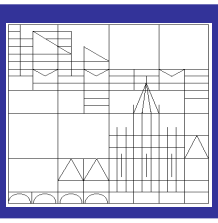
University of Konstanz
Department of Economics



A Simple and Successful Method to Shrink the Weight

Winfried Pohlmeier, Ruben R. Seiberlich, and S. Derya Uysal

Working Paper Series
2013-05

http://www.wiwi.uni-konstanz.de/econdoc/working-paper-series/

A Simple and Successful Method to Shrink the Weight

Winfried Pohlmeier
University of Konstanz

Ruben R. Seiberlich*
University of Konstanz

S. Derya Uysal
IHS Vienna

This draft: March 25, 2013

Abstract

We propose a simple way to improve the efficiency of the average treatment effect on propensity score based estimators. As the weights become arbitrarily large for the propensity scores being close to one or zero, we propose to shrink the propensity scores away from these boundaries. Using a comprehensive Monte Carlo study we show that this simple method substantially reduces the mean squared error of the estimators in finite samples.

JEL classification: C14, C21
Keywords: Econometric evaluation, propensity score, penalizing, shrinkage, average treatment effect

*Corresponding author. Department of Economics, Box D124, University of Konstanz, 78457 Konstanz, Germany. Phone ++49-7531-88-5111, Fax -4450, email: Ruben.Seiberlich@uni-konstanz.de. Financial support by the DFG research group "Psychoeconomics" is gratefully acknowledged. This paper benefited from helpful ideas and discussions with Bertrand Koebel, Fabian Krüger and Bernd Fitzenberger. Furthermore, we thank the participants at the DFH Workshop "Applied Econometrics" 2012, Königsfeld, Germany, the participants at the Workshop of the German Statistical Association (DStatG) 2012, Vienna, Austria as well as the session participants at the European Meeting of the Econometric Society 2012, Málaga, Spain, for their helpful comments.

1 Introduction

In this paper we introduce a simple way of improving propensity score weighting and double robust estimators in terms of mean squared error (MSE) in finite samples. Our approach achieves a lower MSE by shrinking the propensity score towards the share of treated. This Stein-type simple shrinkage substantially mitigates the problems arising from propensity score estimates close to the boundaries. This reduces the variance of the weights and, therefore, the variance of the average treatment effect (ATE) estimators based on propensity score weighting. The shrinkage approach proposed is an attractive alternative to the popular trimming strategies applied to reduce the impact of large weights and can also be used jointly with trimming.

Even though shrinkage methods are very popular in other areas of statistics and econometrics, they have not been combined with weighting estimators yet. A notable exception is Frölich (2004), who uses the ridging method of Seifert and Gasser (1996) for matching estimators of the average treatment effect on the treated. They propose ridging of local polynomials to overcome the problems that arise in estimating a regression function when the conditional variance is unbounded.

The proposed shrinkage method is a linear combination of the conditional mean of the treatment variable and its unconditional mean. Like other shrinkage methods, the degree of shrinkage is determined by a tuning parameter. We propose three different methods to choose this parameter such that certain optimality conditions are satisfied. First, we consider a simple fixed valued tuning parameter, which only depends on the sample size. Second, we minimize the MSE of our linear combination to choose the optimal value. Third, we propose a pure cross validation procedure to obtain the optimal tuning parameter.

We demonstrate the MSE gains in finite samples via a comprehensive Monte Carlo study. To make our results comparable, we design our Monte Carlo study as in the settings of Busso et al. (2009) for poor overlap. We construct 72 settings to capture several possible issues when estimating the treatment effects and consider homogeneous and heterogenous treatment, homoscedastic and heteroscedastic error term as well as different ratios of treated and control group. Moreover, the simulation design captures different functional forms. Since the shrunken propensity scores are constructed in such a way that they converge to the conventional propensity scores,

our proposed method is asymptotically equivalent to standard approaches without shrinkage. Therefore, we focus on sample sizes 100, 200 and 500 only. Additionally, we evaluate the finite sample performance with and without applying trimming rules.

Our results show that the estimators based on the shrunken propensity scores have a lower MSE than the weighting estimators based on the unshrunken propensity scores in all of the settings if we use the fixed valued or the MSE minimizing tuning parameter. For the cross validated tuning parameter, the MSE is reduced in 99.3% of the cases, respectively. If a trimming rule is applied to the proposed approach, we are able to decrease the MSE of the ATE in 99.7% of the cases for the fixed valued tuning parameter. For the MSE minimizing and cross validated tuning parameter, the MSE is reduced in 98.8% and 96.9% of the cases, respectively. In the rare cases where the MSE is not improved, the efficiency loss is very small.

The paper is organized as follows. Section (2) reviews the different weighting and double robust estimators. Section (3) introduces the shrunken propensity score and derives its properties in finite samples. In section (4), we present a Monte Carlo study and compare the MSE of the estimators based on the shrunken and classical propensity score weights. Section (5) concludes.

2 Propensity Score Methods

Consider the case of a binary treatment within Rubin's 1974 potential outcome model.[1] Let Y_{1i} and Y_{0i} be the two potential outcomes for person i if she takes the treatment and if she does not take the treatment, respectively. D_i denotes the binary treatment indicator indicating whether person i participates in the program ($D_i = 1$) or not ($D_i = 0$). The observed outcome variable, Y_i, can than be written as a function of potential outcomes and the treatment variable:

$$Y_i = D_i Y_{1i} + (1 - D_i) Y_{0i} \quad \text{for } i = 1, \ldots, n. \tag{1}$$

The difference between two potential outcomes of an individual, $Y_{1i} - Y_{0i}$, denotes the individual's treatment effect. Depending on the realized treatment status, we only observe one of the two potential outcomes. Hence, the individual treatment

[1] See Imbens and Wooldridge (2009) for advantages of potential outcome model over observed outcome models.

effect cannot be identified from observed data. Under certain assumptions, however, we can still identify various average treatment effects. In this paper, we focus on the average treatment effect (ATE) defined as

$$\Delta_{\text{ATE}} = E[Y_{1i} - Y_{0i}], \tag{2}$$

which measures the expected treatment effect if individuals are randomly assigned to treatment and control groups.

The identification of the ATE crucially depends on two assumptions. The first one is that, conditional on confounding variables, the potential outcomes are stochastically independent of the treatment: $Y_{0i}, Y_{1i} \perp D_i | X_i$, where X_i denotes the confounding variables of individual i. This assumption, known as the Conditional Independence Assumption (CIA), requires that all confounding factors associated with the potential outcomes as well as the participation decision are observed. If the CIA is satisfied, various estimation methods (e.g. weighting, regression and matching methods) are feasible to estimate the ATE.

The second assumption is the overlap assumption. It requires that the probability of receiving the treatment, the so-called propensity score, lies strictly between zero and one. In other words, each unit in a defined population has a positive probability of being treated and of not being treated. Although this type of overlap assumption is standard in the literature (e.g. Rosenbaum and Rubin (1983), Heckman et al. (1997), Hahn (1998), Wooldridge (2002), Imbens (2004)), there is a stronger version of the overlap assumption called "strict overlap" (e.g Robins et al. (1994), Abadie and Imbens (2006), Crump et al. (2009)). Strict overlap requires that the probability of being treated is strictly between ξ and $1 - \xi$ for some $\xi > 0$. Khan and Tamer (2010) point out that a comparable assumption to the strict overlap assumption is needed for $\sqrt{N}$ - convergence of some semiparametric estimators. Busso et al. (2009) provide further evidence on the importance of the (strict) overlap assumption.

Under the assumptions listed above, the ATE can be identified and estimated. There are several estimation methods proposed in the literature. We, however, focus only on the methods which use the propensity scores as weights. The propensity score, i.e. the probability of being treated conditional on the characteristics X_i, is given by

$$p_i = \Pr\left[D_i = 1 \,|X_i\right]. \tag{3}$$

As the propensity score is an unknown probability, it has to estimated. Conventionally, standard parametric maximum likelihood methods are used to obtain the estimated propensity score and are denoted by $\hat{p}_i$.

Following Busso et al. (2009), we write the weighting type estimator for the ATE as follows:

$$\hat{\Delta}_{ATE} = \frac{1}{n_1}\sum_{i=1}^{n} D_i Y_i \hat{\omega}_{i1} - \frac{1}{n_0}\sum_{i=1}^{n}(1-D_i) Y_i \hat{\omega}_{i0}, \tag{4}$$

where n_1 is the number of treated observations and n_0 is the number of controls. $\hat{\omega}_{i0}$ and $\hat{\omega}_{i1}$ are defined differently for different types weighting estimators. We consider here three different weighting schemes proposed in the literature. The first one (IPW1) uses the following weighting functions:

$$\hat{\omega}_{i0}^{(1)} = \frac{n_0}{n}\Big/(1-\hat{p}_i) \tag{5}$$

$$\hat{\omega}_{i1}^{(1)} = \frac{n_1}{n}\Big/\hat{p}_i, \tag{6}$$

where n is the total number of observations. The second weighting function (IPW2) results from an adjustment to force the weights to add up to one and is advocated by Imbens (2004). Formally, they are given by

$$\hat{\omega}_{i0}^{(2)} = \frac{1}{1-\hat{p}_i}\Big/\frac{1}{n_0}\sum_{i=1}^{n}\frac{1-D_i}{1-\hat{p}_i} \tag{7}$$

$$\hat{\omega}_{i1}^{(2)} = \frac{1}{\hat{p}_i}\Big/\frac{1}{n_1}\sum_{i=1}^{n}\frac{D_i}{\hat{p}_i}. \tag{8}$$

The third weighting function (IPW3), which is not so common in the literature, is a combination of the first two methods, where the asymptotic variance of the resulting estimator is minimized for a known propensity score (see Lunceford and Davidian (2004) for details).

$$\hat{\omega}_{i0}^{(3)} = \frac{1}{1-\hat{p}_i}(1-C_{i0})\Big/\frac{1}{n_0}\sum_{i=1}^{n}\frac{(1-D_i)}{1-\hat{p}_i}(1-C_{i0}) \tag{9}$$

$$\hat{\omega}_{i1}^{(3)} = \frac{1}{\hat{p}_i}(1-C_{i1})\Big/\frac{1}{n_1}\sum_{i=1}^{n}\frac{D_i}{\hat{p}_i}(1-C_{i1}) \tag{10}$$

with

$$C_{i0} = \frac{\frac{1}{1-\hat{p}_i}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1-D_i}{1-\hat{p}_i}\hat{p}_i - D_i\right)}{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1-D_i}{1-\hat{p}_i}\hat{p}_i - D_i\right)^2} \tag{11}$$

$$C_{i1} = \frac{\frac{1}{\hat{p}_i}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{D_i}{\hat{p}_i}(1-\hat{p}_i) - (1-D_i)\right)}{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{D_i}{\hat{p}_i}(1-\hat{p}_i) - (1-D_i)\right)^2} \tag{12}$$

In all three cases, $\hat{\omega}_{i0}$ depends on $\frac{1}{1-\hat{p}_i}$ and $\hat{\omega}_{i1}$ on $\frac{1}{\hat{p}_i}$. If the estimated propensity score for individual i is close to one $\hat{\omega}_{i0}$, the weight for individual i, is large compared to the weights of the other observations. Therefore, the estimates are mainly determined by individual i. If the estimated propensity score is close to zero, $\hat{\omega}_{i1}$ is large, which again leads to an ATE estimator which exhibits high variance.

The doubly robust estimator of the ATE we consider here is derived from a weighted regression of the outcome model where the weights are inversely related to the propensity scores. The advantage of doubly robust estimation is that it stays consistent even if the outcome model or the propensity score model is specified incorrectly. It has been shown that doubly robust methods are more efficient than weighting methods (see, for example, Robins and Rotnitzky (1995), Wooldridge (2007)). Here, we consider the doubly robust method used by Hirano and Imbens (2001). They estimate the ATE by a weighted least square regression of the following outcome model with weights based on Equation (5):

$$Y_i = \alpha_0 + \Delta_{ATE} D_i + X_i'\alpha_1 + D_i(X_i - \bar{X})'\alpha_2 + \varepsilon_i \tag{13}$$

$$\hat{\omega}_i^{dr} = \sqrt{\frac{D_i}{\hat{p}_i} + \frac{1-D_i}{1-\hat{p}_i}}, \tag{14}$$

where $\bar{X}$ is the sample average of X_i.

The weight $\hat{\omega}_i^{dr}$ again depends on $\frac{1}{1-\hat{p}_i}$ and $\frac{1}{\hat{p}_i}$, such that propensity scores close to one and zero have a similar effect as for the weighting estimators.

3 Shrunken Weights

A major drawback of the weighting and double robust estimators is that they can exhibit a high variance if the weights of some observations are very large. For the ATE, this is the case if the propensity score is close to one or zero. We propose different variants of Stein-type simple shrinkage methods for the propensity score which help to stabilize the treatment effect estimators by shrinking the propensity scores away from these boundaries.

The basic idea is to shrink the estimated propensity score, $\hat{p}_i = \hat{E}[D_i = 1|X_i = x]$, towards the estimated unconditional mean $\bar{D} = \frac{1}{n}\sum_i^n D_i$, i.e.

$$\hat{p}_i^s = \big(1 - \lambda_i(n)\big)\hat{p}_i + \lambda_i(n)\bar{D}, \tag{15}$$

where $0 \leq \lambda_i(n) \leq 1$ is the tuning parameter, which may depend on the sample size. Equation (15) implies that our proposed shrunken propensity score is always closer to the share of treated and, therefore, the shrunken propensity scores have a lower variance than the conventional propensity scores. This enables us to estimate the treatment effects with a lower MSE.

Shrinking towards the unconditional mean avoids propensity scores close to one or zero. The usual way to deal with propensity scores that are too small or too large is to apply a trimming rule. A trimming rule basically determines an upper and a lower limit for the propensity score. Observations with propensity scores outside of the chosen limits are dropped from the estimation sample.[2] Obviously, dropping observations will cause both a loss of information as well as a loss of efficiency. Since shrinkage pushes the estimated propensity score away from the boundaries, we neither need to apply a trimming rule nor do we have to work with a reduced sample.

Shrinkage could also be applied directly to the parameter estimates of the propensity score by imposing a L_1-norm (Lasso) or L_2-norm (ridging) on the parameter estimates. Besides the higher computational burden, the interpretation of the shrinkage parameter in terms of the shrunken propensity score is, however, not straight forward. Moreover, our approach avoids shrinking a propensity score from one extreme to the other, i.e. we avoid shrinking a large propensity score $\hat{p}_i > \bar{D}$ towards

[2]In the following section, we explain two trimming rules which are often used in the literature. For other trimming rules see Busso et al. (2009).

zero, since $\hat{p}_i^s > \bar{D}$ always holds.

A crucial issue for any shrinkage estimator is the choice of the tuning parameter $\lambda_i(n)$. As we are interested in improving the small sample performance of weighting and double robust estimators, we propose to choose $\lambda_i(n)$ such that the penalty vanishes asymptotically. For $\lambda_i(n) = \mathcal{O}(n^{-\delta})$ with $\delta > 0$, the shrinkage estimator is consistent and converges to the true propensity score. For $\delta > 1/2$, the $\hat{p}_i^s$ has the same asymptotic distribution as the conventional propensity score $\hat{p}_i$.

In the following, we consider three alternative methods of choosing $\lambda_i(n)$. The first method, the fixed tuning parameter method, is based on the functional form $\lambda_i(n) = \frac{c}{n^\delta}$. For a given values of c and δ, this method is easy to implement with no computational cost but is not optimized with respect to any MSE criterion.

In the second method, the MSE minimizing tuning parameter method, $\lambda_i(n)$ is determined my minimizing the MSE of the shrunken propensity score in (15). Thus, the optimal $\lambda_i(n)$ is given by

$$\lambda_i^*(n) = \frac{\mathrm{V}\left[\hat{p}_i\right] - \mathrm{Cov}(\hat{p}_i, \bar{D})}{\mathrm{V}\left[\hat{p}_i\right] + \frac{\mathrm{E}[D_i](1-\mathrm{E}[D_i])}{n} + \left(\mathrm{E}\left[D_i\right] - \mathrm{E}\left[\hat{p}_i\right]\right)^2 - 2\,\mathrm{Cov}(\hat{p}_i, \bar{D})}, \tag{16}$$

where we assume $\mathrm{E}\left[\hat{p}_i\right] \approx p_i$. Since $\lambda_i^*(n)$ depends on unknown parameters, we replace the squared bias, variance and covariances by their the bootstrapped quantities. As for the tuning parameter in the first method, the MSE minimizing tuning parameter (16) also converges to zero as the sample size increases. Note that the MSE minimizing tuning parameter method yields different optimal λ's for each observation in the sample. In order to stabilize the estimation resulting from the estimation noise, the estimated MSE minimizing tuning parameter can be replaced by the mean MSE minimizing tuning parameter, $\bar{\lambda}^*(n) = \frac{1}{n}\sum_{i=1}^n \lambda_i^*(n)$.[3] Additionally, this guarantees that the ordering of the propensity scores does not change.

In the third method, the optimal λ is chosen by means of cross-validation. The idea is to minimize the mean squared prediction error of the estimated propensity score with respect to λ. The mean squared prediction error is calculated by

[3] An alternative would be to choose λ such that the MSE of the vector of the shrunken propensity scores is minimized. The results are comparable to those obtained by using $\bar{\lambda}^*(n)$ and are available upon request.

leave-one-out cross validation for each λ in an equally spaced grid of $k+1$ λ's, i.e. $[0, \lambda_{(1)}, \ldots, \lambda_{(k-1)}, 1]$. The optimal λ is then the one which leads to the smallest cross validated mean squared prediction error. This method again yields a different λ's for each observation, but is computationally less burdensome than the MSE minimizing choice of λ on Method 2.

4 Monte Carlo Study

We demonstrate the efficiency gains due to propensity score shrinkage via a comprehensive Monte Carlo study. We adopt the same data generating processes as Busso et al. (2009) to make our results comparable with theirs. Since our approach shrinks the propensity score towards the treated-to-control ratio, it is especially valuable in situations where the overlap but not the strict overlap assumption is fulfilled. In the following, we, therefore, concentrate on those designs of Busso et al. (2009), which are not consistent with the strict overlap assumption. In settings where the strict overlap is also fulfilled, our approach also improves the MSE of the different estimators.[4] For the simulation study, D and Y are generated as follows:

$$D_i = \mathbb{1}\{\eta + \kappa X_i - u_i > 0\} \tag{17}$$

$$Y_i = D_i + m(p(X_i)) + \gamma D_i m(p(X_i)) + \varepsilon_i, \tag{18}$$

where the error terms u_i and ε_i are independent of each other and the cofounding variable X_i which is assumed to be a standard normally distributed random variable. $p(X_i)$ is the propensity score and $m(\cdot)$ is a function of the propensity score. We use two different functions in the Monte Carlo study given in Table (1).

Table 1: Functional form for $m(q)$

$m(q)$	Formula	Description
$m_1(q)$	$0.15 + 0.7q$	Linear
$m_2(q)$	$0.2 + \sqrt{1-q} - 0.6(0.9-q)^2$	Nonlinear

The error term u_i is drawn from a standard normal distribution leading to the following propensity score function:

$$p(X_i) = \Phi(\eta + \kappa X_i). \tag{19}$$

[4]The results for these settings are available upon request.

We generate various treated-to-control ratios by choosing three different combinations for η and κ. Table (2) summarizes the parameter values and resulting ratios.

Table 2: Treated-to-control ratios

η	κ	Treated-to-control ratio
0	0.95	1:1
0.3	-0.8	3:2
-0.3	0.8	2:3

The error term in the outcome equation, ε_i, is specified as

$$\varepsilon_i = \psi(e_i p(X_i) + e_i D_i) + (1 - \psi)e_i, \tag{20}$$

where e_i is iid standard normal random variable and ψ is a parameter which controls heteroscedasticity, i.e. for $\psi = 0$, ε_i is a homoscedastic error term and if $\psi \neq 0$, ε_i is heteroscedastic. By choosing different values of γ, we specify whether the treatment effect is homogeneous or not. Treatment homogeneity implies that the treatment effect does not vary with different X's. In this case, the causal effect of the treatment is the same for all individuals. As in Busso et al. (2009), we use the following combinations of ψ and γ to create four different settings.

Table 3: Parameter combinations

γ	ψ	Description
0	0	homogeneous treatment, homoscedastic
1	0	heterogeneous treatment, homoscedastic
0	2	homogeneous treatment, heteroscedastic
1	2	heterogeneous treatment, heteroscedastic

Our simulations are based on 10,000, 5,000 and 2,000 Monte Carlo samples for sample sizes $n = 100$, $n = 200$ and $n = 500$, respectively. The choice to make the number of replications proportional to the sample size is motivated by the fact that simulation noise depends negatively on the number of replications and positively on the variance of the estimators (see Huber et al. (2012)), which again depends negatively on the chosen sample size. Hence, the simulation noise is constant if the Monte Carlo samples are chosen proportional to the sample size. Our Monte Carlo study consists of two parts. In the first part, we apply the methods without using

any trimming rules. In the second part, we incorporate two different trimming rules to the conventional as well as shrunken propensity scores.

Propensity Score Shrinkage without Trimming

To provide a reference point for the optimal choices of λ in the different settings without applying any trimming rules, we perform a Monte Carlo study for a hypothetical case where the true ATE is known. Due to the computational burden of this procedure, we do this only for one specification which we believe is the most realistic one and only for sample size $n = 100$. This specification allows heteroscedasticity in the error term ($\psi = 2$) and heterogeneity in the treatment effect ($\gamma = 1$). Furthermore, we also consider the most challenging treated control ratio where we have more control units than treated units ($\eta = -0.3$, $\kappa = 0.8$). Lastly, the outcome equation is chosen to be a nonlinear function of the propensity score ($m_2(q)$). We apply the following procedure to get the optimal λ for known ATE:

1. We draw 10000 Monte Carlo samples for this specification.
2. For each Monte Carlo sample, we estimate the shrunken propensity scores for $\lambda = 0, 0.01, 0.02, ..., 1$ and the ATE by the four methods with each of these shrunken propensity scores.
3. We calculate the MSE over 10000 Monte Carlo samples for each λ and choose the MSE minimizing λ.
4. Steps (1)-(3) are repeated 500 times.

The minimum, mean, maximum and standard error of the mean over 500 optimal λ's are displayed in Table (4).

Table 4: Descriptive statistics for the optimal λ with known ATE

	IPW1	IPW2	IPW3	DR
Min	0.06	0.09	0.16	0.08
Mean	0.82	0.24	0.33	0.35
Max	1.00	0.45	0.50	0.66
Std. Err.	0.028	0.005	0.005	0.009

Note: The MSE minimizing λ^*'s are obtained from a Monte Carlo study for the specification with $n = 100$, $\gamma = 1$, $\psi = 2$, $\eta = -0.3$, $\kappa = 0.8$ and $m_2(q)$. We use 10000 Monte Carlo replications and replicate this procedure 500 times.

The results show that most shrinkage is required for IPW 1 and the least for IPW 2. In all of the 500 replications, λ is never chosen equal to zero, which implies that shrinkage is always optimal.

As in Busso et al. (2009), we estimate the ATE given in Equation (2) for each possible DGP by all three weighting methods and the doubly robust method reviewed in Section 2 using estimated (unshrunken) propensity score, $\hat{p}_i$. $\hat{p}_i$ is obtained by maximum likelihood probit estimation as suggested by the distribution of the error term u_i. Additionally, we estimate the ATEs using the shrunken propensity score, $\hat{p}_i^s$. The optimal tuning parameter λ is chosen in three different ways as introduced in Section 2. The goal is to demonstrate the gains in terms of MSE reduction of the ATE due to propensity score shrinkage as well as to investigate the relative performance of the different shrinkage methods. For the fixed tuning parameter method, we set $c = 1$ and $\delta = 1/2$, i.e. $\lambda_i(n) = 1/\sqrt{n}$. As a summary statistic, we also report the averages over sample sizes in bold letters. The figures in brackets indicate percentage losses due to the bias introduced by shrinkage, $\left(\frac{\text{bias}^2(\text{ATE}(\hat{p}))-\text{bias}^2(\text{ATE}(\hat{p}^s))}{\text{bias}^2(\text{ATE}(\hat{p}))+\text{Var}(\text{ATE}(\hat{p}))}\right)$.

Table 5: Average percentage improvement in MSE for the ATE

	ATE based on $\hat{p}_i$ vs. ATE based on $\hat{p}_i^s$ using											
	$\lambda = 1/\sqrt{n}$				$\lambda =$ argmin MSE				$\lambda =$ cross-validated			
	100	200	500	**avg.**	100	200	500	**avg.**	100	200	500	**avg.**
IPW1	47.4	51.6	37.4	**45.5**	54.9	61.9	42.3	**53.0**	8.3	5.1	2.4	**5.2**
	(-4.0)	(-3.8)	(-6.1)	**(-4.7)**	(-3.1)	(-3.9)	(-6.2)	**(-4.4)**	(-0.8)	(-0.8)	(-0.4)	**(-0.7)**
IPW2	16.5	18.6	18.8	**18.0**	16.6	18.3	21.7	**18.9**	5.3	3.7	2.6	**3.9**
	(-2.1)	(-2.4)	(-3.1)	**(-2.6)**	(-2.4)	(-2.5)	(-2.6)	**(-2.5)**	(-0.7)	(-0.4)	(-0.3)	**(-0.4)**
IPW3	6.7	5.9	4.8	**5.8**	6.8	6.2	5.2	**6.0**	3.0	2.0	1.4	**2.1**
	(-0.9)	(-0.9)	(-1.1)	**(-1.0)**	(-1.1)	(-1.0)	(-0.9)	**(-1.0)**	(-0.4)	(-0.2)	(-0.2)	**(-0.3)**
DR	5.7	6.4	7.5	**6.5**	5.6	6.6	7.8	**6.6**	2.5	2.0	1.7	**2.1**
	(0.0)	(0.0)	(-0.1)	**(0.0)**	(0.0)	(0.0)	(-0.1)	**(-0.0)**	(0.0)	(0.0)	(0.0)	**(0.0)**

Note: avg. denotes the average over the sample sizes. Percentage change which is due to the bias is given in brackets.

Table 5 reveals that, independent of the choice of λ, the improvement turns out to be more pronounced for IPW1 and IPW2, i.e. for those methods which are most vulnerable to very small or very large propensity scores. This result is especially striking since most estimates in the empirical literature are probably based on IPW2 (Busso et al. (2009)). Using the fixed tuning parameter method, the MSE of this estimator could be improved by 18.0% on average. Basically, this improvement comes at no costs due to the simplicity of the linear combination.

The MSE minimizing λ is chosen as in Equation (16) for each individual i. Figure B1 - B3 in Appendix B plot the individual specific $\hat{\lambda}_i^*(n)$. It can be seen that the estimated MSE-minimal tuning parameter exhibits a high variation across observations. In small samples the λ_i's vary strongly over individuals. Therefore, we shrink the propensity score using the average over all observations. The computationally more burdensome MSE-minimizing λ leads to a 18.9% improvement for IPW2. For both choices of λ, the average improvement of IPW3 and DR is still 6.0 to 6.6 percent. We see that the improvement is due to a large reduction of the variance but comes at the expense of introducing a comparatively small bias. For DR, the increase in the squared bias is nearly zero.

If we compare the average results in Table 5 obtained by the fixed tuning parameter method to the ones obtained from MSE minimization, we find that MSE-minimization yields better results for $n = 500$. For $n = 200$ and $n = 100$ this is the case for IPW1. For the other estimators, both methods give about the same result. On average, the cross-validated λ also yields a reduction of the MSE in all cases but is always dominated by the other two choices of λ.

The detailed simulation results for the first part are given in Tables A1-A3 of the Appendix. Tables A1 and A2 show that, in all 288 cases, the use of shrunken propensity scores leads to an improvement of the MSE of the ATE if the fixed valued λ or the MSE-minimizing λ is chosen. Table A3 shows that the use of the shrunken propensity score leads to an improvement in 99.3% of the MSE comparisons if the cross-validated λ is taken.

Propensity Score Shrinkage with Trimming

In the second part of the Monte Carlo study, we evaluate the performance of the proposed shrinkage methods in combination with two trimming rules. Trimming rules are methods for the propensity score are usually applied to avoid the problems occurring if the propensity scores are close to the boundaries. From the various trimming rules proposed in the literature, we consider the two trimming rules which are most commonly used in empirical work and revealed the best performance in the study by Busso et al. (2009). These trimming rules are applied as follows:

1. The first trimming rule goes back to a suggestion by Dehejia and Wahba (2009). Let $T_i^{ATE} = \mathbb{1}\left(\hat{a} < \hat{p}(X_i) < \hat{b}\right)$ setting $\hat{b}$ to be the k^{th} largest propensity score in

the control group and $\hat{a}$ to be the k^{th} smallest propensity score in the treatment group. Then the estimators are computed based on the subsample for which $T_i^{ATE} = 1$.

2. In the second trimming rule suggested by Crump et al. (2009), all units with an estimated propensity score outside the interval $[0.1; 0.9]$ for the ATE are discarded.

As in the first part, we estimate the propensity scores by probit and shrink the propensity scores with the optimal λ's chosen by the three different methods we propose. Different from the first part, we apply the trimming rules 1 and 2 to the conventional and trimming rule 2 to the shrunken propensity scores before estimating the ATEs by weighting and doubly robust methods. Finally, we compare the results based on the shrunken propensity score combined with trimming rule 2 with the results based on: (i) conventional propensity score, (ii) conventional propensity score combined with trimming rule 1, and (iii) conventional propensity score combined with trimming rule 2. As mentioned before, applying the trimming rules to the shrunken propensity score leads to a smaller reduction in the sample size since less observations lie outside the limits of the two trimming rules. For example, if we apply trimming rule 2 to the shrunken propensity score with fixed tuning parameter $\lambda(n) = 1/\sqrt{100} = 0.1$ in the setting where the treated-to-control ratio is 1:1, we use all observations with an conventional propensity score in the interval $[0.05\bar{5}; 0.94\bar{4}]$ instead of only those in the interval $[0.1; 0.9]$. We, therefore, still throw less information away than in the case where the unshrunken propensity scores are trimmed. The estimators based on this procedure converge to the estimators based on the conventional propensity scores which are then trimmed using trimming rule 2.

Again, we perform a Monte Carlo experiment for the most realistic scenario to see which λ would be chosen by cross validation if the true ATE would be known. The results are displayed in Table (6).

Table 6: Descriptive statistics for the optimal λ for known ATE with trimming rule 2

	IPW1	IPW2	IPW3	DR
Min	0.12	0.14	0.16	0.07
Mean	0.16	0.26	0.32	0.17
Max	0.19	0.46	0.50	0.42
St. Dev.	0.01	0.05	0.06	0.03

Note: The MSE minimizing λ^*'s are obtained from a Monte Carlo study for the specification with $n = 100$, $\gamma = 1$, $\psi = 2$, $\eta = -0.3$, $\kappa = 0.8$ and $m_2(q)$. We use 10000 Monte Carlo replications and replicate this procedure 500 times.

In this case, most shrinkage is required for IPW 3 and the least for IPW 1. Again in none of the 500 replications λ is chosen equal to zero, implying that, for this estimation procedure, it also is always optimal to have shrinkage. If we compare the maximum λ's in Table 6 to Table 4, we see that, especially for IPW 1, the degree of shrinkage is a lot smaller if we use trimming rule 2 after shrinking the propensity score.

The results for the fixed tuning parameter method, $\lambda_i(n) = 1/\sqrt{n}$, are given in Tables A4-A6 and summarized in Table 7 below, which contains the average MSE improvements. If we compare our estimation procedure to the estimators based on the conventional propensity scores, the largest percentage improvement in MSE can be obtained for the simple IPW1 estimator. For this weighting estimator, we obtain an improvement of up to 75.7%. When averaging over all settings, the MSE of this estimator is improved by 46.2%. The second largest improvement is obtained for the popular estimator IPW2. The weights of this estimator, compared to IPW1, are forced to add up to one and the average improvement here is 18.7%. IPW3, the estimator that minimizes the asymptotic variance for a known propensity score, can, on average, still be improved by 8.8% in terms of MSE. The MSE of the DR estimator can be reduced by 11.0% if propensity score shrinkage with trimming rule 2 is applied instead of using the unshrunken propensity scores.

Table 7: Average percentage improvement in MSE, fixed valued λ

	ATE based on the shrunken propensity scores $\hat{p}_i^s$ + trimming rule 2 vs.											
	(a) ATE based on $\hat{p}_i$				**(b) ATE based on $\hat{p}_i$ + trimming rule 1**				**(c) ATE based on $\hat{p}_i$ + trimming rule 2**			
	100	200	500	**avg.**	100	200	500	**avg.**	100	200	500	**avg.**
IPW1	45.0	51.6	42.1	**46.2**	25.8	16.3	13.0	**18.4**	20.5	15.1	9.9	**15.2**
	(-0.6)	(-0.5)	(-0.6)	**(-0.6)**	(-1.0)	(-0.9)	(-0.6)	**(-0.8)**	(-0.9)	(-0.3)	(0.7)	**(-0.2)**
IPW2	14.2	18.6	23.2	**18.7**	13.1	8.2	7.7	**9.7**	9.0	5.9	3.9	**6.3**
	(-0.5)	(-0.7)	(-1.2)	**(-0.8)**	(-0.4)	(-0.5)	(-1.1)	**(-0.7)**	(-0.4)	(-0.2)	(0.1)	**(-0.1)**
IPW3	6.7	8.2	11.5	**8.8**	12.8	7.6	6.7	**9.1**	6.6	4.3	2.9	**4.6**
	(-0.2)	(-0.5)	(-1.2)	**(-0.6)**	(-0.1)	(-0.3)	(-0.9)	**(-0.4)**	(-0.1)	(0.1)	(0.3)	**(0.1)**
DR	8.9	10.2	14.0	**11.0**	10.7	5.7	5.5	**7.3**	3.8	2.7	1.8	**2.8**
	(-0.1)	(-0.4)	(-1.2)	**(-0.6)**	(0.1)	(-0.1)	(-0.8)	**(-0.3)**	(0.1)	(0.2)	(0.4)	**(0.2)**

Note: avg. denotes the average over the sample sizes. Percentage change which is due to the bias is given in brackets.

Table 7 part (b) shows that the improvements for IPW1 and IPW2 are smaller than in part (a). For IPW3 and DR, the improvements are smaller in part (b) for sample sizes 200 and 500. Nevertheless, the estimators based on the shrunken propensity scores combined with trimming rule 2 improve IPW1 by 18.4%, IPW2 by 9.7%, IPW3 by 9.0% and DR by 7.3% on average.

The results in Table 7 part (c) reveal that the improvements for all four estimators are smaller than the improvement in part (a) and (b). The average improvements for the four estimators is between 2.8% and 15.2%. Even though we obtain a smaller improvement of the MSE, the suggested procedure reduces the variance and bias for all four estimators for $n = 500$. This even holds for all sample sizes for the double robust estimator. Moreover, the improvement is smaller for larger sample sizes. This is expected since, for large n, the estimator based on the shrunken weights combined with trimming rule 2 converge to the estimators based on the conventional estimators and trimming rule 2.

All in all, the four estimators, based on the conventional propensity score, never have a lower MSE than estimators using shrunken propensity scores combined with trimming rule 2 (see Table A4) in 72 settings. The estimators based on the conventional propensity scores combined with trimming rule 1 yield a lower MSE than the estimators based on our procedure (see Table A5) only once. However, the increase in MSE was only 0.04%. Only in 2 out of 288 cases was the MSE of the estimator based on the conventional propensity scores combined with trimming rule 2 smaller than our suggested procedure. Furthermore, the detailed results in Table A6 show that the losses in MSE in those two cases are only 0.1% and 0.5% .

For the fixed tuning parameter method, the effects of propensity score shrinkage on the distribution of the estimated ATE's can be seen from the boxplots given in Figures B4-B6 of Appendix B. We compare our method to the estimators based on the conventional propensity score. The introduction of propensity score shrinkage does not significantly change the interquartile ranges of the four estimators compared to the estimates without shrinkage. However, the number of outliers is substantially reduced by shrinkage. This holds in particular for the IPW1 estimator, which suffers from a high number of very large outliers and explains why the MSE gains due to propensity score shrinkage are largest for this estimator. Moreover, note that even for small sample sizes, propensity score shrinkage hardly generates any additional bias compared to the estimators without shrinkage.

Thus far, we simply set $\lambda_i(n) = 1/\sqrt{n}$. This choice of $\lambda_i(n)$ yields $\lambda_i(100) = 0.100, \lambda_i(200) = 0.071$ and $\lambda_i(500) = 0.045$ for all four estimators. If we compare those with optimal λ's for known ATE in Table 6, we see that the improvements are obtained with $\lambda_i(n)$'s which are considerably lower than the optimal λ's.

In this part, we use the average over the $\lambda_i(n)$, which minimize the MSEs of the shrunken propensity scores, for each setting. The results based on these $\bar{\lambda}^*(n)$'s are given in Tables A7-A9 in the Appendix. The chosen $\bar{\lambda}^*(n)$ depend on the sample size but, as they minimize the MSE of the shrunken propensity score, and, therefore, do not consider the second stage of the estimation procedure, they are equal for all four estimators. Table 8 below summarizes the average percentage improvements obtained by estimating the ATE based on the shrunken propensity combined with trimming rule 2.

Table 8: Average percentage improvement in MSE, MSE($\hat{p}_i^s$)-minimizing λ

	ATE based on the shrunken propensity scores $\hat{p}_i^s$ + trimming rule 2 vs.											
	(a) ATE based on $\hat{p}_i$				**(b) ATE based on $\hat{p}_i$ + trimming rule 1**				**(c) ATE based on $\hat{p}_i$ + trimming rule 2**			
	100	200	500	**avg.**	100	200	500	**avg.**	100	200	500	**avg.**
IPW1	52.1	62.1	46.1	**53.4**	26.3	15.2	12.7	**18.1**	20.9	13.4	10.8	**15.0**
	(-0.8)	(-0.5)	(-0.6)	**(-0.6)**	(-1.2)	(-1.2)	(-0.8)	**(-1.0)**	(-1.0)	(-0.7)	(0.5)	**(-0.4)**
IPW2	14.5	18.6	25.1	**19.4**	13.9	7.9	6.3	**9.4**	8.6	5.5	3.3	**5.8**
	(-0.7)	(-0.6)	(-1.3)	**(-0.9)**	(-0.6)	(-0.6)	(-1.4)	**(-0.9)**	(-0.6)	(-0.2)	(0.0)	**(-0.3)**
IPW3	6.8	8.9	11.5	**9.1**	13.9	7.1	5.3	**8.8**	6.5	3.8	2.3	**4.2**
	(-0.3)	(-0.4)	(-1.3)	**(-0.7)**	(-0.3)	(-0.3)	(-1.1)	**(-0.6)**	(-0.2)	(0.0)	(0.3)	**(0.0)**
DR	8.8	10.8	14.1	**11.3**	11.9	5.0	4.0	**7.0**	3.8	2.0	1.1	**2.3**
	(-0.2)	(-0.3)	(-1.2)	**(-0.6)**	(0.1)	(-0.1)	(-1.0)	**(-0.3)**	(0.2)	(0.2)	(0.4)	**(0.2)**

Note: avg. denotes the average over the sample sizes. Percentage change which is due to the bias is given in brackets.

Table 8 shows that MSE($\hat{p}_i^s$)-minimizing λ leads to a considerable reduction in the MSE of the treatment effect. Table 8 part (a) shows that the MSE's of the ATE estimators based on MSE($\hat{p}_i^s$)-minimizing λ are on average between 6.8% and 62.1% smaller than those based on conventional propensity score.

Comparing the average results from Table 8 to those obtained by the fixed value λ in Table 7, we see that both choices of λ give about the same result. For $n = 100$, the average value of the MSE-minimizing λ over the 10000 Monte Carlo samples was 0.113. In 0.05% of the cases, the restriction $0 \leq \bar{\lambda}^*(n) \leq 1$ is binding for $n = 100$. For $n = 200$ ($n = 500$), the average value is 0.076 (0.046), the minimum is 0 (0.025) and the maximum 0.429 (0.100). For $n = 200$, $\bar{\lambda}^*(n)$ is set to zero in 0.01% of the cases and never set to one. For $n = 500$, it is never set to zero or one. These numbers highlight that shrinkage can be a useful tool especially for small sample sizes.

Since the average $\lambda^*(n)$ over the 72 different settings is larger than the fixed value λ for each sample size (Table 9), this choice of λ implies on average more shrinkage. If we compare the two resulting λs for $n = 100$ to the optimal λ for known ATE in Table 6, we see that the MSE minimizing λ is closer to these true λ's which explains the slightly higher MSE gains.

The pattern of the MSE reductions by our procedure with respect to the conventional propensity scores combined with trimming rule 2 is analogous to the pattern with respect to the conventional propensity scores combined with trimming rule 1.

The detailed results in Tables A7-A9 show that IPW1 is improved by up to 86.0%, IPW2 up to 38.6%, IPW3 up to 20.9%. For DR, the largest decrease in MSE is 24.1%. Out of the 864 cases (72 settings for four estimators compared to three alternatives), our procedure yields an improvement in the MSE 854 times. In the other 10 cases, the average increase in MSE is only 0.925%.

Next, we use the cross-validated alternative to choose the optimal λ. Table (9) summarizes the λ's chosen by all three methods proposed:

Table 9: Average values for λ obtained through the different methods.

	100	200	500	**avg.**
$\lambda = 1/\sqrt{n}$	0.100	0.071	0.045	**0.072**
$\lambda =$ argmin MSE	0.113	0.076	0.046	**0.078**
$\lambda =$ cross-validated	0.061	0.029	0.013	**0.034**

Note: avg. denotes the average over the sample sizes.

For the fixed valued λ method, the values are independent of the different designs except for the sample size. For the MSE minimizing λ and the cross-validated λ, we obtain different values for each setting. In these cases, Table 9 reports the average values of λ over the different settings for each sample size. In 7.3% of the Monte Carlo samples with $n = 100$, cross-validated λ is equal to 0 and the largest optimal λ is 0.89. For $n = 200$ ($n = 500$), the respective values are 15.3% (31.1%) and 0.13 (0.06). As for the other two methods, these numbers indicated that less shrinkage is optimal for larger sample sizes.

It turns out that the optimal choice of λ by cross-validation is lower than the λ's chosen by the other methods. Thus, the cross-validated λ is not as close to the optimal λ for the known ATE given in Table 6 and lower MSE gains are expected.

The results based on these λ's are given in Tables A10-A12 in the Appendix. Table 10 below summarizes the average percentage improvements. The results show that, on average, the cross-validation method also leads to a gain in the MSE of the ATEs for all four estimators. However, we see that the fixed valued λ's (Table 7) and the MSE-minimizing method (Table 8) provide larger average improvements in the MSEs of the ATEs.

Table 10: Average percentage improvement in MSE, cross-validated λ

	ATE based on the shrunken propensity scores $\hat{p}_i^s$ + trimming rule 2 vs.											
	(a) ATE based on $\hat{p}_i$				**(b) ATE based on $\hat{p}_i$ + trimming rule 1**				**(c) ATE based on $\hat{p}_i$ + trimming rule 2**			
	100	200	500	**avg.**	100	200	500	**avg.**	100	200	500	**avg.**
IPW1	41.0	40.0	43.5	**41.5**	17.4	10.5	5.3	**11.1**	12.1	7.3	4.0	**7.8**
	(-0.1)	(0.0)	(-0.6)	**(-0.2)**	(-0.2)	(-0.1)	(-0.7)	**(-0.4)**	(0.1)	(0.2)	(0.7)	**(0.3)**
IPW2	12.1	15.5	20.5	**16.0**	9.5	5.6	4.4	**6.5**	4.4	2.8	1.7	**3.0**
	(-0.2)	(-0.3)	(-1.2)	**(-0.6)**	(-0.1)	(-0.2)	(-1.0)	**(-0.4)**	(0.0)	(0.2)	(0.2)	**(0.1)**
IPW3	5.0	5.9	8.0	**6.3**	10.3	5.9	3.7	**6.7**	3.0	1.9	1.1	**2.0**
	(-0.1)	(-0.3)	(-1.3)	**(-0.6)**	(0.0)	(-0.2)	(-1.0)	**(-0.4)**	(0.0)	(0.1)	(0.2)	**(0.1)**
DR	7.9	8.9	11.4	**9.4**	9.2	4.8	3.0	**5.7**	1.4	1.0	0.7	**1.0**
	(-0.2)	(-0.4)	(-1.4)	**(-0.7)**	(0.1)	(-0.2)	(-1.1)	**(-0.4)**	(0.1)	(0.1)	(0.1)	**(0.1)**

Note: avg. denotes the average over the sample sizes. Percentage change which is due to the bias is given in brackets.

Using the cross validation method to determine the optimal λ, the improvement in MSEs can be split into the cases where $\lambda = 0$, e.g. no shrinkage, is chosen and into the cases where $\lambda > 0$ is chosen. If we only look at the Monte Carlo samples where at least some shrinkage is chosen, we obtain larger improvements in MSEs. In those cases where λ is set to zero, the MSEs of the estimators are obviously equal.

As reported in Table 10 part (c), even though the gains are less pronounced, our procedure not only reduces the variance but, on average, also leads to a lower squared bias for all four estimators and all sample sizes.

The detailed results in Tables A10-A12 show that IPW1 is improved by up to 81.3%, IPW2 up to 42.0%, IPW3 up to 18.8%. For DR, the largest MSE-reduction is 23.8%. In 18 out of the 288 cases, propensity score shrinkage fails to outperform the estimators based on conventional propensity score estimates. Traditional estimates based on the conventional propensity score combined with trimming rule 1 outperform our shrinkage approaches in only 9 out of the 288 cases. Moreover, failure of MSE reduction due to shrinkage are not only rare, but the losses are also small in magnitude. The average MSE increase over these 27 cases where our procedure is outperformed is only 1.3%. If we compare our procedure to the conventional propensity score combined with trimming rule 2, we see from Table A12 that in all of the 288 cases our procedure yields a lower MSE.

To summarize our findings, we explicitly look at the most realistic setting described before. We focus, thereby, on our suggested procedure and its asymptotic equivalent. The results are given in Table 11:

Table 11: Average percentage improvement in MSE for the most realistic setting and different λs.

	ATE based on $\hat{p}_i$ + trimming rule 2 vs. ATE based on $\hat{p}_i^s$ + trimming rule 2 using											
	$\lambda = 1/\sqrt{n}$				$\lambda =$ argmin MSE				$\lambda =$ cross-validated			
	100	200	500	**avg.**	100	200	500	**avg.**	100	200	500	**avg.**
IPW 1	20.5	15.2	10.4	**15.4**	20.5	14.4	10.3	**15.1**	13.6	9.1	6.1	**9.6**
	(-1.2)	(-0.4)	(0.5)	**(-0.4)**	(-1.7)	(-1.2)	(0.1)	**(-0.9)**	(-0.2)	(-0.1)	(1.1)	**(0.3)**
IPW 2	6.3	4.2	3.2	**4.6**	5.6	4.0	3.1	**4.2**	2.7	2.3	1.1	**2.0**
	(0.1)	(0.4)	(0.6)	**(0.4)**	(0.1)	(0.3)	(0.7)	**(0.4)**	(0.2)	(0.1)	(0.2)	**(0.2)**
IPW 3	5.3	3.7	2.8	**3.9**	4.8	3.2	2.6	**3.5**	2.2	1.9	0.7	**1.6**
	(0.1)	(0.3)	(0.5)	**(0.3)**	(0.1)	(0.3)	(0.6)	**(0.3)**	(0.1)	(0.1)	(0.2)	**(0.1)**
DR	4.1	2.9	2.3	**3.1**	3.8	2.6	1.7	**2.7**	1.7	1.6	0.6	**1.3**
	(0.1)	(0.2)	(0.2)	**(0.2)**	(0.1)	(0.2)	(0.3)	**(0.2)**	(0.1)	(0.0)	(0.1)	**(0.1)**

Note: avg. denotes the average over the sample sizes. Simulation for the specification with $\delta = 1$, $\psi = 2$, $\eta = -0.3$, $\kappa = 0.8$ and $m_2(q)$. Average percentage change which is due to the bias is given in brackets.

Table 11 shows that, like for the average results, the cross validated λ leads to the smallest improvements. Comparing the results based on the fixed valued λ to those based on the MSE minimizing λ, we again find that the improvements are very similar. Moreover, for this setting, the MSE of the ATE is not only reduced due to a variance reduction but also due to a lower squared bias with the exception of IPW1 for $n = 100$ and $n = 200$, .

Like for the average results, we also see that the gain of the MSE of the ATE is largest for small samples and that the sample size increases as the gains decrease. Asymptotically, the ATE based on the shrunken propensity score is equal to the ATE based on the conventional propensity score.

Table 12 below summarizes our results by means of a linear regression. We regress the average percentage MSE improvement on the features of our data generating process represented by a set of dummy variables. The results show that the improvement is less pronounced if the error term is heteroscedastic (Set3 and Set4). The MSE improvement of the weighting estimators is larger when the outcome equation depends on the propensity score in a nonlinear way. Moreover, there is a tendency towards the improvement being higher when the treated-to-control ratio is balanced (Ratio1). If we look at the influence of the sample size on the MSE improvement, we see that compared to part (a) the improvement is higher for larger sample sizes. If the conventional propensity score is combined with a trimming rule, the improvement is significantly lower for larger sample sizes.

Table 12: Regression of the average percentage improvement in MSE for $N = 72$ settings on a set of dummy variables describing the setting.

	Average percentage MSE improvement for ATE based on $\hat{p}_i^s$ + trimming rule 2 vs.											
	(a) ATE based on $\hat{p}_i$				(b) ATE based on $\hat{p}_i$ + trimming rule 1				(c) ATE based on $\hat{p}_i$ + trimming rule 2			
Variables	IPW1	IPW2	IPW3	DR	IPW1	IPW2	IPW3	DR	IPW1	IPW2	IPW3	DR
Cons	54.2	23.6	12.2	14.9	28.3	15.7	15.0	12.1	21.8	9.8	6.5	2.9
	(2.3)	(0.6)	(0.6)	(0.6)	(0.5)	(0.5)	(0.5)	(0.5)	(0.7)	(0.6)	(0.4)	(0.3)
Set2	2.8	-1.0	-1.6	-1.2	3.8	-0.2	-0.2	-0.5	3.5	0.7	0.7	0.5
	(2.2)	(0.5)	(0.6)	(0.6)	(0.5)	(0.5)	(0.5)	(0.5)	(0.7)	(0.5)	(0.4)	(0.3)
Set3	-13.7	-13.1	-9.6	-9.4	-5.6	-4.6	-3.7	-1.2	-3.9	-2.9	-1.1	1.5
	(2.2)	(0.5)	(0.6)	(0.6)	(0.5)	(0.5)	(0.5)	(0.5)	(0.7)	(0.5)	(0.4)	(0.3)
Set4	-10.0	-13.7	-10.6	-10.2	-2.7	-4.7	-3.8	-1.5	-1.3	-2.4	-0.6	1.8
	(2.2)	(0.5)	(0.6)	(0.6)	(0.5)	(0.5)	(0.5)	(0.5)	(0.7)	(0.5)	(0.4)	(0.3)
Curve2	4.1	1.0	0.7	0.2	0.9	1.3	0.8	0.0	2.8	1.6	1.1	0.1
	(1.5)	(0.4)	(0.4)	(0.4)	(0.3)	(0.4)	(0.3)	(0.3)	(0.5)	(0.4)	(0.3)	(0.2)
Ratio2	-14.7	-3.3	-0.5	-2.4	-4.1	-0.7	-0.8	-1.6	-4.4	0.1	0.2	-0.3
	(1.9)	(0.5)	(0.5)	(0.5)	(0.4)	(0.4)	(0.4)	(0.4)	(0.6)	(0.5)	(0.3)	(0.2)
Ratio3	-3.3	-5.5	-0.8	-0.3	-1.5	-1.9	-1.2	-0.2	-2.3	-1.6	-0.8	-0.1
	(1.9)	(0.5)	(0.5)	(0.5)	(0.4)	(0.4)	(0.4)	(0.4)	(0.6)	(0.5)	(0.3)	(0.2)
N2	6.6	4.4	1.5	1.4	-9.5	-4.9	-5.2	-5.0	-5.4	-3.0	-2.2	-1.1
	(1.9)	(0.5)	(0.5)	(0.5)	(0.4)	(0.4)	(0.4)	(0.4)	(0.6)	(0.5)	(0.3)	(0.2)
N3	-2.9	9.0	4.8	5.1	-12.8	-5.4	-6.2	-5.2	-10.6	-5.1	-3.7	-2.0
	(1.9)	(0.5)	(0.5)	(0.5)	(0.4)	(0.4)	(0.4)	(0.4)	(0.6)	(0.5)	(0.3)	(0.2)
R^2	0.74	0.96	0.91	0.9	0.96	0.86	0.87	0.79	0.9	0.79	0.73	0.66

Note: OLS standard errors in brackets. Set1 = homogenous treatment and homoscedasticity, Set2 = heterogenous treatment and homoscedasticity, Set3 = homogenous treatment and heteroscedasticity, Set4 = heterogenous treatment and heteroscedasticity, Curve1 = Outcome equation depends linearly on the propensity score, Cuve2 = Outcome equation depends on the propensity score in a nonlinear way, Ratio1 = Treated-to-Control Ratio 1:1, Ratio2 = Treated-to-Control Ratio 3:2, Ratio3 = Treated-to-Control Ratio 2:3, N1 = sample size 100, N2 = sample size 200, N3 = sample size 500.

5 Conclusion

Estimators that rely on propensity score weighting are among the most popular methods used in the literature on estimation of causal treatment effects. In this paper, we propose a simple and easy-to-implement method to improve those estimators in terms of MSE. The considerable gains in terms of MSE are demonstrated by a comprehensive Monte Carlo simulation study.

The methods we consider here require the first step estimation of the propensity scores. We show that the MSE improvements of the first step estimation lead to MSE reduction of the treatment effect estimator in finite samples. To improve the first step propensity score estimation, we propose a simple shrinkage towards the unconditional mean of the treatment variable. Since the shrinkage parameter is a choice parameter, we suggest three different methods for choosing a λ which satisfies certain optimality conditions.

In the Monte Carlo study, we evaluate the finite sample properties with and without applying trimming rules. All three suggested choices of λ lead to an average improvement in the MSE of the average treatment effects for all four estimators.

In the first part of the Monte Carlo study we compare the estimators based on the shrunken propensity scores to the estimators based on the conventional propensity scores. We obtain a lower MSE in all of the settings if we use the fixed valued or the MSE minimizing tuning parameter. For the cross validated tuning parameter the MSE is reduced in 99.3% of the cases, respectively.

In the second part we base the estimators on the shrunken propensity scores combined with trimming rule proposed by Crump et al. (2009). If we compare this procedure to the estimators based on the conventional propensity score, conventional propensity score combined with trimming rule 1 as well as the conventional propensity score combined with trimming rule we find the following: With the fixed tuning parameter method our procedure leads in 99.7% of the cases to a lower MSE. With the MSE minimizing λ and the cross-validated λ our procedure outperforms in 98.8% and 96.9% cases, respectively.

Given this insight and the fact that the MSE minimizing choice of λ has a much higher computational cost than the fixed valued λ, the latter provides the best trade off between MSE gain and computational burden.

The main advantage of our approach is that it is very simple and can be implemented at basically no cost. Since the shrunken propensity scores are a simple linear combination of the conventional propensity scores and the share of treated, every improvement can be obtained without computational burdens.

References

Abadie, A. and G. W. Imbens (2006): "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, 74, pp. 235–267.

Busso, M., J. DiNardo, and J. McCrary (2009): "Finite Sample Properties of Semiparametric Estimators of Average Treatment Effects," Unpublished manuscript.

Crump, R. K., V. J. Hotz, G. W. Imbens, and O. A. Mitnik (2009): "Dealing with Limited Overlap in Estimation of Average Treatment Effects," *Biometrika*, 96, 187–199.

Dehejia, R. H. and S. Wahba (2009): "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs," *Journal of the American Statistical Association*, 94, 1053–1062.

Frölich, M. (2004): "Finite-sample properties of propensity-score matching and weighting estimators," *Review of Economics and Statistics*, 86, 77–90.

Hahn, J. (1998): "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects," *Econometrica*, 66, pp. 315–331.

Heckman, J., H. Ichimura, and P. Todd (1997): "Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme," *The Review of Economic Studies*, 64, 605–654.

Hirano, K. and G. Imbens (2001): "Estimation of Causal Effects using Propensity Score Weighting: An Application to Data on Right Heart Catheterization," *Health Services and Outcomes Research Methodology*, 2, 259–278.

Huber, M., M. Lechner, and C. Wunsch (2012): "The performance of estimators based on the propensity score," *Journal of Econometrics*, forthcoming.

Imbens, G. (2004): "Nonparametric Estimation of Average Treatment Effects Under Exogeneity," *Review of Economics and Statistics*, 86, 4–29.

Imbens, G. W. and J. M. Wooldridge (2009): "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, 47, 5–86.

Khan, S. and E. Tamer (2010): "Irregular Identification, Support Conditions, and Inverse Weight Estimation," *Econometrica*, 78, 2021–2042.

Lunceford, J. and M. Davidian (2004): "Stratification and Weighting via the Propensity Score in Estimation of Causal Treatment Effects: A Comparative Study," *Statistics in Medicine*, 23, 2937–2960.

Robins, J. M. and A. Rotnitzky (1995): "Semiparametric Efficiency in Multivariate Regression Models with Missing Data," *Journal of the American Statistical Association*, 90, 122–129.

Robins, J. M., A. Rotnitzky, and L. P. Zhao (1994): "Estimation of Regression Coefficients when Some Regressors are not Always Observed," *Journal of the American Statistical Association*, 89, 846–866.

Rosenbaum, P. R. and D. B. Rubin (1983): "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70, 41–55.

Rubin, D. B. (1974): "Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies," *Journal of Educational Psychology*, 66, 688–701.

Seifert, B. and T. Gasser (1996): "Finite-sample variance of local polynomials: analysis and solutions." *Journal of the American Statistical Association*, 91, 267–275.

Wooldridge, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: MIT Press.

——— (2007): "Inverse Probability Weighted Estimation for General Missing Data Problems," *Journal of Econometrics*, 141, 1281 – 1301.

A Tables

Table A1: Percentage improvement in MSE for the ATE if the shrunken propensity scores are used instead of the conventional propensity scores, fixed valued λ

			linear			nonlinear			linear			nonlinear	
	N	100	200	500	100	200	500	100	200	500	100	200	500
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 1:1	IPW 1	59.8	60.9	57.0	56.4	58.9	52.1	60.7	61.6	57.9	58.4	58.2	49.5
		(-0.2)	(-0.3)	(-0.1)	(-6.3)	(-7.0)	(-8.5)	(-0.5)	(-0.7)	(-0.5)	(-11.5)	(-13.1)	(-17.1)
	IPW 2	24.4	26.7	25.3	26.8	30.5	33.8	20.5	22.6	19.4	26.0	29.7	33.3
		(-3.9)	(-4.6)	(-6.8)	(-1.0)	(-1.0)	(-0.5)	(-8.5)	(-10.0)	(-13.7)	(-1.4)	(-1.4)	(-0.8)
	IPW 3	10.3	9.0	6.4	11.4	10.2	8.6	8.2	6.9	3.8	11.1	10.0	8.5
		(-1.6)	(-1.6)	(-2.4)	(-0.4)	(-0.4)	(-0.1)	(-3.5)	(-3.7)	(-4.8)	(-0.6)	(-0.5)	(0.0)
	DR	8.5	9.9	11.0	8.5	10.0	11.3	8.5	9.9	11.1	8.5	10.0	11.0
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.1)	(0.0)	(0.0)	(-0.2)	(-0.3)	(-0.6)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 1:1	IPW 1	45.7	48.1	46.0	43.9	48.2	42.9	46.9	49.2	47.2	48.3	49.9	42.7
		(-0.2)	(-0.3)	(-0.1)	(-5.0)	(-5.6)	(-6.8)	(-0.4)	(-0.5)	(-0.4)	(-9.7)	(-11.1)	(-14.2)
	IPW 2	12.7	14.5	15.6	14.2	17.4	21.9	10.3	11.9	11.5	13.7	16.9	21.6
		(-2.5)	(-3.2)	(-4.9)	(-0.6)	(-0.7)	(-0.4)	(-5.7)	(-7.0)	(-10.0)	(-0.9)	(-0.9)	(-0.5)
	IPW 3	4.8	4.4	2.8	5.5	5.2	4.3	3.5	3.1	1.1	5.3	5.1	4.3
		(-1.0)	(-1.1)	(-1.6)	(-0.3)	(-0.2)	(0.0)	(-2.2)	(-2.4)	(-3.3)	(-0.4)	(-0.3)	(0.0)
	DR	5.5	6.5	7.2	5.6	6.6	7.4	5.5	6.5	7.3	5.7	6.6	7.2
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(-0.1)	(-0.2)	(-0.4)
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 3:2	IPW 1	55.9	47.9	38.1	52.5	43.2	35.2	56.1	48.3	38.7	52.8	43.1	35.2
		(-0.4)	(-0.5)	(-1.0)	(-1.1)	(-1.6)	(-1.4)	(-0.3)	(-0.3)	(-0.7)	(-3.4)	(-4.6)	(-4.8)
	IPW 2	20.6	22.6	22.0	22.2	23.8	23.0	18.8	21.2	20.6	21.5	22.9	22.3
		(-2.0)	(-2.0)	(-2.6)	(-1.1)	(-1.5)	(-1.4)	(-3.9)	(-4.0)	(-4.9)	(-1.6)	(-2.1)	(-1.9)
	IPW 3	8.6	7.6	6.7	9.0	7.6	7.0	7.6	6.6	5.7	8.7	7.3	6.8
		(-0.9)	(-0.8)	(-1.0)	(-0.6)	(-0.7)	(-0.5)	(-1.8)	(-1.7)	(-1.9)	(-0.8)	(-1.0)	(-0.7)
	DR	5.9	6.6	8.3	5.9	6.7	8.4	5.9	6.7	8.2	6.0	6.9	8.6
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 3:2	IPW 1	40.6	33.9	27.4	37.1	29.3	24.7	41.0	34.3	27.9	38.4	30.1	25.4
		(-0.3)	(-0.4)	(-0.7)	(-0.8)	(-1.1)	(-1.0)	(-0.2)	(-0.2)	(-0.5)	(-2.6)	(-3.4)	(-3.5)
	IPW 2	10.7	12.9	13.8	11.7	13.7	14.2	9.6	12.2	12.9	11.2	13.2	13.8
		(-1.3)	(-1.3)	(-1.8)	(-0.7)	(-1.0)	(-1.0)	(-2.5)	(-2.6)	(-3.4)	(-1.1)	(-1.4)	(-1.3)
	IPW 3	3.6	3.5	3.5	3.8	3.6	3.7	3.0	2.9	2.8	3.6	3.4	3.6
		(-0.6)	(-0.5)	(-0.6)	(-0.4)	(-0.4)	(-0.3)	(-1.2)	(-1.0)	(-1.3)	(-0.5)	(-0.6)	(-0.4)
	DR	3.9	4.5	5.8	3.9	4.6	5.9	3.9	4.5	5.8	4.0	4.7	6.0
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 2:3	IPW 1	47.7	65.5	36.3	46.8	66.1	33.0	48.7	66.5	36.8	48.1	67.4	31.6
		(-2.5)	(-2.0)	(-4.0)	(-9.2)	(-6.8)	(-14.6)	(-3.2)	(-2.6)	(-5.4)	(-15.4)	(-10.9)	(-24.6)
	IPW 2	19.7	21.9	18.8	21.6	23.2	21.7	17.0	19.5	15.8	21.5	23.2	21.8
		(-2.0)	(-2.3)	(-3.2)	(-0.1)	(-0.1)	(0.0)	(-5.2)	(-5.9)	(-7.6)	(-0.2)	(-0.1)	(0.0)
	IPW 3	8.9	7.5	5.7	9.7	8.4	7.0	7.4	6.0	4.0	9.6	8.3	7.0
		(-0.9)	(-0.9)	(-1.2)	(-0.1)	(0.0)	(0.0)	(-2.3)	(-2.4)	(-2.8)	(-0.1)	(0.0)	(0.1)
	DR	6.2	6.3	7.3	6.2	6.4	7.4	6.2	6.3	7.3	6.1	6.3	7.1
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(-0.1)	(-0.1)	(-0.3)	(-0.4)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 2:3	IPW 1	36.5	54.8	28.4	37.1	57.2	27.0	37.6	56.3	29.2	40.2	60.6	27.3
		(-1.8)	(-1.7)	(-2.9)	(-7.1)	(-5.8)	(-10.9)	(-2.4)	(-2.2)	(-3.9)	(-12.5)	(-9.7)	(-19.6)
	IPW 2	9.9	11.8	11.2	11.0	12.4	13.2	8.3	10.4	9.3	11.0	12.4	13.3
		(-1.3)	(-1.5)	(-2.1)	(-0.1)	(-0.1)	(0.0)	(-3.3)	(-3.9)	(-5.2)	(-0.1)	(-0.1)	(0.0)
	IPW 3	4.4	3.7	2.8	4.9	4.2	3.7	3.5	2.8	1.8	4.8	4.2	3.6
		(-0.5)	(-0.6)	(-0.8)	(0.0)	(0.0)	(0.0)	(-1.4)	(-1.5)	(-1.8)	(0.0)	(0.0)	(0.0)
	DR	4.0	4.1	5.0	4.0	4.2	5.0	4.0	4.1	5.0	4.0	4.1	4.8
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(-0.1)	(-0.2)	(-0.3)

Note: Percentage change which is due to the bias $\left(\frac{\text{bias}^2(\text{ATE}(\hat{p}))-\text{bias}^2(\text{ATE}(\hat{p}^s))}{\text{bias}^2(\text{ATE}(\hat{p}))+\text{Var}(\text{ATE}(\hat{p}))}\right)$ is given in brackets.

Table A2: Percentage improvement in MSE for the ATE if the shrunken propensity scores are used instead of the conventional propensity scores, MSE($\hat{p}_i^s$)-minimizing λ

		linear			nonlinear			linear			nonlinear		
	N	100	200	500	100	200	500	100	200	500	100	200	500
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 1:1	IPW 1	60.9	66.3	66.3	57.5	63.3	58.4	61.6	67.3	66.3	58.1	64.6	56.9
		(-0.2)	(-0.3)	(-0.2)	(-6.2)	(-6.2)	(-7.3)	(-0.4)	(-0.6)	(-0.5)	(-11.8)	(-11.0)	(-13.9)
	IPW 2	24.6	26.5	31.3	27.7	29.0	34.3	20.5	22.5	28.0	27.0	28.3	33.8
		(-4.3)	(-3.9)	(-5.3)	(-0.9)	(-1.3)	(-0.9)	(-9.2)	(-8.9)	(-10.7)	(-1.3)	(-1.7)	(-1.3)
	IPW 3	9.9	9.3	7.3	11.4	10.0	8.7	7.6	7.3	5.3	11.2	9.8	8.6
		(-1.8)	(-1.4)	(-1.7)	(-0.4)	(-0.5)	(-0.3)	(-4.0)	(-3.2)	(-3.6)	(-0.5)	(-0.6)	(-0.3)
	DR	8.5	9.6	11.2	8.6	9.8	11.7	8.4	9.6	11.1	8.6	9.9	11.7
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(-0.2)	(-0.2)	(-0.4)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 1:1	IPW 1	46.4	54.1	56.5	44.8	52.6	49.7	47.5	55.5	56.8	47.9	56.1	50.2
		(-0.2)	(-0.3)	(-0.1)	(-4.8)	(-5.1)	(-6.1)	(-0.3)	(-0.5)	(-0.4)	(-9.7)	(-9.5)	(-12.2)
	IPW 2	11.9	15.8	21.4	13.7	17.2	23.2	9.5	13.4	19.7	13.3	16.8	22.8
		(-2.7)	(-2.6)	(-3.9)	(-0.6)	(-0.9)	(-0.7)	(-5.9)	(-5.9)	(-7.9)	(-0.9)	(-1.2)	(-0.9)
	IPW 3	4.1	4.5	3.9	5.0	4.9	5.0	2.7	3.3	2.5	4.9	4.8	4.9
		(-1.1)	(-0.8)	(-1.2)	(-0.3)	(-0.3)	(-0.2)	(-2.4)	(-2.0)	(-2.5)	(-0.3)	(-0.4)	(-0.2)
	DR	4.8	6.6	7.7	4.9	6.7	8.3	4.9	6.6	7.7	5.0	6.8	8.4
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(-0.1)	(-0.1)	(-0.3)
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 3:2	IPW 1	51.9	86.8	42.0	47.7	71.8	41.0	52.4	86.7	42.9	48.8	70.6	42.8
		(-0.5)	(-0.2)	(-0.7)	(-1.4)	(-0.7)	(-1.8)	(-0.4)	(-0.1)	(-0.5)	(-4.3)	(-2.4)	(-5.0)
	IPW 2	21.0	21.3	22.2	22.0	22.6	24.3	19.1	19.3	20.6	21.1	21.8	23.5
		(-2.4)	(-2.5)	(-2.0)	(-1.2)	(-1.4)	(-2.0)	(-4.7)	(-4.8)	(-4.1)	(-1.8)	(-2.0)	(-2.6)
	IPW 3	9.5	8.1	6.9	9.9	8.6	6.7	8.3	7.0	6.0	9.6	8.3	6.4
		(-1.2)	(-1.0)	(-0.6)	(-0.6)	(-0.6)	(-0.8)	(-2.3)	(-2.1)	(-1.4)	(-0.9)	(-0.9)	(-1.0)
	DR	6.1	6.7	8.3	6.2	6.7	8.5	6.1	6.6	8.3	6.3	6.9	8.8
		(0.0)	(0.0)	(-0.1)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 3:2	IPW 1	36.1	79.4	30.5	32.0	59.6	29.5	36.6	79.3	31.3	34.2	59.1	31.9
		(-0.4)	(-0.2)	(-0.5)	(-1.0)	(-0.6)	(-1.3)	(-0.3)	(-0.1)	(-0.3)	(-3.1)	(-2.1)	(-3.8)
	IPW 2	10.1	11.8	13.9	10.7	12.9	15.7	9.0	10.5	12.8	10.1	12.4	15.1
		(-1.5)	(-1.7)	(-1.4)	(-0.8)	(-0.9)	(-1.4)	(-3.0)	(-3.3)	(-2.9)	(-1.2)	(-1.3)	(-1.8)
	IPW 3	3.8	3.9	3.7	4.1	4.2	3.6	3.1	3.1	3.1	3.8	4.0	3.4
		(-0.7)	(-0.7)	(-0.4)	(-0.4)	(-0.4)	(-0.5)	(-1.4)	(-1.4)	(-1.0)	(-0.6)	(-0.6)	(-0.7)
	DR	3.8	4.5	5.9	3.9	4.6	6.0	3.8	4.5	5.9	4.0	4.7	6.2
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 2:3	IPW 1	72.5	56.0	38.6	74.2	56.0	33.9	73.5	57.1	38.8	75.6	56.9	34.5
		(-1.3)	(-2.6)	(-4.7)	(-4.9)	(-9.2)	(-16.0)	(-1.7)	(-3.4)	(-6.2)	(-7.9)	(-15.0)	(-25.2)
	IPW 2	21.3	21.7	24.5	23.6	24.0	24.8	18.3	18.7	22.9	23.5	23.9	25.0
		(-2.7)	(-2.6)	(-2.0)	(-0.1)	(-0.1)	(-0.1)	(-6.4)	(-6.5)	(-5.7)	(-0.1)	(-0.1)	(-0.1)
	IPW 3	9.2	8.0	6.5	10.5	9.1	7.1	7.3	6.3	5.1	10.5	9.0	7.1
		(-1.3)	(-1.1)	(-0.7)	(0.0)	(0.0)	(-0.1)	(-3.1)	(-2.7)	(-2.0)	(0.0)	(0.0)	(-0.1)
	DR	6.4	7.1	7.6	6.4	7.1	7.8	6.3	7.1	7.5	6.3	7.0	7.8
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(-0.1)	(-0.1)	(-0.2)	(-0.2)	(-0.2)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 2:3	IPW 1	61.1	44.9	29.9	64.9	46.6	27.1	62.7	46.3	30.3	68.7	49.7	29.1
		(-1.1)	(-2.0)	(-3.6)	(-4.2)	(-7.3)	(-12.6)	(-1.4)	(-2.6)	(-4.8)	(-7.1)	(-12.7)	(-21.2)
	IPW 2	9.7	11.8	15.6	11.1	13.2	15.5	8.1	10.1	15.1	11.1	13.2	15.7
		(-1.6)	(-1.6)	(-1.4)	(-0.1)	(-0.1)	(-0.1)	(-4.0)	(-4.1)	(-4.0)	(-0.1)	(-0.1)	(-0.1)
	IPW 3	3.8	3.7	3.2	4.6	4.3	3.6	2.7	2.6	2.4	4.6	4.3	3.6
		(-0.7)	(-0.7)	(-0.5)	(0.0)	(0.0)	(-0.1)	(-1.8)	(-1.7)	(-1.3)	(0.0)	(0.0)	(-0.1)
	DR	3.5	4.6	4.9	3.6	4.6	5.1	3.5	4.6	4.8	3.6	4.6	5.2
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(-0.1)	(-0.2)	(-0.1)

Note: Percentage change which is due to the bias $\left(\frac{\text{bias}^2(\text{ATE}(\hat{p}))-\text{bias}^2(\text{ATE}(\hat{p}^s))}{\text{bias}^2(\text{ATE}(\hat{p}))+\text{Var}(\text{ATE}(\hat{p}))}\right)$ is given in brackets.

Table A3: Percentage improvement in MSE for the ATE if the shrunken propensity scores are used instead of the conventional propensity scores, data-driven λ

		linear			nonlinear			linear			nonlinear		
	N	100	200	500	100	200	500	100	200	500	100	200	500
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 1:1	IPW 1	9.1	6.4	1.9	7.7	4.9	1.7	9.1	6.4	2.0	6.0	3.3	1.3
		(0.0)	(-0.1)	(0.0)	(-1.1)	(-1.2)	(-0.5)	(-0.1)	(-0.2)	(-0.1)	(-2.2)	(-2.1)	(-0.9)
	IPW 2	7.0	4.8	3.5	8.1	5.8	4.5	5.6	3.5	2.8	8.1	5.6	4.5
		(-1.1)	(-0.4)	(-0.4)	(-0.2)	(-0.3)	(-0.1)	(-2.3)	(-1.1)	(-0.9)	(-0.3)	(-0.4)	(-0.1)
	IPW 3	4.0	2.6	1.7	4.4	2.9	2.2	3.3	2.0	1.2	4.6	2.9	2.2
		(-0.6)	(-0.2)	(-0.3)	(-0.1)	(-0.2)	(-0.1)	(-1.3)	(-0.6)	(-0.6)	(-0.1)	(-0.2)	(0.0)
	DR	3.4	2.7	2.5	3.4	2.8	2.7	3.6	2.7	2.4	3.6	2.7	2.7
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(-0.1)	(0.0)	(-0.1)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 1:1	IPW 1	7.4	5.2	1.8	6.4	4.0	1.5	7.5	5.2	1.8	5.3	2.9	1.2
		(0.0)	(-0.1)	(0.0)	(-0.9)	(-1.0)	(-0.4)	(-0.1)	(-0.1)	(-0.1)	(-1.9)	(-1.8)	(-0.8)
	IPW 2	3.1	2.3	2.2	3.9	2.9	2.8	2.2	1.5	1.8	4.0	2.8	2.7
		(-0.7)	(-0.3)	(-0.3)	(-0.1)	(-0.2)	(-0.1)	(-1.5)	(-0.7)	(-0.7)	(-0.2)	(-0.3)	(-0.1)
	IPW 3	1.6	1.2	1.0	1.9	1.3	1.3	1.2	0.8	0.7	2.1	1.3	1.2
		(-0.4)	(-0.1)	(-0.2)	(-0.1)	(-0.1)	(0.0)	(-0.8)	(-0.4)	(-0.4)	(-0.1)	(-0.1)	(0.0)
	DR	2.1	1.6	1.7	2.1	1.7	1.8	2.2	1.6	1.7	2.2	1.7	1.8
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 3:2	IPW 1	13.2	8.3	5.4	13.1	8.3	5.5	13.4	8.5	5.6	12.8	8.1	5.4
		(-0.1)	(0.0)	(-0.1)	(-0.5)	(-0.3)	(-0.2)	(-0.1)	(0.0)	(0.0)	(-1.5)	(-0.8)	(-0.5)
	IPW 2	7.6	4.9	3.0	8.0	5.1	3.1	6.7	4.3	2.5	7.7	4.9	2.9
		(-0.9)	(-0.4)	(-0.3)	(-0.4)	(-0.3)	(-0.1)	(-1.7)	(-0.8)	(-0.6)	(-0.6)	(-0.4)	(-0.2)
	IPW 3	4.5	2.8	1.7	4.7	2.8	1.7	4.0	2.5	1.5	4.6	2.7	1.7
		(-0.5)	(-0.3)	(-0.2)	(-0.3)	(-0.2)	(-0.1)	(-1.0)	(-0.5)	(-0.4)	(-0.4)	(-0.3)	(-0.1)
	DR	3.0	2.3	1.7	3.0	2.4	1.8	3.0	2.3	1.7	3.0	2.4	1.9
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 3:2	IPW 1	8.7	5.7	3.8	8.4	5.6	3.5	8.9	5.9	3.9	8.6	5.7	3.5
		(-0.1)	(0.0)	(0.0)	(-0.4)	(-0.2)	(-0.1)	(0.0)	(0.0)	(0.0)	(-1.1)	(-0.6)	(-0.4)
	IPW 2	3.2	2.4	1.8	3.4	2.5	1.6	2.7	2.0	1.5	3.2	2.3	1.4
		(-0.5)	(-0.3)	(-0.2)	(-0.2)	(-0.2)	(-0.1)	(-1.1)	(-0.5)	(-0.4)	(-0.4)	(-0.3)	(-0.1)
	IPW 3	1.6	1.2	1.0	1.7	1.2	0.8	1.3	1.0	0.9	1.6	1.1	0.7
		(-0.3)	(-0.2)	(-0.1)	(-0.2)	(-0.1)	(-0.1)	(-0.6)	(-0.3)	(-0.3)	(-0.2)	(-0.2)	(-0.1)
	DR	1.8	1.5	1.1	1.8	1.5	1.1	1.8	1.5	1.1	1.8	1.6	1.2
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 2:3	IPW 1	9.7	6.4	3.8	6.2	2.9	0.1	9.2	6.0	3.4	3.6	0.2	-2.9
		(-0.3)	(-0.5)	(-0.3)	(-1.7)	(-1.8)	(-1.1)	(-0.4)	(-0.6)	(-0.4)	(-2.8)	(-2.9)	(-1.8)
	IPW 2	7.3	5.2	3.2	8.3	5.8	4.0	5.7	4.1	1.9	8.3	5.8	4.0
		(-0.8)	(-0.2)	(-0.4)	(0.0)	(-0.1)	(0.0)	(-2.0)	(-0.8)	(-0.8)	(0.0)	(-0.1)	(0.0)
	IPW 3	4.3	3.0	1.9	4.9	3.2	2.1	3.4	2.5	1.3	4.9	3.2	2.2
		(-0.5)	(-0.1)	(-0.2)	(0.0)	(0.0)	(0.0)	(-1.2)	(-0.5)	(-0.5)	(0.0)	(0.0)	(0.0)
	DR	2.9	2.3	2.0	2.9	2.4	2.0	2.9	2.3	1.9	2.8	2.3	1.9
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(-0.1)	(0.0)	(-0.1)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 2:3	IPW 1	7.7	5.0	2.9	5.3	2.6	0.0	7.4	4.8	2.5	3.3	0.5	-2.5
		(-0.2)	(-0.4)	(-0.2)	(-1.3)	(-1.4)	(-0.8)	(-0.3)	(-0.5)	(-0.3)	(-2.3)	(-2.4)	(-1.5)
	IPW 2	2.9	2.5	1.8	3.6	2.9	2.3	2.0	1.8	1.0	3.6	2.9	2.3
		(-0.5)	(-0.1)	(-0.2)	(0.0)	(-0.1)	(0.0)	(-1.3)	(-0.5)	(-0.5)	(0.0)	(-0.1)	(0.0)
	IPW 3	1.7	1.4	1.0	2.0	1.6	1.2	1.2	1.1	0.7	2.0	1.6	1.2
		(-0.3)	(-0.1)	(-0.1)	(0.0)	(0.0)	(0.0)	(-0.7)	(-0.3)	(-0.3)	(0.0)	(0.0)	(0.0)
	DR	1.7	1.4	1.3	1.7	1.4	1.3	1.7	1.4	1.3	1.6	1.4	1.3
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(-0.1)	(0.0)	(-0.1)

Note: Percentage change which is due to the bias $\left(\frac{\text{bias}^2(\text{ATE}(\hat{p}))-\text{bias}^2(\text{ATE}(\hat{p}^s))}{\text{bias}^2(\text{ATE}(\hat{p}))+\text{Var}(\text{ATE}(\hat{p}))}\right)$ is given in brackets.

Table A4: Percentage improvement in MSE for the ATE if the shrunken propensity scores are used with trimming rule 2 instead of the conventional propensity scores, fixed valued λ

			linear			nonlinear			linear			nonlinear	
	N	100	200	500	100	200	500	100	200	500	100	200	500
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 1:1	IPW 1	53.9	59.0	55.8	55.4	62.2	56.7	54.9	59.8	56.1	64.2	68.3	63.1
		(0.0)	(0.0)	(0.0)	(-1.3)	(-1.0)	(-0.7)	(-0.1)	(0.0)	(0.0)	(-1.2)	(-0.5)	(0.1)
	IPW 2	22.1	29.2	33.6	22.7	31.0	37.3	22.0	29.3	33.6	22.6	30.2	34.0
		(-0.7)	(-0.9)	(-0.7)	(-0.1)	(0.0)	(0.1)	(-1.6)	(-1.9)	(-1.5)	(0.1)	(-0.5)	(-3.1)
	IPW 3	11.2	15.7	17.8	11.8	16.7	18.8	10.5	14.8	16.6	11.8	15.6	14.2
		(-0.2)	(-0.4)	(-0.1)	(0.0)	(0.1)	(0.0)	(-0.5)	(-0.8)	(-0.3)	(0.0)	(-0.9)	(-4.9)
	DR	12.9	18.2	20.7	13.4	18.7	22.6	12.9	18.5	19.9	13.7	17.8	18.6
		(0.1)	(0.1)	(-0.1)	(0.0)	(0.0)	(0.2)	(0.2)	(0.3)	(-0.1)	(-0.6)	(-1.9)	(-5.6)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 1:1	IPW 1	38.2	43.8	42.2	40.5	48.5	43.9	39.6	45.0	43.0	51.2	56.7	51.9
		(0.0)	(0.0)	(0.0)	(-1.0)	(-0.8)	(-0.5)	(0.0)	(0.0)	(0.0)	(-1.0)	(-0.4)	(0.1)
	IPW 2	8.9	13.5	18.5	9.6	15.1	21.1	9.2	13.6	18.6	9.6	14.8	18.7
		(-0.5)	(-0.7)	(-0.6)	(-0.1)	(0.0)	(0.1)	(-1.1)	(-1.4)	(-1.2)	(0.1)	(-0.3)	(-2.4)
	IPW 3	2.1	4.6	5.3	2.7	5.3	5.7	1.8	3.8	4.5	2.7	4.7	2.5
		(-0.1)	(-0.3)	(-0.1)	(0.0)	(0.1)	(0.0)	(-0.4)	(-0.6)	(-0.3)	(0.0)	(-0.6)	(-3.4)
	DR	4.8	7.3	7.6	5.1	7.5	8.9	5.0	7.3	7.0	5.4	7.0	6.1
		(0.1)	(0.1)	(-0.1)	(0.0)	(0.0)	(0.1)	(0.1)	(0.2)	(-0.1)	(-0.4)	(-1.3)	(-4.1)
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 3:2	IPW 1	51.8	45.1	42.9	46.9	38.5	37.2	50.4	42.8	39.0	50.5	43.4	42.6
		(-0.1)	(-0.2)	(-0.1)	(-2.0)	(-2.3)	(-2.1)	(-1.1)	(-2.0)	(-3.4)	(-1.8)	(-1.0)	(0.0)
	IPW 2	19.7	22.8	29.7	21.4	24.7	30.4	20.0	23.6	30.2	20.9	23.9	27.0
		(-0.6)	(-0.5)	(-0.6)	(-0.1)	(-0.1)	(-0.2)	(-0.2)	(0.0)	(0.0)	(0.1)	(-0.4)	(-3.3)
	IPW 3	10.9	10.9	19.0	11.7	11.9	19.3	11.1	10.8	17.8	11.1	10.7	14.8
		(-0.3)	(-0.2)	(-0.2)	(0.1)	(0.1)	(0.0)	(0.2)	(0.1)	(-0.4)	(-0.1)	(-0.7)	(-4.4)
	DR	12.5	12.6	21.2	12.6	12.9	21.4	11.8	11.1	18.0	11.4	10.7	15.4
		(0.1)	(0.1)	(0.0)	(0.0)	(0.1)	(0.0)	(-0.5)	(-1.1)	(-2.6)	(-1.2)	(-2.3)	(-6.5)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 3:2	IPW 1	36.7	31.1	30.9	31.9	25.0	25.4	35.8	29.7	28.4	35.7	29.4	29.8
		(-0.1)	(-0.1)	(-0.1)	(-1.5)	(-1.6)	(-1.5)	(-0.8)	(-1.4)	(-2.5)	(-1.4)	(-0.7)	(0.0)
	IPW 2	7.1	10.9	17.9	7.9	12.0	17.0	7.5	11.8	18.7	7.6	11.4	14.4
		(-0.4)	(-0.4)	(-0.4)	(0.0)	(0.0)	(-0.1)	(-0.1)	(0.0)	(0.0)	(0.0)	(-0.3)	(-2.3)
	IPW 3	1.0	2.7	9.4	1.4	3.4	8.8	1.2	2.8	9.1	1.0	2.5	5.5
		(-0.2)	(-0.2)	(-0.1)	(0.0)	(0.1)	(0.0)	(0.1)	(0.1)	(-0.2)	(0.0)	(-0.5)	(-3.0)
	DR	2.1	4.0	10.3	2.2	4.2	10.5	1.7	3.1	8.5	1.4	2.7	6.2
		(0.0)	(0.1)	(0.0)	(0.0)	(0.0)	(0.0)	(-0.3)	(-0.7)	(-1.7)	(-0.8)	(-1.6)	(-4.5)
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 2:3	IPW 1	45.2	66.2	41.9	50.3	70.7	46.5	46.6	67.4	40.9	58.0	75.7	53.7
		(0.2)	(0.1)	(0.0)	(-0.3)	(-0.2)	(-0.4)	(0.1)	(-0.2)	(-1.6)	(-0.5)	(-0.3)	(-0.1)
	IPW 2	18.2	23.2	26.2	18.7	23.3	27.2	16.3	20.6	20.5	18.8	23.2	27.2
		(-0.5)	(-0.7)	(-0.5)	(-0.1)	(0.0)	(0.0)	(-3.1)	(-4.5)	(-6.9)	(0.0)	(0.0)	(-0.2)
	IPW 3	10.6	12.3	16.6	10.9	13.0	17.6	8.6	9.0	9.8	11.0	12.8	17.3
		(-0.2)	(-0.3)	(-0.1)	(-0.1)	(0.0)	(0.0)	(-2.2)	(-3.6)	(-5.9)	(0.0)	(-0.1)	(-0.3)
	DR	12.5	14.3	18.9	12.6	14.5	19.7	12.7	13.9	15.8	13.1	14.8	19.8
		(0.1)	(0.1)	(0.0)	(0.0)	(0.0)	(0.1)	(0.0)	(-0.7)	(-2.6)	(0.0)	(-0.2)	(-0.4)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 2:3	IPW 1	30.1	52.1	29.8	35.9	58.4	35.3	31.8	53.8	29.8	45.3	65.9	44.0
		(0.1)	(0.1)	(0.0)	(-0.2)	(-0.2)	(-0.3)	(0.1)	(-0.1)	(-1.1)	(-0.4)	(-0.2)	(-0.1)
	IPW 2	7.7	10.1	14.3	8.1	10.1	15.1	6.8	8.4	10.4	8.3	10.2	15.1
		(-0.3)	(-0.4)	(-0.3)	(-0.1)	(0.0)	(0.0)	(-2.0)	(-3.0)	(-4.5)	(0.0)	(0.0)	(-0.1)
	IPW 3	4.0	3.4	7.4	4.2	4.0	8.0	2.8	1.2	3.0	4.3	4.0	7.9
		(-0.1)	(-0.2)	(0.0)	(0.0)	(0.0)	(0.0)	(-1.4)	(-2.3)	(-3.6)	(0.0)	(0.0)	(-0.2)
	DR	8.1	6.1	10.0	8.0	6.3	10.2	8.3	5.6	8.2	8.4	6.7	10.4
		(0.1)	(0.0)	(0.0)	(0.0)	(0.0)	(0.1)	(0.0)	(-0.4)	(-1.6)	(0.0)	(-0.1)	(-0.2)

Note: Percentage change which is due to the bias $\left(\frac{\text{bias}^2(\text{ATE}(\hat{p}))-\text{bias}^2(\text{ATE}(\hat{p}^s))}{\text{bias}^2(\text{ATE}(\hat{p}))+\text{Var}(\text{ATE}(\hat{p}))}\right)$ is given in brackets.

Table A5: Percentage improvement in MSE for the ATE if the shrunken propensity scores are used with trimming rule 2 instead of the conventional propensity scores with trimming rule 1, fixed valued λ

			linear			nonlinear			linear			nonlinear	
	N	100	200	500	100	200	500	100	200	500	100	200	500
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 1:1	IPW 1	26.7	19.3	17.0	29.4	19.4	13.7	34.0	24.3	19.4	34.2	24.7	17.6
		(0.0)	(-0.1)	(0.1)	(-2.2)	(-2.4)	(-1.4)	(-0.1)	(-0.1)	(0.1)	(-3.0)	(-0.8)	(2.1)
	IPW 2	15.0	11.4	10.5	15.4	12.3	11.6	15.1	10.8	9.6	16.7	14.2	11.3
		(-1.0)	(-1.3)	(-1.7)	(-0.2)	(-0.1)	(0.2)	(-2.3)	(-2.8)	(-3.1)	(1.1)	(1.4)	(-0.3)
	IPW 3	13.9	10.3	8.8	14.1	10.9	9.6	14.5	10.2	8.6	15.2	12.4	8.6
		(-0.5)	(-0.6)	(-0.8)	(-0.1)	(0.0)	(0.2)	(-1.1)	(-1.4)	(-1.5)	(1.1)	(1.2)	(-1.0)
	DR	10.2	7.0	7.2	9.9	7.1	7.7	11.1	7.5	7.7	10.3	7.8	5.3
		(0.0)	(0.0)	(0.1)	(0.0)	(0.0)	(0.1)	(0.0)	(0.0)	(0.1)	(0.9)	(0.3)	(-2.7)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 1:1	IPW 1	21.8	13.5	10.6	23.8	14.1	8.6	27.0	16.7	12.0	27.8	18.6	12.0
		(0.0)	(0.0)	(0.1)	(-1.4)	(-1.5)	(-0.8)	(-0.1)	(-0.1)	(0.1)	(-2.0)	(-0.5)	(1.5)
	IPW 2	11.9	6.3	5.4	12.1	7.2	6.1	12.1	5.4	4.4	12.9	8.7	6.3
		(-0.6)	(-0.8)	(-1.1)	(-0.1)	(-0.1)	(0.1)	(-1.4)	(-1.7)	(-2.0)	(0.7)	(0.9)	(-0.2)
	IPW 3	12.1	6.4	4.7	12.2	6.9	5.2	12.6	5.8	4.1	12.8	8.2	4.9
		(-0.3)	(-0.4)	(-0.5)	(-0.1)	(0.0)	(0.1)	(-0.7)	(-0.8)	(-1.0)	(0.7)	(0.8)	(-0.7)
	DR	11.6	6.0	4.8	11.4	6.1	5.1	12.3	5.9	4.5	11.6	6.7	4.0
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.1)	(0.0)	(0.0)	(0.0)	(0.6)	(0.2)	(-1.7)
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 3:2	IPW 1	23.7	15.9	15.0	24.2	14.9	12.3	28.1	17.5	11.9	28.6	20.1	18.1
		(-0.2)	(-0.3)	(-0.2)	(-2.8)	(-3.2)	(-2.9)	(-1.5)	(-2.6)	(-4.8)	(-1.9)	(-0.7)	(1.3)
	IPW 2	15.0	10.0	10.6	14.8	10.7	12.2	16.7	11.3	10.9	16.2	12.4	11.2
		(-0.7)	(-0.6)	(-0.8)	(-0.2)	(-0.3)	(-0.2)	(-0.3)	(0.1)	(0.0)	(1.2)	(0.6)	(-2.3)
	IPW 3	14.3	8.9	9.5	14.1	9.3	10.4	15.8	9.8	8.8	15.1	10.5	8.4
		(-0.4)	(-0.3)	(-0.3)	(-0.1)	(-0.1)	(-0.1)	(0.0)	(0.2)	(-0.5)	(1.0)	(0.3)	(-3.1)
	DR	10.3	6.0	8.0	10.1	6.0	8.2	10.5	5.0	4.7	10.1	5.5	3.9
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(-0.4)	(-1.0)	(-2.8)	(0.4)	(-0.9)	(-5.2)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 3:2	IPW 1	19.0	10.1	8.0	19.3	9.5	5.9	22.1	11.2	6.2	22.5	12.9	10.0
		(-0.1)	(-0.2)	(-0.1)	(-1.8)	(-1.9)	(-1.8)	(-1.0)	(-1.6)	(-3.1)	(-1.2)	(-0.3)	(0.9)
	IPW 2	11.0	5.0	3.7	10.7	5.5	4.2	12.0	5.8	4.0	11.7	6.7	3.8
		(-0.4)	(-0.4)	(-0.5)	(-0.1)	(-0.1)	(-0.1)	(-0.2)	(0.0)	(0.0)	(0.7)	(0.4)	(-1.5)
	IPW 3	11.2	4.8	3.1	11.0	5.1	3.3	12.1	5.4	2.8	11.7	5.9	2.3
		(-0.2)	(-0.2)	(-0.2)	(-0.1)	(-0.1)	(0.0)	(0.0)	(0.1)	(-0.3)	(0.7)	(0.2)	(-2.0)
	DR	10.2	4.2	2.4	10.1	4.3	2.5	10.3	3.5	0.3	10.3	4.1	0.0
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(-0.2)	(-0.5)	(-1.7)	(0.3)	(-0.6)	(-3.4)
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 2:3	IPW 1	23.9	16.3	16.7	28.1	18.5	15.8	30.3	20.7	16.9	33.1	22.2	17.5
		(0.2)	(0.0)	(0.0)	(-0.5)	(-0.9)	(-0.7)	(-0.1)	(-0.6)	(-2.1)	(-1.8)	(-1.4)	(-0.2)
	IPW 2	13.4	9.1	11.1	13.1	9.6	11.2	11.7	5.1	3.4	13.5	10.5	12.1
		(-0.6)	(-0.8)	(-0.7)	(-0.1)	(0.0)	(0.0)	(-3.2)	(-5.1)	(-8.2)	(0.4)	(0.8)	(1.0)
	IPW 3	12.7	8.5	10.3	12.5	8.9	10.3	11.8	5.6	4.3	12.9	9.7	11.0
		(-0.3)	(-0.4)	(-0.3)	(-0.1)	(0.0)	(0.0)	(-2.2)	(-3.7)	(-6.2)	(0.5)	(0.7)	(0.9)
	DR	9.4	6.0	8.8	9.1	5.9	8.6	10.1	5.7	6.5	9.4	6.4	9.0
		(0.0)	(0.0)	(0.1)	(0.0)	(0.0)	(0.1)	(-0.2)	(-0.8)	(-2.5)	(0.6)	(0.6)	(0.7)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 2:3	IPW 1	19.1	10.1	11.2	22.3	12.1	11.1	23.6	13.1	11.6	26.6	15.5	13.1
		(0.1)	(0.0)	(0.0)	(-0.3)	(-0.6)	(-0.5)	(-0.1)	(-0.4)	(-1.3)	(-1.2)	(-1.0)	(-0.2)
	IPW 2	11.2	4.6	6.3	11.0	5.3	6.3	10.3	2.2	1.6	11.3	5.9	6.9
		(-0.4)	(-0.5)	(-0.4)	(-0.1)	(0.0)	(0.0)	(-1.8)	(-3.0)	(-4.8)	(0.3)	(0.5)	(0.6)
	IPW 3	11.5	5.0	6.1	11.4	5.5	6.0	11.1	3.1	2.4	11.6	6.1	6.5
		(-0.2)	(-0.3)	(-0.2)	(0.0)	(0.0)	(0.0)	(-1.2)	(-2.2)	(-3.6)	(0.3)	(0.4)	(0.6)
	DR	11.8	5.1	5.9	11.6	5.1	5.7	12.3	4.7	4.5	11.7	5.5	6.1
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.1)	(-0.1)	(-0.5)	(-1.4)	(0.4)	(0.4)	(0.5)

Note: Percentage change which is due to the bias $\left(\frac{\text{bias}^2(\text{ATE}(\hat{p}))-\text{bias}^2(\text{ATE}(\hat{p}^s))}{\text{bias}^2(\text{ATE}(\hat{p}))+\text{Var}(\text{ATE}(\hat{p}))}\right)$ is given in brackets.

Table A6: Percentage improvement in MSE for the ATE if the shrunken propensity scores are used with trimming rule 2 instead of the conventional propensity scores with trimming rule 2, fixed valued λ

		linear			nonlinear			linear			nonlinear		
	N	100	200	500	100	200	500	100	200	500	100	200	500
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 1:1	IPW 1	24.0	18.0	10.2	24.7	18.5	9.9	25.4	19.0	11.0	30.1	26.1	19.6
		(-0.1)	(0.0)	(0.1)	(-2.8)	(-2.4)	(-1.3)	(-0.2)	(-0.1)	(0.1)	(-1.9)	(1.7)	(6.9)
	IPW 2	12.0	6.6	2.7	12.5	7.4	4.1	11.2	5.0	1.4	13.7	10.0	8.9
		(-1.1)	(-1.5)	(-2.0)	(-0.3)	(-0.1)	(0.2)	(-2.5)	(-3.2)	(-3.7)	(1.4)	(2.8)	(5.4)
	IPW 3	7.5	4.0	1.2	7.8	4.3	1.9	7.0	3.1	0.7	8.8	6.4	5.9
		(-0.5)	(-0.7)	(-1.0)	(-0.2)	(0.0)	(0.2)	(-1.3)	(-1.6)	(-1.8)	(1.2)	(2.3)	(4.4)
	DR	3.0	1.4	0.1	2.8	1.1	-0.1	3.2	1.4	0.3	3.1	2.0	2.1
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.7)	(1.2)	(2.3)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 1:1	IPW 1	19.2	13.6	6.9	19.9	14.6	7.0	20.2	14.4	7.3	24.2	20.5	13.9
		(-0.1)	(0.0)	(0.1)	(-1.7)	(-1.4)	(-0.7)	(-0.1)	(0.0)	(0.1)	(-1.2)	(1.2)	(4.9)
	IPW 2	8.5	4.0	1.9	8.8	5.2	3.2	8.0	3.0	1.1	9.6	6.8	6.1
		(-0.7)	(-0.9)	(-1.2)	(-0.2)	(-0.1)	(0.2)	(-1.5)	(-1.9)	(-2.3)	(0.9)	(1.7)	(3.3)
	IPW 3	7.2	3.6	1.7	7.4	4.2	2.4	6.9	3.0	1.3	8.0	5.5	4.8
		(-0.3)	(-0.4)	(-0.6)	(-0.1)	(0.0)	(0.1)	(-0.8)	(-0.9)	(-1.1)	(0.7)	(1.4)	(2.7)
	DR	6.1	3.7	2.1	5.9	3.6	2.2	6.2	3.8	2.1	5.9	4.1	3.5
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.4)	(0.7)	(1.4)
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 3:2	IPW 1	19.3	13.9	8.0	18.0	12.4	6.3	18.7	13.0	6.5	23.6	20.1	18.2
		(-0.2)	(-0.3)	(-0.2)	(-3.1)	(-3.4)	(-3.2)	(-1.0)	(-1.5)	(-1.6)	(-1.1)	(1.7)	(7.2)
	IPW 2	10.2	7.2	4.0	10.0	7.6	4.6	11.8	9.8	8.5	11.6	10.5	8.9
		(-0.7)	(-0.7)	(-0.8)	(-0.3)	(-0.3)	(-0.2)	(0.4)	(1.6)	(3.5)	(1.7)	(2.9)	(4.5)
	IPW 3	7.2	4.8	2.6	7.0	4.9	2.9	8.3	6.7	5.8	8.3	7.3	6.3
		(-0.4)	(-0.3)	(-0.4)	(-0.2)	(-0.2)	(-0.1)	(0.5)	(1.4)	(2.9)	(1.5)	(2.4)	(3.6)
	DR	2.6	1.8	0.8	2.6	1.8	0.7	2.8	2.3	1.6	3.1	2.9	2.1
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.2)	(0.3)	(0.8)	(0.7)	(1.2)	(1.4)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 3:2	IPW 1	15.0	10.1	5.9	14.2	9.2	5.0	14.7	9.4	5.2	18.1	14.2	13.0
		(-0.1)	(-0.2)	(-0.1)	(-1.9)	(-2.0)	(-1.9)	(-0.7)	(-0.9)	(-1.0)	(-0.7)	(1.3)	(4.7)
	IPW 2	6.1	4.6	1.4	5.9	5.0	1.8	7.0	6.2	4.2	6.9	6.8	4.5
		(-0.4)	(-0.4)	(-0.5)	(-0.2)	(-0.2)	(-0.1)	(0.2)	(0.8)	(2.2)	(1.0)	(1.8)	(2.7)
	IPW 3	5.1	3.9	1.1	4.9	4.1	1.2	5.7	5.0	3.0	5.7	5.5	3.4
		(-0.2)	(-0.2)	(-0.2)	(-0.1)	(-0.1)	(-0.1)	(0.3)	(0.8)	(1.7)	(0.9)	(1.5)	(2.2)
	DR	4.2	3.5	0.9	4.1	3.5	0.8	4.2	3.6	1.4	4.5	4.2	1.5
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.1)	(0.1)	(0.5)	(0.5)	(0.8)	(0.9)
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 2:3	IPW 1	21.0	14.6	11.1	23.0	16.0	11.7	22.6	16.3	12.6	26.9	19.6	15.0
		(0.0)	(0.0)	(0.0)	(-1.2)	(-0.9)	(-0.7)	(0.4)	(1.1)	(1.4)	(-1.8)	(-0.6)	(0.8)
	IPW 2	10.1	6.4	5.1	10.0	6.9	6.1	7.7	3.2	0.2	10.2	7.5	7.0
		(-0.6)	(-0.9)	(-0.9)	(-0.1)	(0.0)	(0.0)	(-3.2)	(-4.2)	(-5.6)	(0.1)	(0.6)	(1.0)
	IPW 3	6.7	4.3	4.1	6.6	4.6	4.7	5.2	2.3	1.1	6.7	5.0	5.5
		(-0.3)	(-0.5)	(-0.5)	(-0.1)	(0.0)	(0.0)	(-1.9)	(-2.4)	(-3.3)	(0.1)	(0.5)	(0.8)
	DR	2.4	1.7	2.9	2.2	1.6	3.0	2.7	2.2	3.5	2.1	1.7	3.2
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.2)	(0.5)	(0.6)	(0.1)	(0.3)	(0.3)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 2:3	IPW 1	15.1	10.6	7.2	17.0	12.1	7.7	16.4	12.0	8.4	20.5	15.2	10.4
		(0.0)	(0.0)	(0.0)	(-0.7)	(-0.6)	(-0.5)	(0.3)	(0.7)	(0.9)	(-1.2)	(-0.4)	(0.5)
	IPW 2	6.0	3.2	2.2	6.2	3.9	2.5	4.5	1.3	-0.5	6.3	4.2	3.2
		(-0.4)	(-0.5)	(-0.5)	(-0.1)	(0.0)	(0.0)	(-1.9)	(-2.4)	(-3.2)	(0.1)	(0.4)	(0.6)
	IPW 3	5.1	2.9	2.1	5.2	3.4	2.3	4.2	1.8	0.5	5.3	3.7	2.8
		(-0.2)	(-0.3)	(-0.2)	(0.0)	(0.0)	(0.0)	(-1.1)	(-1.3)	(-1.8)	(0.1)	(0.3)	(0.5)
	DR	4.3	3.0	2.2	4.2	2.9	2.2	4.5	3.4	2.5	4.1	2.9	2.3
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.2)	(0.4)	(0.4)	(0.1)	(0.2)	(0.2)

Note: Percentage change which is due to the bias $\left(\frac{\text{bias}^2(\text{ATE}(\hat{p}))-\text{bias}^2(\text{ATE}(\hat{p}^s))}{\text{bias}^2(\text{ATE}(\hat{p}))+\text{Var}(\text{ATE}(\hat{p}))}\right)$ is given in brackets.

Table A7: Percentage improvement in MSE for the ATE if the shrunken propensity scores are used with trimming rule 2 instead of the conventional propensity scores, MSE($\hat{p}_i^s$)-minimizing λ

			linear			nonlinear			linear			nonlinear	
	N	100	200	500	100	200	500	100	200	500	100	200	500
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 1:1	IPW 1	55.2	64.1	66.9	55.8	64.9	64.5	55.7	64.9	66.9	62.8	71.7	70.2
		(-0.1)	(-0.1)	(0.0)	(-1.6)	(-1.3)	(-0.6)	(-0.1)	(-0.2)	(0.0)	(-1.7)	(-0.7)	(0.1)
	IPW 2	23.3	28.6	37.3	23.9	29.2	38.6	23.0	28.7	38.4	23.8	29.0	35.0
		(-0.8)	(-0.5)	(-1.2)	(-0.1)	(-0.2)	(0.0)	(-1.9)	(-1.3)	(-2.1)	(0.1)	(-0.1)	(-3.1)
	IPW 3	11.9	15.3	19.0	12.5	15.8	20.9	11.0	14.8	18.2	12.5	15.4	15.9
		(-0.3)	(-0.1)	(-0.5)	(0.0)	(0.0)	(0.0)	(-0.7)	(-0.3)	(-0.9)	(0.0)	(-0.3)	(-4.6)
	DR	13.8	17.6	22.6	14.5	18.4	24.1	13.5	17.6	22.7	14.9	18.3	19.5
		(0.1)	(0.1)	(0.0)	(0.0)	(0.0)	(-0.1)	(0.1)	(0.3)	(0.1)	(-0.7)	(-1.2)	(-5.8)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 1:1	IPW 1	38.1	50.3	53.0	39.9	52.2	50.4	39.1	51.6	53.3	48.9	61.0	58.0
		(-0.1)	(-0.1)	(0.0)	(-1.3)	(-1.1)	(-0.4)	(-0.1)	(-0.1)	(0.0)	(-1.4)	(-0.6)	(0.1)
	IPW 2	7.8	15.9	22.8	8.3	16.2	23.0	7.9	16.1	25.0	8.4	16.1	20.3
		(-0.5)	(-0.3)	(-1.0)	(-0.1)	(-0.1)	(0.0)	(-1.2)	(-0.9)	(-1.7)	(0.1)	(0.0)	(-2.4)
	IPW 3	1.1	6.0	7.0	1.6	6.3	8.4	0.6	5.6	6.9	1.7	6.1	5.1
		(-0.2)	(-0.1)	(-0.5)	(0.0)	(0.0)	(0.0)	(-0.4)	(-0.2)	(-0.7)	(0.0)	(-0.2)	(-3.4)
	DR	4.2	8.3	9.4	4.8	8.9	11.3	4.0	8.3	9.4	5.3	8.9	8.7
		(0.0)	(0.1)	(-0.1)	(0.0)	(0.0)	(-0.1)	(0.1)	(0.2)	(0.0)	(-0.4)	(-0.8)	(-4.3)
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 3:2	IPW 1	47.9	86.0	43.7	41.7	69.1	40.2	46.9	85.3	40.0	46.1	70.0	47.1
		(-0.1)	(0.0)	(-0.2)	(-2.4)	(-1.1)	(-2.4)	(-0.9)	(-0.4)	(-3.7)	(-2.6)	(-0.6)	(-0.1)
	IPW 2	20.8	22.0	28.3	21.5	23.4	31.1	21.5	22.7	28.8	20.9	22.5	28.2
		(-0.9)	(-0.7)	(-0.4)	(-0.2)	(-0.2)	(-0.2)	(-0.6)	(-0.1)	(-0.1)	(0.1)	(-0.5)	(-2.4)
	IPW 3	12.3	11.7	17.1	12.9	12.9	17.9	12.6	11.9	16.5	12.3	11.6	14.0
		(-0.4)	(-0.3)	(-0.2)	(0.0)	(0.0)	(0.2)	(0.0)	(0.2)	(-0.7)	(0.0)	(-0.8)	(-3.3)
	DR	12.8	12.8	19.5	13.1	13.3	20.1	12.3	11.7	16.6	11.8	11.1	15.1
		(0.0)	(0.1)	(0.2)	(0.0)	(0.0)	(0.1)	(-0.4)	(-1.0)	(-2.6)	(-1.1)	(-2.3)	(-5.3)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 3:2	IPW 1	31.9	78.0	31.0	26.5	55.8	28.4	31.4	77.3	28.7	31.1	57.1	35.0
		(0.0)	(0.0)	(-0.1)	(-1.6)	(-1.0)	(-1.7)	(-0.6)	(-0.4)	(-2.6)	(-1.8)	(-0.6)	(-0.1)
	IPW 2	6.4	8.6	15.9	7.0	9.8	18.5	7.0	9.3	16.3	6.5	9.2	16.3
		(-0.6)	(-0.5)	(-0.4)	(-0.1)	(-0.1)	(-0.1)	(-0.4)	(-0.1)	(-0.1)	(0.1)	(-0.3)	(-1.8)
	IPW 3	1.2	1.7	7.8	1.6	2.5	8.6	1.4	1.9	7.5	1.2	1.6	5.8
		(-0.3)	(-0.2)	(-0.2)	(0.0)	(0.0)	(0.1)	(0.0)	(0.1)	(-0.4)	(0.0)	(-0.5)	(-2.4)
	DR	1.9	2.7	10.1	2.1	3.1	10.4	1.6	2.0	8.2	1.3	1.7	6.9
		(0.0)	(0.0)	(0.1)	(0.0)	(0.0)	(0.1)	(-0.2)	(-0.7)	(-1.7)	(-0.8)	(-1.5)	(-3.9)
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 2:3	IPW 1	71.1	55.9	44.6	75.9	61.0	48.2	72.2	57.0	43.7	80.3	67.1	56.7
		(0.0)	(0.0)	(0.4)	(-0.4)	(-0.6)	(-0.2)	(0.0)	(-0.2)	(-1.0)	(-0.7)	(-0.8)	(-0.2)
	IPW 2	20.3	23.0	29.7	21.3	24.1	29.1	17.8	19.7	24.9	21.3	24.2	28.8
		(-1.1)	(-0.8)	(-0.9)	(-0.1)	(-0.1)	(0.2)	(-4.5)	(-4.9)	(-7.7)	(0.0)	(0.0)	(-0.2)
	IPW 3	11.4	12.8	15.7	12.1	13.6	16.4	8.6	9.2	8.9	12.2	13.6	15.7
		(-0.6)	(-0.4)	(-0.6)	(0.0)	(0.0)	(0.1)	(-3.2)	(-3.8)	(-7.1)	(0.0)	(0.0)	(-0.5)
	DR	12.7	14.9	17.9	13.1	15.2	18.5	12.4	14.3	15.2	13.8	15.8	18.3
		(0.0)	(0.0)	(0.2)	(0.0)	(0.0)	(0.0)	(-0.2)	(-0.6)	(-2.3)	(0.0)	(-0.1)	(-1.0)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 2:3	IPW 1	57.0	42.1	29.6	64.1	48.3	33.7	59.0	43.7	29.2	71.2	56.5	43.8
		(0.0)	(0.0)	(0.3)	(-0.3)	(-0.5)	(-0.1)	(0.0)	(-0.1)	(-0.9)	(-0.6)	(-0.7)	(-0.1)
	IPW 2	7.4	12.3	17.8	7.9	12.8	16.9	6.0	10.5	15.4	8.0	12.9	16.8
		(-0.6)	(-0.5)	(-0.7)	(0.0)	(-0.1)	(0.1)	(-2.8)	(-3.1)	(-5.5)	(0.0)	(0.0)	(-0.2)
	IPW 3	2.9	5.9	6.8	3.3	6.2	7.2	1.3	3.8	2.6	3.4	6.3	6.9
		(-0.3)	(-0.2)	(-0.5)	(0.0)	(0.0)	(0.1)	(-1.9)	(-2.3)	(-4.8)	(0.0)	(0.0)	(-0.4)
	DR	6.9	8.7	8.4	7.1	8.9	9.5	6.7	8.5	6.4	7.5	9.3	9.8
		(0.0)	(0.0)	(0.1)	(0.0)	(0.0)	(0.0)	(-0.1)	(-0.3)	(-1.7)	(0.0)	(0.0)	(-0.7)

Note: Percentage change which is due to the bias $\left(\frac{\text{bias}^2(\text{ATE}(\hat{p}))-\text{bias}^2(\text{ATE}(\hat{p}^s))}{\text{bias}^2(\text{ATE}(\hat{p}))+\text{Var}(\text{ATE}(\hat{p}))}\right)$ is given in brackets.

Table A8: Percentage improvement in MSE for the ATE if the shrunken propensity scores are used with trimming rule 2 instead of the conventional propensity scores with trimming rule 1, MSE($\hat{p}_i^s$)-minimizing λ

			linear			nonlinear			linear			nonlinear	
	N	100	200	500	100	200	500	100	200	500	100	200	500
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 1:1	IPW 1	28.6	18.1	18.6	30.4	17.0	18.1	35.4	23.2	22.2	34.7	20.9	24.1
		(-0.1)	(-0.2)	(0.1)	(-2.6)	(-3.2)	(-1.5)	(-0.1)	(-0.3)	(0.0)	(-3.4)	(-2.4)	(1.7)
	IPW 2	17.2	10.7	9.5	17.8	10.8	11.7	16.7	10.6	9.5	19.4	12.5	10.0
		(-1.3)	(-0.8)	(-1.7)	(-0.2)	(-0.3)	(0.0)	(-2.8)	(-2.0)	(-3.1)	(1.4)	(1.0)	(-1.6)
	IPW 3	16.5	8.8	8.3	16.8	8.8	9.5	16.6	9.3	8.6	18.1	10.2	7.1
		(-0.6)	(-0.3)	(-0.9)	(-0.1)	(-0.2)	(0.0)	(-1.4)	(-0.8)	(-1.6)	(1.3)	(0.9)	(-2.2)
	DR	12.5	4.8	7.4	12.1	4.9	7.3	13.2	5.7	8.7	12.8	5.4	3.4
		(0.0)	(0.0)	(-0.1)	(0.0)	(0.0)	(-0.1)	(0.0)	(0.0)	(-0.1)	(1.0)	(0.2)	(-3.9)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 1:1	IPW 1	21.6	11.7	9.5	23.2	11.2	9.3	26.4	15.2	12.7	26.7	14.3	14.1
		(-0.1)	(-0.1)	(0.0)	(-1.7)	(-2.0)	(-0.9)	(-0.1)	(-0.2)	(0.0)	(-2.3)	(-1.6)	(1.2)
	IPW 2	12.4	5.3	2.8	12.8	5.5	4.0	12.2	5.3	3.5	13.8	6.6	2.8
		(-0.7)	(-0.5)	(-1.2)	(-0.2)	(-0.2)	(0.0)	(-1.6)	(-1.2)	(-2.2)	(0.7)	(0.6)	(-1.3)
	IPW 3	13.3	4.9	2.6	13.4	5.0	3.4	13.3	5.2	3.4	14.2	5.9	1.7
		(-0.3)	(-0.2)	(-0.7)	(-0.1)	(-0.1)	(0.0)	(-0.8)	(-0.5)	(-1.1)	(0.7)	(0.5)	(-1.7)
	DR	13.4	3.9	2.3	13.2	4.1	2.4	13.8	4.3	3.5	13.5	4.5	-0.2
		(0.0)	(0.0)	(-0.1)	(0.0)	(0.0)	(-0.1)	(0.0)	(0.0)	(-0.1)	(0.5)	(0.1)	(-2.8)
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 3:2	IPW 1	24.1	16.0	13.8	24.4	13.8	10.1	28.5	17.7	11.4	28.4	17.8	15.9
		(-0.1)	(-0.2)	(-0.4)	(-3.1)	(-3.2)	(-3.6)	(-1.1)	(-2.2)	(-5.3)	(-2.7)	(-0.7)	(1.0)
	IPW 2	15.2	9.8	8.7	15.7	9.8	9.3	16.3	11.4	10.1	17.1	11.1	7.9
		(-1.1)	(-0.9)	(-0.5)	(-0.3)	(-0.3)	(-0.2)	(-0.9)	(-0.2)	(0.1)	(1.2)	(0.8)	(-2.0)
	IPW 3	14.7	8.8	7.5	15.0	8.6	7.7	15.8	10.0	7.9	16.2	9.3	5.5
		(-0.6)	(-0.5)	(-0.2)	(-0.2)	(-0.1)	(-0.1)	(-0.3)	(0.1)	(-0.5)	(1.1)	(0.5)	(-2.6)
	DR	11.2	5.3	5.8	11.0	5.4	5.7	11.4	4.8	3.7	11.0	4.6	1.1
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(-0.2)	(-0.9)	(-2.8)	(0.4)	(-0.8)	(-4.7)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 3:2	IPW 1	18.9	11.1	8.0	19.1	9.5	5.7	22.1	12.5	6.6	21.8	12.0	9.5
		(-0.1)	(-0.1)	(-0.2)	(-1.8)	(-2.1)	(-2.2)	(-0.7)	(-1.4)	(-3.3)	(-1.7)	(-0.6)	(0.6)
	IPW 2	10.4	5.3	3.4	10.5	5.0	3.9	11.2	6.6	4.4	11.3	5.6	3.0
		(-0.6)	(-0.5)	(-0.4)	(-0.1)	(-0.2)	(-0.1)	(-0.5)	(-0.1)	(0.1)	(0.7)	(0.5)	(-1.5)
	IPW 3	10.6	5.1	2.9	10.6	4.8	3.2	11.3	6.1	3.2	11.2	5.1	1.7
		(-0.4)	(-0.3)	(-0.2)	(-0.1)	(-0.1)	(0.0)	(-0.2)	(0.1)	(-0.2)	(0.6)	(0.3)	(-1.9)
	DR	9.6	4.0	2.6	9.5	4.1	2.6	9.8	3.9	1.2	9.4	3.5	-0.4
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(-0.1)	(-0.5)	(-1.6)	(0.2)	(-0.5)	(-3.2)
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 2:3	IPW 1	26.8	15.9	13.7	30.4	16.9	14.0	33.0	20.1	14.3	34.7	19.4	16.9
		(0.1)	(0.1)	(0.0)	(-1.1)	(-1.3)	(-0.9)	(-0.2)	(-0.5)	(-2.1)	(-2.8)	(-2.5)	(-0.9)
	IPW 2	15.1	9.8	9.6	15.6	10.1	10.9	12.6	5.8	1.6	16.0	10.7	11.4
		(-1.2)	(-1.0)	(-1.3)	(-0.1)	(-0.1)	(0.0)	(-4.6)	(-5.7)	(-9.8)	(0.6)	(0.6)	(0.3)
	IPW 3	14.9	8.8	8.5	15.1	9.0	9.3	13.4	6.0	1.8	15.5	9.6	9.6
		(-0.7)	(-0.5)	(-0.7)	(0.0)	(0.0)	(0.0)	(-3.1)	(-4.0)	(-7.5)	(0.6)	(0.5)	(0.2)
	DR	11.7	5.6	7.0	11.4	5.7	7.4	12.2	5.4	4.1	11.5	6.2	7.4
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(-0.2)	(-0.7)	(-3.0)	(0.7)	(0.5)	(-0.2)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 2:3	IPW 1	19.7	11.1	8.2	22.5	12.0	8.6	24.0	14.1	9.1	26.3	14.3	11.2
		(0.1)	(0.0)	(0.0)	(-0.7)	(-0.9)	(-0.5)	(-0.1)	(-0.3)	(-1.5)	(-1.9)	(-1.7)	(-0.6)
	IPW 2	11.1	5.7	4.4	11.3	5.7	4.9	9.6	3.5	-0.2	11.6	6.1	5.1
		(-0.7)	(-0.6)	(-0.8)	(0.0)	(-0.1)	(0.0)	(-2.6)	(-3.3)	(-6.3)	(0.3)	(0.3)	(0.2)
	IPW 3	11.8	5.8	4.0	11.9	5.8	4.4	10.9	4.3	0.3	12.1	6.2	4.5
		(-0.4)	(-0.3)	(-0.5)	(0.0)	(0.0)	(0.0)	(-1.8)	(-2.3)	(-4.8)	(0.3)	(0.3)	(0.1)
	DR	12.6	5.7	3.6	12.4	5.8	3.9	12.9	5.6	2.2	12.5	6.1	3.8
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(-0.1)	(-0.4)	(-2.0)	(0.4)	(0.3)	(-0.2)

Note: Percentage change which is due to the bias $\left(\frac{\text{bias}^2(\text{ATE}(\hat{p}))-\text{bias}^2(\text{ATE}(\hat{p}^s))}{\text{bias}^2(\text{ATE}(\hat{p}))+\text{Var}(\text{ATE}(\hat{p}))}\right)$ is given in brackets.

Table A9: Percentage improvement in MSE for the ATE if the shrunken propensity scores are used with trimming rule 2 instead of the conventional propensity scores with trimming rule 2, MSE($\hat{p}_i^s$)-minimizing λ

			linear			nonlinear			linear			nonlinear	
	N	100	200	500	100	200	500	100	200	500	100	200	500
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 1:1	IPW 1	24.6	14.4	12.4	24.6	13.3	13.0	25.7	15.0	13.4	29.4	19.6	23.1
		(-0.1)	(-0.3)	(0.1)	(-3.1)	(-3.6)	(-1.5)	(-0.2)	(-0.5)	(0.1)	(-2.3)	(-0.6)	(6.5)
	IPW 2	12.4	7.5	3.6	13.2	7.6	5.5	11.1	6.7	2.0	14.5	9.8	9.5
		(-1.4)	(-0.7)	(-1.8)	(-0.3)	(-0.4)	(0.1)	(-3.2)	(-2.0)	(-3.4)	(1.6)	(2.1)	(4.6)
	IPW 3	8.7	4.3	2.3	9.0	4.3	3.3	7.9	3.9	1.5	10.2	6.1	6.7
		(-0.7)	(-0.3)	(-0.9)	(-0.2)	(-0.2)	(0.1)	(-1.7)	(-0.9)	(-1.6)	(1.4)	(1.7)	(3.8)
	DR	4.0	0.9	1.6	3.7	1.0	1.2	4.1	0.9	1.8	4.0	1.8	2.8
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.8)	(0.8)	(1.9)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 1:1	IPW 1	18.3	11.0	9.2	18.7	10.7	9.7	19.2	11.4	9.9	22.5	15.2	17.0
		(-0.1)	(-0.2)	(0.1)	(-1.9)	(-2.2)	(-0.8)	(-0.1)	(-0.3)	(0.1)	(-1.6)	(-0.4)	(4.5)
	IPW 2	7.3	4.6	1.9	7.8	5.1	2.9	6.7	4.2	1.0	8.6	6.5	5.5
		(-0.7)	(-0.4)	(-1.2)	(-0.2)	(-0.2)	(0.1)	(-1.7)	(-1.2)	(-2.2)	(0.9)	(1.3)	(2.8)
	IPW 3	6.5	3.7	1.6	6.7	4.0	2.1	6.1	3.5	1.1	7.4	5.1	4.2
		(-0.4)	(-0.2)	(-0.6)	(-0.1)	(-0.1)	(0.1)	(-0.9)	(-0.5)	(-1.1)	(0.8)	(1.0)	(2.2)
	DR	6.1	3.4	1.6	5.9	3.5	1.3	6.2	3.4	1.6	6.1	4.0	2.2
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.4)	(0.5)	(1.0)
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 3:2	IPW 1	21.5	13.4	9.4	20.0	11.7	7.3	21.3	12.7	7.5	24.8	19.1	20.3
		(-0.1)	(-0.2)	(-0.4)	(-3.3)	(-3.3)	(-3.8)	(-0.7)	(-1.3)	(-2.6)	(-1.8)	(1.7)	(7.6)
	IPW 2	10.4	6.1	2.8	10.8	6.7	3.0	11.7	8.6	6.8	12.5	9.4	8.4
		(-1.1)	(-1.0)	(-0.6)	(-0.3)	(-0.3)	(-0.4)	(-0.3)	(1.3)	(3.3)	(2.0)	(3.0)	(5.5)
	IPW 3	7.1	3.8	1.0	7.3	4.2	1.0	8.1	5.7	3.8	8.8	6.4	5.6
		(-0.7)	(-0.5)	(-0.2)	(-0.2)	(-0.2)	(-0.3)	(0.1)	(1.3)	(2.5)	(1.8)	(2.5)	(4.7)
	DR	2.5	1.0	-1.0	2.4	1.1	-1.3	2.9	1.6	-0.6	2.9	1.9	1.0
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.2)	(0.4)	(0.1)	(0.9)	(1.1)	(2.5)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 3:2	IPW 1	16.0	9.8	6.5	15.2	8.7	5.3	16.1	9.4	5.4	18.5	13.6	14.1
		(-0.1)	(-0.1)	(-0.2)	(-1.9)	(-2.1)	(-2.2)	(-0.4)	(-0.8)	(-1.6)	(-1.1)	(1.0)	(5.3)
	IPW 2	6.2	3.0	1.2	6.4	3.3	1.5	7.0	4.5	3.7	7.4	5.0	4.9
		(-0.7)	(-0.5)	(-0.4)	(-0.2)	(-0.2)	(-0.2)	(-0.2)	(0.8)	(2.0)	(1.2)	(1.7)	(3.5)
	IPW 3	5.2	2.3	0.7	5.3	2.5	0.8	5.8	3.5	2.4	6.1	3.8	3.7
		(-0.4)	(-0.3)	(-0.2)	(-0.1)	(-0.1)	(-0.1)	(0.1)	(0.8)	(1.5)	(1.0)	(1.5)	(3.0)
	DR	4.6	2.1	0.6	4.5	2.1	0.5	4.9	2.4	0.8	4.8	2.6	1.8
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.1)	(0.3)	(0.1)	(0.5)	(0.6)	(1.5)
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 2:3	IPW 1	21.2	13.9	9.1	23.3	15.4	10.0	23.2	15.8	10.4	27.2	18.9	13.3
		(0.0)	(0.0)	(0.0)	(-1.5)	(-1.5)	(-1.0)	(0.9)	(1.1)	(1.3)	(-2.5)	(-1.8)	(0.0)
	IPW 2	8.1	5.7	2.9	9.0	6.9	4.5	5.2	2.2	-2.3	9.2	7.4	5.6
		(-1.4)	(-1.1)	(-1.3)	(-0.1)	(-0.1)	(0.0)	(-4.8)	(-4.8)	(-6.4)	(0.3)	(0.5)	(1.2)
	IPW 3	5.4	3.5	2.0	5.9	4.1	2.8	3.6	1.4	-1.2	6.0	4.5	3.7
		(-0.8)	(-0.6)	(-0.7)	(0.0)	(-0.1)	(0.0)	(-2.9)	(-2.8)	(-3.6)	(0.3)	(0.5)	(1.0)
	DR	1.1	0.7	0.9	0.9	0.7	0.8	1.7	1.3	1.6	0.7	0.8	1.2
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.4)	(0.5)	(0.8)	(0.2)	(0.2)	(0.5)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 2:3	IPW 1	15.6	10.3	6.7	17.2	11.5	7.6	17.1	11.7	7.7	20.5	14.4	10.3
		(0.0)	(0.0)	(0.0)	(-1.0)	(-1.0)	(-0.6)	(0.5)	(0.7)	(0.8)	(-1.7)	(-1.2)	(0.1)
	IPW 2	5.1	3.0	1.5	5.5	3.6	2.5	3.5	1.1	-1.8	5.6	4.0	3.1
		(-0.8)	(-0.6)	(-0.8)	(0.0)	(-0.1)	(0.0)	(-2.7)	(-2.7)	(-4.0)	(0.1)	(0.3)	(0.7)
	IPW 3	4.6	2.6	1.6	4.8	2.8	2.0	3.6	1.5	-0.4	4.8	3.2	2.6
		(-0.4)	(-0.3)	(-0.4)	(0.0)	(0.0)	(0.0)	(-1.6)	(-1.6)	(-2.3)	(0.1)	(0.3)	(0.6)
	DR	4.1	2.4	1.6	3.9	2.4	1.5	4.3	2.9	1.9	3.8	2.6	1.7
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.2)	(0.3)	(0.4)	(0.1)	(0.2)	(0.3)

Note: Percentage change which is due to the bias $\left(\frac{\text{bias}^2(\text{ATE}(\hat{p}))-\text{bias}^2(\text{ATE}(\hat{p}^s))}{\text{bias}^2(\text{ATE}(\hat{p}))+\text{Var}(\text{ATE}(\hat{p}))}\right)$ is given in brackets.

Table A10: Percentage improvement in MSE for the ATE if the shrunken propensity scores are used with trimming rule 2 instead of the conventional propensity scores, data-driven λ

			linear			nonlinear			linear			nonlinear	
	N	100	200	500	100	200	500	100	200	500	100	200	500
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 1:1	IPW 1	55.4	56.9	81.3	53.5	53.4	73.0	55.4	56.8	80.8	58.9	59.1	73.5
		(0.0)	(0.1)	(0.1)	(0.0)	(0.2)	(0.1)	(0.0)	(0.2)	(0.1)	(0.4)	(0.6)	(-0.6)
	IPW 2	18.2	24.1	41.0	19.4	22.7	36.9	18.9	25.5	42.0	18.5	21.4	32.9
		(0.1)	(0.0)	(0.1)	(0.1)	(0.2)	(0.0)	(0.1)	(0.1)	(0.3)	(-0.6)	(-0.9)	(-3.8)
	IPW 3	9.0	10.5	18.0	9.8	10.8	18.8	9.1	10.5	17.6	8.9	9.4	13.8
		(0.1)	(0.1)	(0.4)	(0.1)	(0.3)	(0.0)	(0.3)	(0.3)	(0.7)	(-0.8)	(-1.2)	(-5.0)
	DR	13.0	14.4	22.4	13.8	15.1	23.8	13.2	14.4	22.0	13.6	14.6	20.4
		(0.0)	(0.2)	(0.0)	(0.0)	(0.0)	(0.0)	(0.1)	(0.4)	(0.1)	(-1.2)	(-1.9)	(-4.8)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 1:1	IPW 1	39.2	41.9	71.7	37.6	38.8	61.4	39.6	42.2	71.5	45.0	46.3	63.4
		(0.0)	(0.1)	(0.1)	(-0.1)	(0.2)	(0.1)	(0.0)	(0.1)	(0.1)	(0.3)	(0.5)	(-0.6)
	IPW 2	6.0	10.5	24.9	6.3	9.0	21.9	6.8	11.7	26.6	5.9	8.2	18.7
		(0.1)	(0.0)	(0.1)	(0.0)	(0.1)	(0.0)	(0.2)	(0.0)	(0.2)	(-0.3)	(-0.6)	(-2.9)
	IPW 3	-0.8	0.0	6.8	-0.6	0.0	7.8	-0.6	-0.1	6.8	-0.9	-0.8	4.2
		(0.1)	(0.0)	(0.2)	(0.1)	(0.2)	(0.0)	(0.3)	(0.2)	(0.4)	(-0.4)	(-0.8)	(-3.5)
	DR	1.8	2.7	10.0	2.4	3.2	11.6	2.0	2.7	9.9	2.6	3.1	9.4
		(0.0)	(0.2)	(0.0)	(0.0)	(0.0)	(0.0)	(0.1)	(0.3)	(0.1)	(-0.7)	(-1.3)	(-3.5)
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 3:2	IPW 1	35.8	39.6	35.3	32.4	36.2	28.3	34.1	37.9	32.6	36.1	41.2	27.8
		(0.0)	(0.0)	(0.0)	(-1.0)	(-0.5)	(-0.1)	(-0.8)	(-1.3)	(-1.9)	(-0.2)	(-0.1)	(-3.5)
	IPW 2	18.5	23.3	20.9	20.0	24.1	21.1	19.1	23.5	20.5	18.8	21.7	12.8
		(-0.1)	(-0.1)	(-0.1)	(0.0)	(0.1)	(0.0)	(0.2)	(-0.1)	(-1.1)	(-0.3)	(-1.7)	(-7.5)
	IPW 3	10.8	12.9	10.7	11.3	13.8	11.5	11.0	12.1	9.4	10.0	11.1	2.1
		(0.0)	(0.0)	(0.0)	(0.1)	(0.2)	(0.1)	(0.3)	(-0.2)	(-1.2)	(-0.4)	(-2.1)	(-8.5)
	DR	13.0	16.2	13.7	13.4	16.6	14.0	12.4	14.6	11.7	11.8	13.7	4.7
		(0.0)	(0.1)	(0.0)	(0.0)	(0.0)	(0.0)	(-0.6)	(-1.1)	(-2.3)	(-1.5)	(-3.1)	(-9.1)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 3:2	IPW 1	20.2	23.9	19.7	17.6	21.2	13.8	19.2	23.0	17.8	21.3	26.5	14.6
		(0.0)	(0.0)	(0.0)	(-0.7)	(-0.3)	(0.0)	(-0.7)	(-1.0)	(-1.4)	(-0.2)	(-0.1)	(-2.7)
	IPW 2	4.6	9.0	6.7	5.9	9.8	7.6	5.2	9.3	6.5	5.2	8.2	2.0
		(-0.1)	(-0.1)	(-0.1)	(0.0)	(0.1)	(0.0)	(0.2)	(-0.1)	(-0.7)	(-0.2)	(-1.1)	(-5.2)
	IPW 3	-0.6	1.5	-1.1	-0.2	2.2	0.1	-0.5	1.1	-2.1	-0.9	0.4	-6.0
		(0.0)	(0.0)	(0.0)	(0.0)	(0.1)	(0.0)	(0.2)	(-0.1)	(-0.8)	(-0.2)	(-1.3)	(-5.8)
	DR	0.3	3.3	1.0	0.6	3.7	1.2	-0.1	2.4	-0.3	-0.2	1.8	-5.0
		(0.0)	(0.1)	(0.0)	(0.0)	(0.0)	(0.0)	(-0.4)	(-0.8)	(-1.5)	(-0.9)	(-1.9)	(-6.3)
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 2:3	IPW 1	48.8	41.4	37.6	56.1	46.3	42.1	50.4	42.1	35.7	64.1	53.5	50.1
		(0.1)	(0.2)	(0.0)	(-0.1)	(0.1)	(0.1)	(0.0)	(0.1)	(-2.2)	(0.0)	(0.4)	(0.0)
	IPW 2	16.9	19.1	24.5	17.8	20.0	24.8	15.6	18.2	22.6	17.7	20.2	24.1
		(-0.2)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(-2.1)	(-1.7)	(-3.8)	(0.0)	(0.1)	(-0.8)
	IPW 3	9.7	10.2	13.8	10.2	10.8	14.3	8.1	8.4	9.1	10.1	11.1	13.4
		(-0.1)	(0.0)	(0.1)	(0.0)	(0.0)	(0.0)	(-1.8)	(-1.5)	(-4.1)	(-0.1)	(0.1)	(-1.0)
	DR	12.6	13.0	17.4	12.8	13.5	18.1	12.7	12.7	14.3	13.2	14.3	18.0
		(0.0)	(0.2)	(0.0)	(0.0)	(0.0)	(0.1)	(-0.2)	(0.0)	(-3.1)	(-0.1)	(-0.1)	(-0.9)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 2:3	IPW 1	33.4	27.5	23.4	41.7	33.0	28.3	35.6	28.6	22.4	52.1	41.6	37.7
		(0.0)	(0.1)	(0.0)	(-0.1)	(0.0)	(0.1)	(0.0)	(0.0)	(-1.7)	(0.0)	(0.3)	(0.0)
	IPW 2	6.3	7.8	13.0	6.8	8.2	13.7	5.9	7.4	12.4	6.8	8.4	13.4
		(-0.1)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(-1.2)	(-1.1)	(-2.6)	(0.0)	(0.0)	(-0.5)
	IPW 3	2.0	1.4	5.9	2.1	1.8	6.6	1.2	0.4	3.2	2.1	2.1	6.1
		(0.0)	(0.0)	(0.1)	(0.0)	(0.0)	(0.0)	(-1.0)	(-0.9)	(-2.6)	(0.0)	(0.0)	(-0.6)
	DR	5.8	4.3	8.8	5.9	4.6	9.7	6.0	4.2	6.9	6.2	5.3	9.9
		(0.0)	(0.1)	(0.0)	(0.0)	(0.0)	(0.0)	(-0.1)	(0.0)	(-2.0)	(0.0)	(-0.1)	(-0.5)

Note: Percentage change which is due to the bias $\left(\frac{\text{bias}^2(\text{ATE}(\hat{p}))-\text{bias}^2(\text{ATE}(\hat{p}^s))}{\text{bias}^2(\text{ATE}(\hat{p}))+\text{Var}(\text{ATE}(\hat{p}))}\right)$ is given in brackets.

Table A11: Percentage improvement in MSE for the ATE if the shrunken propensity scores are used with trimming rule 2 instead of the conventional propensity scores with trimming rule 1, data-driven λ

		linear			nonlinear			linear			nonlinear		
	N	100	200	500	100	200	500	100	200	500	100	200	500
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 1:1	IPW 1	14.5	11.2	8.8	17.6	13.4	7.5	22.1	16.5	11.9	21.6	16.6	7.1
		(0.0)	(0.0)	(0.0)	(-0.3)	(-0.3)	(0.0)	(0.0)	(-0.1)	(0.0)	(0.2)	(0.2)	(-1.9)
	IPW 2	8.8	6.7	7.7	8.3	7.6	7.4	9.9	7.2	9.6	9.2	8.6	4.8
		(-0.3)	(0.0)	(0.0)	(0.0)	(-0.1)	(-0.1)	(-0.6)	(-0.1)	(0.0)	(1.0)	(0.1)	(-3.1)
	IPW 3	9.9	7.6	7.2	9.7	8.2	7.2	11.1	8.2	8.6	10.4	8.9	4.4
		(-0.2)	(0.0)	(0.0)	(0.0)	(-0.1)	(0.0)	(-0.4)	(0.0)	(0.0)	(0.9)	(0.0)	(-3.1)
	DR	8.3	6.7	6.2	8.0	6.4	6.6	9.4	7.7	7.4	8.2	6.6	3.9
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.7)	(-0.1)	(-3.1)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 1:1	IPW 1	11.4	7.6	3.9	13.5	9.0	3.1	16.3	11.0	6.6	16.5	11.5	2.6
		(0.0)	(0.0)	(0.0)	(-0.3)	(-0.2)	(0.0)	(0.0)	(0.0)	(0.0)	(0.1)	(0.1)	(-1.5)
	IPW 2	7.5	4.9	3.7	7.4	5.3	3.8	8.2	5.1	5.7	8.1	6.0	1.7
		(-0.1)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(-0.3)	(-0.1)	(0.0)	(0.6)	(0.0)	(-2.1)
	IPW 3	8.5	5.4	3.8	8.5	5.6	4.0	9.2	5.6	5.3	9.1	6.1	1.8
		(-0.1)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(-0.2)	(0.0)	(0.0)	(0.5)	(0.0)	(-2.1)
	DR	8.9	5.0	3.5	8.8	4.9	3.9	9.5	5.4	4.8	9.1	5.1	1.8
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.4)	(-0.1)	(-2.1)
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 3:2	IPW 1	16.6	9.4	7.9	17.8	9.9	6.7	21.4	12.2	7.7	21.3	12.2	3.2
		(0.0)	(-0.1)	(0.1)	(-1.2)	(-0.7)	(0.0)	(-1.0)	(-1.8)	(-2.7)	(0.5)	(1.3)	(-2.1)
	IPW 2	12.7	7.1	7.1	12.1	7.4	7.7	14.1	8.1	7.2	13.2	7.4	2.1
		(-0.3)	(-0.1)	(-0.1)	(-0.1)	(0.0)	(0.1)	(0.0)	(-0.2)	(-1.4)	(1.0)	(-0.5)	(-5.3)
	IPW 3	13.2	7.5	5.8	12.8	7.6	6.2	14.4	8.1	5.5	13.7	7.2	0.2
		(-0.2)	(-0.1)	(0.0)	(-0.1)	(0.0)	(0.1)	(0.0)	(-0.3)	(-1.6)	(1.0)	(-0.6)	(-5.6)
	DR	10.6	5.8	4.9	10.5	5.8	4.9	10.6	5.1	3.5	10.6	4.5	-2.0
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(-0.5)	(-1.2)	(-2.5)	(0.5)	(-1.2)	(-6.4)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 3:2	IPW 1	12.2	4.8	1.0	12.7	5.3	0.4	15.3	6.8	0.9	14.7	7.0	-1.7
		(-0.1)	(0.0)	(0.1)	(-0.8)	(-0.4)	(0.0)	(-0.7)	(-1.1)	(-1.7)	(0.2)	(0.8)	(-1.4)
	IPW 2	9.1	2.6	0.8	8.4	3.0	1.8	10.0	3.2	0.9	9.0	3.1	-1.5
		(-0.1)	(-0.1)	(0.0)	(-0.1)	(0.0)	(0.1)	(0.0)	(-0.1)	(-0.8)	(0.5)	(-0.3)	(-3.4)
	IPW 3	9.5	3.0	0.0	9.1	3.3	0.7	10.3	3.4	-0.3	9.6	3.1	-2.9
		(-0.1)	(0.0)	(0.0)	(-0.1)	(0.0)	(0.0)	(0.0)	(-0.2)	(-1.0)	(0.5)	(-0.3)	(-3.5)
	DR	8.7	2.3	-0.2	8.7	2.3	-0.1	8.7	2.0	-1.2	8.7	1.6	-4.2
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(-0.3)	(-0.7)	(-1.6)	(0.2)	(-0.7)	(-4.0)
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 2:3	IPW 1	16.4	10.1	7.1	20.6	12.5	8.0	22.9	14.6	6.9	25.4	15.8	10.4
		(0.1)	(0.1)	(0.0)	(-0.3)	(-0.1)	(0.0)	(-0.2)	(-0.3)	(-3.3)	(-0.7)	(-0.3)	(-0.2)
	IPW 2	9.7	6.1	6.7	9.4	6.0	6.6	8.7	4.9	3.5	9.9	6.7	6.4
		(-0.4)	(0.0)	(-0.1)	(0.0)	(-0.1)	(0.0)	(-2.2)	(-1.8)	(-4.7)	(0.6)	(0.4)	(-0.1)
	IPW 3	10.5	6.3	5.9	10.3	6.3	6.0	9.9	5.6	2.3	10.7	6.9	5.7
		(-0.3)	(0.0)	(-0.1)	(0.0)	(-0.1)	(0.0)	(-1.8)	(-1.5)	(-4.5)	(0.6)	(0.4)	(-0.2)
	DR	8.1	5.0	4.9	7.9	4.8	5.2	8.5	5.5	2.4	8.2	5.2	5.1
		(0.0)	(0.0)	(0.0)	(0.0)	(-0.1)	(0.0)	(-0.3)	(-0.5)	(-3.5)	(0.6)	(0.5)	(-0.1)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 2:3	IPW 1	13.4	6.7	3.1	16.4	8.3	3.9	17.8	9.7	3.4	20.3	11.0	5.8
		(0.0)	(0.0)	(0.0)	(-0.2)	(-0.1)	(0.0)	(-0.1)	(-0.2)	(-2.1)	(-0.5)	(-0.2)	(-0.1)
	IPW 2	8.5	4.4	3.4	8.5	4.2	3.5	8.0	3.8	2.0	8.8	4.6	3.5
		(-0.2)	(0.0)	(-0.1)	(0.0)	(-0.1)	(0.0)	(-1.2)	(-1.0)	(-2.8)	(0.3)	(0.3)	(0.0)
	IPW 3	9.4	4.6	3.2	9.4	4.5	3.5	9.2	4.2	1.5	9.7	4.9	3.4
		(-0.1)	(0.0)	(0.0)	(0.0)	(-0.1)	(0.0)	(-0.9)	(-0.8)	(-2.7)	(0.3)	(0.3)	(0.0)
	DR	9.9	4.3	3.0	9.8	4.1	3.4	10.2	4.5	1.9	10.1	4.4	3.4
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(-0.1)	(-0.3)	(-2.0)	(0.4)	(0.3)	(0.0)

Note: Percentage change which is due to the bias $\left(\frac{\text{bias}^2(\text{ATE}(\hat{p}))-\text{bias}^2(\text{ATE}(\hat{p}^s))}{\text{bias}^2(\text{ATE}(\hat{p}))+\text{Var}(\text{ATE}(\hat{p}))}\right)$ is given in brackets.

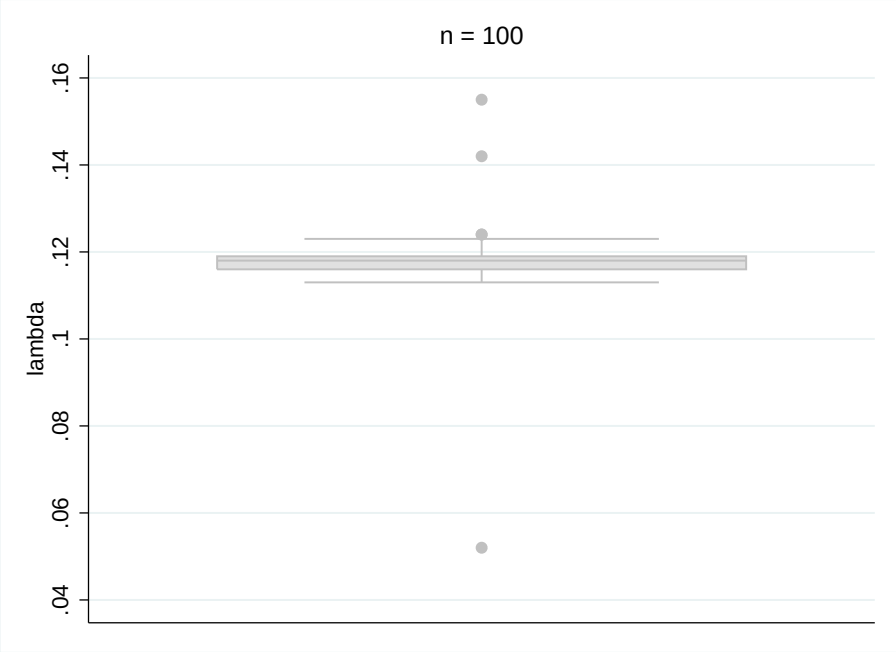
Table A12: Percentage improvement in MSE for the ATE if the shrunken propensity scores are used with trimming rule 2 instead of the conventional propensity scores with trimming rule 2, data-driven λ

		linear			nonlinear			linear			nonlinear		
	N	100	200	500	100	200	500	100	200	500	100	200	500
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 1:1	IPW 1	11.5	7.2	2.2	13.0	8.5	3.6	12.2	7.6	2.4	17.9	13.3	8.4
		(0.0)	(-0.1)	(0.0)	(-0.4)	(-0.6)	(-0.1)	(0.0)	(-0.1)	(0.0)	(1.8)	(2.1)	(3.2)
	IPW 2	4.9	3.3	1.0	5.0	3.2	1.4	4.9	3.2	1.1	5.9	4.3	2.9
		(-0.4)	(0.0)	(0.0)	(0.0)	(-0.1)	(-0.1)	(-0.9)	(-0.1)	(0.0)	(1.1)	(1.0)	(1.5)
	IPW 3	2.6	1.8	0.6	2.7	1.8	0.8	2.5	1.6	0.5	3.4	2.6	2.0
		(-0.3)	(0.0)	(0.0)	(0.0)	(-0.1)	(0.0)	(-0.6)	(-0.1)	(0.0)	(0.9)	(0.8)	(1.1)
	DR	0.5	0.8	0.4	0.3	0.6	0.3	0.6	0.7	0.4	0.5	0.8	0.9
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.4)	(0.4)	(0.6)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 1:1	IPW 1	8.2	5.3	2.6	9.4	6.2	3.2	8.7	5.5	2.7	13.1	9.7	6.4
		(0.0)	(0.0)	(0.0)	(-0.4)	(-0.3)	(0.0)	(0.0)	(-0.1)	(0.0)	(1.1)	(1.4)	(2.2)
	IPW 2	2.6	2.1	1.4	2.7	2.2	1.3	2.5	2.0	1.5	3.3	2.8	2.0
		(-0.2)	(0.0)	(0.0)	(0.0)	(-0.1)	(0.0)	(-0.4)	(-0.1)	(0.0)	(0.6)	(0.6)	(0.8)
	IPW 3	2.0	1.6	1.1	2.1	1.7	1.0	1.9	1.5	1.1	2.5	2.2	1.5
		(-0.1)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(-0.3)	(0.0)	(0.0)	(0.5)	(0.4)	(0.6)
	DR	1.8	1.7	1.1	1.7	1.5	1.0	1.8	1.6	1.1	1.8	1.6	1.2
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.2)	(0.2)	(0.3)
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 3:2	IPW 1	12.6	6.0	2.7	12.3	5.7	2.3	12.3	5.4	1.9	16.8	10.6	7.0
		(-0.1)	(-0.1)	(0.0)	(-1.3)	(-0.7)	(0.0)	(-0.4)	(-0.6)	(-0.5)	(1.3)	(3.3)	(4.7)
	IPW 2	6.5	3.1	2.3	6.2	2.9	1.9	7.8	4.8	4.2	7.4	4.3	3.3
		(-0.3)	(-0.2)	(-0.2)	(-0.2)	(-0.1)	(0.0)	(0.8)	(1.3)	(1.3)	(1.4)	(1.6)	(1.8)
	IPW 3	4.2	1.6	1.4	4.1	1.6	1.2	5.1	2.6	2.6	5.0	2.7	2.4
		(-0.2)	(-0.1)	(-0.1)	(-0.1)	(0.0)	(0.0)	(0.6)	(0.9)	(0.8)	(1.2)	(1.3)	(1.4)
	DR	1.2	0.1	0.6	1.0	0.1	0.6	1.4	0.1	0.7	1.3	0.5	1.2
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.2)	(0.1)	(0.1)	(0.5)	(0.5)	(0.6)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 3:2	IPW 1	8.8	4.4	2.2	8.5	4.3	1.8	8.7	4.1	1.8	11.2	7.2	4.6
		(0.0)	(0.0)	(0.0)	(-0.8)	(-0.4)	(0.0)	(-0.2)	(-0.3)	(-0.3)	(0.7)	(1.9)	(3.0)
	IPW 2	3.6	1.9	1.5	3.3	1.9	0.9	4.5	3.0	2.7	3.9	2.6	1.7
		(-0.1)	(-0.1)	(-0.1)	(-0.1)	(0.0)	(0.0)	(0.5)	(0.8)	(0.8)	(0.8)	(0.8)	(1.0)
	IPW 3	2.9	1.4	1.0	2.7	1.4	0.6	3.4	2.0	1.8	3.2	1.9	1.2
		(-0.1)	(-0.1)	(-0.1)	(-0.1)	(0.0)	(0.0)	(0.4)	(0.5)	(0.5)	(0.6)	(0.7)	(0.8)
	DR	2.4	1.2	0.5	2.3	1.2	0.6	2.5	1.2	0.7	2.4	1.4	0.7
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.1)	(0.1)	(0.1)	(0.2)	(0.2)	(0.3)
		homogenous, homoscedastic						heterogenous, homoscedastic					
Ratio 2:3	IPW 1	12.9	7.0	4.1	14.9	8.3	5.6	14.3	7.9	5.1	18.8	11.4	8.9
		(0.0)	(-0.1)	(0.1)	(-0.5)	(-0.6)	(0.2)	(0.7)	(0.4)	(0.9)	(-0.1)	(-0.1)	(1.6)
	IPW 2	5.4	3.3	2.1	5.4	2.8	1.7	4.1	2.2	0.9	5.6	3.1	2.2
		(-0.5)	(0.1)	(-0.1)	(0.0)	(-0.1)	(0.0)	(-2.1)	(-1.3)	(-1.5)	(0.3)	(0.2)	(0.5)
	IPW 3	3.3	2.0	1.1	3.3	1.8	0.9	2.5	1.3	0.4	3.5	2.0	1.3
		(-0.3)	(0.0)	(-0.1)	(0.0)	(-0.1)	(0.0)	(-1.3)	(-0.8)	(-0.9)	(0.3)	(0.1)	(0.4)
	DR	0.6	0.5	0.4	0.4	0.4	0.3	0.9	0.8	0.6	0.3	0.4	0.5
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.3)	(0.2)	(0.4)	(0.1)	(0.1)	(0.2)
		homogenous, heteroscedastic						heterogenous, heteroscedastic					
Ratio 2:3	IPW 1	9.0	6.0	2.7	10.5	6.9	3.7	10.0	6.7	3.3	13.6	9.1	6.1
		(0.0)	(-0.1)	(0.0)	(-0.4)	(-0.4)	(0.1)	(0.4)	(0.2)	(0.5)	(-0.2)	(-0.1)	(1.1)
	IPW 2	2.5	2.5	1.1	2.5	2.1	0.9	1.9	1.9	0.4	2.7	2.3	1.1
		(-0.2)	(0.0)	(-0.1)	(0.0)	(-0.1)	(0.0)	(-1.2)	(-0.7)	(-1.0)	(0.2)	(0.1)	(0.2)
	IPW 3	2.1	2.1	0.6	2.1	1.8	0.5	1.6	1.7	0.1	2.2	1.9	0.7
		(-0.1)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(-0.7)	(-0.4)	(-0.6)	(0.1)	(0.1)	(0.2)
	DR	1.8	1.7	0.5	1.7	1.6	0.5	2.0	1.9	0.6	1.7	1.6	0.6
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.2)	(0.1)	(0.1)	(0.1)	(0.0)	(0.1)

Note: Percentage change which is due to the bias $\left(\frac{\text{bias}^2(\text{ATE}(\hat{p}))-\text{bias}^2(\text{ATE}(\hat{p}^s))}{\text{bias}^2(\text{ATE}(\hat{p}))+\text{Var}(\text{ATE}(\hat{p}))}\right)$ is given in brackets.

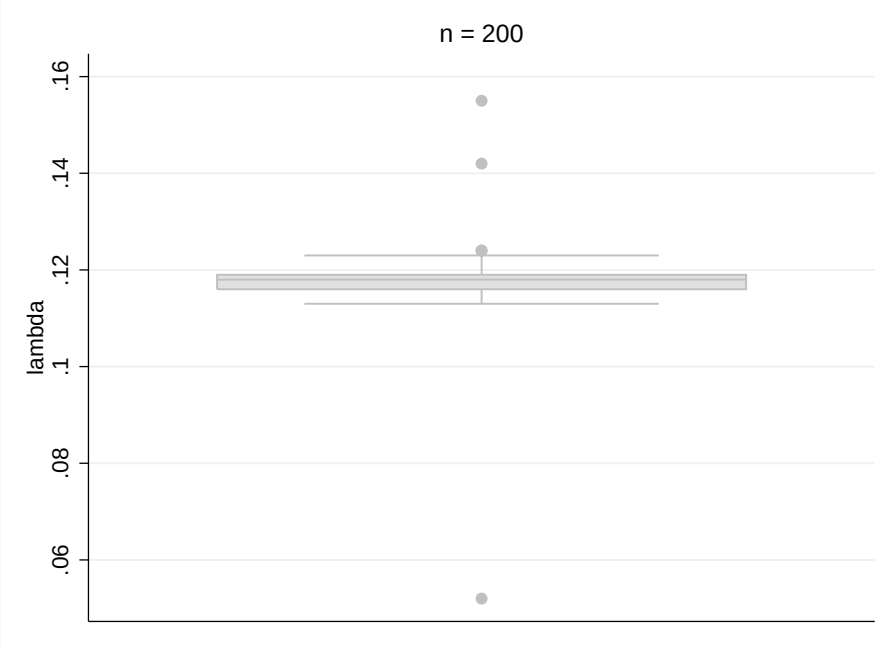
B Figures

Figure B1: Individual MSE minimizing λs for $n = 100$, $\delta = 1$, $\psi = 2$, $\nu = -0.3$, $\kappa = 0.8$ and $m_2(q)$.



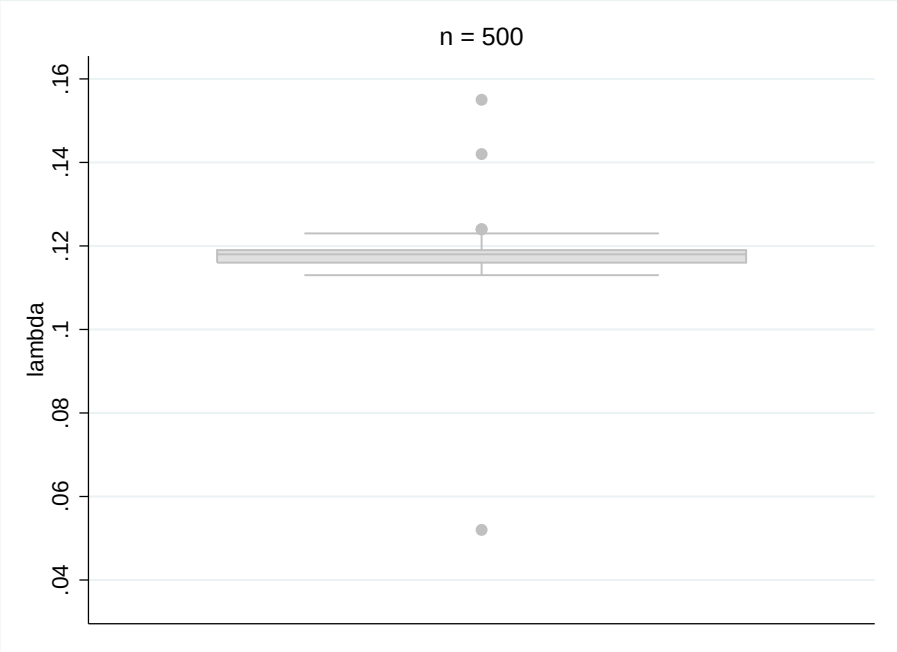
Note: For each individual the average is taken over 10000 Monte Carlo samples.

Figure B2: Individual MSE minimizing λs for $n = 200$, $\delta = 1$, $\psi = 2$, $\nu = -0.3$, $\kappa = 0.8$ and $m_2(q)$.



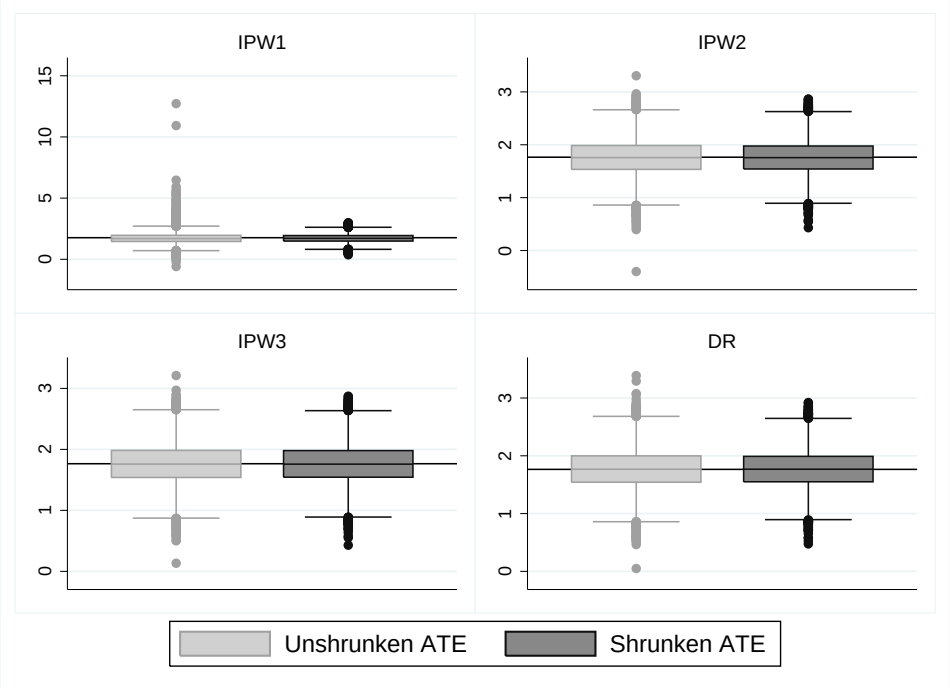
Note: For each individual the average is taken over 5000 Monte Carlo samples.

Figure B3: Individual MSE minimizing λs for $n = 500$, $\delta = 1$, $\psi = 2$, $\nu = -0.3$, $\kappa = 0.8$ and $m_2(q)$.



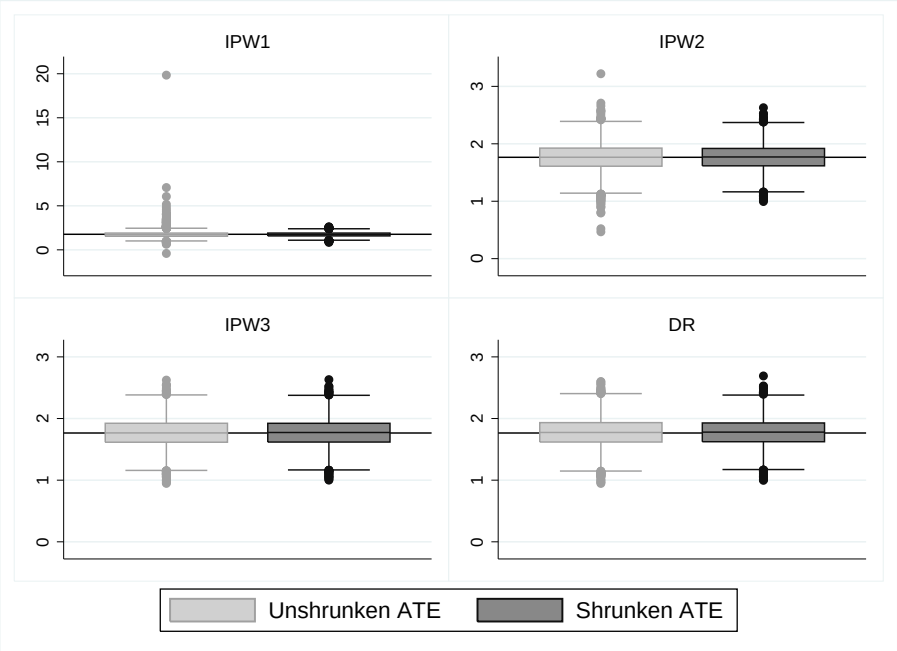
Note: For each individual the average is taken over 2000 Monte Carlo samples.

Figure B4: ATE's for $n = 100$, $\delta = 1$, $\psi = 2$, $\nu = -0.3$, $\kappa = 0.8$ and $m_2(q)$.



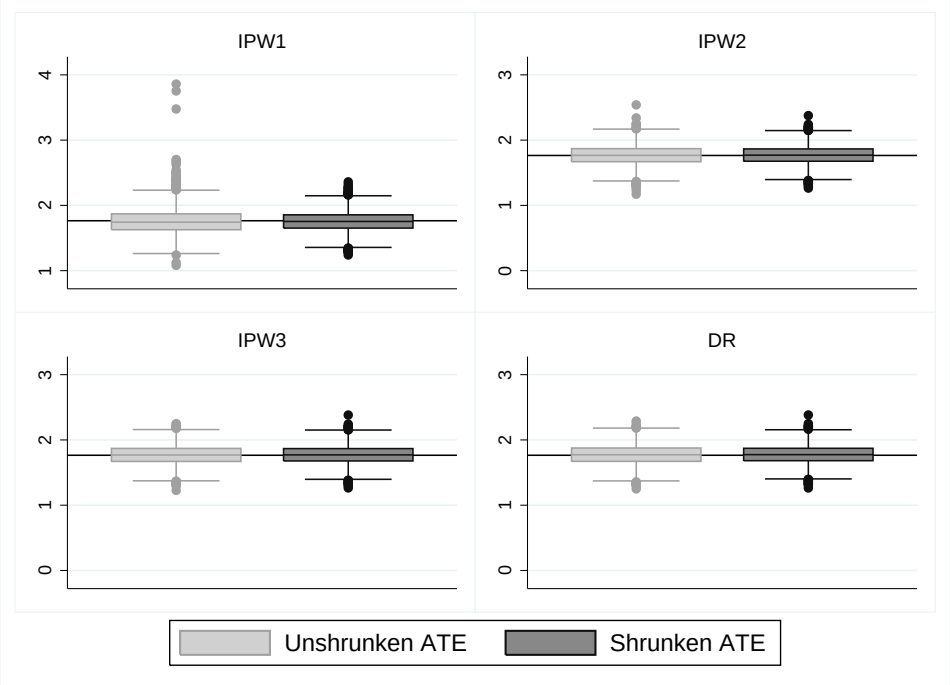
Note: For shrunken ATE's fixed valued λ's are used.

Figure B5: ATE's for $n = 200$, $\delta = 1$, $\psi = 2$, $\nu = -0.3$, $\kappa = 0.8$ and $m_2(q)$.



Note: For shrunken ATE's fixed valued λ's are used.

Figure B6: ATE's for $n = 500$, $\delta = 1$, $\psi = 2$, $\nu = -0.3$, $\kappa = 0.8$ and $m_2(q)$.



Note: For shrunken ATE's fixed valued λ's are used.