Causal Inference using Generalized Empirical Likelihood Methods

Pierre Chaussé * George Luta †

December 7, 2017

Abstract

In this paper, we propose a one step method for estimating the average treatment effect, when the assignment to treatment is not random. We use a misspecified generalized empirical likelihood setup in which we constrain the sample to be balanced. We show that the implied probabilities that we obtain play a similar role as the weights from the weighting methods based on the propensity score. In Monte Carlo simulations, we show that GEL dominates many existing methods in terms of bias and root mean squared errors. We then apply our method to the training program studied by Lalonde (1986).

Classification JEL: C21, C13, J01

1 Introduction

In this paper, we propose a one step method based on the generalized empirical likelihood (GEL), to estimate the average treatment effect (ATE) or the treatment effect on the treated (ATT). The method is valid whether the assumption of random assignment is satisfied or not. In the former case, GEL is correctly specified and its properties are well known. We can therefore easily obtain standard errors, test hypotheses or build confidence intervals. For the latter case, it is misspecified, but the properties of some of the GEL methods have also been derived. The method applies to the case of multiple treatments, and allows outcomes to be either continuous or discrete.

Let the random variable $Y(1)$ be some outcome of interest for individuals when they receive a given treatment, and $Y(0)$ be the outcome when they receive the control. The problem of estimating the ATE, defined as $\mathrm{E}[Y(1)] - \mathrm{E}[Y(0)]$, comes from the fact that only one of the outcomes $Y(1)$ and $Y(0)$ is observed. If we have a sample of n individuals, each of them is either in the treated group, in which case only $Y(1)$ is observed, or in the control one, in which case we only observe $Y(0)$. In other words, we only observe $Y_i = Z_iY_i(1) + (1 - Z_i)Y_i(0)$, where $Z_i = 1$ if individual i is treated and 0 otherwise. The way this issue has been approached, is by realizing that it is a missing value problem. In fact, for each individual, we are missing one of the outcome. The missing value problem is not an issue when observations are missing completely at random. This

*Department of Economics, University of Waterloo, Ontario, Canada (pchausse@uwaterloo.ca)

†Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University

assumption is satisfied in randomized experiments because the treatment is randomly assigned, which makes treatment assignments and, as a result, missing values independent of individual characteristics. This is the ideal situation for which a semi-parametric efficiency bound has been derived for the ATE estimator (see Robins et al. (1994)). In particular, it was shown that the difference between the sample mean of the outcome between the treated and the control, although unbiased and consistent, is not in general semi-parametric efficient. In fact, a method that would use the information about the independence between the treatment indicator Z and any vector of individual characteristics X should dominate the difference in sample means. This idea is explored by Wu and Ying (2011) who incorporate this information in the form of estimating equations and estimate the ATE by Empirical Likelihood (EL). The authors show that the estimator is semi-parametric efficient if we let the number of estimating equations grow with the sample size. An alternative approach is to control for individual characteristics X_i in a regression of Y_i on Z_i. This method is likely to improve efficiency if X_i is correlated with Y_i. However, failing to correctly model the dependence between the outcome and the covariates may bias the estimator of the ATE as noted by Freedman (2008).

In many cases, which is common in social science, treatment does not come from a controlled experiment in which assignment to treatment is random. It is true in particular in observational studies. In those cases, assignment to treatment is potentially endogenous and depends on individual characteristics. In general, methods to estimate the ATE in observational studies rely on the propensity score, $\pi(x) = P(Z = 1|X)$, and are consistent under some identification assumptions (Rosenbaum and Robin, 1983, see). We can divide the literature in three families of estimators. In the first, treated individuals are matched to individuals from the control group with similar propensity scores (see Heckman et al., 1997; Dehejia and Wahba, 1999, 2002, among others). In the second, the propensity score is used to construct a weighted estimator (see Hirano et al., 2003; Hahn, 1998, among others). In particular, the usual identification assumptions imply that ATE can be written as $\mathrm{E}[ZY/\pi(x)] - \mathrm{E}[(1-Z)Y/(1-\pi(x))]$, which leads to a natural estimator once $\pi(x)$ has been estimated. In the last family, the ATE estimator does not rely on the propensity score. Instead, the treated and control groups are balanced using a non-parametric approach (see Hainmueller, 2012; Zubizarreta, 2014; Chan et al., 2016, among others).

What we propose in this paper belongs to the third family. We use the implied probabilities of GEL to make a selected set of moments of the covariates identical in both the treated and the control groups. It is like redefining the distribution used to draw observations from the population so that our sample behaves as if it was generated by a randomized trial. To illustrate the idea behind our method, let p_i be the implied probability assigned to observation i, and $\{Y_i^*, Z_i^*\}$ be obtained by drawing observations from our sample using the distribution $\{p_i\}_{i=1}^n$, with replacement. Then, our estimator of the treatment effect is comparable to the estimator of α_1 in the regression $Y_i^* = \alpha_0 + \alpha_1 Z_i^* + u_i$, which is the estimator used in randomized trials. We propose a particular set of estimating equations, so that one of the parameters is either the ATT or the ATE. It is therefore a one step method for which the asymptotic properties are well known. The GEL method is based on the Cressie and Read family of discrepancies (see Newey and Smith, 2004). We could have defined our estimator based on a more general class of discrepancies, but we decided to restrict ourselves to the ones that belong to the GEL family of estimators.

In the case of randomized trials, we do not need to change our estimating equations. Wu and Ying (2011) show that if we balance the covariates using Empirical Likelihood (EL), the ATE estimator is consistent, asymptotically normal and semi-parametric efficient. It must also be true for GEL, since all members of its family share the same first order asymptotic properties. In the case of observational data, we are incorrectly forcing the distribution of the covariates to be the same for the two groups. Our estimating equations are therefore misspecified even in large samples. However, the properties of GEL have also been studied in the case of misspecification (see Schennach, 2007; Chen et al., 2007; Lee, 2016, among others). In particular, the Exponential Tilting estimator (ET) is robust to misspecification, which means that it is root-n consistent and asymptotically normal even if the population moments are not satisfied. For the other members of the GEL family, it can also be true if we are willing to make stronger assumptions[1].

Notice that our family of methods shares some similarities with the three-step method of Chan et al. (2016). In the first two steps, they obtain weights for the two groups separately. In the third step, they construct an estimator of the ATT or ATE using those weights. Then, an expression for the standard errors is derived using a GMM type argument. We show in Section 2 that we can express their method as a special case of ours, by reformulating our method as a constrained GEL. In fact we offer a larger set of balancing options. Furthermore, our method is more flexible in the sense that it allows to control for characteristics that are not included in the set of balancing covariates. Their argument for not presenting their method as being related to GEL is that the theoretical results derived for GEL are not applicable to misspecified models, and that GEL does not allow the number of moment conditions to grow with the sample size (see Chan et al., 2016, Section 2.2). We showed above that a literature on misspecified GEL does exist, and the asymptotic theory for GEL with growing number of moment conditions has been derived by Donald et al. (2003).

In Section 2, we present the estimation method, Sections 3.1 to 3.3 present the simulation experiments of Frolich (2004), Busso et al. (2014) and Chan et al. (2016) respectively, and we apply our method to the training program study of Lalonde (1986) in Section 4. In Section 5 we conclude.

2 Generalized empirical likelihood

2.1 ATE under randomized trials

In randomized trials, the average treatment effect is defined as $\tau_{ate} = \mathrm{E}(Y_i(1)|Z_i = 1) - \mathrm{E}(Y_i(0)|Z_i = 0)$. The OLS estimator of α_1 in the following model is a consistent estimator of τ_{ate}.

$$Y_i = \alpha_0 + \alpha_1 Z_i + \varepsilon_i \tag{1}$$

Although this estimator is unbiased and consistent under weak regularity conditions, it is not the most efficient estimator of the ATE. In theory, controling for individual characteristics that are correlated with the outcome Y_i can lower the variance of the estimator. However, it may also increase its bias. There is a debate in the statistics literature with respect to whether we should control for observed covariates, because some could be tempted to pick the covariates that would produce the desired results (see Freedman, 2008; Yang and Tsiatis, 2001; Lin, 2013, among

[1]See Schennach (2007), Assumption 3 for ET, Chen et al. (2007), Assumption 3 for EL and Gospodinov et al. (2013) Assumption B for CUE. For ET, conditions are not more restrictive than those required for correctly specified GEL. For the other two methods, we need to assume bounded moment conditions and the existence of an interior solution with positive implied probabilities.

others). The properties of the ATE estimate from a regression approach that controls for covariates depend on the form of dependence between the outcome and the covariates, which is unknown. What is known if the experiment has been properly set, however, is that the distribution of the vector of covariates X_i in the treated and control groups is the same. We can therefore construct moment conditions that reflect that property. More precisely, random assignment implies that the indicator of treatment Z_i is independent of the vector X_i. We can therefore augment the moment conditions implied by the OLS estimator of Equation (1) with conditions that are compatible with the independence assumption. In particular, the following moment conditions are valid under random assignment:

$$\mathrm{E}(g_i(\theta)) = \mathrm{E}\begin{pmatrix} Y_i - \theta_1 - \theta_2 Z_i \\ (Y_i - \theta_1 - \theta_2 Z_i)Z_i \\ (Z_i - \theta_3) \\ (Z_i - \theta_3)u_k(X_i) \end{pmatrix} = 0, \tag{2}$$

where $u_k(X_i)$ is a $k \times 1$ vector of functions of X_i. If we just want to impose the conditions $\mathrm{E}(X_i|Z_i = 0) = \mathrm{E}(X_i|Z_i = 1)$, then $u_k(X_i) = X_i$ and k corresponds to the number of covariates.

Wu and Ying (2011) suggest to use empirical likelihood (EL) to estimate the parameter of interest, θ_2. The authors show consistency and asymptotic normality for fixed k, and show that the estimator is semi-parametric efficient, if we let k grow to infinity at the rate $o(n^{1/3})$. In fact, this result can easily be generalized to all methods that belong to the generalized empirical likelihood family (GEL) and also to any generalized method of moments (GMM) method, which include the two-step, the iterated and continuously updated GMM (CUE). The asymptotic theory for these methods when the number of moment conditions is allowed to grow with the sample size is well developed (see Donald et al., 2003, 2008).

The GEL estimator of $\theta = \{\theta_1, \theta_2, \theta_3\}'$ using the moment conditions (2) is defined as

$$\begin{aligned} &\arg\min_{\theta, p_i} \sum_{i=1}^n CR_\gamma(p_i, 1/n) \\ &\text{subject to} \\ &\sum_{i=1}^n p_i g_i(\theta) = 0 \\ &\sum_{i=1}^n p_i = 1 \end{aligned} \tag{3}$$

where $CR_\gamma(p_i, 1/n)$ is a discrepancy function that is a member of the Cressie and Read family (see Newey and Smith (2004)). In particular, $\gamma = -1$ corresponds to the Empirical Likelihood (EL), $\gamma = 0$ is the Exponential Tilting (ET), $\gamma = 1$ is the CUE or Euclidean Empirical Likelihood (EEL), and $\gamma = -1/2$ is the Hellinger Distance (HD) of Kitamura et al. (2013). Newey and Smith (2004) show that the GEL estimator is the solution to the following saddle point problem:

$$\min_\theta \max_\lambda \sum_{i=1}^n \rho_\gamma(\lambda' g_i(\theta)), \tag{4}$$

where the subscript γ indicates that the functional form of $\rho_\gamma()$ depends on the discrepancy function $CR_\gamma()$. In particular $\rho_\gamma(v) = \log(1-v)$ for EL, $\rho_\gamma(v) = -\exp(v)$ for ET, $\rho_\gamma(v) = -v - v^2/2$ for

EEL or CUE, and $\rho_\gamma(v) = -2/(1-v/2)$ for HD. The implied probabilities are defined as[2]

$$p_i = \frac{\rho'_\gamma(\lambda' g_i(\theta))}{\sum_{j=1}^n \rho'_\gamma(\lambda' g_j(\theta))}, \tag{5}$$

which is clear from the inner first order condition of problem (4)

$$\sum_{i=1}^n \rho'_\gamma(\lambda' g_i(\theta)) g_i(\theta) = 0$$

and the condition $\sum_i p_i = 1$. Once the implied probabilities are obtained, the estimator of θ, $\hat\theta$, satisfies the following system of equations (we omitted the hat on p_i for clarity):

$$\sum_{i=1}^n p_i \begin{pmatrix} Y_i - \hat\theta_1 - \hat\theta_2 Z_i) \\ (Y_i - \hat\theta_1 - \hat\theta_2 Z_i) Z_i \\ (Z_i - \hat\theta_3) \\ (Z_i - \hat\theta_3) u_k(X_i) \end{pmatrix} = 0. \tag{6}$$

It follows that the estimator of the ATE is:

$$\begin{aligned} \hat\theta_2 &= \sum_{i=1}^n \left(\frac{p_i}{\hat\theta_3}\right) Z_i Y_i - \sum_{i=1}^n \left(\frac{p_i}{1-\hat\theta_3}\right)(1-Z_i) Y_i \\ &= \sum_{Z_i=1} \left(\frac{p_i}{\hat\theta_3}\right) Y_i - \sum_{Z_i=0} \left(\frac{p_i}{1-\hat\theta_3}\right) Y_i \end{aligned} \tag{7}$$

Since θ_3 is the unconditional probability of receiving the treatment, the weights in the first and second terms are respectively the implied distribution of the observations given $Z_i = 1$ and $Z_i = 0$. Under some regularity conditions, $\hat\theta$ converges in probability to the true vector θ_0, and $\sqrt{n}(\hat\theta - \theta_0) \to N(0, (G'\Omega^{-1}G)^{-1})$, where $G = \mathrm{E}(\partial g_i(\theta_0)/\partial\theta)$ and Ω is the asymptotic covariance matrix of $\sqrt{n}\bar g(\theta_0)$. A confidence interval for the ATE can therefore be derived from its asymptotic distribution. Alternatively, we can construct a confidence interval using the Owen (2001) approach. As shown by Newey and Smith (2004),

$$2n\left(\frac{1}{n}\sum_{i=1}^n \rho_\gamma(\hat\lambda' g_i(\hat\theta)) - \rho_\gamma(0)\right) \to \chi^2_k, \tag{8}$$

where k is the number of over-identifying restrictions in our model. If we define $\mathcal{R}(\theta_2)$ as

$$\mathcal{R}(\theta_2) = \min_{\theta_1,\theta_3}\left[\max_\lambda 2\sum_{i=1}^n \rho_\gamma(\lambda' g_i(\theta))\right], \tag{9}$$

then $[\mathcal{R}(\theta_{2,0}) - \mathcal{R}(\hat\theta_2)] \to \chi^2_1$. A $(1-\alpha)$ GEL confidence interval is therefore defined as $\{\theta_2 | [\mathcal{R}(\theta_2) - \mathcal{R}(\hat\theta_2)] \le C_{1-\alpha}\}$, where $C_{1-\alpha}$ is the $(1-\alpha)$-quantile of the χ^2_1 distribution. Although the confidence

[2] Notice that for EEL, the implied probabilities are not always positive. However, Antoine et al. (2007) provide a method for dealing with negative probabilities. This issue is more likely to appear when the model is misspecified.

interval is also based on an asymptotic distribution, Owen (2001) shows, using EL, that it often produces more accurate coverage. A test that uses $[\mathcal{R}(\theta_2^*) - \mathcal{R}(\hat{\theta}_2)]$ to test the null hypothesis $H_0 : \theta_2 = \theta_2^*$, is a likelihood ratio (LR) test. The confidence interval is therefore obtained by inverting the LR statistics.

The asymptotic properties of the EL method proposed by Wu and Ying (2011) for estimating the ATE have been derived, but there are very few studies which have analyzed its behaviour in small samples. Chaussé et al. (2016) have studied the small sample properties of EL, CUE and ET using the moment conditions (2), with $u_k(X_i) = X_i$. They compare the bias, the root mean squares errors (RMSE) and the coverage of the confidence intervals. One of the main results is that the empirical coverage obtained by inverting the likelihood ratio test of CUE and ET is less sensitive to the distribution of the outcome Y_i and more accurate than the one obtained using an OLS regression, especially when the correlation between the covariates and the outcome is high.

2.2 ATE and ATT for observational studies

When the assignment is not random, we can still use GEL as a re-weighting approach. There are many re-weighting methods for estimating the ATE and ATT (see Graham et al., 2012; Hainmueller, 2012; Chan et al., 2016; Zubizarreta, 2014), but none of them propose a one step procedure with well known asymptotic properties. Without random assignment, the GEL moment conditions are no longer valid. According to Schennach (2007), only ET is robust to global misspecification when the moment conditions have unbounded support. Being robust to misspecification means that the estimator $\hat{\theta}$ is root-n consistent and converges to a unique pseudo true value. In our case, we can see the pseudo true value of θ_2 as the population treatment effect when the control and treated groups are perfectly balanced. In fact, the moment conditions (2) will force the two groups to be perfectly balanced even asymptotically. If we are willing to make stronger assumptions on $g_i(\theta)$, all GEL methods become $\sqrt{n}$-consistent and an asymptotic variance exists. For example, Chen et al. (2007) show that EL is $\sqrt{n}$-consistent if $g_i(\theta)$ is uniformly bounded. Also, Lee (2016) presents a Bootstrap testing procedure for GEL that is robust to misspecification in case the asymptotic approximation does not perform well.

We can also generalize our family of methods to multiple treatments. Let Z_i be a $l \times 1$ vector of treatment indicators. Then $Z_{ij} = 1$ if individual i is assigned to treatment j for $j = 1, ..., l$. Then θ_2 and θ_3 are $l \times 1$ vectors, and the moment conditions that balance the l treated groups with the control one are

$$\mathrm{E}(g_i(\theta)) = \mathrm{E}\begin{pmatrix} Y_i - \theta_1 - \theta_2' Z_i \\ (Y_i - \theta_1 - \theta_2' Z_i) Z_i \\ (Z_i - \theta_3) \\ (Z_i - \theta_3) \otimes u_k(X_i) \end{pmatrix} = 0, \tag{10}$$

where $\otimes$ is the Kronecker product operator. Furthermore, if we have the information about the population value of $\mathrm{E}(u_k(X_i)) \equiv u_k$, from a census for example, we can use that information to target a specific group for the treatment effect. We would then augment the vector of moment conditions to include that information:

$$\mathrm{E}(g_i(\theta)) = \mathrm{E}\begin{pmatrix} Y_i - \theta_1 - \theta_2' Z_i \\ (Y_i - \theta_1 - \theta_2' Z_i) Z_i \\ (Z_i - \theta_3) \\ (Z_i - \theta_3) \otimes u_k(X_i) \\ u_k(X_i) - u_k \end{pmatrix} = 0, \tag{11}$$

If u_k is the population moments of X_i, θ_2 is the vector of ATE. On the other hand, if u_k is the moments of X_i for a subsample of the population targeted by the treatments, θ_2 becomes the average treatment effect of the treated (ATT). For this general model, the number of coefficients to estimate is $(2l+1)$ and the number of moment conditions is $(1+2l+k(l+1))$. The specification test (8), that we can use to test the null hypothesis of random assignment, would be asymptotically χ^2_{kl} for the model (10) and $\chi^2_{k(l+1)}$ for model (11). The same method as the one described above can also be use to construct GEL confidence intervals.

It is also worth noticing that the method can also be applied to cases in which Y_i is a binary response. Since the method is just a way to estimate the model $Y_i = \theta_1 + \theta_2' Z_i + \varepsilon_i$, with Z_i being a vector of orthogonal dummy variables, we are simply measuring differences in proportions.

2.3 Inference

We saw in Section 2.1 that GEL confidence intervals for θ_2 can be constructed without using any estimate of the standard error, by inverting the LR statistics. It turns out that this approach is also valid under misspecification. Even if the specification test (8) converges to a non-central χ^2, the LR statistics $[\mathcal{R}(\theta_2^*) - \mathcal{R}(\hat{\theta}_2)]$, where θ_2^* is the pseudo true θ_2, converges to a χ^2_1, if the regularity conditions that are required for GEL to be $\sqrt{n}$-consistent are satisfied[3]. For confidence intervals based on the asymptotic distribution, we need standard errors that are robust to misspecification.

In order to derive the asymptotic variance for misspecified GEL, we need to write the FOC of problem (4) as a just-identified GMM. Following Lee (2016), we can define $\hat{\beta} \equiv \{\hat{\theta}', \hat{\lambda}'\}'$ as the solution to the following system of equations:

$$\frac{1}{n}\sum_{i=1}^{n} \psi_i(\beta) = 0, \tag{12}$$

where

$$\psi_i(\beta) = \begin{pmatrix} \rho_1(\lambda' g_i(\theta)) G_i(\theta)' \lambda \\ \rho_1(\lambda' g_i(\theta)) g_i(\theta) \end{pmatrix} \tag{13}$$

where $\rho_1(v) = \partial\rho(v)/\partial v$. Notice that the subscript γ is omitted from $\rho_\gamma(v)$ for clarity. Let $\beta^* = \{\theta^{*'}, \lambda^{*'}\}'$ be the pseudo true value, which is the unique solution to $E(\psi_i(\beta)) = 0$. If we assume some regularity conditions, which are basically the conditions for just identified GMM based on $\mathrm{E}(\psi_i(\beta)) = 0$ (see Hansen, 1982) [4], we have the following results:

$$\sqrt{n}(\hat{\beta} - \beta^*) \xrightarrow{d} N(0, \Gamma^{-1}\Omega(\Gamma')^{-1})$$

[3]See Schennach (2007), Assumption 3 for ET, Chen et al. (2007), Assumption 3 for EL and Gospodinov et al. (2013) Assumption B for CUE.

[4]It is not obvious to link the assumptions on $\psi_i(\beta)$ to the ones that are required for $g_i(\theta)$. For the latter, we refer the reader to the literature on misspecified GEL that we described above.

where $\Gamma = E[(\partial/\partial\beta)\psi_i(\beta^*)]$ and $\Omega = E[\psi_i(\beta^*)\psi_i(\beta^*)']$. More precisely, we have

$$\Gamma_{11} = \mathrm{E}\left[\rho_1(\lambda^{*'}g_i(\theta^*))(I_p\otimes\lambda^{*'})G_i^{(2)}(\theta^*) + \rho_2(\lambda^{*'}g_i(\theta^*))G_i(\theta^*)'\lambda^*\lambda^{*'}G_i(\theta^*)\right]$$

where I_p is the $p\times p$ identity matrix, $p=\dim(\theta)$, $G_i(\theta)^{(2)} = (\partial/\partial\theta)\mathrm{Vec}(G_i(\theta))$ and $\rho_2(v) = (\partial^2/\partial v^2)\rho(v)$,

$$\Gamma_{12} = \Gamma_{21}' = \mathrm{E}\left[\rho_1(\lambda^{*'}g_i(\theta^*))G_i(\theta^*)' + \rho_2(\lambda^{*'}g_i(\theta^*))G_i(\theta^*)'\lambda^* g_i(\theta^*)'\right],$$

$$\Gamma_{22} = \mathrm{E}\left[\rho_2(\lambda^{*'}g_i(\theta^*))g_i(\theta^*)g_i(\theta^*)'\right],$$

and

$$\Omega = \mathrm{E}\left[\rho_1(\lambda^{*'}g_i(\theta^*))^2\begin{pmatrix} G_i(\theta^*)'\lambda^*\lambda^{*'}G_i(\theta^*) & G_i(\theta^*)'\lambda^* g_i(\theta^*)' \\ g_i(\theta^*)\lambda^{*'}G_i(\theta^*) & g_i(\theta^*)g_i(\theta^*)'\end{pmatrix}\right].$$

All matrices can be estimated using sample means and by replacing β^* by $\hat\beta$. If the model is correctly specified, $\lambda^*=0$ and $\rho_1(0)=\rho_2(0)=-1$ (see Newey and Smith (2004)), which implies

$$\Gamma = -\mathrm{E}\begin{pmatrix} 0 & G_i(\theta_0)' \\ G_i(\theta_0) & g_i(\theta_0)g_i(\theta_0)'\end{pmatrix}$$

$$\Omega = \mathrm{E}\begin{pmatrix} 0 & 0 \\ 0 & g_i(\theta_0)g_i(\theta_0)'\end{pmatrix}$$

where 0 stands for a matrix of zeros with the appropriate dimension. It follows that the partition associated with $\hat\theta$ reduces to the GEL covariance matrix
$[E(G_i(\theta^*))'E(g_i(\theta^*)g_i(\theta^*)')^{-1}E(G_i(\theta^*))]^{-1}$. If we use the set of moment conditions (10), we have

$$G_i(\theta) = \begin{pmatrix} -1 & -Z_i' & 0 \\ -Z_i & -\mathrm{diag}(Z_i) & 0 \\ 0 & 0 & -I_l \\ 0 & 0 & -I_l\otimes u_k(X_i)\end{pmatrix}, \tag{14}$$

where $\mathrm{diag}(x)$ is a diagonal matrix with the main diagonal being equal to x, and if we use the conditions (11), we have

$$G_i(\theta) = \begin{pmatrix} -1 & -Z_i' & 0 \\ -Z_i & -\mathrm{diag}(Z_i) & 0 \\ 0 & 0 & -I_l \\ 0 & 0 & -I_l\otimes u_k(X_i) \\ 0 & 0 & 0\end{pmatrix}. \tag{15}$$

Notice that for both cases, $G_i(\theta)$ does not depend on θ, which implies that $G^{(2)}(\theta)=0$. The expression of Γ_{11} can therefore be written as:

$$\Gamma_{11} = \mathrm{E}\left[\rho_2(\lambda^{*'}g_i(\theta^*))G_i(\theta^*)'\lambda^*\lambda^{*'}G_i(\theta^*)\right].$$

If we do not have access to a large survey to measure $E(u_k(X_i))$, it is possible to replace u_k by $\bar{u}_k = \sum_{i=1}^n u_k(X_i)/n$ in order to compute the ATE, or by the sample mean of the treated group to obtain the ATT. Using this approach would produce estimates that are identical to what we would get by using the method proposed by Chan et al. (2016). Although the estimates of θ and λ are not affected by whether u_k is obtained from the population or the sample, the standard errors are. In fact, the GEL standard errors, whether it is robust or not to misspecification, will underestimate the true standard errors because the derivation of the asymptotic variance assumes that u_k is not random.

If instead we estimate u_k using our current sample, the above estimator of the GEL covariance matrix is biased and inconsistent. If we rewrite the moment conditions for ATE or ATT by replacing the sample mean by a vector of coefficients θ_4, we obtain

$$g_i(\theta) = \begin{pmatrix} Y_i - \theta_1 - \theta_2 Z_i \\ (Y_i - \theta_1 - \theta_2 Z_i) Z_i \\ (Z_i - \theta_3) \\ (Z_i - \theta_3) u_k(X_i) \\ u_k(X_i) - \theta_4 \end{pmatrix} \tag{16}$$

where θ_4 is either $E(u_k(X_i))$ for ATE or $E(u_k(X_i)|Z_i = 1)$ for ATT. Notice that we don't need to multiply $(u_k(X_i) - \theta_4)$ by Z_i since the fourth set of conditions balance the moments between the treated and control groups. Let Ξ_{ij} be an indicator that individual i belongs to group j. Then, we can define $\hat{\theta}_4$ as the vector that satisfies the sample moment condition:

$$\frac{1}{n}\sum_{i=1}^n \Xi_{ij}(u_k(X_i) - \theta_4) = 0$$

For ATE, group j is the whole sample, which implies that we set $\Xi_{ij} = 1$ for all i, and for ATT, group j is the treated, which implies that we set $\Xi_{ij} = Z_i$. For the latter case, we only consider a single treatment so that $l = 1$. Clearly, $\hat{\theta}_4 = \frac{1}{n}\sum_{i=1}^n u_k(X_i)$ for the ATE, and it is equal to $\frac{1}{n_1}\sum_{Z_i=1} u_k(X_i)$ for the ATT, where $n_1 = \sum_{i=1}^n Z_i$.

We can think of the problem as being the following constrained GEL estimator:

$$\hat{\theta} = \arg\min_{\lambda_2,\theta} \left\{ \left[\arg\max_{\lambda_1} \sum_{i=1}^n \rho(\lambda_1' g_i(\theta))\right] + \lambda_2' \left(\sum_{i=1}^n \Xi_{ij}(u_k(X_i) - \theta_4)\right)\right\} \tag{17}$$

The first order conditions for λ_1 and $\theta_{(1)} = \{\theta_1, \theta_2, \theta_3\}'$ are

$$\sum_{i=1}^n \rho_1(\lambda_1' g_i(\theta)) g_i(\theta) = 0$$

$$\sum_{i=1}^n \rho_1(\lambda_1' g_i(\theta)) G_{i1}(\theta)' \lambda_1 = 0$$

where $G_{i1}(\theta)$ is given by Equation (15). If we replace θ_4 by $\hat{\theta}_4 = \sum_i \Xi_{ij} u_k(X_i)/n_j$, where $n_j = \sum_i \Xi_{ij}$, we get the same moment conditions as given by Equation (13). We are, however, missing the first order condition with respect to θ_4 and λ_2. For λ_2, we have

$$\sum_{i=1}^{n} \Xi_{ij}(u_k(X_i) - \theta_4) = 0,$$

and for θ_4, we get

$$\sum_{i=1}^{n} \rho_1(\lambda_1' g_i(\theta)) G_{i2}(\theta)'\lambda_1 - n_j \lambda_2 = 0$$

where

$$G_{i2}(\theta) = \begin{pmatrix} 0_{(1+l(k+2))\times k} \\ -I_k \end{pmatrix}$$

It follows that

$$\lambda_2 = \frac{1}{n_j} \sum_{i=1}^{n} \rho_1(\lambda_1' g_i(\theta)) G_{i2}(\theta)'\lambda_1$$

The extra vector λ_2 is therefore directly linked to λ_1. If the assignment is random, $\lambda_1^* = 0$, which also implies $\lambda_2^* = 0$. If we combine the additional first order conditions, $\hat{\lambda}_1$ and $\hat{\theta}$ solves the sample moments:

$$\sum_{i=1}^{n} \begin{pmatrix} \rho_1(\lambda_1' g_i(\theta)) G_{i1}(\theta)'\lambda_1 \\ \rho_1(\lambda_1' g_i(\theta)) G_{i2}(\theta)'\lambda_1 - \lambda_2 \Xi_{ij} \\ \rho_1(\lambda_1' g_i(\theta)) g_i(\theta) \\ \Xi_{ij}(u_k(X_i) - \theta_4) \end{pmatrix} = 0$$

It follows that the augmented $\psi(\beta)$ is

$$\psi_{aug,i}(\beta) = \begin{pmatrix} \rho_1(\lambda_1' g_i(\theta)) G_{i1}(\theta)'\lambda_1 \\ \rho_1(\lambda_1' g_i(\theta)) G_{i2}(\theta)'\lambda_1 - \lambda_2 \Xi_{ij} \\ \rho_1(\lambda_1' g_i(\theta)) g_i(\theta) \\ \Xi_{ij}(u_k(X_i) - \theta_4) \end{pmatrix}, \tag{18}$$

Let $\beta = \{\theta', \lambda'\}'$, where $\theta = \{\theta_1, \theta_2', \theta_3', \theta_4'\}'$, $\lambda = \{\lambda_1', \lambda_2'\}'$ and $G_i(\theta) = \{G_{1i}(\theta), G_{2i}(\theta)\}$. Then, the augmented Γ becomes

$$\Gamma_{aug,11} = \mathrm{E}\left(\rho_2(\lambda_1^{*'} g_i(\theta^*)) G_i(\theta^*)'\lambda_1^* \lambda_1^{*'} G_i(\theta^*)\right),$$

$$\Gamma_{aug,12} = \Gamma_{aug,21}' = \mathrm{E}\begin{pmatrix} \rho_1(\lambda_1^{*'} g_i(\theta^*)) G_i(\theta^*)' + \rho_2(\lambda_1^{*'} g_i(\theta^*)) G_i(\theta^*)'\lambda_1^* g_i(\theta^*)' & \begin{matrix} 0 \\ -\Xi_{ij} I_k \end{matrix} \end{pmatrix},$$

and

$$\Gamma_{aug,22} = \mathrm{E}\begin{pmatrix} \rho_2(\lambda_1^{*'} g_i(\theta^*)) g_i(\theta^*) g_i(\theta^*)' & 0 \\ 0 & 0 \end{pmatrix}.$$

The asymptotic covariance matrix is $\Gamma_{aug}^{-1}\Omega_{aug}\Gamma_{aug}^{-1}$, and its first $l+1$ diagonal elements are the asymptotic variances of $\{\hat{\theta}_1, \hat{\theta}_2'\}'$. Note that the standard errors obtained from this method are identical to the ones obtained by Chan et al. (2016), but we provide a constrained GEL formulation.

Table 1: Different DGP's used in the simulations

$m(x) =$	$0.15 + 0.7x$
	$0.1 + x/2 + \exp(-200(x-0.7)^2)/2$
	$0.8 - 2(x-0.9)^2 - 5(x-0.7)^3 - 10(x-0.6)^{10}$
	$0.2 + \sqrt{1-x} - 0.6(0.9-x)^2$
	$0.2 + \sqrt{1-x} - 0.6(0.9-x)^2 - 0.1x\cos(30x)$
	$0.4 + 0.25\sin(8x - 5 + 0.4\exp(-16(4x-2.5)^2)$
$\{\alpha,\beta\} =$	$\{0,1\}$
	$\{0.15, 0.7\}$
	$\{0.3, 0.4\}$
	$\{0, 0.4\}$
	$\{0.6, 0.4\}$

3 Numerical Study

3.1 First Study

The first numerical experiment we consider is based on Frolich (2004) who analyzes the finite sample properties of several matching and re-weighting estimators. Busso et al. (2014) update the results by considering more recent methods. The model is

$$Y_i(0) = m(W_i) + \sigma\varepsilon_i,$$

$$Z_i = 1\{\alpha + \beta W_i - U_i > 0\},$$

where $X_i \sim N(0,1)$, $W_i = \Lambda(\sqrt{2}X_i)$, where $\Lambda()$ is the CDF of the logistic distribution, and both ε_i and U_i are iid and distributed as a $U(-\sqrt{3},\sqrt{3})$. Also, $Y_i(1) = Y_i(0)$, so that the true ATE is 0. Frolich (2004) considers six different functions $m()$ and five different sets of parameters $\{\alpha,\beta\}'$. Table 1 shows the details of the 30 different specifications. The sample size is set to $n = 100$, $\sigma = 0.01$ and the number of iterations is 10,000. All estimations are done using the function ATEgel() from the R package gmm of Chaussé (2010). For EL, the algorithm proposed by Wu (2005) is used to compute the vector λ.

In order to have results comparable with those from Busso et al. (2014), Tables 2 to 5 show respectively 1,000 times the bias, 100 times the variance, 100 times the MSE and the coverage of 95% confidence intervals, on average across all functions $m(x)$, for each set of parameters. For detailed results for each $m(x)$, see Appendix A.[5] In the tables, we refer to METi, for i=1,2,3, to $u_k(X_i) = X_i$, $u_k(X_i) = \{X_i, X_i^2\}'$ and $u_k(X_i) = \{X_i, X_i^2, X_i^3\}$ respectively, with MET = ET or EL. Although we have also applied the CUE method in our simulations, we do not present the results because for a large proportion of the iterations, it fails to balance the groups with strictly positive implied probabilities. If we use the approach proposed by Antoine et al. (2007) to transform the negative probabilities, the resulting set of implied probabilities do not balance the groups. Since it is not clear how to interpret the GEL estimate of the treatment effect when the balancing

[5]Notice that some EL estimations failed. They failed when the time to estimate the model was over a certain threshold. The number of failures depends on the model and the choice of $u_k(X_i)$. It is, however, negligible.

probabilities are negative, we can hardly compare the results with the other methods[6]. For each method, we consider unconstrained moment balancing using moment conditions (2) (METi), balancing based on the sample moments (METi Bal Sample), based on known population moments of the treated (METi Bal Pop), or based on the sample moments of the treated (METi ATT). We present the properties for both the ATE and ATT, but we can only compare the results reported by Busso et al. (2014) with our two ATT methods (METi Bal Pop and METi ATT), because the authors only considered the ATT estimation.

Table 2: Average (1000xBias) across all functions $m(x)$

	$\{\alpha,\beta\}_1$	$\{\alpha,\beta\}_2$	$\{\alpha,\beta\}_3$	$\{\alpha,\beta\}_4$	$\{\alpha,\beta\}_5$	Average
ET1	2.524	3.563	1.391	4.298	3.818	3.119
ET2	6.028	0.9677	0.68	3.366	7.54	3.716
ET3	4.657	1.888	0.523	1.559	**2.267**	**2.179**
ET1 Bal Sample	4.853	2.375	1.134	4.47	4.467	3.46
ET2 Bal Sample	22.81	3.563	**0.4635**	4.059	6.016	7.382
ET3 Bal Sample	13.98	5.38	1.186	2.731	6.41	5.937
ET1 ATT	10.88	0.9921	1.013	3.705	5.968	4.513
ET2 ATT	11.72	3.49	1.42	0.842	5.342	4.562
ET3 ATT	14.17	5.5	1.636	**0.7723**	7.154	5.846
ET1 Bal Pop	11.36	**0.9653**	1.057	4.081	5.933	4.68
ET2 Bal Pop	14.17	4.022	0.9876	1.661	5.453	5.258
ET3 Bal Pop	16.05	5.987	1.129	2.822	8.759	6.949
EL1	**1.861**	6.236	1.784	3.026	2.866	3.155
EL2	7.637	1.622	0.8379	3.644	6.496	4.047
EL3	13.35	4.259	0.9416	2.212	3.452	4.842
EL1 Bal Sample	3.774	3.964	1.359	3.35	2.862	3.062
EL2 Bal Sample	7.862	1.809	0.5885	4.716	3.614	3.718
EL3 Bal Sample	15.69	6.439	1.389	4.875	6.365	6.952
EL1 ATT	21.66	7.55	2.463	8.715	4.205	8.919
EL2 ATT	12.45	4.751	1.598	2.676	6.174	5.53
EL3 ATT	17.37	6.587	1.236	2.207	8.482	7.177
EL1 Bal Pop	22.41	6.813	2.155	8.328	4.09	8.758
EL2 Bal Pop	14.29	4.579	1.393	4.409	6.6	6.254
EL3 Bal Pop	20.19	7.942	1.37	5.487	9.483	8.893

Table 3: Average (nxVariance) across all functions $m(x)$

	$\{\alpha,\beta\}_1$	$\{\alpha,\beta\}_2$	$\{\alpha,\beta\}_3$	$\{\alpha,\beta\}_4$	$\{\alpha,\beta\}_5$	Average
ET1	0.1482	0.09925	0.06443	0.103	0.1226	0.1075
ET2	0.1228	0.09287	0.05728	0.1268	0.1656	0.1131
ET3	**0.09766**	**0.07527**	**0.05187**	0.1109	0.131	**0.09334**
ET1 Bal Sample	0.1498	0.09908	0.06451	0.1025	0.1228	0.1077
ET2 Bal Sample	0.2248	0.1218	0.06216	0.125	0.1503	0.1368
ET3 Bal Sample	0.1813	0.1129	0.06404	0.1165	0.1391	0.1228
ET1 ATT	0.2181	0.1355	0.06936	0.08001	0.141	0.1288
ET2 ATT	0.2038	0.1457	0.07474	**0.07553**	0.1591	0.1318
ET3 ATT	0.1803	0.1226	0.06881	0.07785	0.1501	0.1199
ET1 Bal Pop	0.222	0.1335	0.06871	0.08247	0.1395	0.1292
ET2 Bal Pop	0.2142	0.1534	0.0762	0.0851	0.1609	0.138
ET3 Bal Pop	0.1927	0.1167	0.06958	0.09558	0.1489	0.1247
EL1	0.2043	0.114	0.0669	0.1085	0.1271	0.1242
EL2	0.1612	0.1075	0.06031	0.1308	0.1648	0.1249
EL3	0.1517	0.09076	0.05538	0.1318	0.1594	0.1178
EL1 Bal Sample	0.1789	0.1074	0.06483	0.1129	0.1246	0.1177
EL2 Bal Sample	0.1974	0.1116	0.0614	0.1156	0.1138	0.12
EL3 Bal Sample	0.168	0.1086	0.06252	0.09435	**0.09425**	0.1056
EL1 ATT	0.285	0.1793	0.07704	0.09229	0.1417	0.1551
EL2 ATT	0.1928	0.1281	0.07168	0.0758	0.1213	0.1179
EL3 ATT	0.1868	0.1261	0.06857	0.07771	0.09644	0.1111
EL1 Bal Pop	0.2834	0.1804	0.07425	0.0901	0.1417	0.1539
EL2 Bal Pop	0.1957	0.1241	0.0724	0.08071	0.1254	0.1197
EL3 Bal Pop	0.1924	0.1211	0.06749	0.08802	0.1003	0.1139

Notice that for the sets of parameters 1 and 5, the strict overlap assumption is violated, and for the sets 2 to 4 it is not violated[7]. One of the conclusion of Busso et al. (2014) is that weighting methods are much more biased in designs 1 and 5, but outperform the matching methods when

[6]The results for CUE are available upon request

[7]The overlap assumption implies that the probability of being treated, conditional on the covariates, is strictly between 0 and 1. For strict overlap, it has to be between c and $(1-c)$ for some strictly positive constant c.

Table 4: Average (nxMSE) across all functions $m(x)$

	$\{\alpha,\beta\}_1$	$\{\alpha,\beta\}_2$	$\{\alpha,\beta\}_3$	$\{\alpha,\beta\}_4$	$\{\alpha,\beta\}_5$	Average
ET1	0.1492	0.1015	0.06492	0.1073	0.1255	0.1097
ET2	0.1286	0.09303	0.05734	0.129	0.1766	0.1169
ET3	**0.1022**	**0.07611**	**0.0519**	0.1114	0.1317	**0.09466**
ET1 Bal Sample	0.1536	0.1001	0.06483	0.1067	0.1261	0.1103
ET2 Bal Sample	0.3033	0.1239	0.06219	0.1287	0.1562	0.1549
ET3 Bal Sample	0.2172	0.1193	0.06425	0.1177	0.1496	0.1336
ET1 ATT	0.2345	0.1357	0.06958	0.0825	0.146	0.1337
ET2 ATT	0.2309	0.1473	0.07511	**0.07575**	0.1633	0.1385
ET3 ATT	0.2163	0.1293	0.0694	0.07794	0.1626	0.1311
ET1 Bal Pop	0.2398	0.1336	0.06895	0.08535	0.1447	0.1345
ET2 Bal Pop	0.2506	0.1557	0.07639	0.08587	0.1654	0.1468
ET3 Bal Pop	0.2348	0.1242	0.06981	0.09713	0.1674	0.1387
EL1	0.2049	0.121	0.06771	0.1102	0.1284	0.1265
EL2	0.1697	0.1082	0.06041	0.1334	0.1721	0.1287
EL3	0.1881	0.09519	0.0555	0.1331	0.1615	0.1267
EL1 Bal Sample	0.1808	0.1103	0.0653	0.115	0.1256	0.1194
EL2 Bal Sample	0.2059	0.1123	0.06144	0.1191	0.1156	0.1229
EL3 Bal Sample	0.2156	0.1177	0.06283	0.09774	**0.1027**	0.1193
EL1 ATT	0.3484	0.1889	0.07884	0.1039	0.1437	0.1727
EL2 ATT	0.2167	0.1316	0.07215	0.0775	0.1264	0.1249
EL3 ATT	0.2223	0.1338	0.06891	0.07877	0.1079	0.1223
EL1 Bal Pop	0.3503	0.1879	0.0757	0.1008	0.1436	0.1717
EL2 Bal Pop	0.225	0.1271	0.07274	0.08521	0.1314	0.1283
EL3 Bal Pop	0.2386	0.1326	0.06782	0.09344	0.1157	0.1296

Table 5: Average coverage of 95% confidence intervals across all functions $m(x)$

	$\{\alpha,\beta\}_1$	$\{\alpha,\beta\}_2$	$\{\alpha,\beta\}_3$	$\{\alpha,\beta\}_4$	$\{\alpha,\beta\}_5$
ET1	**0.9166**	**0.9324**	**0.9397**	0.8801	0.885
ET2	0.9139	0.9284	0.9367	0.8271	0.8192
ET3	0.9153	0.9288	0.9351	0.8272	0.8252
ET1 Bal Sample	0.9107	0.9307	0.9392	0.876	0.8802
ET2 Bal Sample	0.8056	0.9041	0.9311	0.8059	0.81
ET3 Bal Sample	0.8233	0.8969	0.9253	0.8347	0.8391
ET1 ATT	0.866	0.9135	0.9372	0.9275	0.8565
ET2 ATT	0.847	0.9011	0.9272	0.9352	0.7973
ET3 ATT	0.8734	0.9053	0.9285	**0.9363**	0.8379
ET1 Bal Pop	0.8452	0.9042	0.9342	0.9142	0.8429
ET2 Bal Pop	0.7901	0.8759	0.9105	0.8915	0.7606
ET3 Bal Pop	0.7157	0.8308	0.8882	0.8526	0.7095
EL1	0.8664	0.9173	0.9373	0.8617	0.8719
EL2	0.8742	0.9112	0.9328	0.8052	0.8085
EL3	0.8478	0.9043	0.9289	0.7993	0.8012
EL1 Bal Sample	0.8814	0.9199	0.9393	0.849	0.8731
EL2 Bal Sample	0.8196	0.9029	0.9299	0.8041	0.8634
EL3 Bal Sample	0.7763	0.8787	0.9221	0.8535	**0.89**
EL1 ATT	0.6953	0.839	0.9241	0.9022	0.84
EL2 ATT	0.8413	0.9023	0.9245	0.9361	0.8524
EL3 ATT	0.8066	0.8754	0.9203	0.9344	0.8881
EL1 Bal Pop	0.6774	0.8339	0.9243	0.8954	0.8313
EL2 Bal Pop	0.7922	0.8849	0.9102	0.8947	0.8238
EL3 Bal Pop	0.7174	0.8196	0.8874	0.8626	0.8293

overlap is satisfied. Our method seems to behave in the same way. Overall, it is less biased than any methods that are not bias-corrected, and if we only consider designs 2 to 4, the bias of GEL is only slightly higher than the bias-corrected methods, with ET being less biased than EL. It is not clear, however, whether it is better to balance 1, 2 or 3 moments. For example, ET1 ATT is better for the second design and ET3 ATT is the least biased for the fourth one.

For the variance, our method is comparable to best methods studied by Busso et al. (2014) if we balance more than one moment. In fact, the effect of balancing more moments on efficiency is more stable than it is on the bias. Also, ET seems better than EL in terms of the variance for most DGP's. Table 4 shows that it is also the case with the MSE. Finally, Table 5 shows that the coverage of the confidence intervals are close to 95% when overlap is good, which is a good sign given the small sample size.

3.2 Second Study

The previous DGP's represent the ideal cases because they imply that the propensity score methods are correctly specified. Since our method does not rely on the propensity score, it should be robust to misspecification. To compare its properties with the ones studied by Busso et al. (2014), we consider the following set of DGP's used by the authors:

$$Y_i(0) = \begin{cases} DGP_1\text{:} \ -\sum_{j=1}^4 X_{ji} + \varepsilon_i \\ DGP_2 \text{ and } DGP_4\text{: } 0.65 + \sum_{j\neq l} X_{ji}X_{li} + \varepsilon_i \\ DGP_3\text{:} \ -0.5\sum_{j=1}^4 X_{ji} + 0.5(0.65 + \sum_{j\neq l} X_{ji}X_{li}) + \varepsilon_i \end{cases}$$

$$Z_i^* = \begin{cases} DGP_1 \text{ and } DGP_4\text{: } \sum_{j=1}^4 X_{ji} - U_i \\ DGP_2\text{: } 0.65 + \sum_{j\neq l} X_{ji}X_{li} - U_i \\ DGP_3\text{: } 0.5\sum_{j=1}^4 X_{ji} + 0.5(0.65 + \sum_{j\neq l} X_{ji}X_{li}) - U_i \end{cases}$$

where U_i has a logistic distribution with location and scale parameters equal to 0 and 1 respectively, $\varepsilon_i \sim N(0,1)$, and $\{X_1, X_2, X_3, X_4\}'$ are jointly uniform with zero mean and covariance matrix equals to

$$\Sigma_x = \frac{1}{3}\begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix}.$$

The treatment indicator is $Z_i = 1\{Z_i^* \geq 0\}$ and $Y_i(1) = Z_i + Y_i(0)$, which implies that the ATE and ATT are equal to one. For each method, we balance the groups with respect to the vector of four moments $u_4(X) = \{X_1, X_2, X_3, X_4\}'$, the vector of 8 moments $u_8(X) = \{u_4(X)', X_1^2, X_2^2, X_3^2, X_4^2\}'$, the vector of 10 moments $u_{10}(X) = \{u_4, X_1X_2, X_1X_3, X_1X_4, X_2X_3, X_2X_4, X_3X_4\}$, and the vector of 14 moments $u_{14}(X) = \{u_{10}(X)', X_1^2, X_2^2, X_3^2, X_4^2\}'$. The sample size is set to 400, and the number of iterations to 10,000. The properties of the different methods are presented in Tables 6 and 7.

For all DGP, ET and EL are less biased when the number of moments that we balance is greater than four. It is especially true for DGP_2 and DGP_3. Also, ET outperforms EL both in terms of bias and variance. If we only focus on $u_{14}(X)$ and the ATT, ET is less biased than most methods studied by Busso et al. (2014) (see Table 3 on page 894) and more efficient. Table 7 shows once more that our inference method produces accurate confidence intervals, ET being better than EL in that category as well.

3.3 Third Study

In the last study, we want to reproduce the experiment of Chan et al. (2016) to compare our method with theirs, and analyze the properties of both the ATE and ATT estimators. The DGP is: $Y(1) \sim N(210 + b'W, 1)$, $Y(0) \sim N(200 - 0.5b'W, 1)$ with $b = \{27.4, 13.7, 13.7, 13.7\}$ and $W \sim N(0, I)$. Also, the treatment assignment Z is such that $P(Z = 1|W) = \Lambda(\eta'W)$ with $\eta = \{-1, 0.5, -0.25, -0.1\}$, where $\Lambda()$ is the CDF of the logistic distribution. We observe $Y = ZY(1) + (1 - Z)Y(0)$, the treatment indicator Z and the four covariates $\{X_1, X_2, X_3, X_4\}$, where $X_1 = \exp(W_1/2)$, $X_2 = W_2/(1+\exp(W_1))$, $X_3 = (W_1W_3/25+0.6)^3$ and $X_4 = (W_2+W_4+20)^2$. Given this setup, the true ATE is 10 and the ATT is -5.

Table 6: 1000xBias and nxVariance for each misspecified DGP

	1000xBias				nxVariance			
	DGP_1	DGP_2	DGP_3	DGP_4	DGP_1	DGP_2	DGP_3	DGP_4
ET4	0.5126	450.6	123.1	1.148	4.779	5.941	4.75	7.181
ET8	0.4522	142.1	36.49	0.6646	4.845	5.045	4.513	5.603
ET10	0.3588	1.527	0.7549	0.3593	4.866	4.665	4.414	4.866
ET14	0.3309	1.493	0.7443	0.3302	4.93	4.712	4.465	4.93
ET4 Bal Sample	0.4619	450.6	123.2	1.052	4.778	5.942	4.751	7.195
ET8 Bal Sample	0.3068	143.8	36.65	0.51	4.862	5.129	4.537	5.664
ET10 Bal Sample	0.4571	1.492	0.6554	0.4583	5.014	4.749	4.445	5.014
ET14 Bal Sample	0.327	1.322	0.5659	0.3274	5.185	4.85	4.524	5.185
ET4 ATT	-0.129	449.8	121.8	2.053	5.838	5.969	5.016	8.488
ET8 ATT	**0.008443**	132.4	35.15	8.735	6.042	5.252	4.899	6.919
ET10 ATT	0.172	1.113	0.3454	0.1697	6.21	5.329	5.033	6.21
ET14 ATT	0.1916	0.9663	0.1598	0.2589	6.524	5.53	5.216	6.532
ET4 Bal Pop	-0.1233	451.5	122.4	2.1	5.838	6.007	5.056	8.479
ET8 Bal Pop	-0.1823	132.6	35.1	8.143	5.996	5.283	4.934	6.868
ET10 Bal Pop	0.2607	1.608	0.6004	0.2608	6.134	5.28	5.045	6.135
ET14 Bal Pop	-0.08746	1.741	0.4038	-0.09051	6.433	5.481	5.212	6.435
EL4	0.505	451.2	124.2	0.9455	5.081	5.946	4.776	7.977
EL8	0.4321	144.4	38.45	0.71	5.174	5.282	4.628	6.42
EL10	0.4526	1.274	0.5468	0.4549	5.226	4.953	4.537	5.226
EL14	0.4803	1.475	0.5772	0.4808	5.318	4.994	4.599	5.318
EL4 Bal Sample	0.4448	449.6	121.7	0.9235	5.051	5.933	4.756	7.421
EL8 Bal Sample	0.2816	143.4	36.23	0.5111	5.146	5.256	4.604	5.955
EL10 Bal Sample	0.3423	1.325	0.5536	0.3416	5.244	4.954	4.535	5.244
EL14 Bal Sample	0.2158	1.265	0.463	0.2149	5.401	5.04	4.618	5.401
EL4 ATT	-1.157	450.2	107.1	-87.77	9.263	5.96	5.353	13.04
EL8 ATT	-0.6647	132.3	17.8	-92.73	8.658	5.664	5.532	9.79
EL10 ATT	0.14	**0.02852**	0.2163	0.1378	7.635	6.807	5.991	7.635
EL14 ATT	0.8783	0.4378	0.3396	1.015	7.616	6.74	6.007	7.629
EL4 Bal Pop	-1.191	451.1	106.7	-92.46	9	6.008	5.297	12.58
EL8 Bal Pop	-1.085	134.9	18.22	-95.68	8.262	5.543	5.376	9.325
EL10 Bal Pop	-0.0409	0.8392	0.3317	**-0.04085**	7.206	6.43	5.8	7.207
EL14 Bal Pop	-0.02165	1.55	0.2281	0.06255	7.127	6.383	5.788	7.143

We want to compare ET and EL for the estimation of the ATE using unconditional balancing, balancing based on the sample moments and on the population ones. The latter is estimated using a sample size of 10 millions. Also, we want to estimate the ATT using balancing based on the sample moments and the population ones, which is also estimated using the same large simulated sample. For each case, we consider $u_4(X) = \{X_1, X_2, X_3, X_4\}'$, $u_8(X) = \{u_4(X)', X_1^2, X_2^2, X_3^2, X_4^2\}'$, and $u_{14}(X) = \{u_8(X)', X_1X_2, X_1X_3, X_1X_4, X_2X_3, X_2X_4, X_3X_4\}'$. To compare our results with Chan et al. (2016), we set the sample size to 200 and 1,000, and the number of iterations to 5,000.

Tables 8 and 9 present the results for ATE, and Tables 10 and 11 for ATT. The tables present the same properties as for the first two studies, with the standard error bias. The standard error bias is presented so that we can see how accurate the robust-to-misspecification standard error is. It compares the estimated standard errors on average with the standard error of the coefficient over the 5,000 iterations. We can see that the bias is small, which implies that the inaccuracy of the confidence intervals is mostly due to the bias.

The meaning of the labels are as for the first two simulation studies, with the exception that $METi$ for $i = 4, 8, 14$ refers to the three different $u_i(X)$ described above. Notice that METi Bal Sample refers to the Chan et al. (2016) method. The bold entries refer to the best methods. For ATE, EL4 is the least biased for n=200, and ET4 Bal Pop is the least biased when n=1000[8]. For both sample sizes, balancing using the population moments dominates in terms of RMSE. This result is mainly driven by the fact that using the population moments reduces the standard error substantially. For coverage, the Bal. Sample seems the most accurate. However, it is meaningless to compare coverage when the estimators are biased. If we omit the methods that rely on population moments, which will often be unfeasible, ELi, for $i = 4, 8, 14$, perform very well to compute the ATE. The RMSE is small and the coverage is close to 95%. For ATT, EL with

[8]CUE14 Bal Sample and CUE 14Bal Pop are even less biased but as we explained above, most balancing is done with negative implied probabilities which does not constitute a proper balancing

Table 7: nxMSE and Coverage of 95% confidence intrervals for each misspecified DGP

	nxMSE				Coverage			
	DGP_1	DGP_2	DGP_3	DGP_4	DGP_1	DGP_2	DGP_3	DGP_4
ET4	4.778	87.14	10.81	7.181	0.9467	0.0405	0.7903	0.9441
ET8	4.844	13.12	5.045	5.603	0.9437	0.7403	0.9308	0.9419
ET10	4.866	4.665	4.414	4.866	0.9429	0.9435	0.9426	0.9428
ET14	4.93	4.712	4.465	4.93	0.9401	0.9418	0.9417	0.9402
ET4 Bal Sample	4.778	87.15	10.82	7.195	0.9466	0.0405	0.7898	0.9434
ET8 Bal Sample	4.861	13.4	5.074	5.663	0.9427	0.7317	0.9295	0.9388
ET10 Bal Sample	5.014	4.749	4.445	5.014	0.9404	0.9408	0.9409	0.9404
ET14 Bal Sample	5.185	4.85	4.523	5.185	0.9316	0.9369	0.9384	0.9315
ET4 ATT	5.837	86.89	10.95	8.489	0.94	0.0432	0.8014	0.9365
ET8 ATT	6.041	12.26	5.393	6.949	0.9361	0.7837	0.9329	0.9337
ET10 ATT	6.21	5.329	5.033	6.21	0.9327	0.9408	0.9387	0.9326
ET14 ATT	6.523	5.53	5.215	6.531	0.9224	0.9336	0.9369	0.9222
ET4 Bal Pop	5.838	87.53	11.04	8.48	0.9367	0.0407	0.7931	0.9352
ET8 Bal Pop	5.995	12.32	5.426	6.894	0.9299	0.7728	0.9269	0.9253
ET10 Bal Pop	6.134	5.28	5.045	6.135	0.9266	0.9328	0.9326	0.9266
ET14 Bal Pop	6.432	5.482	5.212	6.434	0.9116	0.9238	0.9258	0.9115
EL4	5.081	87.37	10.94	7.977	0.9435	0.0394	0.7874	0.9397
EL8	5.174	13.62	5.218	6.42	0.9419	0.7336	0.9275	0.9367
EL10	5.226	4.954	4.536	5.226	0.9385	0.939	0.9401	0.9384
EL14	5.318	4.995	4.598	5.318	0.9337	0.9361	0.9384	0.9337
EL4 Bal Sample	5.05	86.78	10.68	7.42	0.9437	0.0411	0.795	0.9458
EL8 Bal Sample	5.146	13.48	5.128	5.954	0.9387	0.7312	0.9274	0.9396
EL10 Bal Sample	5.243	4.954	4.535	5.243	0.9351	0.9376	0.9384	0.9352
EL14 Bal Sample	5.401	5.041	4.617	5.401	0.9252	0.9324	0.9363	0.9252
EL4 ATT	9.262	87.03	9.943	16.12	0.8953	0.0428	0.8347	0.879
EL8 ATT	8.657	12.67	5.658	13.23	0.8901	0.789	0.9326	0.8441
EL10 ATT	7.635	6.806	5.991	7.635	0.9034	0.9137	0.9217	0.9034
EL14 ATT	7.616	6.74	6.006	7.629	0.9021	0.912	0.9214	0.902
EL4 Bal Pop	9	87.41	9.852	16	0.8933	0.04	0.8295	0.8767
EL8 Bal Pop	8.262	12.82	5.508	12.99	0.8871	0.7712	0.9309	0.8337
EL10 Bal Pop	7.205	6.43	5.8	7.206	0.901	0.9144	0.917	0.9011
EL14 Bal Pop	7.126	6.383	5.788	7.142	0.8953	0.907	0.9126	0.8947

population moments performs in general better than the other methods. When we don't have the population moments, EL is still the method with the smallest RMSE and better coverage.

4 Training Program

To illustrate the methods, we consider the experiment analyzed first by Lalonde (1986) and used later by Dehejia and Wahba (1999, 2002) to illustrate how to use the matching methods in non-experimental studies. The objective of the original paper was to measure the effect of a training program on the real income. The dependent variable is the real income in 1978 and the covariates used for matching the treated group to the control are age, education, 1974 real income, 1975 real income and dummy variables for race, marital status, and academic achievement. The 1974 income was added by Dehejia and Wahba (1999), which resulted in a smaller sample size than the one used by Lalonde (1986), because that year income was not available for all participants. In the original sample, the treatment has been randomly assigned and Dehejia and Wahba (1999) argue that the missing observations did not affect the randomization. Table 12 compares the sample means of the covariates, where the income variables re75 and re74 are expressed in thousands of dollars of 1982. None of them are statistically different at 5% except "nodegree".

Because of random assignment, the difference in the average 1978 income between the two groups is a consistent estimate of the ATT. There is no need to set the moments to the ones from the treated group, but assuming perfect balancing (moment conditions (2)) may result in more efficient estimates. Table 13 presents the results[9]. The first column is OLS regression

[9]The treatment effect is expressed in thousands, so we need to multiply it by one thousand to compare our estimates with Lalonde (1986) or Dehejia and Wahba (2002).

Table 8: Properties of ATE estimation in the third study, n=200

	Bias	RMSE	SD	Coverage	Bias.sd
ET4	-1.7300	4.7947	4.4722	0.9052	-0.3480
ET8	-0.9562	4.7974	4.7017	0.9092	-0.5204
ET14	-0.7247	4.9385	4.8856	0.8926	-0.8036
ET4 Bal Sample	**0.4659**^{4}	4.3363	4.3116	0.9364	-0.2374
ET8 Bal Sample	-1.2680	4.3781	4.1909	0.9090	-0.3597
ET14 Bal Sample	**-0.3856**^{3}	4.1812	4.1638	**0.9395**^{2}	-0.1113
ET4 Bal Pop	0.6245	**2.5371**^{3}	2.4592	0.8692	-0.4659
ET8 Bal Pop	-1.5028	2.8254	**2.3929**^{4}	0.7269	-0.7909
ET14 Bal Pop	-0.5289	**2.2954**^{2}	**2.2338**^{2}	0.6524	-1.1239
EL4	**0.3478**^{1}	4.5754	4.5627	**0.9386**^{4}	**-0.0894**^{3}
EL8	**0.3854**^{2}	4.5588	4.5430	**0.9408**^{1}	**-0.0765**^{2}
EL14	0.6389	4.6130	4.5690	**0.9394**^{3}	**-0.0721**^{1}
EL4 Bal Sample	-0.7884	4.3777	4.3065	0.9292	-0.2254
EL8 Bal Sample	-1.2821	4.3126	4.1180	0.9124	-0.2796
EL14 Bal Sample	-0.5376	3.9591	3.9228	0.9386	**0.0958**^{4}
EL4 Bal Pop	-0.6590	**2.5909**^{4}	2.5059	0.8880	-0.4121
EL8 Bal Pop	-1.5017	2.7666	**2.3238**^{3}	0.7384	-0.7043
EL14 Bal Pop	-0.6516	**2.2920**^{1}	**2.1976**^{1}	0.6526	-1.1188

Bold numbers are the best methods and the superscripts represent their rank

Table 9: Properties of ATE estimation in the third study, n=1000

	Bias	RMSE	SD	Coverage	Bias.sd
ET4	-1.5500	2.5112	1.9760	0.8670	-0.0619
ET8	-0.9111	2.2328	2.0387	0.9174	-0.0622
ET14	-0.6934	2.2072	2.0957	0.9264	-0.1060
ET4 Bal Sample	**-0.0961**^{2}	1.9479	1.9457	**0.9372**^{3}	-0.0814
ET8 Bal Sample	-0.7894	1.9822	1.8184	0.9164	-0.0656
ET14 Bal Sample	**-0.2615**^{4}	1.8586	1.8403	**0.9410**^{2}	**-0.0579**^{4}
ET4 Bal Pop	**-0.0570**^{1}	**1.2000**^{3}	1.1988	0.9090	-0.1817
ET8 Bal Pop	-0.8431	**1.2463**^{4}	**0.9180**^{4}	0.7848	-0.1245
ET14 Bal Pop	-0.2940	**0.9361**^{1}	**0.8889**^{3}	0.8146	-0.2423
EL4	1.1426	2.4420	2.1584	**0.9538**^{1}	0.2432
EL8	**0.2487**^{3}	2.0160	2.0008	0.9684	0.2576
EL14	0.6233	2.0270	1.9290	0.9790	0.4580
EL4 Bal Sample	-2.1600	2.9673	2.0346	0.8008	-0.0752
EL8 Bal Sample	-1.0029	2.0358	1.7718	0.9134	**-0.0001**^{1}
EL14 Bal Sample	-0.6764	1.9071	1.7833	**0.9330**^{4}	**-0.0097**^{2}
EL4 Bal Pop	-2.1306	2.5348	1.3733	0.6256	-0.1317
EL8 Bal Pop	-1.0522	1.3196	**0.7965**^{2}	0.7666	**0.0392**^{3}
EL14 Bal Pop	-0.7060	**1.0510**^{2}	**0.7787**^{1}	0.7424	-0.1386

Bold numbers are the best methods and the superscripts represent their rank

of re78 on the treatment indicator and the second is OLS in which the covariates from Tables 12 are controlled for. The columns 3 to 5 are the results for EL, ET and CUE with $u_k = \{age, educ, black, hispan, married, nodegree, re74, re75\}$ (e.g. ET8 stands for ET with 8 balancing moments) and the last three are the results when $u_k = \{age, age^2, educ, educ^2, black, hispan, married, nodegree, re74, re74^2, re75, re75^2\}$. The coefficients of the covariates in the second column have been omitted. We can see that balancing using more moments reduces the standard errors of the estimated ATT. EL seems to be the most efficient, but the difference is small. We also obtain a lower ATT when balancing the groups. It is possible that removing observations to introduce the 1974 income had an impact on the randomization of the experiment, and that forcing the covariates to be perfectly balanced produces a more accurate estimate. The p-values of the overidentification test are between 0.06 and 0.12, which does suggest that the groups may not be well balanced.

To follow Dehejia and Wahba (2002), we drop the control group from the National Supported Work (NSW) and use either the Population Survey of Income Dynamics (PSID) or Current Population Survey (CPS) sample as the control group and use GEL to balance the covariates with the treated group. Since the method that matches the moments to the treated is the same as the method proposed by Chan et al. (2016), we also look at the PSID2, CPS2, PSID3 and CPS3, which are subsamples built by Lalonde (1986) in an attempt to build samples with individuals who experience similar economic conditions as the ones from the treated groups. The same two

Table 10: Properties of ATT estimation in the third study, n=200

	Bias	RMSE	SD	Coverage	Bias.sd
ET4 Bal Sample	2.4917	5.7173	5.1463	0.9170	**-0.0008**[1]
ET8 Bal Sample	2.7697	5.5505	4.8106	**0.9177**[4]	0.2808
ET14 Bal Sample	3.8605	5.8275	4.3665	0.9149	0.8457
ET4 Bal Pop	2.5486	3.1087	**1.7802**[3]	0.6396	**-0.1491**[3]
ET8 Bal Pop	1.9487	2.7735	1.9737	0.6832	-0.4785
ET14 Bal Pop	**1.0101**[2]	**2.0072**[2]	**1.7349**[2]	0.7365	-0.5885
EL4 Bal Sample	**1.2001**[4]	5.4129	5.2787	**0.9242**[3]	-0.2951
EL8 Bal Sample	2.3105	5.2906	4.7598	**0.9256**[2]	0.2571
EL14 Bal Sample	1.5186	3.6463	3.3154	**0.9636**[1]	1.6649
EL4 Bal Pop	**1.1024**[3]	**2.1154**[3]	**1.8056**[4]	0.8846	**-0.0673**[2]
EL8 Bal Pop	1.6468	**2.5869**[4]	1.9952	0.7310	-0.4981
EL14 Bal Pop	**0.4573**[1]	**1.4718**[1]	**1.3991**[1]	0.8578	**-0.1785**[4]

Bold numbers are the best methods and the superscripts represent their rank

Table 11: Properties of ATT estimation in the third study, n=1000

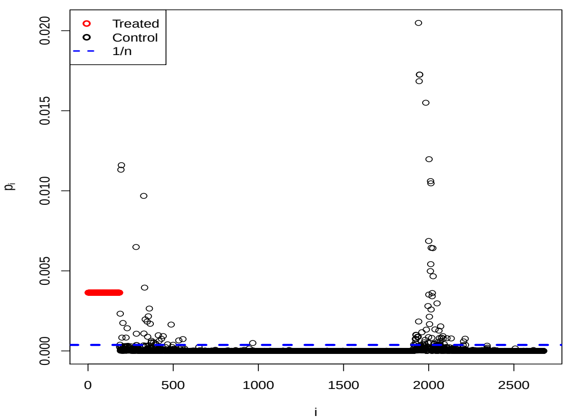
	Bias	RMSE	SD	Coverage	Bias.sd
ET4 Bal Sample	2.1923	3.1796	2.3031	0.8416	0.0136
ET8 Bal Sample	1.8540	2.9442	2.2873	0.8694	0.0103
ET14 Bal Sample	1.1463	2.5956	2.3290	**0.9216**[4]	**0.0074**[2]
ET4 Bal Pop	2.1837	2.3104	0.7547	0.1734	**-0.0004**[1]
ET8 Bal Pop	1.7425	1.8888	**0.7290**[4]	0.2960	-0.0402
ET14 Bal Pop	1.0736	**1.2619**[3]	**0.6633**[1]	0.5658	-0.0720
EL4 Bal Sample	**0.2998**[2]	2.4552	2.4371	**0.9374**[2]	-0.1143
EL8 Bal Sample	1.3029	2.6451	2.3021	0.9120	**-0.0082**[3]
EL14 Bal Sample	**0.9566**[4]	2.5010	2.3111	**0.9314**[3]	**0.0091**[4]
EL4 Bal Pop	**0.2624**[1]	**0.9189**[1]	0.8807	**0.9522**[1]	0.0902
EL8 Bal Pop	1.1913	**1.3745**[4]	**0.6856**[3]	0.6110	0.0219
EL14 Bal Pop	**0.8783**[3]	**1.1081**[2]	**0.6756**[2]	0.6690	-0.0828

Bold numbers are the best methods and the superscripts represent their rank

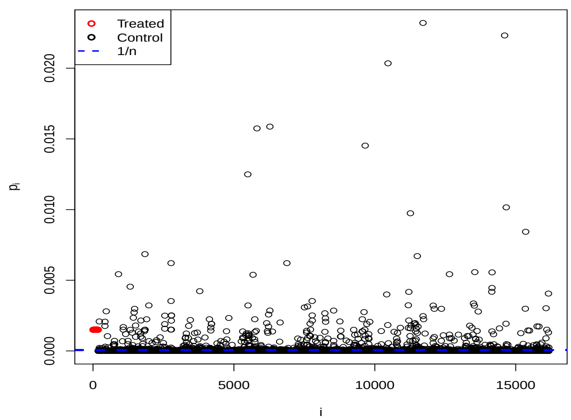
sets of covariates are used (8 and 12). Tables 14 to 16 present the results for ET, EL and CUE with 8 balancing moments and Tables 17 to 19 show the results with 12 balancing moments.

The coefficient estimates $\hat{\theta}_1$ to $\hat{\theta}_3$ are the ones from the moment conditions (16). Therefore, $\hat{\theta}_2$ is the treatment effect estimate and $\hat{\theta}_3$ is the estimated probability of being assigned to treatment using the implied probabilities as probability distribution function. Notice that the ET and CUE estimates of θ_3 are much higher then the actual proportions of treated, which are respectively 7.07%, 42.24% and 59.11% for the PSID, PSID2 and PSID3, and 1.14%, 7.24% and 30.13% for the CPS, CPS2 and CPS3. Figure 1 shows how the implied probabilities are estimated by ET and EL in order to match the moments of the treated. Although both ET and EL balance the moments perfectly, they assign probabilities in a very different way. For ET, many observations that belong to the control group have probabilities lower than $1/n$, and the assigned probabilities for the treated are constant and much higher than $1/n$. On the other hand, the probabilities assigned by EL on the treated is much closer to $1/n$, which explains why EL estimates of θ_3 are much closer to the proportions of treated.

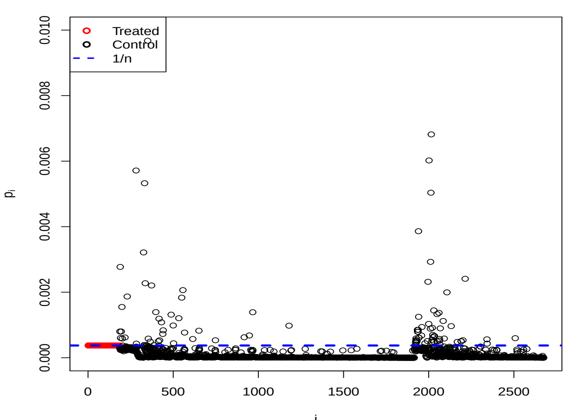
If we consider the treatment effect estimates, CUE and EL seem much more unstable than ET as we change the control group. As we saw in Section 3, the CUE implied probabilities need to be transformed in order to avoid negative values, and these transformed implied probabilities do not balance the moments perfectly. This property may explain the instability that we observe here. With respect to EL, its instability may come from the fact that it is not a method that is robust to misspecification. The method that produces the best results is clearly ET with the full PSID as control group. In fact, it produces an estimate of the ATT that is very close to the experimental estimate, with standard errors smaller than any method analyzed by Dehejia and Wahba (2002). All three methods, however, underestimate the ATT when the CPS is used as the control group.



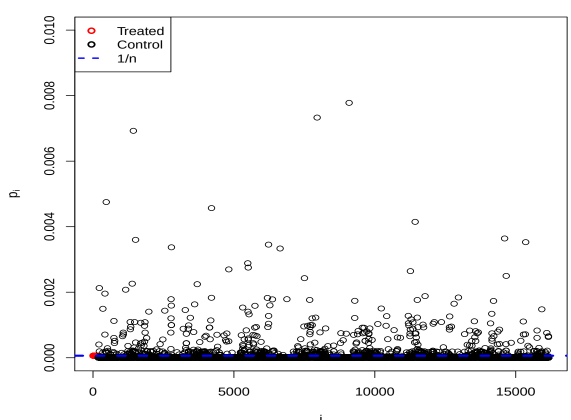
(a) ET-PSID



(b) ET-CPS



(c) EL-PSID



(d) EL-CPS

Figure 1: Implied probabilities using either PSID or CPS as control group

Table 12: Sample mean of the covariates for each group

	N	age	educ	black	hispan	married	nodegree	re74	re75
Control	260	25.054	10.088	0.827	0.108	0.154	0.835	2.107	1.267
Treated	185	25.816	10.346	0.843	0.059	0.189	0.708	2.096	1.532

Table 13: Estimated ATT using OLS, OLS with control variables, EL, ET and CUE with unconditional balancing

	OLS	OLScont	EL8	ET8	CUE8	EL12	ET12	CUE12
$\hat{\theta}_1$	4.5548***	0.7851	4.5519***	4.5368***	4.5230***	4.5607***	4.5397***	4.5253***
	(0.3407)	(3.3546)	(0.3375)	(0.3363)	(0.3356)	(0.3382)	(0.3386)	(0.3389)
$\hat{\theta}_2$	1.7943**	1.6763*	1.6374*	1.6294*	1.6266*	1.6279*	1.6033*	1.5835*
	(0.6727)	(0.6853)	(0.6710)	(0.6730)	(0.6761)	(0.6642)	(0.6643)	(0.6677)
$\hat{\theta}_3$			0.4157***	0.4127***	0.4085***	0.4157***	0.4112***	0.4050***
			(0.0243)	(0.0243)	(0.0242)	(0.0248)	(0.0247)	(0.0244)
R^2	0.0178	0.0548						
Adj. R^2	0.0156	0.0353						
Num. obs.	445	445	445	445	445	445	445	445
RMSE	6.5795	6.5135						
LR test			17.2582	17.3217	16.7238	24.1274	23.7643	22.0489
LR test p-value			0.1005	0.0987	0.1163	0.0630	0.0692	0.1065

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. For OLS methods, the standard errors are robust to heteroscedasticity.

Table 14: Estimated ATT by ET using the three different PSID and CPS as control groups and 8 covariates

	ET-PSID	ET-PSID2	ET-PSID3	ET-CPS	ET-CPS2	ET-CPS3
$\hat{\theta}_1$	4.2886***	3.8401***	3.9599***	5.0808***	5.2049***	5.0730***
	(0.6267)	(0.6971)	(0.8671)	(0.3272)	(0.3894)	(0.5597)
$\hat{\theta}_2$	2.0612*	2.5095**	2.3897*	1.2675*	1.1427	1.2750
	(0.8249)	(0.8833)	(1.0254)	(0.6449)	(0.6782)	(0.7894)
$\hat{\theta}_3$	0.6500***	0.7855***	0.8298***	0.1789***	0.3293***	0.5648***
	(0.0506)	(0.0398)	(0.0350)	(0.0229)	(0.0321)	(0.0370)
LR test	4780.8155	404.9884	180.1020	30285.3152	3984.6115	572.8247
LR test p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Num. obs.	2675	438	313	16177	2554	614

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 15: Estimated ATT by CUE using the three different PSID and CPS as control groups and 8 covariates

	CUE-PSID	CUE-PSID2	CUE-PSID3	CUE-CPS	CUE-CPS2	CUE-CPS3
$\hat{\theta}_1$	5.5623***	4.2318***	5.1145***	5.6658***	5.1427***	4.7004***
	(0.6780)	(1.0611)	(1.0435)	(0.2634)	(0.4002)	(0.8687)
$\hat{\theta}_2$	0.7858	2.1175	1.2345	0.6847	1.2063	1.6484
	(0.8956)	(1.2018)	(1.1934)	(0.6213)	(0.6852)	(1.0174)
$\hat{\theta}_3$	0.3567***	0.7320***	0.8318***	0.1176***	0.3496***	0.6174***
	(0.0217)	(0.0282)	(0.0253)	(0.0086)	(0.0217)	(0.0288)
LR test	2156.3921	185.2555	90.5904	14603.4424	2024.9106	314.3652
LR test p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Num. obs.	2675	438	313	16177	2554	614

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 16: Estimated ATT by EL using the three different PSID and CPS as control groups and 8 covariates

	EL-PSID	EL-PSID2	EL-PSID3	EL-CPS	EL-CPS2	EL-CPS3
$\hat{\theta}_1$	3.3314***	2.8754***	2.5785***	4.9415***	5.0369***	5.0362***
	(0.3730)	(0.3885)	(0.6361)	(0.4770)	(0.4414)	(0.5509)
$\hat{\theta}_2$	3.0177***	3.4737***	3.7706***	1.4077	1.3123	1.3130
	(0.6498)	(0.6626)	(0.8402)	(0.7347)	(0.7097)	(0.7836)
$\hat{\theta}_3$	0.0692*	0.4224***	0.5910***	0.0114***	0.0724***	0.3013***
	(0.0302)	(0.0838)	(0.0736)	(0.0020)	(0.0135)	(0.0414)
LR test	10500.4382	715.9939	272.3546	63046.2390	7610.4213	973.1549
LR test p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Num. obs.	2675	438	313	16177	2554	614

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 17: Estimated ATT by ET using the three different PSID and CPS as control groups and 12 covariates

	ET-PSID	ET-PSID2	ET-PSID3	ET-CPS	ET-CPS2	ET-CPS3
$\hat{\theta}_1$	4.5858***	4.3368***	4.7357***	5.0036***	4.8370***	5.2799***
	(0.6752)	(0.7302)	(0.7893)	(0.4445)	(0.5476)	(0.6325)
$\hat{\theta}_2$	1.7611*	2.0122*	1.6136	1.3449	1.5134*	1.0702
	(0.8564)	(0.9160)	(0.9499)	(0.7108)	(0.7679)	(0.8178)
$\hat{\theta}_3$	0.6735***	0.8142***	0.8677***	0.2770***	0.5006***	0.6866***
	(0.0488)	(0.0370)	(0.0340)	(0.0375)	(0.0517)	(0.0486)
LR test	4800.5340	421.5520	199.5875	31018.3559	4368.8791	689.0368
LR test p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Num. obs.	2675	438	313	16177	2554	614

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 18: Estimated ATT by CUE using the three different PSID and CPS as control groups and 12 covariates

	CUE-PSID	CUE-PSID2	CUE-PSID3	CUE-CPS	CUE-CPS2	CUE-CPS3
$\hat{\theta}_1$	4.8098***	3.6592*	4.0007**	5.3230***	5.4828	4.9784***
	(0.6939)	(1.4888)	(1.3665)	(0.2796)	(24.4489)	(0.6620)
$\hat{\theta}_2$	1.5380	2.6895	2.3485	1.0269	0.8657	1.3714
	(0.8725)	(1.6263)	(1.4635)	(0.6287)	(24.4668)	(0.8393)
$\hat{\theta}_3$	0.3897***	0.7558***	0.8490***	0.1316***	0.3963***	0.6637***
	(0.0229)	(0.0278)	(0.0246)	(0.0094)	(0.0235)	(0.0290)
LR test	2200.2249	193.2103	95.1032	14771.5064	2087.1901	335.2450
LR test p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Num. obs.	2675	438	313	16177	2554	614

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 19: Estimated ATT by EL using the three different PSID and CPS as control groups and 12 covariates

	EL-PSID	EL-PSID2	EL-PSID3	EL-CPS	EL-CPS2	EL-CPS3
$\hat{\theta}_1$	3.8445***	4.8756***	5.6087***	3.8795***	5.6018***	6.1149***
	(0.5309)	(0.7809)	(0.6920)	(0.4548)	(0.6550)	(0.6549)
$\hat{\theta}_2$	2.5047**	1.4736	0.7404	2.4697***	0.7474	0.2342
	(0.7689)	(0.9515)	(0.8671)	(0.6978)	(0.8600)	(0.8323)
$\hat{\theta}_3$	0.0692*	0.4224***	0.5910***	0.0114**	0.0724**	0.3013***
	(0.0292)	(0.0854)	(0.1099)	(0.0044)	(0.0247)	(0.0786)
LR test	14377.8726	868.4966	422.1832	75996.6112	8152.4069	1056.4872
LR test p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Num. obs.	2675	438	313	16177	2554	614

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

5 Conclusion and potential extensions

We presented a GEL method to estimate treatment effects that is valid in both randomized trials and observational studies. By selecting the appropriate estimating equations, we are able to exactly balance the treated and control groups with respect to observed characteristics. The balancing is non-parametric and does not rely on the estimation of a propensity score. Also, being part of the GEL family of estimators, we can use tools for inference and re-sampling that have been developed and tested for GEL.

We studied the small sample properties of our family of methods using three sets of simulation experiments. The first two sets are used by Frolich (2004) and Busso et al. (2014) to compare the properties of different ATT estimation methods. The first set is based on DGP's that imply correctly specified propensity score models, which are likely to favor any methods that rely on its estimation. For this set of DGP's, ET is the best GEL method in terms of bias and variance, for most DGP's. In comparison with the methods analyzed by Busso et al. (2014), ET has a bias comparable to the methods that are not bias-corrected when the strict overlap assumption is violated, and is comparable to the bias-corrected methods when overlap is good. Also, ET competes well with the best methods studied in terms of efficiency. In the second set of DPG's in which the propensity score models are misspecified, ET performs as well as the best methods in terms of both the bias and variance, when the number of balancing moments is high enough. In a final simulation study, we considered the DPG's from Chan et al. (2016). This experiment is used to compare all of our proposed methods applied to the estimation of both the ATT and ATE. Overall, ET and EL are comparable, but ET seems to have properties that are more stable across different DGP's. It is worth noticing that CUE failed to balance the moments with strictly positive implied probabilities in a large proportion of the iterations for most DGP's studied. We therefore suggest to avoid it.

In the case of observational studies, the estimating equations are misspecified. Although some have studied the properties of GEL in misspecified models, we do not know what happens in such models if we let the number of estimating equations grow with the sample size. This extension would be of interest since the semi-parametric efficiency bound can only be reached when k grows with n. Another extension would be to derive specification tests for constrained GEL. We do have a mean to test hypotheses on coefficients using the asymptotic distribution, but it would be a good contribution to derive a non-parametric confidence interval like the one we can construct by inverting the LR test of GEL.

A Detailed tables

Table 20: 1000×Bias for all models and $u_k(X_i) = X_i$

		ET				EL			
			Bal Sample	ATT	Bal Pop		Bal Sample	ATT	Bal Pop
$m_1(x)$	$\{\alpha,\beta\}_1$	-1.678	-2.676	-5.283	-5.721	**-1.162**	-3.665	6.918	6.337
	$\{\alpha,\beta\}_2$	1.963	1.372	**0.5761**	0.6332	3.216	1.917	4.322	3.913
	$\{\alpha,\beta\}_3$	0.9221	0.7769	**0.7129**	0.7281	1.152	0.8963	1.503	1.33
	$\{\alpha,\beta\}_4$	**-0.2022**	-0.2902	1.745	2.035	-0.5872	-0.568	5.133	5.006
	$\{\alpha,\beta\}_5$	**-0.2776**	-1.1	-2.38	-2.46	-2.165	-2.83	-3.224	-3.501
$m_2(x)$	$\{\alpha,\beta\}_1$	-4.028	-8.152	-14.75	-15.65	**3.413**	-4.694	47.26	48.3
	$\{\alpha,\beta\}_2$	7.512	5.076	1.49	**1.471**	13.17	7.823	20.51	17.61
	$\{\alpha,\beta\}_3$	-0.4059	**-0.353**	-0.5081	-0.6972	-0.5614	-0.5391	-10.18	-9.142
	$\{\alpha,\beta\}_4$	0.1954	0.1013	1.549	1.849	**-0.06534**	-0.4766	3.924	3.915
	$\{\alpha,\beta\}_5$	1.147	3.972	8.827	9.477	**0.2474**	1.005	-2.144	-1.196
$m_3(x)$	$\{\alpha,\beta\}_1$	-3.639	-7.059	-14.74	-15.39	**-0.1439**	-7.795	-13.2	-15.68
	$\{\alpha,\beta\}_2$	3.601	2.417	0.695	0.7367	6.127	3.632	0.4245	**0.01486**
	$\{\alpha,\beta\}_3$	1.283	1.036	0.8819	**0.8183**	1.642	1.247	1.118	0.8249
	$\{\alpha,\beta\}_4$	-13.99	-13.04	-11.13	-11.8	-7.33	**-3.423**	-22.46	-21.61
	$\{\alpha,\beta\}_5$	-10.2	-10.71	-12.03	-12.05	-6.842	-5.185	**-3.07**	-3.423
$m_4(x)$	$\{\alpha,\beta\}_1$	-0.36	-0.2861	-6.529	-6.562	-0.3573	**-0.1667**	18.3	18.78
	$\{\alpha,\beta\}_2$	-0.3933	-0.199	-0.03093	**0.01627**	-0.6408	-0.2624	7.045	7.233
	$\{\alpha,\beta\}_3$	-0.3894	-0.3434	-0.326	**-0.2898**	-0.4737	-0.3975	0.4523	0.4289
	$\{\alpha,\beta\}_4$	2.383	2.563	2.471	2.772	**1.388**	2.151	5.693	5.743
	$\{\alpha,\beta\}_5$	-7.735	-7.55	-7.765	-7.578	-4.597	**-3.345**	-4.727	-4.549
$m_5(x)$	$\{\alpha,\beta\}_1$	**0.5589**	0.766	-2.21	-2.337	1.567	2.304	26.15	25.92
	$\{\alpha,\beta\}_2$	-0.09461	**-0.06491**	-0.6103	-0.4352	-0.1585	-0.341	7.002	6.909
	$\{\alpha,\beta\}_3$	-0.1674	-0.1135	-0.1306	-0.1608	-0.2064	**-0.03529**	1.004	1.041
	$\{\alpha,\beta\}_4$	**1.897**	2.072	3.126	3.466	2.066	3.945	8.776	8.559
	$\{\alpha,\beta\}_5$	-3.197	-1.919	-0.7615	**0.1104**	-2.079	-1.232	-6.23	-5.776
$m_6(x)$	$\{\alpha,\beta\}_1$	-4.882	-10.18	-21.79	-22.51	4.522	**-4.016**	-18.16	-19.41
	$\{\alpha,\beta\}_2$	7.815	5.122	2.55	**2.499**	14.1	9.809	5.999	5.201
	$\{\alpha,\beta\}_3$	5.18	4.181	3.519	3.65	6.667	5.039	0.5197	**0.1647**
	$\{\alpha,\beta\}_4$	7.123	8.753	**-2.209**	-2.564	6.719	9.538	-6.298	-5.142
	$\{\alpha,\beta\}_5$	**-0.349**	-1.557	-4.046	-3.919	-1.266	-3.573	-5.832	-6.096

Table 21: $n\times$Variance for all models and $u_k(X_i) = X_i$

		ET				EL			
			Bal Sample	ATT	Bal Pop		Bal Sample	ATT	Bal Pop
$m_1(x)$	$\{\alpha,\beta\}_1$	0.06773	**0.0659**	0.1114	0.1116	0.09343	0.08799	0.1918	0.1959
	$\{\alpha,\beta\}_2$	0.05063	**0.05038**	0.06657	0.06504	0.05715	0.05548	0.09562	0.09227
	$\{\alpha,\beta\}_3$	0.04428	**0.04427**	0.04789	0.04777	0.04503	0.04492	0.05291	0.05111
	$\{\alpha,\beta\}_4$	0.09748	0.09701	**0.07429**	0.07663	0.1042	0.1134	0.08896	0.08712
	$\{\alpha,\beta\}_5$	0.09814	0.09868	0.119	0.1172	**0.09757**	0.09974	0.1298	0.1273
$m_2(x)$	$\{\alpha,\beta\}_1$	0.2009	**0.1978**	0.4211	0.4382	0.249	0.2578	0.3426	0.3076
	$\{\alpha,\beta\}_2$	0.1494	**0.1463**	0.3093	0.3046	0.1814	0.1744	0.4436	0.4628
	$\{\alpha,\beta\}_3$	0.06973	0.06988	0.07781	0.07879	0.07464	**0.06947**	0.08378	0.08106
	$\{\alpha,\beta\}_4$	0.09527	0.09446	**0.07514**	0.07695	0.1006	0.1119	0.08954	0.08692
	$\{\alpha,\beta\}_5$	0.194	0.1895	0.2012	0.199	0.2005	0.1866	0.186	**0.1843**
$m_3(x)$	$\{\alpha,\beta\}_1$	0.1323	**0.1282**	0.1637	0.161	0.2198	0.1756	0.2671	0.2736
	$\{\alpha,\beta\}_2$	0.06655	**0.06598**	0.07654	0.0744	0.0792	0.0733	0.1104	0.1068
	$\{\alpha,\beta\}_3$	**0.04482**	0.04484	0.04785	0.04733	0.04594	0.04567	0.05309	0.05119
	$\{\alpha,\beta\}_4$	0.1198	0.1166	**0.0915**	0.09474	0.1246	0.12	0.1024	0.0995
	$\{\alpha,\beta\}_5$	**0.1087**	0.1121	0.1341	0.1326	0.1136	0.113	0.1381	0.1399
$m_4(x)$	$\{\alpha,\beta\}_1$	**0.08753**	0.09153	0.1338	0.1346	0.1439	0.1109	0.2394	0.2418
	$\{\alpha,\beta\}_2$	**0.05452**	0.05491	0.06533	0.0643	0.06176	0.05806	0.08809	0.08549
	$\{\alpha,\beta\}_3$	**0.04347**	0.04351	0.04693	0.04629	0.04391	0.04389	0.05028	0.04879
	$\{\alpha,\beta\}_4$	0.09447	0.09442	**0.07384**	0.07661	0.09907	0.09906	0.08687	0.08451
	$\{\alpha,\beta\}_5$	**0.1046**	0.1053	0.1231	0.1204	0.1106	0.1077	0.1248	0.1272
$m_5(x)$	$\{\alpha,\beta\}_1$	**0.09403**	0.09828	0.1608	0.1632	0.145	0.1212	0.284	0.2915
	$\{\alpha,\beta\}_2$	**0.06084**	0.06111	0.07734	0.07636	0.07162	0.06634	0.1138	0.1098
	$\{\alpha,\beta\}_3$	**0.04861**	0.04869	0.05249	0.05247	0.04909	0.04893	0.05727	0.05569
	$\{\alpha,\beta\}_4$	0.09811	0.09752	**0.07821**	0.08031	0.1025	0.1003	0.09325	0.08916
	$\{\alpha,\beta\}_5$	0.1262	**0.126**	0.142	0.1428	0.1308	0.1304	0.1374	0.137
$m_6(x)$	$\{\alpha,\beta\}_1$	**0.307**	0.3172	0.3179	0.3236	0.3745	0.3197	0.3851	0.3898
	$\{\alpha,\beta\}_2$	**0.2135**	0.2157	0.218	0.2163	0.233	0.2166	0.2244	0.225
	$\{\alpha,\beta\}_3$	**0.1357**	0.1358	0.1432	0.1396	0.1428	0.1361	0.1649	0.1576
	$\{\alpha,\beta\}_4$	0.1129	0.1147	**0.08708**	0.08961	0.1201	0.1329	0.09277	0.09338
	$\{\alpha,\beta\}_5$	**0.1041**	0.1055	0.1264	0.1251	0.1098	0.1099	0.134	0.1344

Table 22: $n\times$MSE for all models and $u_k(X_i) = X_i$

		ET				EL			
			Bal Sample	ATT	Bal Pop		Bal Sample	ATT	Bal Pop
$m_1(x)$	$\{\alpha,\beta\}_1$	0.06801	**0.06661**	0.1142	0.1149	0.09355	0.08932	0.1965	0.1999
	$\{\alpha,\beta\}_2$	0.05101	**0.05056**	0.0666	0.06507	0.05817	0.05585	0.09748	0.09379
	$\{\alpha,\beta\}_3$	0.04436	**0.04433**	0.04794	0.04781	0.04516	0.04499	0.05313	0.05128
	$\{\alpha,\beta\}_4$	0.09748	0.09701	**0.07459**	0.07704	0.1042	0.1134	0.09159	0.08961
	$\{\alpha,\beta\}_5$	0.09814	0.09879	0.1195	0.1178	**0.09803**	0.1005	0.1308	0.1285
$m_2(x)$	$\{\alpha,\beta\}_1$	**0.2025**	0.2044	0.4428	0.4626	0.2501	0.26	0.5659	0.5409
	$\{\alpha,\beta\}_2$	0.1551	**0.1489**	0.3095	0.3048	0.1987	0.1805	0.4856	0.4937
	$\{\alpha,\beta\}_3$	0.06974	0.06989	0.07783	0.07883	0.07466	**0.0695**	0.09413	0.08941
	$\{\alpha,\beta\}_4$	0.09527	0.09445	**0.07537**	0.07728	0.1006	0.1119	0.09107	0.08844
	$\{\alpha,\beta\}_5$	0.1941	0.191	0.209	0.2079	0.2004	0.1867	0.1865	**0.1845**
$m_3(x)$	$\{\alpha,\beta\}_1$	0.1336	**0.1332**	0.1854	0.1847	0.2198	0.1817	0.2845	0.2981
	$\{\alpha,\beta\}_2$	0.06784	**0.06656**	0.07659	0.07444	0.08294	0.07461	0.1104	0.1068
	$\{\alpha,\beta\}_3$	0.04498	**0.04494**	0.04792	0.04739	0.0462	0.04582	0.05321	0.05126
	$\{\alpha,\beta\}_4$	0.1394	0.1336	**0.1039**	0.1087	0.1299	0.1211	0.1528	0.1462
	$\{\alpha,\beta\}_5$	0.1191	0.1235	0.1485	0.1471	0.1182	**0.1157**	0.139	0.141
$m_4(x)$	$\{\alpha,\beta\}_1$	**0.08753**	0.09153	0.1381	0.1389	0.1439	0.1109	0.2728	0.277
	$\{\alpha,\beta\}_2$	**0.05453**	0.05491	0.06533	0.06429	0.0618	0.05806	0.09305	0.09071
	$\{\alpha,\beta\}_3$	**0.04348**	0.04352	0.04694	0.04629	0.04393	0.0439	0.0503	0.0488
	$\{\alpha,\beta\}_4$	0.09503	0.09506	**0.07444**	0.07737	0.09925	0.09951	0.0901	0.0878
	$\{\alpha,\beta\}_5$	0.1106	0.111	0.1291	0.1261	0.1127	**0.1088**	0.127	0.1292
$m_5(x)$	$\{\alpha,\beta\}_1$	**0.09405**	0.09833	0.1613	0.1637	0.1452	0.1218	0.3523	0.3587
	$\{\alpha,\beta\}_2$	**0.06083**	0.06111	0.07737	0.07637	0.07162	0.06634	0.1187	0.1145
	$\{\alpha,\beta\}_3$	**0.04861**	0.04869	0.05249	0.05247	0.04909	0.04892	0.05737	0.05579
	$\{\alpha,\beta\}_4$	0.09846	0.09794	**0.07918**	0.0815	0.1029	0.1019	0.1009	0.09647
	$\{\alpha,\beta\}_5$	0.1272	**0.1264**	0.142	0.1428	0.1312	0.1305	0.1413	0.1403
$m_6(x)$	$\{\alpha,\beta\}_1$	**0.3094**	0.3275	0.3654	0.3742	0.3765	0.3213	0.418	0.4275
	$\{\alpha,\beta\}_2$	0.2196	0.2183	0.2186	**0.2169**	0.2528	0.2262	0.2279	0.2277
	$\{\alpha,\beta\}_3$	0.1384	**0.1376**	0.1444	0.1409	0.1472	0.1386	0.1649	0.1576
	$\{\alpha,\beta\}_4$	0.118	0.1224	**0.08756**	0.09026	0.1246	0.142	0.09672	0.09601
	$\{\alpha,\beta\}_5$	**0.1041**	0.1057	0.128	0.1266	0.1099	0.1112	0.1373	0.1381

Table 23: Coverage of 95% confidence intervals for all models and $u_k(X_i) = X_i$

		ET				EL			
			Bal Sample	ATT	Bal Pop		Bal Sample	ATT	Bal Pop
	$\{\alpha,\beta\}_1$	**0.9173**	0.9163	0.8688	0.847	0.8682	0.8666	0.6768	0.6559
	$\{\alpha,\beta\}_2$	**0.9355**	0.9346	0.9139	0.9094	0.9248	0.9236	0.8404	0.8404
$m_1(x)$	$\{\alpha,\beta\}_3$	**0.9359**	0.9353	0.9357	0.9303	0.9349	0.9339	0.9224	0.9218
	$\{\alpha,\beta\}_4$	0.8847	0.8814	**0.9321**	0.9193	0.8587	0.835	0.9098	0.8998
	$\{\alpha,\beta\}_5$	**0.884**	0.8793	0.8501	0.837	0.8765	0.8694	0.8184	0.8103
	$\{\alpha,\beta\}_1$	**0.9235**	0.9163	0.8995	0.8875	0.8838	0.8728	0.7175	0.7288
	$\{\alpha,\beta\}_2$	**0.9253**	0.9236	0.8909	0.8692	0.9047	0.9014	0.7722	0.7509
$m_2(x)$	$\{\alpha,\beta\}_3$	0.9403	0.9397	0.9366	0.9335	0.9363	**0.9436**	0.9243	0.9271
	$\{\alpha,\beta\}_4$	0.8878	0.8848	**0.9267**	0.9162	0.8619	0.832	0.9046	0.8998
	$\{\alpha,\beta\}_5$	**0.9057**	0.9019	0.8775	0.8561	0.8948	0.8947	0.8738	0.87
	$\{\alpha,\beta\}_1$	**0.9043**	0.8945	0.8398	0.8052	0.8362	0.851	0.6786	0.6434
	$\{\alpha,\beta\}_2$	**0.9315**	0.9302	0.915	0.9062	0.9138	0.9164	0.8368	0.8374
$m_3(x)$	$\{\alpha,\beta\}_3$	**0.9416**	0.9414	0.9389	0.9362	0.9384	0.9392	0.9253	0.925
	$\{\alpha,\beta\}_4$	0.8505	0.8553	**0.9097**	0.8948	0.8562	0.8661	0.8612	0.8632
	$\{\alpha,\beta\}_5$	**0.8678**	0.8605	0.8337	0.8217	0.8542	0.86	0.8244	0.8129
	$\{\alpha,\beta\}_1$	**0.9177**	0.9137	0.8565	0.8395	0.8596	0.9018	0.6884	0.6577
	$\{\alpha,\beta\}_2$	**0.932**	0.9299	0.9183	0.913	0.9201	0.9223	0.8551	0.8506
$m_4(x)$	$\{\alpha,\beta\}_3$	**0.9423**	0.9423	0.9371	0.9374	0.9421	0.9406	0.9279	0.9288
	$\{\alpha,\beta\}_4$	0.8885	0.8831	**0.9332**	0.9209	0.8669	0.8648	0.9093	0.902
	$\{\alpha,\beta\}_5$	0.8721	0.8689	0.8492	0.8422	0.8598	**0.8723**	0.8409	0.8313
	$\{\alpha,\beta\}_1$	**0.9217**	0.915	0.8674	0.8513	0.8694	0.8968	0.6606	0.6427
	$\{\alpha,\beta\}_2$	**0.9353**	0.9339	0.9173	0.9107	0.9174	0.9259	0.8407	0.8404
$m_5(x)$	$\{\alpha,\beta\}_3$	**0.9416**	0.9401	0.941	0.9352	0.9395	0.9384	0.9285	0.9267
	$\{\alpha,\beta\}_4$	0.8852	0.8805	**0.9272**	0.9147	0.8639	0.8657	0.8982	0.8892
	$\{\alpha,\beta\}_5$	**0.8888**	0.8853	0.868	0.8521	0.8732	0.8725	0.8506	0.8449
	$\{\alpha,\beta\}_1$	**0.9153**	0.9084	0.8638	0.8407	0.8811	0.8998	0.7501	0.736
	$\{\alpha,\beta\}_2$	**0.9348**	0.9321	0.9255	0.9166	0.9231	0.9298	0.8888	0.8839
$m_6(x)$	$\{\alpha,\beta\}_3$	0.9367	0.9366	0.9339	0.9326	0.9328	**0.9399**	0.9163	0.9163
	$\{\alpha,\beta\}_4$	0.8837	0.8711	**0.9361**	0.9194	0.8627	0.8301	0.93	0.9182
	$\{\alpha,\beta\}_5$	**0.8918**	0.8854	0.8603	0.8482	0.8727	0.8699	0.832	0.8182

Table 24: 1000×Bias for all models and $u_k(X_i) = \{X_i, X_i^2\}'$

		ET				EL			
			Bal Sample	ATT	Bal Pop		Bal Sample	ATT	Bal Pop
	$\{\alpha,\beta\}_1$	-3.504	-15.59	**-3.439**	-4.241	-6.083	-8.757	3.687	4.179
	$\{\alpha,\beta\}_2$	0.4563	-1.992	-1.627	-2.009	**0.3753**	-0.4242	-1.213	-1.546
$m_1(x)$	$\{\alpha,\beta\}_3$	0.355	**-0.08948**	-0.2331	-0.1901	0.357	0.1303	-0.3697	-0.2472
	$\{\alpha,\beta\}_4$	-2.762	-1.478	-0.5864	-0.9853	-2.942	**0.1762**	-1.619	-2.505
	$\{\alpha,\beta\}_5$	**-3.206**	-6.016	-6.046	-6.404	-4.838	-4.314	-3.416	-3.689
	$\{\alpha,\beta\}_1$	-9.81	-36.08	**-1.924**	-5.863	-10.66	-7.955	11.67	15.54
	$\{\alpha,\beta\}_2$	**0.6323**	-9.128	-7.065	-8.352	0.9951	-2.561	-10.94	-9.947
$m_2(x)$	$\{\alpha,\beta\}_3$	**-0.6622**	-0.7169	-2.439	-1.208	-1.061	-1.254	-3.135	-2.487
	$\{\alpha,\beta\}_4$	-1.765	-0.9709	**0.08242**	-0.2733	-1.862	0.1811	-0.7315	-1.421
	$\{\alpha,\beta\}_5$	22.95	15.98	11.71	12.5	18.34	**-2.052**	-5.572	-6.63
	$\{\alpha,\beta\}_1$	**-9.568**	-38.38	-22.62	-26.29	-14.64	-16.51	-16.77	-17.96
	$\{\alpha,\beta\}_2$	0.9057	-3.567	-3.782	-4.289	1.013	**-0.3191**	-4.299	-4.524
$m_3(x)$	$\{\alpha,\beta\}_3$	0.634	**-0.0673**	-0.4607	-0.3675	0.7078	0.3469	-0.6083	-0.4628
	$\{\alpha,\beta\}_4$	3.018	7.079	**-0.2438**	1.708	5.567	8.877	1.675	4.461
	$\{\alpha,\beta\}_5$	-2.037	-3.592	-3.686	-3.652	**-1.386**	-2.07	-1.864	-1.996
	$\{\alpha,\beta\}_1$	**0.05923**	1.059	1.455	1.914	0.1358	0.533	-2.662	-2.332
	$\{\alpha,\beta\}_2$	-0.1293	0.5051	1.431	1.211	**-0.06621**	0.0929	0.5889	0.8287
$m_4(x)$	$\{\alpha,\beta\}_3$	-0.1565	**-0.003957**	0.1864	0.1151	-0.1583	-0.09748	-0.0247	0.08342
	$\{\alpha,\beta\}_4$	-0.9858	0.9771	**0.008446**	0.2129	-0.7846	4.884	0.3742	0.8518
	$\{\alpha,\beta\}_5$	2.449	1.413	**-0.9037**	-0.9582	3.045	-2.044	-7.394	-7.381
	$\{\alpha,\beta\}_1$	**1.037**	4.904	8.968	10.15	2.638	4.451	9.995	15.05
	$\{\alpha,\beta\}_2$	**-0.8599**	-1.173	-2.376	-2.464	-1.337	-1.957	-4.589	-5.275
$m_5(x)$	$\{\alpha,\beta\}_3$	**0.4631**	0.7666	1.514	1.089	0.6548	0.972	1.499	1.674
	$\{\alpha,\beta\}_4$	-1.159	0.8583	-0.5265	-0.2743	**-0.2432**	5.484	-2.048	-1.761
	$\{\alpha,\beta\}_5$	9.478	2.721	-2.086	**-1.699**	6.566	-8.292	-13.32	-14.68
	$\{\alpha,\beta\}_1$	-12.19	-40.83	-31.9	-36.53	-11.67	**-8.966**	-29.91	-30.67
	$\{\alpha,\beta\}_2$	**2.822**	-5.013	-4.658	-5.806	5.944	5.498	-6.876	-5.352
$m_6(x)$	$\{\alpha,\beta\}_3$	1.809	-1.137	-3.686	-2.955	2.089	**0.73**	-3.951	-3.405
	$\{\alpha,\beta\}_4$	10.51	12.99	**3.604**	6.513	10.47	8.696	9.611	15.45
	$\{\alpha,\beta\}_5$	-5.121	-6.377	-7.617	-7.5	-4.802	**-2.914**	-5.48	-5.218

Table 25: $n\times$Variance for all models and $u_k(X_i) = \{X_i, X_i^2\}'$

		ET				EL			
			Bal Sample	ATT	Bal Pop		Bal Sample	ATT	Bal Pop
$m_1(x)$	$\{\alpha,\beta\}_1$	**0.06476**	0.1076	0.1129	0.1174	0.08625	0.1234	0.1279	0.1334
	$\{\alpha,\beta\}_2$	**0.05142**	0.06052	0.07297	0.07334	0.05757	0.06268	0.0797	0.07904
	$\{\alpha,\beta\}_3$	**0.04341**	0.04497	0.0496	0.04904	0.04462	0.04558	0.0521	0.05085
	$\{\alpha,\beta\}_4$	0.1211	0.1148	**0.0738**	0.08341	0.1283	0.1191	0.0774	0.08517
	$\{\alpha,\beta\}_5$	0.1249	0.1273	0.1453	0.1476	0.1235	**0.1029**	0.1226	0.1268
$m_2(x)$	$\{\alpha,\beta\}_1$	**0.185**	0.363	0.2919	0.293	0.2497	0.2933	0.3315	0.2985
	$\{\alpha,\beta\}_2$	**0.1473**	0.2216	0.3009	0.3255	0.1874	0.2072	0.2347	0.2175
	$\{\alpha,\beta\}_3$	**0.04966**	0.05241	0.06148	0.06023	0.05161	0.05283	0.06484	0.06217
	$\{\alpha,\beta\}_4$	0.1209	0.1117	**0.07355**	0.08249	0.1267	0.1202	0.07684	0.08412
	$\{\alpha,\beta\}_5$	0.3063	0.2439	0.2306	0.2371	0.2934	0.1684	0.1649	**0.1646**
$m_3(x)$	$\{\alpha,\beta\}_1$	**0.09931**	0.243	0.1877	0.2029	0.151	0.193	0.1423	0.1489
	$\{\alpha,\beta\}_2$	**0.06116**	0.07882	0.09133	0.09455	0.07174	0.0751	0.08601	0.08768
	$\{\alpha,\beta\}_3$	**0.04547**	0.04747	0.05214	0.05186	0.04698	0.04782	0.0542	0.05338
	$\{\alpha,\beta\}_4$	0.126	0.1273	0.07583	0.0835	0.1272	0.1223	**0.07273**	0.0776
	$\{\alpha,\beta\}_5$	0.1264	0.1286	0.1398	0.1435	0.1295	**0.0987**	0.1098	0.1165
$m_4(x)$	$\{\alpha,\beta\}_1$	**0.06438**	0.09518	0.1211	0.1254	0.08205	0.1129	0.1336	0.1422
	$\{\alpha,\beta\}_2$	**0.04816**	0.05409	0.06825	0.06838	0.05287	0.05752	0.07201	0.07135
	$\{\alpha,\beta\}_3$	**0.04298**	0.04425	0.04853	0.04877	0.04404	0.045	0.0499	0.05017
	$\{\alpha,\beta\}_4$	0.1245	0.1229	**0.07632**	0.08674	0.1274	0.1018	0.07677	0.0786
	$\{\alpha,\beta\}_5$	0.1245	0.1186	0.1305	0.1306	0.1285	**0.09341**	0.09452	0.1031
$m_5(x)$	$\{\alpha,\beta\}_1$	**0.07598**	0.1173	0.1669	0.1831	0.1001	0.1443	0.1734	0.195
	$\{\alpha,\beta\}_2$	**0.05467**	0.06101	0.08046	0.07905	0.0598	0.06524	0.08978	0.08842
	$\{\alpha,\beta\}_3$	**0.04914**	0.05044	0.05706	0.05683	0.05009	0.05114	0.05831	0.05779
	$\{\alpha,\beta\}_4$	0.1254	0.1267	**0.07647**	0.08805	0.128	0.1003	0.0781	0.08018
	$\{\alpha,\beta\}_5$	0.1731	0.1424	0.1498	0.1463	0.1734	**0.111**	0.1216	0.1246
$m_6(x)$	$\{\alpha,\beta\}_1$	**0.2471**	0.4228	0.3419	0.3637	0.2982	0.3173	0.2483	0.2565
	$\{\alpha,\beta\}_2$	**0.1945**	0.2549	0.2605	0.2798	0.2157	0.2022	0.2064	0.2006
	$\{\alpha,\beta\}_3$	**0.113**	0.1334	0.1797	0.1905	0.1245	0.126	0.1507	0.1601
	$\{\alpha,\beta\}_4$	0.1428	0.1463	0.07724	0.08644	0.1472	0.1301	**0.07298**	0.0786
	$\{\alpha,\beta\}_5$	0.1384	0.1411	0.1589	0.1602	0.1406	**0.1086**	0.1146	0.1171

Table 26: $n\times$MSE for all models and $u_k(X_i) = \{X_i, X_i^2\}'$

		ET				EL			
			Bal Sample	ATT	Bal Pop		Bal Sample	ATT	Bal Pop
$m_1(x)$	$\{\alpha,\beta\}_1$	**0.06598**	0.1319	0.1141	0.1192	0.08994	0.1311	0.1292	0.1351
	$\{\alpha,\beta\}_2$	**0.05143**	0.06091	0.07323	0.07373	0.05758	0.06269	0.07984	0.07927
	$\{\alpha,\beta\}_3$	**0.04342**	0.04497	0.0496	0.04904	0.04462	0.04558	0.05211	0.05085
	$\{\alpha,\beta\}_4$	0.1218	0.1151	**0.07383**	0.08349	0.1292	0.1191	0.07766	0.08578
	$\{\alpha,\beta\}_5$	0.1259	0.1309	0.1489	0.1517	0.1258	**0.1047**	0.1238	0.1281
$m_2(x)$	$\{\alpha,\beta\}_1$	**0.1946**	0.4931	0.2923	0.2964	0.2611	0.2996	0.3451	0.3226
	$\{\alpha,\beta\}_2$	**0.1473**	0.2299	0.3059	0.3325	0.1875	0.2078	0.2466	0.2274
	$\{\alpha,\beta\}_3$	**0.0497**	0.05246	0.06207	0.06037	0.05171	0.05298	0.06582	0.06278
	$\{\alpha,\beta\}_4$	0.1212	0.1117	**0.07355**	0.08249	0.1271	0.1202	0.07689	0.08431
	$\{\alpha,\beta\}_5$	0.359	0.2694	0.2443	0.2527	0.327	0.1688	**0.168**	0.169
$m_3(x)$	$\{\alpha,\beta\}_1$	**0.1085**	0.3903	0.2389	0.272	0.1724	0.2203	0.1704	0.1811
	$\{\alpha,\beta\}_2$	**0.06124**	0.08009	0.09275	0.09638	0.07183	0.0751	0.08785	0.08972
	$\{\alpha,\beta\}_3$	**0.0455**	0.04746	0.05215	0.05187	0.04703	0.04783	0.05423	0.0534
	$\{\alpha,\beta\}_4$	0.1269	0.1323	0.07582	0.08379	0.1303	0.1301	**0.073**	0.07958
	$\{\alpha,\beta\}_5$	0.1268	0.1298	0.1411	0.1448	0.1297	**0.09911**	0.1101	0.1168
$m_4(x)$	$\{\alpha,\beta\}_1$	**0.06438**	0.09529	0.1213	0.1257	0.08204	0.1129	0.1343	0.1427
	$\{\alpha,\beta\}_2$	**0.04816**	0.05411	0.06845	0.06852	0.05287	0.05752	0.07203	0.07141
	$\{\alpha,\beta\}_3$	**0.04298**	0.04424	0.04853	0.04877	0.04404	0.045	0.04989	0.05017
	$\{\alpha,\beta\}_4$	0.1246	0.123	**0.07631**	0.08674	0.1274	0.1042	0.07677	0.07867
	$\{\alpha,\beta\}_5$	0.125	0.1188	0.1305	0.1307	0.1294	**0.09381**	0.09996	0.1086
$m_5(x)$	$\{\alpha,\beta\}_1$	**0.07608**	0.1197	0.175	0.1934	0.1008	0.1462	0.1834	0.2176
	$\{\alpha,\beta\}_2$	**0.05474**	0.06114	0.08101	0.07965	0.05997	0.06561	0.09188	0.09119
	$\{\alpha,\beta\}_3$	**0.04916**	0.05049	0.05728	0.05694	0.05013	0.05123	0.05853	0.05806
	$\{\alpha,\beta\}_4$	0.1255	0.1268	**0.07649**	0.08804	0.128	0.1033	0.07851	0.08048
	$\{\alpha,\beta\}_5$	0.182	0.1431	0.1502	0.1465	0.1777	**0.1178**	0.1393	0.1461
$m_6(x)$	$\{\alpha,\beta\}_1$	**0.262**	0.5895	0.4437	0.4971	0.3118	0.3253	0.3377	0.3505
	$\{\alpha,\beta\}_2$	**0.1953**	0.2574	0.2626	0.2832	0.2192	0.2052	0.2111	0.2035
	$\{\alpha,\beta\}_3$	**0.1133**	0.1335	0.181	0.1913	0.1249	0.1261	0.1523	0.1612
	$\{\alpha,\beta\}_4$	0.1538	0.1632	**0.07853**	0.09067	0.1581	0.1376	0.0822	0.1025
	$\{\alpha,\beta\}_5$	0.141	0.1451	0.1647	0.1659	0.1429	**0.1094**	0.1175	0.1198

Table 27: Coverage of 95% confidence intervals for all models and $u_k(X_i) = \{X_i, X_i^2\}'$

		ET				EL			
			Bal Sample	ATT	Bal Pop		Bal Sample	ATT	Bal Pop
$m_1(x)$	$\{\alpha,\beta\}_1$	**0.9205**	0.8256	0.8612	0.8153	0.878	0.7873	0.8177	0.7809
	$\{\alpha,\beta\}_2$	**0.932**	0.912	0.9038	0.8807	0.9176	0.9007	0.8829	0.8659
	$\{\alpha,\beta\}_3$	**0.9386**	0.935	0.9313	0.9212	0.9351	0.932	0.9253	0.9121
	$\{\alpha,\beta\}_4$	0.8321	0.8176	**0.9345**	0.8892	0.8048	0.783	0.9315	0.8871
	$\{\alpha,\beta\}_5$	0.8284	0.8015	0.7769	0.7315	0.8154	**0.8458**	0.8121	0.785
$m_2(x)$	$\{\alpha,\beta\}_1$	**0.9156**	0.8023	0.902	0.8892	0.8782	0.8585	0.8443	0.8367
	$\{\alpha,\beta\}_2$	0.9182	0.8776	0.8909	0.8502	0.8936	0.886	**0.9186**	0.9117
	$\{\alpha,\beta\}_3$	**0.9388**	0.9321	0.9273	0.9126	0.9348	0.9305	0.9183	0.9064
	$\{\alpha,\beta\}_4$	0.8318	0.8258	**0.9359**	0.8927	0.8091	0.7798	0.9351	0.8867
	$\{\alpha,\beta\}_5$	0.7796	0.8181	0.8275	0.7881	0.8066	**0.9062**	0.8994	0.8783
$m_3(x)$	$\{\alpha,\beta\}_1$	**0.8953**	0.7064	0.8159	0.7162	0.8304	0.81	0.8763	0.802
	$\{\alpha,\beta\}_2$	**0.9261**	0.898	0.8945	0.8637	0.9063	0.8994	0.8958	0.8637
	$\{\alpha,\beta\}_3$	**0.9365**	0.9328	0.934	0.9222	0.9333	0.9309	0.9264	0.9164
	$\{\alpha,\beta\}_4$	0.8298	0.7964	0.9363	0.8968	0.806	0.7854	**0.9462**	0.9076
	$\{\alpha,\beta\}_5$	0.8279	0.8046	0.7909	0.7484	0.8067	**0.8546**	0.8438	0.8096
$m_4(x)$	$\{\alpha,\beta\}_1$	**0.9234**	0.8656	0.8476	0.8006	0.8924	0.8047	0.8159	0.7693
	$\{\alpha,\beta\}_2$	**0.9365**	0.922	0.9083	0.8898	0.9251	0.9078	0.9018	0.8815
	$\{\alpha,\beta\}_3$	**0.9404**	0.9369	0.9312	0.921	0.9372	0.9338	0.9267	0.914
	$\{\alpha,\beta\}_4$	0.8283	0.8049	0.9314	0.8865	0.8051	0.8345	**0.9342**	0.9071
	$\{\alpha,\beta\}_5$	0.8275	0.8078	0.7934	0.7601	0.8015	0.8585	**0.8643**	0.8297
$m_5(x)$	$\{\alpha,\beta\}_1$	**0.9193**	0.8572	0.8393	0.7749	0.8884	0.7878	0.8258	0.7578
	$\{\alpha,\beta\}_2$	**0.9353**	0.918	0.9053	0.8899	0.921	0.902	0.8869	0.8663
	$\{\alpha,\beta\}_3$	**0.9381**	0.934	0.9293	0.9215	0.9345	0.9306	0.9243	0.9169
	$\{\alpha,\beta\}_4$	0.8289	0.8013	0.935	0.8858	0.8087	0.8408	**0.9362**	0.9035
	$\{\alpha,\beta\}_5$	0.8223	0.83	0.8221	0.7987	0.81	**0.8678**	0.8448	0.8131
$m_6(x)$	$\{\alpha,\beta\}_1$	**0.9092**	0.7766	0.8161	0.7447	0.8776	0.8693	0.8677	0.8064
	$\{\alpha,\beta\}_2$	0.9226	0.8967	0.9036	0.8813	0.9037	0.9214	**0.9279**	0.9201
	$\{\alpha,\beta\}_3$	**0.9276**	0.9157	0.9102	0.8644	0.9217	0.9218	0.9262	0.8951
	$\{\alpha,\beta\}_4$	0.8114	0.7895	**0.9384**	0.8977	0.7974	0.8014	0.9336	0.8758
	$\{\alpha,\beta\}_5$	0.8292	0.7981	0.7732	0.7367	0.8105	0.8472	**0.8502**	0.8269

Table 28: 1000×Bias for all models and $u_k(X_i) = \{X_i, X_i^2, X_i^3\}'$

		ET				EL			
			Bal Sample	ATT	Bal Pop		Bal Sample	ATT	Bal Pop
$m_1(x)$	$\{\alpha,\beta\}_1$	**0.7751**	2.325	5.857	6.301	1.97	2.446	11.95	13.63
	$\{\alpha,\beta\}_2$	-0.2476	0.7295	0.5258	0.7588	**0.2161**	0.769	0.9917	1.488
	$\{\alpha,\beta\}_3$	0.287	0.4366	0.464	0.338	0.3755	0.4639	0.2868	**0.2739**
	$\{\alpha,\beta\}_4$	-0.3271	0.8798	**0.07912**	-1.64	-0.4522	2.303	-0.2324	-2.502
	$\{\alpha,\beta\}_5$	**-0.518**	-1.675	-1.414	-0.5293	-2.11	-2.773	-1.24	-1.514
$m_2(x)$	$\{\alpha,\beta\}_1$	**8.628**	30.01	34.16	33.5	26.37	34.1	31.6	31.47
	$\{\alpha,\beta\}_2$	**2.207**	7.389	12.26	11.82	5.043	10.19	14.8	16.89
	$\{\alpha,\beta\}_3$	**-0.9937**	-2.288	-5.39	-3.108	-1.776	-2.599	-4.186	-3.493
	$\{\alpha,\beta\}_4$	-0.09835	1.779	0.2512	-0.9018	-0.05267	2.27	**0.01653**	-1.336
	$\{\alpha,\beta\}_5$	**-4.965**	-22.15	-23.52	-28.88	-9.523	-19.37	-20.08	-23.98
$m_3(x)$	$\{\alpha,\beta\}_1$	**4.722**	14.39	12.94	15.25	12.5	17.25	10.9	12.02
	$\{\alpha,\beta\}_2$	**1.407**	3.773	3.266	3.54	3.114	4.489	3.719	4.322
	$\{\alpha,\beta\}_3$	**0.3227**	0.6111	0.4732	0.5499	0.5376	0.7102	0.5466	0.7067
	$\{\alpha,\beta\}_4$	4.751	6.969	**1.605**	5.203	7.603	8.439	5.563	11.24
	$\{\alpha,\beta\}_5$	2.395	2.298	**1.778**	1.897	3.331	-1.892	-2.269	-2.937
$m_4(x)$	$\{\alpha,\beta\}_1$	**0.348**	0.8804	-2.041	-2.39	0.4385	0.6534	-13.15	-15.52
	$\{\alpha,\beta\}_2$	-0.4219	-0.4447	**0.1804**	-0.2407	-0.4752	-0.2762	-2.914	-3.115
	$\{\alpha,\beta\}_3$	0.01978	0.02207	0.3242	0.01928	**-0.01261**	0.07477	-0.4129	-0.4155
	$\{\alpha,\beta\}_4$	-1.03	0.3679	-0.2867	0.2233	-0.9152	6.877	**0.1344**	1.555
	$\{\alpha,\beta\}_5$	1.025	**0.2621**	-2.109	-2.488	1.785	-2.451	-8.773	-9.952
$m_5(x)$	$\{\alpha,\beta\}_1$	**0.2798**	4.265	-1.999	-5.817	2.419	2.537	-14.58	-19.84
	$\{\alpha,\beta\}_2$	**-0.4467**	-2.468	-1.276	-2.44	-1.588	-2.613	-2.49	-2.653
	$\{\alpha,\beta\}_3$	0.751	1.422	2.386	1.729	1.186	1.475	**0.2942**	0.8205
	$\{\alpha,\beta\}_4$	2.65	4.249	1.136	**1.107**	3.85	8.685	1.817	2.829
	$\{\alpha,\beta\}_5$	1.48	-11.65	-13.82	-16.18	**-0.2491**	-10.66	-13.42	-15.1
$m_6(x)$	$\{\alpha,\beta\}_1$	**13.19**	32	28.01	33.04	36.38	37.15	22.06	28.63
	$\{\alpha,\beta\}_2$	**6.596**	17.48	15.5	17.12	15.12	20.3	14.6	19.18
	$\{\alpha,\beta\}_3$	**0.7634**	2.339	-0.7786	1.028	1.762	3.014	1.69	2.511
	$\{\alpha,\beta\}_4$	0.501	-2.143	1.276	7.856	**0.3987**	0.6734	5.477	13.46
	$\{\alpha,\beta\}_5$	-3.22	0.4294	**-0.2893**	2.584	-3.716	-1.048	-5.112	-3.414

Table 29: $n\times$Variance for all models and $u_k(X_i) = \{X_i, X_i^2, X_i^3\}'$

		ET				EL			
			Bal Sample	ATT	Bal Pop		Bal Sample	ATT	Bal Pop
$m_1(x)$	$\{\alpha,\beta\}_1$	**0.05915**	0.1234	0.1041	0.1107	0.09155	0.1259	0.1053	0.1106
	$\{\alpha,\beta\}_2$	**0.04955**	0.06745	0.0766	0.08225	0.05721	0.07154	0.08098	0.08631
	$\{\alpha,\beta\}_3$	**0.04421**	0.04978	0.05623	0.05725	0.04624	0.05164	0.05911	0.06056
	$\{\alpha,\beta\}_4$	0.1087	0.1121	**0.07719**	0.09398	0.131	0.1023	0.08107	0.09456
	$\{\alpha,\beta\}_5$	0.1111	0.1199	0.13	0.1307	0.1291	**0.07597**	0.08668	0.08696
$m_2(x)$	$\{\alpha,\beta\}_1$	**0.1445**	0.211	0.3435	0.3658	0.2291	0.2205	0.3784	0.3807
	$\{\alpha,\beta\}_2$	**0.1101**	0.1956	0.1988	0.159	0.1455	0.171	0.2242	0.1853
	$\{\alpha,\beta\}_3$	**0.04836**	0.05831	0.06967	0.07137	0.0514	0.05937	0.06879	0.07129
	$\{\alpha,\beta\}_4$	0.1103	0.1113	**0.07773**	0.09323	0.1323	0.1057	0.08233	0.09621
	$\{\alpha,\beta\}_5$	0.1844	0.2076	0.2288	0.2256	0.2334	0.1487	**0.1471**	0.1567
$m_3(x)$	$\{\alpha,\beta\}_1$	**0.06736**	0.148	0.1347	0.15	0.1048	0.145	0.1248	0.1407
	$\{\alpha,\beta\}_2$	**0.05118**	0.07168	0.07916	0.08322	0.05967	0.07524	0.08253	0.08627
	$\{\alpha,\beta\}_3$	**0.04393**	0.0493	0.05678	0.0561	0.04572	0.05081	0.05925	0.05866
	$\{\alpha,\beta\}_4$	0.1092	0.1195	0.07459	0.09158	0.1281	0.09668	**0.07188**	0.08282
	$\{\alpha,\beta\}_5$	0.1111	0.1168	0.124	0.1269	0.1308	**0.07328**	0.08286	0.08219
$m_4(x)$	$\{\alpha,\beta\}_1$	**0.06292**	0.1299	0.102	0.1076	0.09622	0.1054	0.09239	0.08941
	$\{\alpha,\beta\}_2$	**0.04994**	0.06793	0.0748	0.07741	0.05738	0.06588	0.06778	0.07093
	$\{\alpha,\beta\}_3$	**0.04377**	0.04911	0.05503	0.05574	0.04535	0.04866	0.05373	0.05441
	$\{\alpha,\beta\}_4$	0.1093	0.117	0.07924	0.09923	0.1291	0.07971	**0.07839**	0.08176
	$\{\alpha,\beta\}_5$	0.1117	0.1158	0.1204	0.1208	0.1328	0.07622	**0.07488**	0.07988
$m_5(x)$	$\{\alpha,\beta\}_1$	**0.07172**	0.1496	0.1213	0.1205	0.1134	0.1186	0.1078	0.1019
	$\{\alpha,\beta\}_2$	**0.05464**	0.07569	0.08651	0.09631	0.06316	0.07753	0.08116	0.08849
	$\{\alpha,\beta\}_3$	**0.04925**	0.05682	0.06329	0.06634	0.05182	0.05526	0.0592	0.05999
	$\{\alpha,\beta\}_4$	0.1127	0.1216	0.08328	0.09995	0.1324	0.0833	**0.0799**	0.08328
	$\{\alpha,\beta\}_5$	0.1461	0.1475	0.1652	0.1686	0.183	0.1056	**0.1045**	0.1165
$m_6(x)$	$\{\alpha,\beta\}_1$	**0.1803**	0.326	0.2763	0.3017	0.2752	0.2929	0.3122	0.3312
	$\{\alpha,\beta\}_2$	**0.1362**	0.1994	0.22	0.202	0.1616	0.1907	0.2199	0.2093
	$\{\alpha,\beta\}_3$	**0.08171**	0.1209	0.1118	0.1107	0.09173	0.1094	0.1113	0.1
	$\{\alpha,\beta\}_4$	0.1153	0.1177	0.07507	0.09551	0.1381	0.09834	**0.07268**	0.08947
	$\{\alpha,\beta\}_5$	0.1213	0.1269	0.132	0.1208	0.1474	0.08578	0.08259	**0.07951**

Table 30: $n\times$MSE for all models and $u_k(X_i) = \{X_i, X_i^2, X_i^3\}'$

		ET				EL			
			Bal Sample	ATT	Bal Pop		Bal Sample	ATT	Bal Pop
$m_1(x)$	$\{\alpha,\beta\}_1$	**0.0592**	0.1239	0.1075	0.1146	0.09193	0.1264	0.1195	0.1292
	$\{\alpha,\beta\}_2$	**0.04955**	0.06749	0.07662	0.0823	0.05721	0.0716	0.08107	0.08652
	$\{\alpha,\beta\}_3$	**0.04421**	0.04979	0.05624	0.05725	0.04625	0.05166	0.05912	0.06056
	$\{\alpha,\beta\}_4$	0.1087	0.1121	**0.07718**	0.09424	0.131	0.1028	0.08107	0.09518
	$\{\alpha,\beta\}_5$	0.1111	0.1201	0.1302	0.1307	0.1295	**0.0767**	0.08679	0.08714
$m_2(x)$	$\{\alpha,\beta\}_1$	**0.1519**	0.3011	0.4601	0.478	0.2986	0.3367	0.4781	0.4796
	$\{\alpha,\beta\}_2$	**0.1106**	0.201	0.2138	0.1729	0.148	0.1813	0.2461	0.2138
	$\{\alpha,\beta\}_3$	**0.04845**	0.05883	0.07257	0.07233	0.05171	0.06004	0.07054	0.0725
	$\{\alpha,\beta\}_4$	0.1103	0.1115	**0.07773**	0.0933	0.1323	0.1062	0.08232	0.09638
	$\{\alpha,\beta\}_5$	0.1868	0.2566	0.2841	0.3089	0.2424	**0.1861**	0.1874	0.2141
$m_3(x)$	$\{\alpha,\beta\}_1$	**0.06958**	0.1687	0.1514	0.1732	0.1204	0.1747	0.1366	0.1551
	$\{\alpha,\beta\}_2$	**0.05137**	0.07309	0.08021	0.08447	0.06063	0.07725	0.0839	0.08812
	$\{\alpha,\beta\}_3$	**0.04394**	0.04933	0.0568	0.05612	0.04574	0.05085	0.05927	0.05871
	$\{\alpha,\beta\}_4$	0.1114	0.1243	**0.07484**	0.09428	0.1339	0.1038	0.07497	0.09543
	$\{\alpha,\beta\}_5$	0.1117	0.1173	0.1243	0.1272	0.1319	**0.07361**	0.08333	0.08301
$m_4(x)$	$\{\alpha,\beta\}_1$	**0.06292**	0.1299	0.1024	0.1081	0.09623	0.1055	0.1096	0.1134
	$\{\alpha,\beta\}_2$	**0.04995**	0.06794	0.07479	0.0774	0.0574	0.06588	0.06862	0.07189
	$\{\alpha,\beta\}_3$	**0.04376**	0.04911	0.05503	0.05574	0.04534	0.04866	0.05374	0.05442
	$\{\alpha,\beta\}_4$	0.1094	0.117	0.07924	0.09922	0.1292	0.08441	**0.07838**	0.08199
	$\{\alpha,\beta\}_5$	0.1118	0.1158	0.1208	0.1213	0.1331	**0.07679**	0.08253	0.08973
$m_5(x)$	$\{\alpha,\beta\}_1$	**0.07172**	0.1514	0.1216	0.1239	0.1139	0.1192	0.129	0.1412
	$\{\alpha,\beta\}_2$	**0.05466**	0.07629	0.08667	0.0969	0.06341	0.0782	0.08177	0.08919
	$\{\alpha,\beta\}_3$	**0.0493**	0.05702	0.06385	0.06663	0.05196	0.05547	0.0592	0.06006
	$\{\alpha,\beta\}_4$	0.1134	0.1234	0.0834	0.1001	0.1339	0.09081	**0.08022**	0.08407
	$\{\alpha,\beta\}_5$	0.1463	0.1611	0.1843	0.1947	0.183	**0.1169**	0.1225	0.1393
$m_6(x)$	$\{\alpha,\beta\}_1$	**0.1977**	0.4284	0.3547	0.4108	0.4075	0.4309	0.3607	0.413
	$\{\alpha,\beta\}_2$	**0.1406**	0.2299	0.244	0.2313	0.1845	0.2319	0.2412	0.2461
	$\{\alpha,\beta\}_3$	**0.08176**	0.1214	0.1119	0.1108	0.09203	0.1103	0.1116	0.1007
	$\{\alpha,\beta\}_4$	0.1153	0.1181	**0.07522**	0.1017	0.1381	0.09836	0.07566	0.1076
	$\{\alpha,\beta\}_5$	0.1224	0.1269	0.132	0.1214	0.1488	0.08586	0.08516	**0.08063**

Table 31: Coverage of 95% confidence intervals for all models and $u_k(X_i) = \{X_i, X_i^2, X_i^3\}'$

		ET				EL			
			Bal Sample	ATT	Bal Pop		Bal Sample	ATT	Bal Pop
	$\{\alpha, \beta\}_1$	**0.9267**	0.8272	0.8718	0.7538	0.869	0.7618	0.8167	0.737
	$\{\alpha, \beta\}_2$	**0.9334**	0.9019	0.9057	0.8322	0.9136	0.88	0.8755	0.8138
$m_1(x)$	$\{\alpha, \beta\}_3$	**0.9327**	0.9223	0.9257	0.8912	0.9253	0.9143	0.9105	0.8776
	$\{\alpha, \beta\}_4$	0.8352	0.8438	**0.9387**	0.85	0.8091	0.8407	0.9318	0.8468
	$\{\alpha, \beta\}_5$	0.8268	0.8325	0.8259	0.7188	0.8041	**0.9052**	0.8805	0.8442
	$\{\alpha, \beta\}_1$	**0.9089**	0.8258	0.85	0.6508	0.8211	0.7394	0.7359	0.6463
	$\{\alpha, \beta\}_2$	**0.9176**	0.8816	0.9002	0.8407	0.8862	0.8806	0.834	0.7933
$m_2(x)$	$\{\alpha, \beta\}_3$	**0.9353**	0.928	0.9287	0.8721	0.9296	0.9223	0.9269	0.8669
	$\{\alpha, \beta\}_4$	0.8243	0.8378	**0.9362**	0.8538	0.7927	0.8261	0.9288	0.8432
	$\{\alpha, \beta\}_5$	0.8255	0.8608	0.8614	0.6378	0.8058	0.8705	**0.877**	0.743
	$\{\alpha, \beta\}_1$	**0.9148**	0.822	0.8689	0.6867	0.8554	0.7593	0.8142	0.7031
	$\{\alpha, \beta\}_2$	**0.9345**	0.9007	0.9075	0.8337	0.9101	0.8751	0.8813	0.8173
$m_3(x)$	$\{\alpha, \beta\}_3$	**0.9406**	0.9327	0.9296	0.8975	0.935	0.9227	0.9158	0.8833
	$\{\alpha, \beta\}_4$	0.8306	0.8351	**0.9425**	0.864	0.8027	0.8511	0.9421	0.863
	$\{\alpha, \beta\}_5$	0.8236	0.834	0.8287	0.7204	0.7984	**0.9012**	0.8864	0.8547
	$\{\alpha, \beta\}_1$	**0.9177**	0.8224	0.8677	0.7715	0.8673	0.8057	0.8374	0.7683
	$\{\alpha, \beta\}_2$	**0.9323**	0.9026	0.902	0.8421	0.9144	0.8882	0.9	0.8591
$m_4(x)$	$\{\alpha, \beta\}_3$	**0.9385**	0.9319	0.9304	0.8979	0.9356	0.928	0.9241	0.8987
	$\{\alpha, \beta\}_4$	0.8253	0.8313	0.932	0.8429	0.8008	0.8787	**0.934**	0.8907
	$\{\alpha, \beta\}_5$	0.8253	0.8283	0.8302	0.7366	0.7964	0.8984	**0.9016**	0.8596
	$\{\alpha, \beta\}_1$	**0.922**	0.831	0.8752	0.7657	0.872	0.8191	0.8345	0.7638
	$\{\alpha, \beta\}_2$	**0.9312**	0.9017	0.9031	0.8323	0.9138	0.8814	0.9008	0.8408
$m_5(x)$	$\{\alpha, \beta\}_3$	**0.9372**	0.9277	0.9279	0.8913	0.9322	0.925	0.9279	0.9044
	$\{\alpha, \beta\}_4$	0.8253	0.8299	0.9281	0.845	0.7948	0.8729	**0.9309**	0.8844
	$\{\alpha, \beta\}_5$	0.8225	0.8393	0.8404	0.6953	0.8023	0.8715	**0.8777**	0.799
	$\{\alpha, \beta\}_1$	0.9014	0.8116	**0.9069**	0.6655	0.8018	0.7728	0.8012	0.686
	$\{\alpha, \beta\}_2$	**0.9236**	0.8929	0.9131	0.8036	0.888	0.8671	0.8608	0.7933
$m_6(x)$	$\{\alpha, \beta\}_3$	0.9261	0.909	**0.9285**	0.879	0.9157	0.9202	0.9164	0.8937
	$\{\alpha, \beta\}_4$	0.8225	0.8303	**0.9401**	0.8597	0.7958	0.8513	0.9385	0.8476
	$\{\alpha, \beta\}_5$	0.8274	0.8398	0.8405	0.7482	0.8004	0.8933	**0.9054**	0.8756

References

B. Antoine, H. Bonnal, and E. Renault. On the efficient use of the informational content of estimating equations: Implied probabilities and euclidean empirical likelihood. *Journal of Econometrics*, 138:461–487, 2007.

M. Blackwell. A framework for dynamic causal inference in political science. *Americal Journal of Political Science*, 57(2), 2013.

M. Busso, J. DiNardo, and J. McCrary. New evidence on the finite sample properties of propensity score reweighting and matching estimators. *Review of Economics and Statistics*, 96(5):885–897, 2014.

K.C.G. Chan, S.C.P. Yam, and Z. Zhang. Globally efficient nonparametric inference of average treatment effects by empirical balancing calibration weighting. *Journal of the Royal Statistics Society*, 78(3):673–700, 2016.

P. Chaussé, J. Liu, and G. Luta. A simulation-based comparison of covariate adjustment methods for the analysis of randomized controlled trials. *International Journal of Environmental Research and Public Health*, 13(4), 2016.

P. Chaussé. Computing generalized method of moments and generalized empirical likelihood with R. *Journal of Statistical Software*, 34(11):1–35, 2010. URL `http://www.jstatsoft.org/v34/i11/`.

X. Chen, H. Hing, and M. Shum. Nonparametric likelihood ratio model selection tests between parametric likelihood and moment condition models. *Journal of Econometrics*, 141(1):109–140, 2007.

R.H. Dehejia and S. Wahba. Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association*, 94(448), 1999.

R.H. Dehejia and S. Wahba. Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics*, 84(1):151–161, 2002.

S. Donald, G. Imbens, and W. Newey. Empirical likelihood estimation and consistent tests with conditional moment restrictions. *Journal of Econometrics*, 117:55–93, 2003.

S.G. Donald, G. Imbens, and W. Newey. Choosing the number of moments in conditional moment restrition models. *MIT working paper*, 2008.

D.A. Freedman. On regression adjustment to experimental data. *Advances in Applied Mathematics*, 40:180–193, 2008.

M. Frolich. Finite sample properties of propensity-score matching and weighting estimators. *Review of Economics and Statistics*, 86(1):77–90, 2004.

N. Gospodinov, R. Kan, and C. Robotti. Misspecification-robust standard errors for the continuously-updated gmm estimator. *Working Paper*, 2013.

B.S. Graham, C.C. de Xavier Pinto, and D. Egel. Inverse probability tilting for moment condition models with missing data. *Review of Economic Studies*, 79:1053–1079, 2012.

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

J. Hainmueller. Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational study. *Political Analysis*, 20(1), 2012.

L. P. Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50:1029–1054, 1982.

J. Heckman, H. Ichimura, and P. Todd. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic Studies*, 64(4):605–654, 1997.

K. Hirano, G. Imbens, and G. Ridder. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, 71(4):1161–1189, 2003.

Y. Kitamura, T. Otsu, and K. Evdokimov. Robustness, infinitesimal neighborhoods, and moment restrictions. *Econometrica*, 81(3):1185–1201, 2013.

R. Lalonde. Evaluating the econometric evaluations of training programs. *American Economic Review*, 76:604–620, 1986.

S. Lee. Asymptotic refinements of misspecified-robust bootstrap for gel estimators. *Journal of Econometrics*, 192:86–104, 2016.

W. Lin. Agnostic notes on regression adjustment to experimental data: Reexamining freedman's critique. *The Annals of Applied Statistics*, 7(1):295–318, 2013.

W.K. Newey and R. J. Smith. High order properties of gmm and generalized empirical likelihood estimator. *Econometrica*, 72:219–255, 2004.

A. B. Owen. *Empirical likelihood*. Chapman & Hall/CRC, Boca Raton, Florida, USA, 2001.

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

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

S.M. Schennach. Point estimation with the exponentially tilted empirical likelihood. *The Annals of Statistics*, 35(2):634–672, 2007.

C. Wu. Algorithms and r codes for the pseudo empirical likelihood method in survey sampling. *Survey Methodology*, 31(2):239, 2005.

X. Wu and Z. Ying. An empirical likelihood approach to nonparametric covariate adjustment in randomized clinical trials. *arXiv:1108.0484 [stat.ME]*, 2011.

L. Yang and A. A. Tsiatis. Efficiency study of estimators for a treatment effect in a pretest-posttest trial. *The American Statistician*, 55(4):314–321, 2001.

J.R. Zubizarreta. Stable weights that balance covariates for estimation with incomplete outcome data. *Journal of the American Statistical Association*, 110(511):910–922, 2014.