Survey calibration for causal inference: a simple method to balance covariate distributions

Beręsewicz Maciej*

First version: 2023–10–14

Abstract

This paper proposes a simple method for balancing distributions of covariates for causal inference based on observational studies. The method makes it possible to balance an arbitrary number of quantiles (e.g., medians, quartiles, or deciles) together with means if necessary. The proposed approach is based on the theory of calibration estimators (Deville and Särndal 1992), in particular, calibration estimators for quantiles, proposed by Harms and Duchesne (2006). By modifying the entropy balancing method and the covariate balancing propensity score method, it is possible to balance the distributions of the treatment and control groups. The method does not require numerical integration, kernel density estimation or assumptions about the distributions; valid estimates can be obtained by drawing on existing asymptotic theory. Results of a simulation study indicate that the method efficiently estimates average treatment effects on the treated (ATT), the average treatment effect (ATE), the quantile treatment effect on the treated (QTT) and the quantile treatment effect (QTE), especially in the presence of non-linearity and mis-specification of the models. The proposed methods are implemented in an open source R package, `jointCalib`.

Keywords: calibration estimators, quantile estimation, Heaviside function, entropy balancing, covariate balancing propensity score

Acknowledgements

This work was financed by the National Science Centre in Poland, OPUS 22 grant no. 2020/39/B/HS4/00941. Codes to reproduce the results are freely available from the github repository: https://github.com/ncn-foreigners/paper-note-quantiles-obs-studies. An R package that implements the proposed methods is available at https://github.com/ncn-foreigners/jointCalib. I would like to thank Marcin Szymkowiak, Piotr Chlebicki, Łukasz Chrostowski and Joanna Tyrowicz for their valuable comments on the first draft. I am grateful to Grzegorz Grygiel for his comments and proofreading.

*Corresponding author, maciej.beresewicz@ue.poznan.pl, (1) Poznań University of Economics and Business, Department of Statistics, Al. Niepodległości 10, 61-875 Poznań; (2) Statistical Office in Poznań, Poland.

1 Introduction

Recent literature on causal inference for observational studies includes several approaches to balancing whole distributions of covariates $\mathbf{X}$ rather than moments (e.g. means, variances). In particular, Hazlett (2020) proposes kernel entropy balancing (KEB), which consists in making the multivariate density of covariates approximately equal for the treated and control groups when the same choice of kernel is used to estimate these densities. Zhao (2019) proposes a covariate balancing rule that modifies the balancing propensity score model by reproducing the kernel Hilbert space (RKHS) and is implemented within the framework of the tailored loss function approach. Finally, Sant'Anna et al. (2022) proposes an integrated propensity score model, which aims to minimise imbalances in the joint distribution of covariates.

In this paper we propose a simple method which consists in balancing control and treatment distributions at specific quantiles along with moments, if needed. Drawing on the theory of calibration estimators in survey sampling (Deville and Särndal, 1992) and, in particular, on quantile calibration estimators (Harms and Duchesne, 2006), we develop methods for observational studies. The proposed method involves adding new variables to be balanced based on pre-treatment covariates $\mathbf{X}$ using a modified Heaviside function (or its approximation), thus reducing bias and the root mean square error. The procedure does not require a prior knowledge of distributions, does not involve integration or tuning parameters such as bandwidths, and is computationally simple as it applies local linear approximation.

The paper is structured as follows. Section 2 presents the theory underlying calibration estimators for means/totals and quantiles. Section 3 describes the proposed approach for entropy balancing (Hainmueller, 2012) and the covariate balancing propensity score (Imai and Ratkovic, 2014), which we refer to as *distributional entropy balancing* (DEB) and *distributional propensity score* (DPS) methods respectively. Section 4 presents results of two simulation studies aimed at validating the two methods: DEB is used to estimate the average treatment effect on the treated (ATT) and the quantile treatment effect on the treated (QTT), while DPS is used to estimate the average treatment effect (ATE) and the quantile treatment effect (QTE). The paper ends with a conclusion and additional results are presented in the appendix. Open source software in R (R

Core Team, 2023) is available.

2 Theoretical basis from survey sampling

2.1 Calibration estimator for a total

Let $\mathbf{X}$ be a random auxiliary (benchmark, pre-treatment) variable and Y be the target random variable of interest. In most applications, the goal is to estimate a finite population total $\tau_y = \sum_{k \in U} y_k$ or the mean $\bar{\tau}_y = \tau_y / N$ of the variable of interest y, where U is the population of size N. The Horvitz-Thompson estimator is a well-known estimator of a finite population total, which is expressed as $\hat{\tau}_{y\pi} = \sum_{k=1}^{n} d_k y_k = \sum_{k \in s} d_k y_k$, where s denotes a probability sample of size n, $d_k = 1/\pi_k$ is a design weight, and π_k is the first-order inclusion probability of the i-th element of the population U. This estimator is unbiased for τ_Y i.e. $E\left(\hat{\tau}_{y\pi}\right) = \tau_Y$.

Let $\mathbf{x}_k^{\circ}$ be a J_1-dimensional vector of auxiliary variables (benchmark variables) for which $\tau_{\mathbf{x}} = \sum_{k \in U} \mathbf{x}_k^{\circ} = \left(\sum_{k \in U} x_{k1}, \ldots, \sum_{k \in U} x_{kJ_1}\right)^T$ is assumed to be known. In most cases, in practice the d_k weights do not reproduce known population totals for benchmark variables $\mathbf{x}_k^{\circ}$. It means that the resulting estimate $\hat{\tau}_{\mathbf{x}\pi} = \sum_{k \in s} d_k \mathbf{x}_k^{\circ}$ is not equal to $\tau_{\mathbf{x}}$. The main idea of calibration is to look for new calibration weights w_k that are as close as possible to original design weights d_k and reproduce known population totals $\tau_{\mathbf{x}}$ exactly. In other words, in order to find new calibration weights w_k we have to minimise a distance function $D\left(\mathbf{d}, \mathbf{v}\right) = \sum_{k \in s} d_k \, G\left(\frac{v_k}{d_k}\right) \rightarrow \min$ to fulfil calibration equations $\sum_{k \in s} v_k \mathbf{x}_k^{\circ} = \sum_{k \in U} \mathbf{x}_k^{\circ}$, where $\mathbf{d} = \left(d_1, \ldots, d_n\right)^T$, $\mathbf{v} = \left(v_1, \ldots, v_n\right)^T$ and $G\left(\cdot\right)$ is a function which must satisfy some regularity conditions: $G\left(\cdot\right)$ is strictly convex and twice continuously differentiable, $G\left(\cdot\right) \geq 0$, $G\left(1\right) = 0$, $G'\left(1\right) = 0$ and $G''\left(1\right) = 1$. Examples of $G\left(\cdot\right)$ functions are given by Deville and Särndal (1992). For instance, if $G\left(x\right) = \frac{(x-1)^2}{2}$, then using the method of Lagrange multipliers the final calibration weights w_k can be expressed as $w_k = d_k + d_k \left(\tau_{\mathbf{x}} - \hat{\tau}_{\mathbf{x}\pi}\right)^T \left(\sum_{j \in s} d_j \mathbf{x}_j^{\circ} \mathbf{x}_j^{\circ T}\right)^{-1} \mathbf{x}_k^{\circ}$. It is worth adding that in order to avoid negative or large w_k weights in the process of minimising the $D\left(\cdot\right)$ function, one can consider some boundary constraints $L \leq \frac{w_k}{d_k} \leq U$, where $0 \leq L \leq 1 \leq U$, $k = 1, \ldots, n$. The final calibration estimator of a population total τ_y can be expressed as $\hat{\tau}_{y\mathbf{x}} = \sum_{k \in s} w_k y_k$, where w_k are calibration weights obtained after selecting a given $G(\cdot)$ function.

2.2 Calibration estimator for a quantile

Harms and Duchesne (2006) considered a way of estimating quantiles using the calibration approach, which is very similar to that proposed by Deville and Särndal (1992) for a finite population total τ_y. By analogy, in their approach it is not necessary to know values for all auxiliary variables for all units in the population. It is enough to know the corresponding quantiles for the benchmark variables. Let us briefly discuss the problem of finding calibration weights in this setup.

We want to estimate a quantile $Q_{y,\alpha}$ of order $\alpha \in (0,1)$ of the variable of interest y, which can be expressed as $Q_{y,\alpha} = \inf\{t \,|F_y(t) \geq \alpha\}$, where $F_y(t) = N^{-1}\sum_{k\in U} H(t-y_k)$ and the Heaviside function is given by

$$H(t-y_k) = \begin{cases} 1, & t \geq y_k, \\ 0, & t < y_k. \end{cases} \tag{1}$$

We assume that $\mathbf{Q}_{\mathbf{x},\alpha} = \left(Q_{x_1,\alpha}, \ldots, Q_{x_{J_2},\alpha}\right)^T$ is a vector of known population quantiles of order α for a vector of auxiliary variables $\mathbf{x}_k^*$, where $\alpha \in (0,1)$ and $\mathbf{x}_k^*$ is a J_2-dimensional vector of auxiliary variables. It is worth noting that, in general, the numbers J_1 and J_2 of the auxiliary variables are different. It may happen that for a specific auxiliary variable its population total and the corresponding quantile of order α will be known. However, in most cases, quantiles will be known for continuous auxiliary variables, unlike totals, which will generally be known for categorical variables. In order to find new calibration weights w_k that reproduce known population quantiles in a vector $Q_{\mathbf{x},\alpha}$, an interpolated distribution function estimator of $F_y(t)$ is defined as $\hat{F}_{y,cal}(t) = \frac{\sum_{k\in s} w_k H_{y,s}(t,y_k)}{\sum_{k\in s} w_k}$, where the Heaviside function in formula (1) is replaced by the modified function $H_{y,s}(t,y_k)$ given by

$$H_{y,s}(t,y_k) = \begin{cases} 1, & y_k \leqslant L_{y,s}(t), \\ \beta_{y,s}(t), & y_k = U_{y,s}(t), \\ 0, & y_k > U_{y,s}(t), \end{cases} \tag{2}$$

where $L_{y,s}(t) = \max\{\{y_k, k \in s \mid y_k \leqslant t\} \cup \{-\infty\}\}$, $U_{y,s}(t) = \min\{\{y_k, k \in s \mid y_k > t\} \cup \{\infty\}\}$

and $\beta_{y,s}(t) = \frac{t - L_{y,s}(t)}{U_{y,s}(t) - L_{y,s}(t)}$ for $k = 1, \ldots, n$, $t \in \mathbb{R}$. A calibration estimator of quantile $Q_{y,\alpha}$ of order α for variable y is defined as $\hat{Q}_{y,cal,\alpha} = \hat{F}^{-1}_{y,cal}(\alpha)$, where a vector $\mathbf{w} = (w_1, \ldots, w_n)^T$ is a solution of optimization problem $D(\mathbf{d}, \mathbf{v}) = \sum_{k \in s} d_k \, G\left(\frac{v_k}{d_k}\right) \to \min$ subject to the calibration constraints $\sum_{k \in s} v_k = N$ and $\hat{\mathbf{Q}}_{\mathbf{x},cal,\alpha} = \left(\hat{Q}_{x_1,cal,\alpha}, \ldots, \hat{Q}_{x_{J_2},cal,\alpha}\right)^T = \mathbf{Q}_{\mathbf{x},\alpha}$ or equivalently $\hat{F}_{x_j,cal}\left(Q_{x_j,\alpha}\right) = \alpha$, where $j = 1, \ldots, J_2$.

As in the previous case, if $G(x) = \frac{(x-1)^2}{2}$ then using the method of Lagrange multipliers the final calibration weights w_k can be expressed as $w_k = d_k + d_k\left(\mathbf{T_a} - \sum_{k \in s} d_k \mathbf{a}_k\right)^T \left(\sum_{j \in s} d_j \mathbf{a}_j \mathbf{a}_j^T\right)^{-1} \mathbf{a}_k$, where $\mathbf{T_a} = (N, \alpha, \ldots, \alpha)^T$ and the elements of $\mathbf{a}_k = (1, a_{k1}, \ldots, a_{kJ_2})^T$ are given by

$$a_{kj} = \begin{cases} N^{-1}, & x_{kj} \leq L_{x_j,s}\left(Q_{x_j,\alpha}\right), \\ N^{-1}\beta_{x_j,s}\left(Q_{x_j,\alpha}\right), & x_{kj} = U_{x_j,s}\left(Q_{x_j,\alpha}\right), \\ 0, & x_{kj} > U_{x_j,s}\left(Q_{x_j,\alpha}\right), \end{cases} \tag{3}$$

with $j = 1, \ldots, J_2$.

Alternatively, one can consider the logistic function instead of (3)

$$a_{kj} = \frac{1}{1 + \exp\left(-2l\left(x_{kj} - Q_{x_j,\alpha}\right)\right)} \frac{1}{N},$$

where x_{kj} is the kth row of the auxiliary variable x_j, N is the population size, $Q_{x_j,\alpha}$ is the known population α-th quantile, and l is a constant set to a large value.

In the next sections we describe how this method can be applied to make causal inferences in observational studies. We focus on four causal parameters: ATT, QTT, ATE and QTE.

3 Proposed approaches

3.1 Setup

Let us assume that $\mathcal{D} = \{0, 1\}$ is a treatment indicator variable, sample s_1 denotes the treatment group, s_0 denotes the control group and Y denotes the variable of interest, where $Y(1)$ and $Y(0)$ are the potential outcomes for the treatment group and the control group, respectively. The realised outcome is $Y = \mathcal{D}Y(1) + (1-\mathcal{D})Y(0)$ and $\mathbf{X}$ is an observed vector of pre-treatment covariates. Let $p(\mathbf{x}) = \mathbb{P}(\mathcal{D} = 1|\mathbf{X} = \mathbf{x})$ be the propensity score and for $\delta = \{0, 1\}$ the distribution and the quantile of the potential outcome $Y(\delta)$ is given by $F_{Y(\delta)}(y) = \mathbb{P}(Y(\delta) \leq y)$ and $q_{Y(\delta)}(\alpha) = \inf \left\{t \left| F_{Y(\delta)}(t) \geq \alpha \right.\right\}$.

Let us further assume that the researcher is interested in estimating the average treatment effect on the treated ATT $= \mathbb{E}[Y(1) \mid \mathcal{D} = 1] - \mathbb{E}[Y(0) \mid \mathcal{D} = 1]$, the quantile treatment effect on the treated QTT$(\alpha) = q_{Y(1)|D=1}(\alpha) - q_{Y(0)|D=1}(\alpha)$, an overall average treatment effect ATE $= \mathbb{E}(Y(1) - Y(0))$ or the quantile treatment effect QTE $= q_{Y(1)} - q_{Y(0)}$.

In this paper we follow a commonly used identification strategy in policy evaluation (Rosenbaum and Rubin, 1983; Firpo, 2007, cf.): 1) given $\mathbf{X}$, $(Y(1), Y(0))$ are jointly independent from D (conditional ignorability), 2) $p(\mathbf{x})$ is uniformly bounded away from zero and one, 3) uniqueness of quantiles. The identification strategy is the same as that used in the literature mentioned above, since we adopt the same assumptions.

For the ATT and QTT the counterfactual mean and quantile can be estimated as

$$\mathbb{E}[\widehat{Y(0) \mid \mathcal{D}} = 1] = \frac{\sum_{k \in s_0} w_k y_k}{\sum_{k \in s_0} w_k},$$
$$\mathbb{E}[\widehat{q_{Y(0)|D=1}}(\alpha)] = \frac{\sum_{k \in s_0} w_k H(t - y_k)}{\sum_{k \in s_0} w_k},$$

where w_k is a weight chosen for each control unit.

For ATE and QTE one can use the approach suggested by Rosenbaum (1987), i.e.

$$ATE = \mathbb{E}\left[\left(\frac{\mathcal{D}}{p(\mathbf{X})} - \frac{1-\mathcal{D}}{1-p(\mathbf{X})}\right) Y\right], \tag{4}$$

and for $\delta \in \{0, 1\}$, $F_{Y(\delta)}(y)$ is identified by

$$F_{Y(\delta)}(y) = \mathbb{E}\left[\frac{1\{\mathcal{D} = \delta\}}{\delta p(\mathbf{X}) + (1-\delta)(1-p(\mathbf{X}))} 1\{Y \leqslant y\}\right], \tag{5}$$

where $1\{.\}$ is the indicator function, which means that QTE(τ) can also be written as functionals of the observed data (cf. Sant'Anna et al., 2022).

3.2 Distributional entropy balancing

Hainmueller (2012) proposed entropy balancing to reweight the control group to the known characteristics of the treatment group to estimate ATT and QTT. This method can be summarised as follows

$$
\begin{aligned}
\max_{w} H(v) = -\sum_{k\in s_0} v_k \log\left(v_k/d_k\right) & \\
\text{s.t.} \sum_{k\in s_0} v_k G_{kj} = m_k \text{ for } j \in 1,\dots,J & \\
\sum_{k\in s_0} v_k = 1 \text{ and } v \geq 0 \text{ for all } k \in s_0 &
\end{aligned}
\tag{6}
$$

where v_k is defined as previously, $d_k > 0$ is the base weight for unit k, $H(.)$ is the Kullback-Leibler divergence between the distributions of the solution weights and the base weights, $G_{kj} \in \mathbb{R}^J$ contains J pretreatment covariates, and m_j is the mean of the j-th covariate in the treatment group. As in the case of calibration, w_k are solutions to (6).

The method in (6) can be simply extended to achieve not only the mean balance but also the distributional balance. Instead of using known or estimated population totals $\mathbf{Q}_{\mathbf{x},\alpha}$, we can use treatment group quantiles denoted by $\mathbf{q}_{\mathbf{x},\alpha} = \left(q_{x_1,\alpha},\dots,q_{x_{J_2},\alpha}\right)^T$, and the definition of the vector $\mathbf{a}_k = (1, a_{k1},\dots,a_{kJ_2})^T$ changes to

$$
a_{kj} = \begin{cases} n_1^{-1}, & x_{kj} \leq L_{x_j,s}\left(q_{x_j,\alpha}\right), \\ n_1^{-1}\beta_{x_j,s}\left(q_{x_j,\alpha}\right), & x_{kj} = U_{x_j,s}\left(q_{x_j,\alpha}\right), \\ 0, & x_{kj} > U_{x_j,s}\left(q_{x_j,\alpha}\right), \end{cases}
\tag{7}
$$

with $j = 1,\dots,J_2$ where n_1 is the size of the treatment group. Alternatively, one can use a modified (2.2) given by

$$
a_{kj} = \frac{1}{1+\exp\left(-2l\left(x_{kj} - q_{x_j,\alpha}\right)\right)}\frac{1}{n_1}.
$$

Our proposal, which leads to distributional entropy balancing (hereinafter DEB), consists in extending the original idea by adding additional constraint(s) on the weights on $\mathbf{a}_k$, as presented below.

$$
\begin{aligned}
\max_{w} H(w) = &-\sum_{k\in s_0} v_k \log\left(v_k/q_k\right), \\
\text{s.t.} \; &\sum_{k\in s_0} v_k G_{kj} = m_k \text{ for } j \in 1,\ldots,J_1, \\
&\sum_{k\in s_0} v_k a_{kj} = \frac{\alpha_j}{n_1} \text{ for } j \in 1,\ldots,J_2, \\
&\sum_{k\in s_0} v_k = 1 \text{ and } v \geq 0 \text{ for all } k \in s_0.
\end{aligned}
$$

This approach is similar to that proposed by Hazlett (2020), who extended (6) by replacing the first condition by $\sum_{k\in s_0} v_i \phi\left(x_i\right) = \frac{1}{n_1}\sum_{k\in s_1} \phi\left(x_i\right)$, where $\phi\left(x_i\right)$ are the basis functions for the kernel function (in particular the Gaussian kernel). Like Hazlett (2020), we assume the conditional expectation of $Y(0)$ is linear in the $\mathbf{a}_k$ as we locally approximate this relationship with a linear model on ranges specified by quantiles. This approach is similar to segmented regression.

Remark 1: Instead of modelling the whole distribution, one can start by adjusting the medians or quartiles. The number of quantiles for $\mathbf{x}_j^*$ can vary. In the simulation study we show that even a small number of quantiles significantly improves the estimates, especially in the presence of non-linear relationships.

Remark 2: This approach assumes that the distributions of $\mathbf{x}_k^*$ between the control and treatment groups have the same support, i.e. it is possible to generate a vector $\sum a_{kj} > 0$.

Our approach can be further applied to hierarchically regularised entropy balancing, as proposed by Xu and Yang (2023), or the empirical likelihood method, as recently discussed by Zhang et al. (2022).

3.3 Distributional propensity score method

Imai and Ratkovic (2014) proposed the covariate balancing propensity score (CBPS) to estimate the (4), where unknown parameters of the propensity score model γ are estimated using the generalized method of moments as

$$\mathbb{E}\left[\left(\frac{\mathcal{D}}{p\left(\mathbf{X};\gamma\right)}-\frac{1-\mathcal{D}}{1-p\left(\mathbf{X};\gamma\right)}\right)f(\mathbf{X})\right]=\mathbf{0}, \tag{8}$$

where $p(\dot{)}$ is the propensity score. This balances means of the the $\mathbf{X}$ variables, which may not be sufficient if the variables are highly skewed or we are interested in estimating DTE or QTE.

We propose a simple approach based on the specification of moments and α-quantiles to be balanced. Instead of using the matrix $\mathbf{X}$, we propose using the matrix $\mathcal{X}$, which is constructed as follows

$$\mathcal{X}=\begin{bmatrix}\mathbf{1}^1 & \mathbf{X}^1 & \mathbf{A}^1\\ \mathbf{1}^0 & \mathbf{X}^0 & \mathbf{A}^0\end{bmatrix},$$

where $\mathbf{X}^0,\mathbf{X}^1$ are matrices of size $n_0\times J_1$ and $n_1\times J_1$ with J_1 covariates to be balanced at the means, and $\mathbf{A}^1,\mathbf{A}^0$ are matrices based on J_2 covariates with elements defined as follows

$$a^1_{kj}=\begin{cases} n_1^{-1}, & x^1_{kj}\leq L_{x_j,1}\left(q^1_{x_j,\alpha}\right),\\ n_1^{-1}\beta_{x_j,1}\left(q^1_{x_j,\alpha}\right), & x^1_{kj}=U_{x_j,1}\left(q^1_{x_j,\alpha}\right),\\ 0, & x^1_{kj}>U_{x_j,1}\left(q^1_{x_j,\alpha}\right),\end{cases} \tag{9}$$

and

$$a^0_{kj}=\begin{cases} n_1^{-1}, & x^0_{kj}\leq L_{x_j,0}\left(q^1_{x_j,\alpha}\right),\\ n_1^{-1}\beta_{x_j,0}\left(q^1_{x_j,\alpha}\right), & x^0_{kj}=U_{x_j,0}\left(q^1_{x_j,\alpha}\right),\\ 0, & x^0_{kj}>U_{x_j,0}\left(q^1_{x_j,\alpha}\right),\end{cases} \tag{10}$$

where n_1 is the size of the treatment group, or, alternatively, the logistic function (2.2) can be used.

Note that the elements of $\mathbf{A}^1$ sum up to the selected α orders of the quantiles as $n\to\infty$ (though for small sample sizes, they may not sum up to the specified alpha). As a result, the propensity score weights balance the α orders of the treatment and control groups and, as shown in Harms and Duchesne (2006), the α quantiles of the selected variables.

In our approach we simply plug a matrix $\mathcal{X}$ into (8) and follow procedures to estimate γ

parameters with a just-identified or over-identified set of equations (e.g. generalised method of moments, empirical likelihood). Since we do not change the estimation method itself, any method proposed in the literature can be applied. Furthermore, this matrix can be plugged into a high-dimensional setting with variable selection as in Ning et al. (2020) or the improved CBPS proposed by Fan et al. (2016).

4 Empirical results

4.1 Simulation for the DEB method

To show the effectiveness of our approach we follow the simulation procedure described by Hainmueller (2012). We generate 6 variables: three (X_1, X_2 and X_3) from a multivariate normal distribution MVN($\mathbf{0}, \mathbf{\Sigma}$), where

$$\mathbf{\Sigma} = \begin{bmatrix} 2 & 1 & -1 \\ 1 & 1 & -0.5 \\ -1 & -0.5 & 1 \end{bmatrix},$$

$X_4 \sim \text{Uniform}[-3, 3]$, $X_5 \sim \chi^2(6)$ and $X_6 \sim \text{Bernoulli}(0.5)$. The treatment and control groups are formed using

$$\mathcal{D} = \mathbf{1}[X_1 + 2X_2 - 2X_3 - X_4 - 0.5X_5 + X_6 + \epsilon > 0].$$

We consider three designs: Design 1 (D1): $\epsilon \sim N(0, 30)$, Design 2 (D2): $\epsilon \sim N(0, 100)$ and Design 3 (D3): $\epsilon \sim \chi^2(5)$ scaled to mean 0.5 and variance 67.6; and three outcome designs:

$$\begin{aligned} Y_1 &= X_1 + X_2 + X_3 - X_4 + X_5 + X_6 + \eta, \\ Y_2 &= X_1 + X_2 + 0.2X_3X_4 - \sqrt{X_5} + \eta, \\ Y_3 &= (X_1 + X_2 + X_5)^2 + \eta, \end{aligned}$$

where $\eta \sim N(0, 1)$. In the simulation study we consider equal sample sizes $n_0 = n_1 = 1000$. As the definition of $\mathcal{D}$ can lead to unequal sample sizes, we use simple random sampling with replacement from the simulated treatment and control groups to meet the requirement of $n_0 = n_1 = 1000$.

In the simulation study, we use three methods: entropy balancing (EB), kernel entropy balancing (KEB), distributional entropy balancing (DEB) with balancing means and quartiles (DEB MQ) and DEB with balancing means and deciles (DEB MD) of X_1 to X_5.

Table 1 contains results for ATT and QTT(α) for Y_3 for all designs where $\alpha \in \{0.10, 0.25, 0.5, 0.75, 0.90\}$. Note that we use partially overlapping α for balancing and QTT. Results for Y_1 and Y_2 are presented in the Appendix A. In all studies we report Monte Carlo Bias $= \bar{\hat{\theta}} - \theta$, Variance $= \frac{1}{R-1}\sum_{r=1}^{R}\left(\hat{\theta}_r - \bar{\hat{\theta}}\right)^2$ and root mean square error RMSE $= \sqrt{\text{Bias}^2 + \text{Variance}}$, where $\bar{\hat{\theta}} = \frac{1}{R}\sum_{r=1}^{R}\hat{\theta}_r$, and θ is the known effect (i.e. ATT, QTT, ATE or QTE) and R is the number of simulations set to 500.

For all designs, as expected, KEB yields better results in terms of RMSE (mainly thanks to small variance) for Y3, while results for Y1 and Y2 vary. The proposed estimators are better than KEB for Y2 under all three designs, and for Y1 results obtained using DEB are comparable or slightly better than those for KEB.

Compared with EB, DEB performs better in terms of the RMSE, which is smaller for ATT and QTT(0.25) to QTT(0.90), with a small increase / and slightly higher / for QTT(0.10) in D1 and D3. The proposed approach improves the estimates of ATT by almost halving the variance of EB for D1 and D2 and significantly reducing the bias for the non-linear case (D3). For D3, an increase in variance is observed for DEB with mean and deciles. DEB MQ and MD are more efficient compared to EB as α increases. For the non-linear case, DEB MD yields an almost unbiased estimate of the cost of increasing the variance, since the bias for $\alpha = 0.90$ in D3 is around 0.83, while for EB it is over 3.6 and the variance is 20.7 and 12.7, respectively. For D2, both the bias and variance decrease compared to their corresponding values for EB, while for D1 only the variance decreases leading to a decrease in RMSE. The results suggest that the proposed approach offers more efficient ATT and QTT estimators, especially for non-linear cases in comparison to EB. It should be noted that the DEB approach significantly improves estimates of the upper part of the distribution, which can be beneficial in economic studies.

4.2 Simulation for the DPS method

In the next simulation, we follow Imai and Ratkovic (2014) and Sant'Anna et al. (2022). We generate four variables $\mathbf{X} \sim \text{MVN}(0, \mathbf{I})$ where $\mathbf{I}$ is an 4×4 identity matrix (for *correctly specified models*). Next, we generate $\mathbf{W} = (W_1, W_2, W_3, W_4)'$ with $W_1 = \exp(X_1/2)$, $W_2 = X_2/(1 + \exp(W_1))$, $W_3 = (X_1 X_2/25 + 0.6)^3$ and $W_4 = (X_2 + X_4 + 20)'$ (for *mis-specified models* where instead of $\mathbf{X}$ we observe $\mathbf{W}$). The true propensity score of the treatment status $\mathcal{D}$ is given by

$$p(\mathbf{X}) = \frac{\exp\left(-X_1 + 0.5X_2 - 0.25X_3 - 0.1X_4\right)}{1 + \exp\left(-X_1 + 0.5X_2 - 0.25X_3 - 0.1X_4\right)},$$

and the treatment status $\mathcal{D}$ is generated $\mathcal{D} = 1\{p(\mathbf{X}) > U\}$, where $U \sim \text{Uniform}(0,1)$. The potential outcomes $Y(1)$ and $Y(0)$ are given by $Y(1) = 210 + m(\mathbf{X}) + \varepsilon(1)$ and $Y(0) = 200 - m(\mathbf{X}) + \varepsilon(0)$, where $m(\mathbf{X})$ is defined as $m(\mathbf{X}) = 27.4X_1 + 13.7X_2 + 13.7X_3 + 13.7X_4$ and where $\varepsilon(1)$ and $\varepsilon(0)$ are independent $N(0,1)$ random variables. We focus on ATE and QTE(α) where α is defined as in DEP. The true effect equals 10 for ATE and all QTE. We compare the following approaches:

- IPS – identity, exponential and projection approach (denoted as IPS (ind), IPS (exp) and IPS (proj)),
- CBPS – just- and over-identified with balancing means (denoted as CBPS (j) and CBPS (o)),
- DPS – just- and over-identified with balancing means and quartiles (denoted as DPS (j MQ) and DPS (o MQ)),
- DPS – just- and over-identified with balancing means and deciles (denoted as DPS (j MD) and DPS (o MD)),
- DPS – just- and over-identified with balancing deciles only (denoted as DPS (j D) and DPS (o D)).

To compare balance of distributions we use the same metrics as Sant'Anna et al. (2022), i.e.

- the Cramér-von Mises related statistic (denoted as CVM)

$$\mathrm{CVM}(\gamma) = \sqrt{\frac{1}{n}\sum_{k=1}^{n} \mathrm{DistImb}\left(\mathbf{X}_k^*, \gamma\right)^2},$$

- the Kolmogorov-Smirnov related statistic (denoted as KS)

$$\mathrm{KS}(\gamma) = \sup_{k:1,\ldots,n} \left\| \mathrm{DistImb}\left(\mathbf{X}_k^*, \gamma\right)\right\|,$$

where $\mathrm{DistImb}(\mathbf{x}, \gamma) = \mathbb{E}_n\left[\left(w_1^{ps}\left(\mathcal{D}, \mathbf{X}^*; \gamma\right) - w_0^{ps}\left(\mathcal{D}, \mathbf{X}^*; \gamma\right)\right) 1\left\{\mathbf{X}^* \leqslant \mathbf{x}\right\}\right]$, where $\mathbf{X}^* = \mathbf{W}$ for a mis-specified model, and w_D is the propensity score weight obtained by applying all proposed methods.

Table 2 shows results for a misspecified model based on a Monte Carlo study with 500 replications. Results for correctly specified models are presented in Appendix A3. In all cases, DPS performs better than CBPS and IPS in terms of both bias and variance, resulting in a more efficient estimator. This pattern is observed particularly for the over-identified DPS with decile constraints, since the simulation study only includes continuous variables. The proposed approach leads to nearly unbiased estimates of QTE for all α, especially for the upper part of the distribution.

Table 3 shows a comparison of the mean and median of the CVM and KS statistics. In all cases, the proposed method produces more balanced distributions than IPS.

5 Summary

In this paper we have proposed a simple method for balancing distributions based on the theory of calibration estimators for quantiles. The proposed methods are flexible and allows the researcher to balance an arbitrary number of quantiles that may vary according to pre-treatment variables. In particular, if the researcher is interested in estimating a particular α quantile of treatment effects, they can focus on balancing only these α quantiles for continuous variables.

Furthermore, the proposed methods perform well for linear and especially for nonlinear and misspecified models. In the two simulation studies, we show that DEB and DPS reduce bias and

RMSE for both average and quantile treatment effects. The DEB method is comparable to kernel entropy balancing, but significantly faster and less complicated. DPS outperformed the recently proposed integrated propensity score method. The proposed methods are computationally simple and can be implemented in existing statistical software (e.g. Stata, Python). For the purpose of this study, we developed the `jointCalib` package, which allows the user to run DEB with different distance functions, such as raking, logit, hyperbolic sinus or empirical likelihood. The DPS method is based on the `CBPS` package and uses its estimation techniques to balance means and quantiles.

The main limitation of these methods is the uniqueness of the quantiles. For example, one may be interested in balancing the first and second quartiles, while these values may be exactly the same in the treatment group (e.g. equal to 0). In such cases, the researcher has to carefully study the distribution of the pre-treatment variables in the control and treatment groups in order to select appropriate quantiles.

Further work may involve adapting this approach to synthetic control methods by modifying the set of control variables to account for quantiles rather than means. This approach may be an attractive alternative to Chen (2020) and Gunsilius (2023).

References

Chen, Y. (2020). A distributional synthetic control method for policy evaluation. *Journal of Applied Econometrics*, 35(5):505–525.

Deville, J.-C. and Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American statistical Association*, 87(418):376–382.

Fan, J., Imai, K., Liu, H., Ning, Y., Yang, X., et al. (2016). Improving covariate balancing propensity score: A doubly robust and efficient approach. *URL: https://imai. fas. harvard. edu/research/CBPStheory. html.*

Firpo, S. (2007). Efficient Semiparametric Estimation of Quantile Treatment Effects. *Econometrica*, 75(1):259–276.

Gunsilius, F. F. (2023). Distributional Synthetic Controls. *Econometrica*, 91(3):1105–1117.

Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political analysis*, 20(1):25–46.

Harms, T. and Duchesne, P. (2006). On calibration estimation for quantiles. *Survey Methodology*, 32(1):37–52.

Hazlett, C. (2020). Kernel Balancing: A Flexible Non-Parametric Weighting Procedure for Estimating Causal Effects. *SSRN Electronic Journal*, 30(3).

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

Ning, Y., Sida, P., and Imai, K. (2020). Robust estimation of causal effects via a high-dimensional covariate balancing propensity score. *Biometrika*, 107(3):533–554.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rosenbaum, P. R. (1987). Model-based direct adjustment. *Journal of the American statistical Association*, 82(398):387–394.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Sant'Anna, P. H., Song, X., and Xu, Q. (2022). Covariate distribution balance via propensity scores. *Journal of Applied Econometrics*, 37(6):1093–1120.

Xu, Y. and Yang, E. (2023). Hierarchically Regularized Entropy Balancing. *Political Analysis*, 31(3):457–464.

Zhang, S., Han, P., and Wu, C. (2022). Calibration Techniques Encompassing Survey Sampling, Missing Data Analysis and Causal Inference. *International Statistical Review*, page insr.12518.

Zhao, Q. (2019). Covariate balancing propensity score by tailored loss functions. *The Annals of Statistics*, 47(2).

List of Tables

Tables for the main part of the paper

Table 1: Results of simulation for Y_3 (strong non-linearity) with $n_0 = n_1 = 1000$ based on 500 replicates

Measure	Method	ATT	QTT				
			0.10	0.25	0.50	0.75	0.90
Design 1 (strong separation, normal errors)							
Bias	EB	0.0044	0.0068	0.0004	-0.0031	-0.0216	0.0855
	KEB	0.0747	0.0010	-0.0108	0.0078	0.0213	0.0675
	DEB MQ	0.0141	0.0072	0.0039	-0.0021	-0.0532	0.0689
	DEB MD	0.0098	0.0109	0.0090	0.0052	-0.0345	0.1074
Variance	EB	0.4737	0.0260	0.0366	0.1839	1.1612	6.5252
	KEB	0.0834	0.0270	0.0300	0.0797	0.4980	3.4192
	DEB MQ	0.2999	0.0250	0.0314	0.1326	0.9447	5.9581
	DEB MD	0.2513	0.0264	0.0330	0.1233	0.8677	5.7197
RMSE	EB	0.6883	0.1614	0.1914	0.4289	1.0778	2.5559
	KEB	0.2983	0.1642	0.1736	0.2825	0.7060	1.8503
	DEB MQ	0.5478	0.1584	0.1773	0.3642	0.9734	2.4419
	DEB MD	0.5014	0.1629	0.1819	0.3511	0.9321	2.3940
Design 2 (Weak separation, normal errors)							
Bias	EB	0.0018	-0.0069	-0.0040	-0.0124	-0.0039	0.0618
	KEB	0.0723	-0.0123	-0.0111	-0.0045	0.0524	0.0113
	DEB MQ	0.0006	-0.0057	-0.0035	-0.0080	-0.0326	0.0334
	DEB MD	0.0041	-0.0029	0.0002	0.0025	-0.0182	0.0254
Variance	EB	0.5035	0.0278	0.0355	0.2110	1.0354	4.8602
	KEB	0.0846	0.0290	0.0256	0.0780	0.4865	2.6391
	DEB MQ	0.3527	0.0253	0.0282	0.1320	0.8261	4.6520
	DEB MD	0.2514	0.0265	0.0289	0.1222	0.6690	4.2490
RMSE	EB	0.7096	0.1668	0.1885	0.4596	1.0176	2.2055
	KEB	0.2997	0.1708	0.1603	0.2793	0.6995	1.6246
	DEB MQ	0.5939	0.1591	0.1681	0.3634	0.9095	2.1571
	DEB MD	0.5014	0.1629	0.1699	0.3496	0.8181	2.0615
Design 3 (Medium separation, leptokurtic errors)							
Bias	EB	1.1901	0.0430	0.0977	0.4505	1.7150	3.6745
	KEB	0.1218	0.0834	0.1357	0.0502	-0.3744	0.8978
	DEB MQ	0.6893	0.0447	0.0861	0.2746	0.8790	2.1801
	DEB MD	0.2402	0.0270	0.0416	0.1144	0.3347	0.8294
Variance	EB	0.5373	0.0355	0.0447	0.2679	1.7579	12.6890
	KEB	0.5368	0.0353	0.0479	0.3696	2.9304	14.3232
	DEB MQ	0.4965	0.0357	0.0437	0.2719	1.9107	15.8733
	DEB MD	0.9612	0.0451	0.0556	0.3337	2.3104	20.6418
RMSE	EB	1.3977	0.1933	0.2330	0.6862	2.1678	5.1177
	KEB	0.7427	0.2055	0.2576	0.6100	1.7523	3.8896
	DEB MQ	0.9857	0.1943	0.2262	0.5893	1.6381	4.5416
	DEB MD	1.0094	0.2141	0.2394	0.5889	1.5564	4.6184

Table 2: Simulation results for mis-specified model (design 2) based on 500 replications

Method	ATE	QTE				
		0.10	0.25	0.50	0.75	0.90
			Bias			
IPS (exp)	2.0073	-4.0190	-2.2247	0.7095	5.2301	10.4740
IPS (ind)	2.2844	-3.0931	-1.3918	1.1960	5.1096	9.4856
IPS (proj)	0.3018	-2.6261	-1.8409	-0.4423	1.9471	4.2648
CBPS (j)	2.6558	-3.2478	-1.4447	1.4964	5.8841	10.5782
DPS (j MQ)	0.6931	-1.4792	-0.8639	-0.0642	1.8188	4.1355
DPS (j MD)	0.2633	-1.2441	-0.8092	-0.1933	1.1209	2.5604
DPS (j D)	-0.4888	-0.4936	-0.6743	-0.5461	-0.2158	-0.4357
CBPS (o)	2.9651	-2.9019	-1.1142	1.7327	6.0674	11.2099
DPS (o MQ)	0.6931	-1.1227	-0.6876	-0.1783	1.1199	2.9258
DPS (o MD)	0.2839	-0.7870	-0.5274	-0.0616	0.8369	1.8841
DPS (o D)	0.1041	0.0239	-0.1466	-0.0090	0.3576	0.3468
			Variance			
IPS (exp)	6.7101	13.8096	9.8939	8.0500	13.4849	38.8827
IPS (ind)	7.2561	15.4085	11.1044	8.8530	14.4145	34.6430
IPS (proj)	5.9519	14.6141	10.8652	7.9917	12.8501	31.4331
CBPS (j)	6.5495	13.2423	9.5448	8.4143	14.5832	35.3577
DPS (j MQ)	5.9922	11.2284	8.8770	8.5189	12.2348	24.3055
DPS (j MD)	6.5039	12.2954	9.9444	9.2286	12.8167	25.2888
DPS (j D)	5.7912	11.6075	9.7880	8.7243	12.1644	22.9389
CBPS (o)	7.6237	12.1805	9.4589	8.5988	15.6431	42.3916
DPS (o MQ)	5.9922	10.5597	8.4340	8.1760	11.6608	23.6440
DPS (o MD)	5.4491	10.7479	8.6797	8.0561	10.7197	19.8878
DPS (o D)	5.1791	10.5077	8.5301	7.9814	10.8403	20.0139
			RMSE			
IPS (exp)	3.2771	5.4737	3.8527	2.9246	6.3905	12.1896
IPS (ind)	3.5319	4.9976	3.6113	3.2068	6.3657	11.1633
IPS (proj)	2.4582	4.6380	3.7755	2.8614	4.0794	7.0442
CBPS (j)	3.6882	4.8775	3.4106	3.2640	7.0147	12.1349
DPS (j MQ)	2.5441	3.6628	3.1021	2.9194	3.9424	6.4349
DPS (j MD)	2.5638	3.7207	3.2556	3.0440	3.7514	5.6431
DPS (j D)	2.4556	3.4426	3.2004	3.0037	3.4944	4.8092
CBPS (o)	4.0516	4.5389	3.2711	3.4060	7.2427	12.9635
DPS (o MQ)	2.5441	3.4380	2.9844	2.8649	3.5937	5.6749
DPS (o MD)	2.3515	3.3715	2.9930	2.8390	3.3794	4.8412
DPS (o D)	2.2782	3.2416	2.9243	2.8251	3.3118	4.4871

Table 3: Covariate balancing measures for mis-specified model (design 2) based on 500 replications

Method	CVM		KS	
	Mean	Median	Mean	Median
IPS (ind)	0.28	0.26	2.29	2.22
IPS (exp)	0.37	0.34	2.61	2.50
IPS (proj)	0.75	0.53	3.33	3.08
CBPS (j)	0.40	0.37	2.71	2.67
DPS (j MQ)	0.28	0.24	2.10	2.05
DPS (j MD)	0.32	0.27	2.07	2.00
DPS (j D)	0.31	0.26	2.11	2.06
CBPS (o)	0.41	0.39	2.78	2.73
DPS (o MQ)	0.26	0.22	2.06	2.01
DPS (o MD)	0.26	0.23	1.99	1.96
DPS (o D)	0.26	0.22	1.99	1.94

Appendix for the paper *Survey calibration for causal inference: a simple method to balance covariate distributions*

A Simulation DEB

Table A1: Results of simulation for Y_1 (linear) with $n_0 = n_1 = 1000$ based on 500 replicates

Measure	Method	ATT	QTT				
			0.10	0.25	0.50	0.75	0.90
Design 1 (strong separation, normal errors)							
Bias	EB	-0.0015	0.0029	0.0137	0.0032	0.0049	-0.0033
	KEB	0.0050	0.0173	0.0188	0.0077	0.0011	-0.0044
	DEB MQ	-0.0010	0.0060	0.0138	0.0024	0.0026	0.0005
	DEB MD	-0.0004	0.0104	0.0111	0.0052	0.0067	0.0046
Variance	EB	0.0044	0.0617	0.0325	0.0269	0.0378	0.0967
	KEB	0.0062	0.0477	0.0289	0.0246	0.0383	0.0830
	DEB MQ	0.0045	0.0581	0.0327	0.0274	0.0384	0.0984
	DEB MD	0.0048	0.0587	0.0342	0.0283	0.0402	0.1000
RMSE	EB	0.0664	0.2483	0.1808	0.1642	0.1944	0.3111
	KEB	0.0789	0.2191	0.1711	0.1569	0.1958	0.2881
	DEB MQ	0.0669	0.2412	0.1815	0.1654	0.1960	0.3137
	DEB MD	0.0696	0.2425	0.1852	0.1684	0.2005	0.3162
Design 2 (Weak separation, normal errors)							
Bias	EB	-0.0040	0.0019	0.0021	-0.0083	0.0028	0.0021
	KEB	0.0026	0.0152	0.0056	-0.0082	0.0029	0.0076
	DEB MQ	-0.0037	0.0031	0.0017	-0.0078	0.0016	-0.0016
	DEB MD	-0.0035	0.0048	0.0007	-0.0089	0.0029	0.0097
Variance	EB	0.0040	0.0602	0.0358	0.0260	0.0332	0.0783
	KEB	0.0054	0.0489	0.0288	0.0255	0.0333	0.0622
	DEB MQ	0.0041	0.0579	0.0356	0.0271	0.0347	0.0828
	DEB MD	0.0041	0.0626	0.0368	0.0272	0.0356	0.0809
RMSE	EB	0.0631	0.2453	0.1892	0.1614	0.1821	0.2799
	KEB	0.0734	0.2216	0.1699	0.1598	0.1824	0.2494
	DEB MQ	0.0639	0.2407	0.1886	0.1647	0.1862	0.2878
	DEB MD	0.0642	0.2502	0.1920	0.1653	0.1886	0.2846
Design 3 (Medium separation, leptokurtic errors)							
Bias	EB	0.0066	-0.2357	-0.1322	0.0030	0.1698	0.2624
	KEB	0.0898	0.1145	0.0592	0.0281	0.1019	0.1635
	DEB MQ	0.0053	-0.2230	-0.1213	0.0075	0.1644	0.2337
	DEB MD	0.0062	-0.2437	-0.1266	0.0134	0.1873	0.2254
Variance	EB	0.0081	0.0797	0.0465	0.0400	0.0613	0.1733
	KEB	0.0212	0.0749	0.0536	0.0731	0.1215	0.2628
	DEB MQ	0.0089	0.0887	0.0518	0.0462	0.0695	0.1828
	DEB MD	0.0114	0.1369	0.0731	0.0595	0.0824	0.2349
RMSE	EB	0.0901	0.3678	0.2530	0.2000	0.3003	0.4920
	KEB	0.1710	0.2966	0.2389	0.2718	0.3631	0.5381
	DEB MQ	0.0943	0.3721	0.2578	0.2151	0.3107	0.4873
	DEB MD	0.1071	0.4430	0.2986	0.2443	0.3428	0.5345

Table A2: Results of simulation for Y_2 (medium non-linearity) with $n_0 = n_1 = 1000$ based on 500 replicates

Measure	Method	ATT	QTT				
			0.10	0.25	0.50	0.75	0.90
Design 1 (strong separation, normal errors)							
Bias	EB	-0.0027	-0.0047	0.0003	0.0097	-0.0072	-0.0075
	KEB	0.0053	0.0122	0.0039	0.0096	0.0013	-0.0014
	DEB MQ	-0.0016	0.0001	-0.0016	0.0099	-0.0062	-0.0088
	DEB MD	-0.0014	-0.0048	-0.0029	0.0101	-0.0008	-0.0062
Variance	EB	0.0049	0.0480	0.0302	0.0213	0.0309	0.0713
	KEB	0.0064	0.0377	0.0245	0.0233	0.0283	0.0637
	DEB MQ	0.0049	0.0411	0.0235	0.0200	0.0266	0.0581
	DEB MD	0.0055	0.0384	0.0239	0.0210	0.0264	0.0563
RMSE	EB	0.0700	0.2192	0.1739	0.1462	0.1759	0.2672
	KEB	0.0803	0.1946	0.1564	0.1528	0.1683	0.2523
	DEB MQ	0.0702	0.2026	0.1534	0.1416	0.1632	0.2412
	DEB MD	0.0740	0.1959	0.1546	0.1453	0.1626	0.2373
Design 2 (Weak separation, normal errors)							
Bias	EB	-0.0028	0.0047	0.0034	-0.0039	0.0009	0.0008
	KEB	-0.0027	0.0098	-0.0000	-0.0066	-0.0034	-0.0031
	DEB MQ	-0.0028	0.0055	0.0014	-0.0053	0.0016	-0.0025
	DEB MD	-0.0028	0.0009	-0.0004	-0.0011	0.0044	-0.0004
Variance	EB	0.0045	0.0559	0.0314	0.0194	0.0269	0.0598
	KEB	0.0053	0.0448	0.0268	0.0187	0.0227	0.0511
	DEB MQ	0.0045	0.0468	0.0263	0.0173	0.0209	0.0532
	DEB MD	0.0045	0.0442	0.0254	0.0169	0.0220	0.0472
RMSE	EB	0.0673	0.2364	0.1773	0.1393	0.1641	0.2445
	KEB	0.0726	0.2119	0.1637	0.1370	0.1505	0.2261
	DEB MQ	0.0670	0.2164	0.1621	0.1317	0.1445	0.2306
	DEB MD	0.0668	0.2101	0.1595	0.1300	0.1484	0.2171
Design 3 (Medium separation, leptokurtic errors)							
Bias	EB	0.0388	-0.3582	-0.2149	-0.0179	0.2204	0.5335
	KEB	0.1511	0.0808	0.1121	0.0830	0.0193	0.2120
	DEB MQ	0.0504	-0.2129	-0.0955	-0.0033	0.1351	0.4001
	DEB MD	0.0697	-0.1018	-0.0345	0.0093	0.1056	0.3457
Var	EB	0.0088	0.0553	0.0267	0.0256	0.0505	0.1157
	KEB	0.0178	0.0436	0.0257	0.0353	0.1080	0.2385
	DEB MQ	0.0095	0.0465	0.0242	0.0269	0.0524	0.1252
	DEB MD	0.0150	0.0514	0.0361	0.0387	0.0712	0.1394
RMSE	EB	0.1014	0.4285	0.2700	0.1611	0.3148	0.6327
	KEB	0.2016	0.2239	0.1956	0.2053	0.3292	0.5324
	DEB MQ	0.1099	0.3030	0.1825	0.1642	0.2658	0.5341
	DEB MD	0.1410	0.2485	0.1931	0.1970	0.2870	0.5088

B Simulation for propensity score

Table A3: Simulation results for correctly specified model (design 1) based on 500 replications

Method	ATE	QTE				
		0.10	0.25	0.50	0.75	0.90
			Bias			
IPS (exp)	-0.0378	0.0279	-0.0371	-0.2570	0.0195	0.0706
IPS (ind)	0.3536	0.1618	0.1559	0.1256	0.5079	0.7946
IPS (proj)	-0.0390	0.0177	-0.0340	-0.2510	0.0144	0.0863
CBPS (j)	-0.1158	-0.0334	-0.0620	-0.3772	-0.0748	-0.0619
DPS (j MQ)	-0.0485	-0.0204	-0.0619	-0.2350	0.0113	0.2196
DPS (j MD)	-0.0208	-0.0165	-0.0654	-0.1872	0.0307	0.3950
DPS (j D)	-0.4777	1.0186	0.4195	-0.3893	-1.1203	-2.1375
CBPS (o)	-0.1303	-0.0628	-0.1349	-0.3353	-0.0715	-0.0208
DPS (o MQ)	-0.0485	0.0395	-0.0649	-0.3007	-0.1279	-0.0503
DPS (o MD)	-0.0220	0.0056	-0.0843	-0.2132	0.0577	0.3011
DPS (o D)	0.0084	0.9148	0.4509	-0.0334	-0.2882	-0.8203
			Variance			
IPS (exp)	6.8488	12.8167	9.5348	8.5481	14.6863	36.0200
IPS (ind)	6.4538	13.5540	10.8244	9.1452	14.1257	31.2927
IPS (proj)	6.6207	12.8720	9.5067	8.4839	14.2258	34.4839
CBPS (j)	7.7502	11.2431	9.0807	9.7304	17.0821	39.5007
DPS (j MQ)	7.1257	11.1017	9.4488	9.2409	14.6424	35.2882
DPS (j MD)	7.0747	12.3001	9.9328	9.6852	15.7340	35.5981
DPS (j D)	6.1781	11.1420	9.5545	9.3014	14.5493	29.7210
CBPS (o)	7.6881	11.2360	9.0135	9.6730	16.5078	37.8701
DPS (o MQ)	7.1257	11.0554	9.0681	8.6422	13.2662	28.7042
DPS (o MD)	5.8066	11.2075	9.0012	8.5178	12.4013	24.7986
DPS (o D)	5.4536	10.9623	8.9961	8.4550	12.2454	23.0597
			RMSE			
IPS (exp)	2.6173	3.5802	3.0881	2.9350	3.8323	6.0021
IPS (ind)	2.5649	3.6851	3.2937	3.0267	3.7926	5.6501
IPS (proj)	2.5734	3.5878	3.0835	2.9235	3.7717	5.8729
CBPS (j)	2.7863	3.3532	3.0141	3.1421	4.1337	6.2853
DPS (j MQ)	2.6698	3.3320	3.0745	3.0490	3.8266	5.9444
DPS (j MD)	2.6599	3.5072	3.1523	3.1177	3.9667	5.9795
DPS (j D)	2.5311	3.4899	3.1194	3.0746	3.9755	5.8557
CBPS (o)	2.7758	3.3526	3.0053	3.1282	4.0636	6.1539
DPS (o MQ)	2.6698	3.3252	3.0120	2.9551	3.6445	5.3579
DPS (o MD)	2.4098	3.3478	3.0014	2.9263	3.5220	4.9889
DPS (o D)	2.3353	3.4350	3.0330	2.9079	3.5112	4.8716

Table A4: Covariate balancing measures for correctly specified model (design 1) based on 500 replications

Method	CVM		KS	
	Mean	Median	Mean	Median
IPS (ind)	0.15	0.15	1.95	1.89
IPS (exp)	0.20	0.19	2.09	2.02
IPS (proj)	0.20	0.19	2.08	2.02
CBPS (j)	0.22	0.20	2.15	2.10
DPS (j MQ)	0.19	0.17	1.96	1.92
DPS (j MD)	0.18	0.17	1.87	1.83
DPS (j D)	0.18	0.17	1.92	1.88
CBPS (o)	0.21	0.20	2.15	2.08
DPS (o MQ)	0.20	0.18	2.06	1.98
DPS (o MD)	0.22	0.21	2.18	2.16
DPS (o D)	0.23	0.21	2.24	2.22