **IATE** | *N* | | 0.042 | 0.134 | - | 0.166 | - | - | - | 75 | 54 |
| **IATE eff** | *N* | | 0.042 | 0.127 | - | 0.157 | - | - | - | - | - |
| **ATE** | *1* | *grf* | 0.083 | 0.083 | 0.026 | 0.087 | -0.14 | 0.45 | -0.001 | 7 | 4 |
| **GATE** | *5* | | 0.082 | 0.087 | 0.056 | 0.103 | 0.05 | -0.19 | -0.003 | 66 | 42 |
| **IATE** | *N* | | 0.098 | 0.130 | - | 0.177 | - | - | - | 76 | 61 |
| **ATE** | *1* | *grf* | 0.030 | 0.033 | 0.026 | 0.039 | 0.05 | -0.33 | -0.002 | 72 | 51 |
| **GATE** | *5* | *cent* | 0.030 | 0.052 | 0.054 | 0.066 | -0.03 | 0.07 | -0.001 | 88 | 71 |
| **IATE** | *N* | | 0.049 | 0.116 | - | 0.152 | - | - | - | 80 | 64 |
| **ATE** | *1* | *dml* | 0.012 | 0.024 | 0.028 | 0.030 | -0.05 | -0.23 | 0.003 | 94 | 81 |
| **GATE** | *5* | | 0.011 | 0.053 | 0.065 | 0.067 | 0.03 | -0.09 | 0.030 | 95 | 78 |
| **IATE** | *N* | *ols* | 0.011 | 0.193 | - | 0.238 | - | - | - | 12 | 7 |
| **IATE** | *N* | *rf* | 0.000 | 0.266 | - | 0.371 | - | - | - | - | - |
| **ATE** | *1* | *dml-* | 0.012 | 0.024 | 0.028 | 0.030 | -0.03 | -0.21 | 0.003 | 95 | 80 |
| **GATE** | *5* | *norm* | 0.012 | 0.053 | 0.065 | 0.067 | 0.04 | -0.13 | -0.001 | 95 | 78 |
| **IATE** | *N* | *ols* | 0.012 | 0.193 | - | 0.238 | - | - | - | 11 | 8 |
| **IATE** | *N* | *rf* | 0.001 | 0.264 | - | 0.367 | - | - | - | - | - |
| **ATE** | *1* | *ols* | 0.020 | 0.023 | 0.026 | 0.033 | -0.13 | 0.15 | -0.010 | 62 | 45 |
| **GATE** | *5* | | 0.020 | 0.049 | 0.055 | 0.068 | -0.06 | 0.04 | 0.055 | 66 | 46 |
| **IATE** | *N* | | 0.020 | 0.193 | - | 0.236 | - | - | 0.236 | 43 | 29 |

**Note:** For GATE and IATE the results are averaged over all effects. *CovP(95, 80)* denotes the (average) probability that the true value is part of the 95%/80% confidence interval. 1'000 / 500 replications used for 2'500 / 10'000 obs.

Table C.28: More GATE groups, no selectivity

| | | | Estimation of effects | | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | N = 2'500 | | | | | | |
| **GATE** | *5* | *mcf* | -0.004 | 0.092 | 0.087 | 0.115 | 0.00 | 0.01 | 0.002 | 86 | 67 |
| **GATE** | *10* | | -0.004 | 0.095 | 0.090 | 0.118 | -0.01 | -0.04 | -0.001 | 86 | 86 |
| **GATE** | *20* | | -0.004 | 0.097 | 0.060 | 0.120 | -0.01 | -0.02 | -0.001 | 85 | 65 |
| **GATE** | *40* | | -0.004 | 0.099 | 0.093 | 0.123 | -0.01 | -0.01 | -0.003 | 84 | 64 |
| **GATE** | *5* | *mcf* | 0.002 | 0.090 | 0.091 | 0.114 | 0.02 | 0.09 | -0.010 | 83 | 64 |
| **GATE** | *10* | *cent* | 0.002 | 0.094 | 0.094 | 0.117 | -0.02 | 0.02 | -0.012 | 82 | 63 |
| **GATE** | *20* | | 0.002 | 0.094 | 0.094 | 0.118 | 0.00 | 0.04 | -0.011 | 83 | 63 |
| **GATE** | *40* | | 0.002 | 0.096 | 0.096 | 0.120 | 0.00 | 0.03 | -0.013 | 82 | 62 |
| **GATE** | *5* | *grf* | -0.001 | 0.076 | 0.096 | 0.096 | 0.06 | 0.09 | -0.001 | 95 | 80 |
| **GATE** | *10* | | -0.001 | 0.108 | 0.135 | 0.135 | 0.03 | 0.00 | -0.001 | 95 | 79 |
| **GATE** | *20* | | -0.001 | 0.152 | 0.190 | 0.191 | -0.01 | 0.02 | -0.001 | 95 | 80 |
| **GATE** | *40* | | -0.001 | 0.215 | 0.270 | 0.270 | 0.01 | 0.05 | -0.001 | 95 | 80 |
| **GATE** | *5* | *grf* | 0.004 | 0.075 | 0.094 | 0.094 | 0.01 | 0.01 | 0.000 | 95 | 80 |
| **GATE** | *10* | *cent* | 0.004 | 0.106 | 0.133 | 0.133 | -0.00 | 0.02 | 0.000 | 95 | 80 |
| **GATE** | *20* | | 0.004 | 0.150 | 0.189 | 0.189 | 0.02 | 0.07 | -0.000 | 95 | 80 |
| **GATE** | *40* | | 0.004 | 0.213 | 0.269 | 0.269 | 0.02 | 0.11 | -0.001 | 95 | 80 |
| **GATE** | *5* | *dml* | -0.002 | 0.075 | 0.094 | 0.095 | -0.02 | 0.04 | 0.002 | 95 | 81 |
| **GATE** | *10* | | -0.002 | 0.108 | 0.134 | 0.134 | -0.02 | -0.06 | 0.002 | 95 | 81 |
| **GATE** | *20* | | -0.002 | 0.152 | 0.191 | 0.191 | -0.01 | 0.07 | 0.001 | 95 | 80 |
| **GATE** | *40* | | -0.002 | 0.217 | 0.272 | 0.273 | 0.01 | 0.06 | -0.003 | 95 | 79 |
| **GATE** | *5* | *dml-* | -0.002 | 0.075 | 0.094 | 0.094 | -0.02 | 0.05 | 0.002 | 95 | 80 |
| **GATE** | *10* | *norm* | -0.002 | 0.107 | 0.133 | 0.133 | -0.02 | -0.05 | 0.000 | 95 | 80 |
| **GATE** | *20* | | -0.002 | 0.150 | 0.189 | 0.189 | -0.01 | 0.07 | -0.001 | 95 | 80 |
| **GATE** | *40* | | -0.002 | 0.214 | 0.268 | 0.269 | 0.00 | 0.06 | -0.003 | 94 | 79 |
| **GATE** | *5* | *ols* | -0.002 | 0.073 | 0.092 | 0.093 | -0.04 | 0.09 | -0.027 | 84 | 64 |
| **GATE** | *10* | | -0.002 | 0.105 | 0.131 | 0.131 | -0.01 | -0.03 | -0.039 | 83 | 63 |
| **GATE** | *20* | | -0.002 | 0.149 | 0.187 | 0.187 | -0.01 | 0.07 | -0.056 | 83 | 63 |
| **GATE** | *40* | | -0.002 | 0.213 | 0.268 | 0.268 | 0.01 | 0.07 | -0.082 | 83 | 63 |

**Note:** Table to be continued.

Table C.28 - continued: More GATE groups, no selectivity

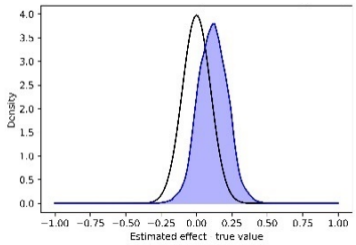
| | | | Estimation of effects | | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | N = 10'000 | | | | | | |
| **GATE** | *5* | *mcf* | 0.001 | 0.049 | 0.045 | 0.063 | 0.07 | -0.21 | 0.002 | 87 | 66 |
| **GATE** | *10* | | 0.001 | 0.053 | 0.048 | 0.066 | 0.01 | -0.23 | 0.000 | 84 | 64 |
| **GATE** | *20* | | 0.001 | 0.053 | 0.049 | 0.067 | 0.04 | -0.25 | 0.000 | 85 | 64 |
| **GATE** | *40* | | 0.001 | 0.054 | 0.050 | 0.068 | 0.04 | -0.18 | -0.001 | 84 | 64 |
| **GATE** | *5* | *mcf* | 0.004 | 0.045 | 0.045 | 0.058 | 0.05 | 0.17 | -0.002 | 87 | 68 |
| **GATE** | *10* | *cent* | 0.004 | 0.050 | 0.049 | 0.063 | 0.07 | -0.22 | -0.005 | 85 | 64 |
| **GATE** | *20* | | 0.004 | 0.049 | 0.049 | 0.062 | 0.09 | -0.16 | -0.004 | 85 | 65 |
| **GATE** | *40* | | 0.004 | 0.050 | 0.050 | 0.063 | 0.09 | -0.11 | -0.005 | 85 | 64 |
| **GATE** | *5* | *grf* | -0.001 | 0.037 | 0.047 | 0.047 | -0.09 | 0.13 | -0.001 | 94 | 80 |
| **GATE** | *10* | | -0.001 | 0.053 | 0.067 | 0.067 | -0.03 | 0.03 | -0.001 | 94 | 80 |
| **GATE** | *20* | | -0.001 | 0.076 | 0.094 | 0.095 | -0.04 | -0.05 | -0.001 | 95 | 79 |
| **GATE** | *40* | | -0.001 | 0.107 | 0.134 | 0.134 | -0.01 | -0.00 | -0.002 | 94 | 79 |
| **GATE** | *5* | *grf* | 0.002 | 0.036 | 0.046 | 0.046 | -0.07 | -0.05 | 0.001 | 95 | 80 |
| **GATE** | *10* | *cent* | 0.002 | 0.052 | 0.065 | 0.065 | -0.04 | -0.13 | 0.001 | 95 | 80 |
| **GATE** | *20* | | 0.002 | 0.072 | 0.091 | 0.091 | 0.02 | -0.02 | 0.002 | 96 | 81 |
| **GATE** | *40* | | 0.002 | 0.104 | 0.130 | 0.130 | 0.03 | -0.05 | 0.002 | 95 | 80 |
| **GATE** | *5* | *dml* | -0.001 | 0.037 | 0.046 | 0.046 | 0.06 | -0.02 | 0.001 | 95 | 82 |
| **GATE** | *10* | | -0.001 | 0.052 | 0.064 | 0.064 | -0.04 | -0.08 | 0.002 | 96 | 81 |
| **GATE** | *20* | | -0.001 | 0.075 | 0.094 | 0.094 | -0.03 | 0.00 | 0.000 | 95 | 80 |
| **GATE** | *40* | | -0.001 | 0.106 | 0.131 | 0.132 | -0.01 | -0.02 | 0.000 | 95 | 80 |
| **GATE** | *5* | *dml-* | -0.001 | 0.037 | 0.046 | 0.046 | 0.05 | -0.03 | 0.001 | 95 | 82 |
| **GATE** | *10* | *norm* | -0.001 | 0.052 | 0.064 | 0.064 | -0.04 | -0.07 | 0.002 | 96 | 81 |
| **GATE** | *20* | | -0.001 | 0.075 | 0.093 | 0.093 | -0.03 | 0.01 | 0.000 | 95 | 80 |
| **GATE** | *40* | | -0.001 | 0.106 | 0.132 | 0.132 | -0.01 | -0.02 | 0.000 | 95 | 80 |
| **GATE** | *5* | *ols* | 0.000 | 0.037 | 0.046 | 0.046 | 0.01 | -0.03 | -0.013 | 85 | 62 |
| **GATE** | *10* | | -0.001 | 0.052 | 0.064 | 0.064 | -0.04 | -0.13 | -0.018 | 84 | 63 |
| **GATE** | *20* | | -0.001 | 0.074 | 0.093 | 0.093 | -0.03 | -0.02 | -0.028 | 83 | 63 |
| **GATE** | *40* | | -0.001 | 0.105 | 0.131 | 0.132 | 0.01 | 0.00 | -0.039 | 83 | 63 |

**Note:** For GATE, the results are averaged over all effects. *CovP(95, 80)* denotes the (average) probability that the true value is part of the 95%/80% confidence interval. 1'000 / 500 replications used for 2'500 / 10'000 obs.

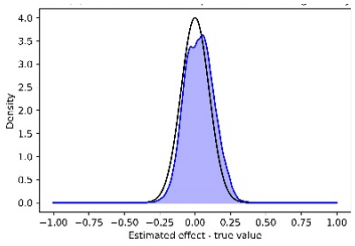
Table C.28 - continued: More GATE groups, no selectivity - Distribution of GATE minus true value of the centred mcf

N = 2'500

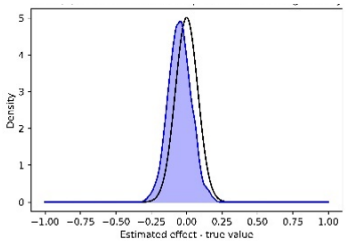
True GATE: 0.50



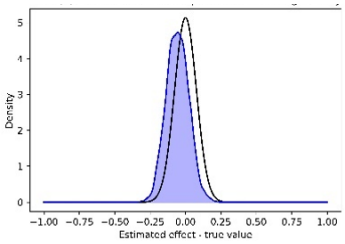

True GATE: 0.77



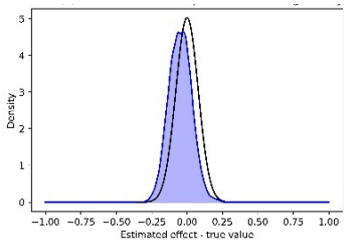

True GATE: 1.20




True GATE: 1.24

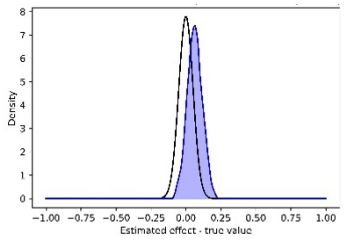



True GATE: 1.24

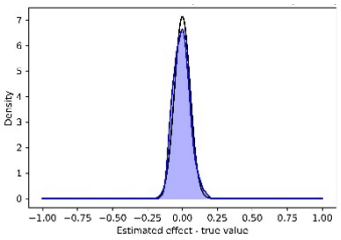



N = 10'000

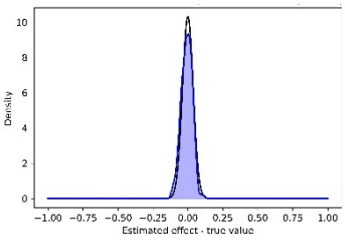
True GATE: 0.50



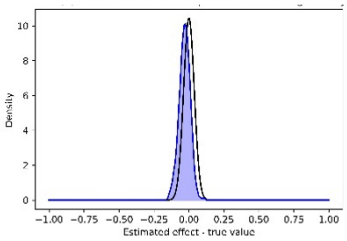

True GATE: 0.77




True GATE: 1.20




True GATE: 1.24




True GATE: 1.24

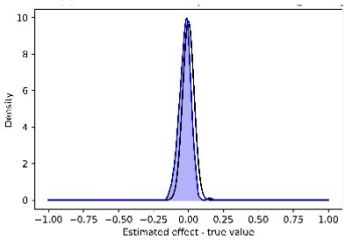


Table C.29: More GATE groups, medium selectivity

| | | | Estimation of effects | | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | N = 2'500 | | | | | | |
| **GATE** | *5* | *mcf* | 0.175 | 0.178 | 0.090 | 0.211 | 0.06 | -0.04 | 0.008 | 53 | 32 |
| **GATE** | *10* | | 0.174 | 0.178 | 0.093 | 0.212 | 0.02 | -0.06 | -0.003 | 55 | 33 |
| **GATE** | *20* | | 0.174 | 0.178 | 0.094 | 0.213 | 0.03 | -0.06 | -0.003 | 56 | 34 |
| **GATE** | *40* | | 0.174 | 0.179 | 0.096 | 0.215 | 0.03 | -0.06 | -0.005 | 56 | 35 |
| **GATE** | *5* | *mcf* | 0.037 | 0.094 | 0.090 | 0.121 | 0.05 | 0.07 | -0.006 | 84 | 65 |
| **GATE** | *10* | *cent* | 0.036 | 0.097 | 0.094 | 0.126 | 0.02 | 0.02 | -0.008 | 83 | 64 |
| **GATE** | *20* | | 0.036 | 0.098 | 0.094 | 0.126 | 0.03 | 0.02 | -0.007 | 83 | 64 |
| **GATE** | *40* | | 0.036 | 0.099 | 0.095 | 0.128 | 0.03 | 0.05 | -0.008 | 83 | 64 |
| **GATE** | *5* | *grf* | 0.097 | 0.113 | 0.096 | 0.137 | 0.03 | -0.00 | -0.003 | 81 | 57 |
| **GATE** | *10* | | 0.097 | 0.134 | 0.134 | 0.166 | 0.02 | 0.04 | -0.002 | 88 | 68 |
| **GATE** | *20* | | 0.097 | 0.171 | 0.189 | 0.213 | -0.01 | 0.05 | -0.002 | 92 | 74 |
| **GATE** | *40* | | 0.097 | 0.228 | 0.268 | 0.285 | -0.01 | 0.09 | -0.002 | 93 | 77 |
| **GATE** | *5* | *grf* | 0.037 | 0.081 | 0.093 | 0.102 | -0.01 | 0.07 | -0.000 | 93 | 76 |
| **GATE** | *10* | *cent* | 0.037 | 0.110 | 0.132 | 0.138 | -0.02 | 0.07 | -0.001 | 94 | 78 |
| **GATE** | *20* | | 0.037 | 0.151 | 0.186 | 0.190 | 0.02 | 0.10 | 0.000 | 95 | 79 |
| **GATE** | *40* | | 0.037 | 0.213 | 0.265 | 0.268 | -0.01 | 0.08 | -0.001 | 95 | 79 |
| **GATE** | *5* | *dml* | 0.019 | 0.081 | 0.099 | 0.101 | -0.04 | -0.05 | 0.001 | 95 | 80 |
| **GATE** | *10* | | 0.019 | 0.113 | 0.140 | 0.142 | -0.04 | -0.06 | 0.001 | 95 | 79 |
| **GATE** | *20* | | 0.019 | 0.160 | 0.200 | 0.201 | -0.02 | 0.05 | -0.001 | 95 | 79 |
| **GATE** | *40* | | 0.019 | 0.227 | 0.284 | 0.285 | -0.02 | 0.07 | -0.005 | 94 | 79 |
| **GATE** | *5* | *dml-* | 0.020 | 0.081 | 0.081 | 0.101 | -0.04 | -0.05 | 0.001 | 95 | 79 |
| **GATE** | *10* | *norm* | 0.020 | 0.113 | 0.140 | 0.141 | -0.04 | -0.07 | 0.000 | 95 | 79 |
| **GATE** | *20* | | 0.020 | 0.160 | 0.199 | 0.200 | -0.02 | 0.06 | -0.002 | 94 | 79 |
| **GATE** | *40* | | 0.020 | 0.226 | 0.283 | 0.284 | -0.02 | 0.07 | -0.005 | 94 | 79 |
| **GATE** | *5* | *ols* | 0.007 | 0.080 | 0.094 | 0.100 | -0.06 | 0.01 | -0.027 | 81 | 61 |
| **GATE** | *10* | | 0.006 | 0.110 | 0.133 | 0.138 | -0.01 | 0.04 | -0.039 | 82 | 62 |
| **GATE** | *20* | | 0.006 | 0.154 | 0.191 | 0.194 | -0.01 | 0.07 | -0.057 | 82 | 62 |
| **GATE** | *40* | | 0.006 | 0.219 | 0.273 | 0.275 | 0.02 | 0.06 | -0.083 | 82 | 62 |

**Note:** Table to be continued.

Table C.29 - continued: More GATE groups, medium selectivity

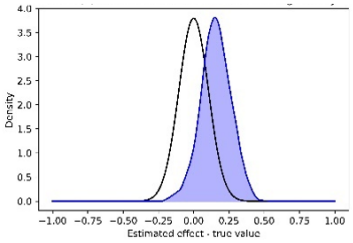
| | | | Estimation of effects | | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | N = 10'000 | | | | | | |
| **GATE** | *5* | *mcf* | 0.144 | 0.144 | 0.048 | 0.158 | 0.16 | -0.03 | -0.001 | 21 | 7 |
| **GATE** | *10* | | 0.143 | 0.143 | 0.050 | 0.159 | 0.05 | -0.06 | -0.002 | 25 | 8 |
| **GATE** | *20* | | 0.142 | 0.143 | 0.051 | 0.158 | 0.07 | -0.07 | -0.001 | 26 | 10 |
| **GATE** | *40* | | 0.142 | 0.143 | 0.52 | 0.159 | 0.09 | -0.08 | -0.002 | 26 | 11 |
| **GATE** | *5* | *mcf* | 0.013 | 0.049 | 0.045 | 0.063 | 0.22 | 0.17 | 0.004 | 86 | 69 |
| **GATE** | *10* | *cent* | 0.013 | 0.054 | 0.049 | 0.069 | 0.09 | 0.10 | -0.002 | 84 | 63 |
| **GATE** | *20* | | 0.012 | 0.054 | 0.050 | 0.068 | 0.10 | 0.06 | -0.001 | 85 | 64 |
| **GATE** | *40* | | 0.012 | 0.055 | 0.051 | 0.069 | 0.12 | 0.03 | -0.002 | 84 | 64 |
| **GATE** | *5* | *grf* | 0.064 | 0.068 | 0.047 | 0.080 | -0.03 | -0.15 | -0.001 | 71 | 46 |
| **GATE** | *10* | | 0.064 | 0.076 | 0.067 | 0.093 | 0.00 | -0.07 | -0.001 | 84 | 60 |
| **GATE** | *20* | | 0.064 | 0.092 | 0.094 | 0.114 | -0.04 | -0.04 | -0.001 | 89 | 69 |
| **GATE** | *40* | | 0.064 | 0.119 | 0.133 | 0.148 | -0.03 | -0.00 | -0.001 | 92 | 75 |
| **GATE** | *5* | *grf* | 0.016 | 0.041 | 0.048 | 0.051 | -0.04 | -0.03 | -0.001 | 92 | 75 |
| **GATE** | *10* | *cent* | 0.016 | 0.055 | 0.065 | 0.068 | -0.05 | -0.09 | -0.000 | 94 | 78 |
| **GATE** | *20* | | 0.016 | 0.075 | 0.092 | 0.094 | -0.01 | 0.02 | 0.000 | 95 | 79 |
| **GATE** | *40* | | 0.016 | 0.106 | 0.131 | 0.132 | -0.01 | -0.04 | 0.000 | 95 | 79 |
| **GATE** | *5* | *dml* | 0.010 | 0.041 | 0.050 | 0.051 | 0.14 | 0.04 | 0.000 | 94 | 80 |
| **GATE** | *10* | | 0.010 | 0.056 | 0.069 | 0.070 | 0.05 | 0.05 | 0.001 | 94 | 80 |
| **GATE** | *20* | | 0.010 | 0.081 | 0.100 | 0.101 | -0.04 | 0.01 | -0.001 | 94 | 79 |
| **GATE** | *40* | | 0.010 | 0.113 | 0.140 | 0.141 | -0.01 | -0.08 | -0.001 | 95 | 79 |
| **GATE** | *5* | *dml-* | 0.010 | 0.041 | 0.050 | 0.051 | 0.13 | 0.04 | 0.000 | 94 | 80 |
| **GATE** | *10* | *norm* | 0.010 | 0.056 | 0.069 | 0.070 | 0.05 | 0.05 | 0.001 | 94 | 80 |
| **GATE** | *20* | | 0.010 | 0.081 | 0.100 | 0.101 | -0.04 | 0.01 | -0.001 | 94 | 79 |
| **GATE** | *40* | | 0.010 | 0.113 | 0.140 | 0.141 | -0.01 | -0.08 | -0.001 | 95 | 79 |
| **GATE** | *5* | *ols* | 0.007 | 0.046 | 0.047 | 0.058 | 0.09 | 0.22 | -0.014 | 74 | 53 |
| **GATE** | *10* | | 0.007 | 0.059 | 0.066 | 0.075 | 0.05 | 0.02 | -0.019 | 75 | 58 |
| **GATE** | *20* | | 0.007 | 0.081 | 0.095 | 0.101 | -0.03 | -0.03 | -0.029 | 80 | 60 |
| **GATE** | *40* | | 0.007 | 0.111 | 0.134 | 0.139 | 0.00 | -0.05 | -0.040 | 81 | 61 |

**Note:** For GATE, the results are averaged over all effects. *CovP(95, 80)* denotes the (average) probability that the true value is part of the 95%/80% confidence interval. 1'000 / 500 replications used for 2'500 / 10'000 obs.

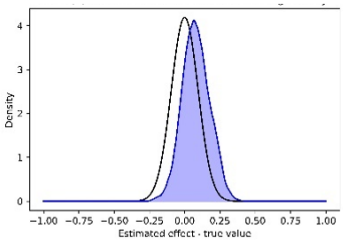
Table C.29 - continued: More GATE groups, medium selectivity - Distribution of GATE minus true value of the centred mcf

N = 2'500

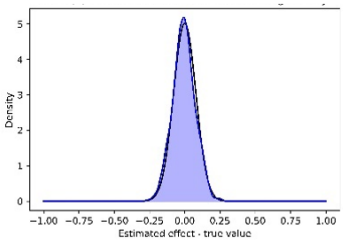
True GATE: 0.50



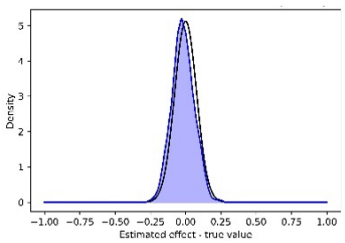

True GATE: 0.77



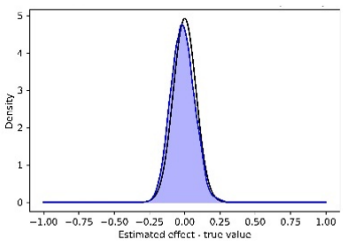

True GATE: 1.20




True GATE: 1.24

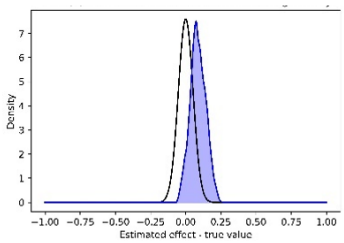



True GATE: 1.24

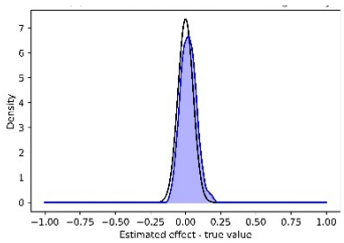



N = 10'000

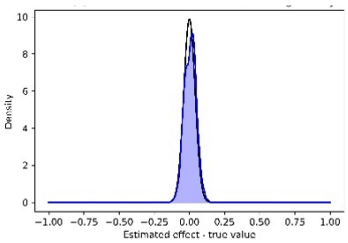
True GATE: 0.50



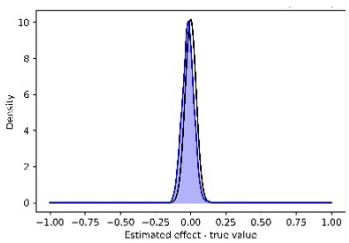

True GATE: 0.77




True GATE: 1.20




True GATE: 1.24




True GATE: 1.24

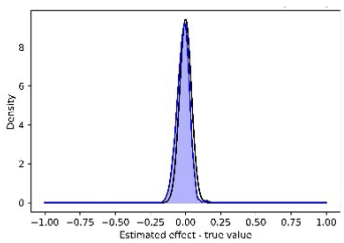


Table C.30: More GATE groups, strong selectivity

| | | | Estimation of effects | | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | N = 2'500 | | | | | | |
| **GATE** | *5* | *mcf* | 0.436 | 0.436 | 0.010 | 0.455 | 0.00 | 0.06 | 0.001 | 1 | 0 |
| **GATE** | *10* | | 0.435 | 0.435 | 0.097 | 0.455 | 0.00 | 0.03 | 0.000 | 2 | 0 |
| **GATE** | *20* | | 0.434 | 0.434 | 0.097 | 0.439 | 0.02 | -0.03 | -0.001 | 0 | 0 |
| **GATE** | *40* | | 0.435 | 0.435 | 0.064 | 0.439 | 0.02 | -0.04 | -0.001 | 0 | 0 |
| **GATE** | *5* | *mcf* | 0.046 | 0.112 | 0.085 | 0.146 | -0.03 | 0.11 | 0.009 | 81 | 62 |
| **GATE** | *10* | *cent* | 0.044 | 0.117 | 0.089 | 0.151 | -0.03 | 0.07 | 0.007 | 80 | 61 |
| **GATE** | *20* | | 0.044 | 0.117 | 0.089 | 0.152 | -0.01 | 0.05 | 0.008 | 80 | 62 |
| **GATE** | *40* | | 0.043 | 0.119 | 0.090 | 0.153 | 0.00 | 0.06 | 0.007 | 80 | 61 |
| **GATE** | *5* | *grf* | 0.284 | 0.284 | 0.101 | 0.309 | -0.06 | 0.06 | -0.011 | 19 | 8 |
| **GATE** | *10* | | 0.283 | 0.287 | 0.136 | 0.322 | -0.03 | 0.02 | -0.009 | 42 | 21 |
| **GATE** | *20* | | 0.284 | 0.297 | 0.187 | 0.346 | -0.04 | 0.03 | -0.006 | 63 | 39 |
| **GATE** | *40* | | 0.283 | 0.324 | 0.260 | 0.391 | -0.06 | 0.09 | -0.004 | 79 | 55 |
| **GATE** | *5* | *grf* | 0.090 | 0.119 | 0.096 | 0.148 | 0.00 | 0.09 | -0.007 | 75 | 56 |
| **GATE** | *10* | *cent* | 0.090 | 0.138 | 0.130 | 0.173 | -0.02 | 0.06 | -0.005 | 84 | 64 |
| **GATE** | *20* | | 0.090 | 0.171 | 0.181 | 0.214 | 0.00 | 0.09 | -0.003 | 89 | 71 |
| **GATE** | *40* | | 0.090 | 0.223 | 0.255 | 0.280 | -0.03 | 0.10 | -0.003 | 92 | 75 |
| **GATE** | *5* | *dml* | 0.147 | 0.165 | 0.127 | 0.195 | -0.61 | 4.66 | -0.014 | 66 | 46 |
| **GATE** | *10* | | 0.147 | 0.187 | 0.173 | 0.228 | -0.78 | 8.17 | -0.017 | 78 | 57 |
| **GATE** | *20* | | 0.147 | 0.226 | 0.241 | 0.284 | -0.99 | 12.63 | -0.025 | 84 | 64 |
| **GATE** | *40* | | 0.148 | 0.286 | 0.338 | 0.372 | -1.07 | 14.96 | -0.040 | 87 | 69 |
| **GATE** | *5* | *dml-* | 0.140 | 0.157 | 0.123 | 0.187 | -0.13 | 0.01 | -0.005 | 74 | 52 |
| **GATE** | *10* | *norm* | 0.140 | 0.182 | 0.170 | 0.221 | -0.15 | 0.08 | -0.005 | 82 | 62 |
| **GATE** | *20* | | 0.140 | 0.225 | 0.239 | 0.278 | -0.21 | 0.34 | -0.010 | 87 | 68 |
| **GATE** | *40* | | 0.140 | 0.293 | 0.338 | 0.367 | -0.29 | 0.76 | -0.020 | 89 | 71 |
| **GATE** | *5* | *ols* | 0.111 | 0.112 | 0.054 | 0.123 | -0.02 | 0.01 | -0.015 | 25 | 13 |
| **GATE** | *10* | | 0.109 | 0.153 | 0.143 | 0.188 | -0.04 | -0.04 | -0.042 | 69 | 49 |
| **GATE** | *20* | | 0.109 | 0.190 | 0.202 | 0.236 | 0.00 | 0.00 | -0.060 | 75 | 55 |
| **GATE** | *40* | | 0.109 | 0.250 | 0.288 | 0.313 | 0.01 | 0.08 | -0.087 | 79 | 58 |

**Note:** Table to be continued.

Table C.30 - continued: More GATE groups, strong selectivity

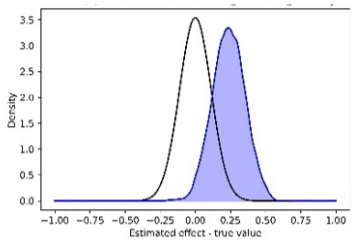
| | | | Estimation of effects | | | | | | Estimation of std. errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *# of groups* | Estim-ator | Bias | Mean abs. error | Std. dev. | RMSE | Skew-ness | Ex. Kurto-sis | Bias (SE) | CovP (95) in % | CovP (80) in % |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| | | | | | N = 10'000 | | | | | | |
| **GATE** | *5* | *mcf* | 0.375 | 0.375 | 0.053 | 0.384 | -0.10 | -0.10 | -0.002 | 0 | 0 |
| **GATE** | *10* | | 0.374 | 0.374 | 0.056 | 0.383 | -0.09 | -0.09 | -0.002 | 0 | 0 |
| **GATE** | *20* | | 0.373 | 0.373 | 0.056 | 0.383 | -0.09 | -0.08 | -0.002 | 0 | 0 |
| **GATE** | *40* | | 0.373 | 0.373 | 0.057 | 0.383 | -0.08 | -0.06 | -0.002 | 0 | 0 |
| **GATE** | *5* | *mcf* | -0.014 | 0.085 | 0.045 | 0.100 | 0.01 | 0.03 | 0.007 | 65 | 41 |
| **GATE** | *10* | *cent* | -0.016 | 0.091 | 0.049 | 0.104 | -0.05 | -0.04 | 0.005 | 62 | 35 |
| **GATE** | *20* | | -0.016 | 0.090 | 0.049 | 0.104 | -0.03 | 0.02 | 0.006 | 63 | 37 |
| **GATE** | *40* | | -0.016 | 0.091 | 0.050 | 0.104 | -0.02 | -0.02 | 0.005 | 63 | 37 |
| **GATE** | *5* | *grf* | 0.211 | 0.211 | 0.052 | 0.220 | 0.00 | -0.04 | -0.005 | 3 | 0 |
| **GATE** | *10* | | 0.211 | 0.211 | 0.071 | 0.225 | 0.01 | 0.02 | -0.006 | 15 | 5 |
| **GATE** | *20* | | 0.211 | 0.212 | 0.096 | 0.234 | -0.01 | 0.01 | -0.004 | 39 | 19 |
| **GATE** | *40* | | 0.211 | 0.219 | 0.134 | 0.252 | -0.06 | 0.04 | -0.004 | 63 | 37 |
| **GATE** | *5* | *grf* | 0.043 | 0.063 | 0.051 | 0.080 | -0.09 | 0.03 | -0.006 | 75 | 56 |
| **GATE** | *10* | *cent* | 0.043 | 0.073 | 0.068 | 0.092 | -0.06 | 0.17 | -0.004 | 83 | 63 |
| **GATE** | *20* | | 0.043 | 0.089 | 0.093 | 0.112 | -0.08 | 0.08 | -0.002 | 89 | 71 |
| **GATE** | *40* | | 0.043 | 0.115 | 0.130 | 0.145 | -0.02 | -0.03 | -0.002 | 92 | 74 |
| **GATE** | *5* | *dml* | 0.088 | 0.099 | 0.091 | 0.119 | 0.10 | 6.78 | -0.015 | 62 | 42 |
| **GATE** | *10* | | 0.088 | 0.110 | 0.104 | 0.138 | 0.13 | 8.59 | -0.017 | 74 | 53 |
| **GATE** | *20* | | 0.088 | 0.132 | 0.144 | 0.173 | -0.21 | 10.79 | -0.024 | 83 | 61 |
| **GATE** | *40* | | 0.088 | 0.164 | 0.198 | 0.224 | -0.62 | 11.75 | -0.032 | 87 | 68 |
| **GATE** | *5* | *dml-* | 0.078 | 0.090 | 0.073 | 0.108 | -0.08 | 0.04 | -0.007 | 71 | 49 |
| **GATE** | *10* | *norm* | 0.079 | 0.103 | 0.098 | 0.127 | -0.09 | 0.30 | -0.006 | 81 | 60 |
| **GATE** | *20* | | 0.079 | 0.129 | 0.138 | 0.160 | -0.22 | 1.02 | -0.010 | 87 | 67 |
| **GATE** | *40* | | 0.079 | 0.164 | 0.192 | 0.209 | -0.35 | 1.75 | -0.015 | 89 | 71 |
| **GATE** | *5* | *ols* | 0.113 | 0.118 | 0.052 | 0.134 | 0.05 | -0.07 | -0.015 | 27 | 17 |
| **GATE** | *10* | | 0.110 | 0.121 | 0.071 | 0.142 | -0.02 | -0.13 | -0.021 | 42 | 27 |
| **GATE** | *20* | | 0.110 | 0.132 | 0.100 | 0.158 | 0.02 | -0.07 | -0.029 | 56 | 38 |
| **GATE** | *40* | | 0.110 | 0.153 | 0.141 | 0.187 | -0.01 | -0.06 | -0.042 | 67 | 47 |

**Note:** For GATE, the results are averaged over all effects. *CovP(95, 80)* denotes the (average) probability that the true value is part of the 95%/80% confidence interval. 1'000 / 500 replications used for 2'500 / 10'000 obs.

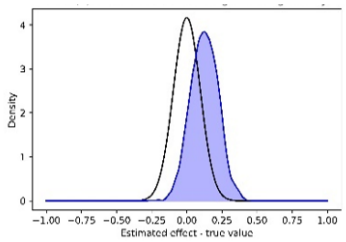
Table C.30 - continued: More GATE groups, strong selectivity - Distribution of GATE minus true value of the centred mcf

N = 2'500

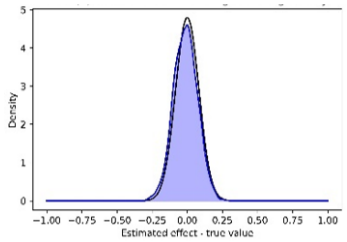
True GATE: 0.50



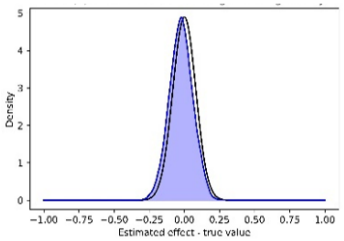

True GATE: 0.77



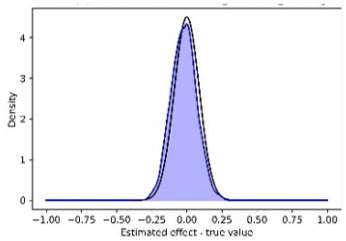

True GATE: 1.20




True GATE: 1.24

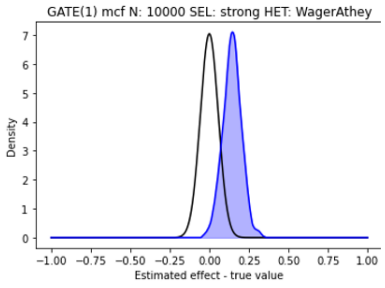



True GATE: 1.24

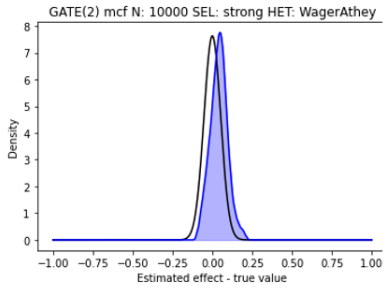



N = 10'000

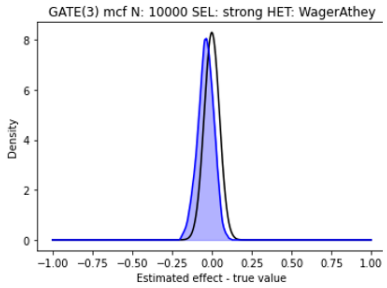
True GATE: 0.50



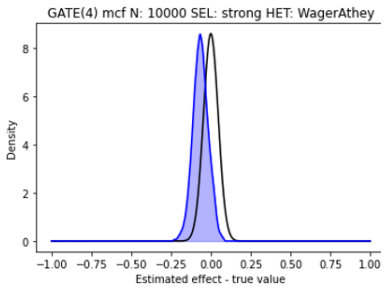

True GATE: 0.77




True GATE: 1.20




True GATE: 1.24




True GATE: 1.24