The Stata Journal (2023)
23, Number 3, pp. 709–729 DOI: 10.1177/1536867X231196291

ebct: Using entropy balancing for continuous treatments to estimate dose–response functions and their derivatives

Stefan Tübbicke
Institute for Employment Research
Nuremberg, Germany
stefan.tuebbicke@iab.de

Abstract. Interest in evaluating dose–response functions of continuous treatments has been increasing recently. To facilitate the estimation of causal effects in this setting, I introduce the `ebct` command for the estimation of dose–response functions and their derivatives using entropy balancing for continuous treatments. First, balancing weights are estimated by numerically solving a globally convex optimization problem. These weights eradicate Pearson correlations between covariates and the treatment variable. Because simple uncorrelatedness may be insufficient to yield consistent estimates in the next step, higher moments of the treatment variable can be rendered uncorrelated with covariates. Second, the weights are used in local linear kernel regressions to estimate the dose–response function or its derivative. To perform statistical inference, I use a bootstrap procedure. The command also provides the option of producing publication-quality graphs for the estimated relationships.

Keywords: st0726, ebct, entropy balancing, continuous treatments, balancing weights, observational studies, dose–response functions

1 Introduction

The number of studies evaluating continuous treatments sharply increased after the seminal articles of Hirano and Imbens (2004) and Imai and van Dyk (2004). In such a setting, all units studied are affected by some treatment but with different intensities. Examples include the evaluation of job training programs with varying durations (Choe, Flores-Lagunes, and Lee 2015; Flores et al. 2012; Kluve et al. 2012; Li and Fraser 2015) or quality (Galdo and Chong 2012); the size of subsidies aimed at firms or regions (Becker, Egger, and von Ehrlich 2012; Bia and Mattei 2012; Mitze, Paloyo, and Alecke 2015); transfer times to specialized hospitals (Kreif et al. 2015); and air quality as measured by particulate matter concentrations (Wu et al. Forthcoming). When evaluating continuous treatments, the researcher typically aims to estimate the dose–response function (DRF), which describes how average population outcomes depend on the dose of the treatment. Alternatively, one may estimate the derivative of the DRF. By doing so, one may infer if and by how much marginal increases in the dose affect average population outcomes.

Most available methods use the generalized propensity score (GPS, Imbens [2000]), that is, the conditional *density* of the treatment variable, to account for nonrandom selection into different treatment intensities based on observed covariates. Approaches include regression adjustment for the GPS (Hirano and Imbens 2004), matching on the GPS (Wu et al. Forthcoming), and weighting on the GPS (Fong, Hazlett, and Imai 2018; Flores et al. 2012; Alejo, Galvao, and Montes-Rojas 2018).[1] Recently, Tübbicke (2022) introduced the so-called entropy balancing for continuous treatments (EBCT) approach, extending the entropy balancing for binary treatments approach by Hainmueller (2012).[2] As the name suggests, EBCT can be used in the estimation of causal effects in the context of continuous treatments. In contrast to many approaches based on the GPS—which typically balance covariates only asymptotically—EBCT directly optimizes in-sample covariate balance algorithmically. As Tübbicke (2022) shows, EBCT displays favorable properties in terms of balancing performance and finite-sample performance of subsequent effect estimates.

In this article, I present the `ebct` command, which implements a two-step estimation procedure. First, covariate balancing weights are estimated using the EBCT methodology. Following the majority of the continuous treatment literature, Pearson correlations are used as balancing indicators. The resulting weights lead to exactly zero correlations between covariates and the treatment variable in the reweighted sample. Because uncorrelatedness does not imply independence between covariates and the treatment, remaining nonlinear imbalances after reweighting may cause bias in resulting estimates. Hence, balancing targets can be chosen such that higher orders of the treatment variable are rendered uncorrelated with covariates to curb potential biases. Because of the convex nature of the optimization problem, `ebct` estimates balancing weights fairly rapidly. Second, the DRF or its derivative is estimated using local linear kernel regressions. Hence, the command estimates these quantities in a completely nonparametric fashion. To obtain standard errors, `ebct` has a built-in bootstrap procedure. Moreover, the command can produce publication-ready graphs of final estimates.

The remainder of the article is organized as follows. Section 2 introduces the reader to causal effects of continuous treatments, identifying assumptions and the use of EBCT to estimate these effects. Section 3 provides details of the `ebct` command. Section 4 applies the command to the estimation of the effects of lottery winnings on subsequent labor earnings. Section 5 concludes.

1. For Stata, the following approaches have been implemented as ado-packages: regression adjustment (`dosereponse`, Bia and Mattei [2008]), weighting on the GPS (`drf`, Bia et al. [2014]), and the weighting approach by Alejo, Galvao, and Montes-Rojas (2018), which can also be used to estimate quantile treatment effects (`qcte`, Alejo, Galvao, and Montes-Rojas [2020]).
2. The approach by Hainmueller (2012) has been implemented in the `ebal` command (Hainmueller and Xu 2013). Similarly, see the `ebalfit` command by Jann (2021a), which also estimates asymptotic standard errors for resulting effect estimates. See Jann (2021b) for details.

2 Estimating the effects of continuous treatments

Before going into detail on the estimation approach of the `ebct` command, let us review causal effects of continuous treatments in terms of the potential-outcomes framework (Roy 1951; Rubin 1974). Assume that we observe an independent and identically distributed sample of n individuals i and k pretreatment covariates denoted as a vector $\mathbf{x}_i$. The posttreatment outcome is y_i, and the treatment intensity is measured by t_i with possible values $t \in [t_{\min}, t_{\max}]$. The set of potential outcomes is given by $y_i(t)$—also often called the unit-dose response—denoting the outcome that would have been observed had the unit received treatment with intensity t. Aggregating these unit-level responses leads to the population DRF $E\{y_i(t)\}$. Along with its derivative $dE\{y_i(t)\}/dt$, which describes the marginal change in average population outcomes due to marginal changes in the treatment intensity, the DRF represents the key relationship to be estimated in practice. If treatment intensities were randomly assigned, comparisons of average outcomes between individuals with different treatment intensities would directly give consistent estimates of these quantities. However, this is mostly not the case even in experimental settings. In the absence of other identifying variations, for example, through a natural experiment, the following three identifying assumptions must be invoked to obtain consistent estimates in observational studies.

2.1 Identifying assumptions

First, the set of covariates $\mathbf{x}_i$ must satisfy the conditional independence assumption (CIA, Lechner [2001]), also known as the selection-on-observables assumption (Heckman and Robb 1985). This means that, conditional on $\mathbf{x}_i$, potential outcomes must be independent of the treatment intensity t_i actually assigned; that is,

$$y_i(t) \perp\!\!\!\perp t_i \mid \mathbf{x}_i \;\forall\; t \in [t_{\min}, t_{\max}]$$

This assumption requires that the researcher observe all covariates $\mathbf{x}_i$ that are related simultaneously to the selection mechanism and the outcome of interest. Thus, the set of covariates $\mathbf{x}_i$ must be sufficiently rich to make this assumption credible. This must be discussed case by case. Second, there is an assumption that requires common support; that is, the conditional density of treatment must be positive over $[t_{\min}, t_{\max}]$:

$$f_{t|\mathbf{x}}(t_i = t \mid \mathbf{x}_i) > 0 \;\forall\; t \in [t_{\min}, t_{\max}]$$

Using the example of lottery winnings from section 4, we see this assumption would be violated if only male individuals received a high treatment intensity, that is, a large prize. If this is the case, the sample must be trimmed to the range of treatment intensities where both genders can be found and the DRF can be estimated only on this subset of observations to avoid extrapolation (Crump et al. 2009; Lechner and Strittmatter 2019). Last, one must assume the so-called stable-unit treatment-value assumption (see Rubin [1980]), essentially ruling out general equilibrium and spillover effects of treatment by requiring that each unit's outcome depend only on its own treatment intensity (see Imbens and Wooldridge [2009] and Manski [2013] for examples).

2.2 Using EBCT to estimate the DRF and its derivative

The proposed approach follows a two-step procedure. First, the EBCT (Tübbicke 2022) algorithm is used to estimate covariate balancing weights that eradicate the Pearson correlation between the treatment variable (and possibly higher moments of it) and covariates among units that actually received the treatment. That is, untreated units must be dropped from the analysis. Second, these weights are used in nonparametric outcome regressions to estimate the DRF and its derivative. To obtain standard errors, we repeat these steps on bootstrap samples.

2.2.1 Estimating balancing weights

Unlike other covariate balancing strategies for continuous treatments, the EBCT approach does not require estimating the GPS. Instead, it directly estimates balancing weights w_i such that certain moment restrictions are met in the reweighted sample. More specifically, the algorithm estimates balancing weights w_i by minimizing the dispersion of weights such that, after weighting, covariates are completely *uncorrelated* with the treatment variable while maintaining unconditional means of covariates and the treatment.[3] Because simple uncorrelatedness may be insufficient to avoid bias in resulting estimates due to lingering imbalances in covariates across the treatment variable distribution, EBCT can also render higher moments of the treatment variable uncorrelated with covariates to help deal with such potential biases. In mathematical terms, EBCT solves the globally convex constrained minimization problem

$$\min_w H(w) = \sum_{i=1}^n w_i \ln(w_i/q_i) \quad \text{such that} \quad \begin{aligned} \sum_{i=1}^n w_i t_i^m \mathbf{x}_i &= \bar{t}^m \bar{\mathbf{x}} \quad \text{for all } m = 1,\dots,p \\ \sum_{i=1}^n w_i t_i^m &= \bar{t}^m \qquad \text{for all } m = 1,\dots,p \\ \sum_{i=1}^n w_i \mathbf{x}_i &= \bar{\mathbf{x}} \\ \sum_{i=1}^n w_i &= 1 \\ w_i &> 0 \end{aligned}$$

where $h(w_i) = w_i \ln(w_i/q_i)$ is the Kullback (1997) entropy metric of dispersion, q_i are user-specified base weights with default $= 1/n$, and $\bar{t}^m$ and $\bar{\mathbf{x}}$ are unconditional means of t_i^m and $\mathbf{x}_i$, respectively. Moreover, let p be the highest order of t_i up to which covariate balance is sought. If $p = 1$, then EBCT requires the conditional mean of t_i to be the

3. Note that the original approach by Hainmueller (2012) aims for mean (or higher-order) balance of covariates across the treatment and comparison groups. In the context of continuous treatments, there are essentially infinitely many treatment groups, requiring a new measure of covariate balance. Hence, most of the continuous treatment literature has adopted Pearson correlations as their main measures of covariate balance.

same in the reweighted sample as in the unweighted sample, and the weights render t_i uncorrelated with $\mathbf{x}_i$. If $p = 2$ is chosen, then the mean of t_i^2 is also retained, and t_i^2 is required to be uncorrelated with $\mathbf{x}_i$, and so forth. Note that the choice of p is likely to be subject to a bias-variance tradeoff: The larger p is, the more likely it is that t_i and $\mathbf{x}_i$ can be considered independent of one another. However, a value of larger p will also likely lead to an increase in the variance of resulting effect estimates. Hence, one should flexibly check covariate balance in the reweighted sample and increase p only if balance is found to be insufficient. Section 4 will come back to this topic when going through the example application. Independent of the choice of p, the w_i are obtained by numerically minimizing the corresponding Lagrange function of the minimization problem (see Tübbicke [2022] for details).

2.2.2 Estimation of the DRF and its derivative

Once EBCT weights are obtained and balance is found to be sufficient, the DRF $E\{y_i(t)\}$ and its derivative $dE\{y_i(t)\}/dt$ can be estimated using weighted parametric or nonparametric regression techniques. To avoid potential misspecification bias due to wrong functional form assumptions, the `ebct` command uses local linear kernel regressions based on the Epanechnikov kernel because of its well-known optimality properties (Fan 1992; Heckman, Ichimura, and Todd 1997). In particular, estimates of $E\{y_i(t)\}$ and $dE\{y_i(t)\}/dt$ are obtained for a set of evenly spaced gridpoints $t \in [t_{\min}, t_{\max}]$ by minimizing the local weighted sum of squares as

$$\min_{\alpha,\beta} \sum_{i=1}^{n} w_i k\left(\frac{t_i - t}{h}\right) \{y_i - \alpha - \beta(t_i - t)\}^2$$

where $k(\cdot)$ denotes the kernel function evaluated at t with bandwidth h and

$$\begin{pmatrix} \widehat{E\{y_i(t)\}} \\ \widehat{dE\{y_i(t)\}}/dt \end{pmatrix} = \begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{pmatrix}$$

The derivative of the DRF is thus estimated using a Taylor-based approach, which tends to yield a lower mean squared error compared with analytic derivatives; see Racine (2016) for details. To choose the bandwidth h, the `ebct` command provides two data-driven options. The default is to use an algorithm following Fan (1992) to minimize the asymptotic mean integrated squared error. Second, the command provides a cross-validation routine that randomly splits the sample in half and uses out-of-sample predictions to find the bandwidth that minimizes the mean squared prediction error. To obtain standard errors, the estimation of balancing weights and the fit of the local linear regressions is repeated numerous times using bootstrap samples (Efron and Tibshirani 1986), and confidence intervals are computed using the normal approximation. As MacKinnon (2006) notes, an advantage of this approach is that the number of replications does not have to be very large to obtain reasonably precise estimates of confidence intervals. For a more general overview of bootstrap methods, see Cameron and Trivedi (2022).

2.3 Syntax

The command syntax is

`ebct` *covariates* [*if*] [*in*], `treatvar(`*varname*`)` [`basew(`*varname*`)` `samplew(`*varname*`)` `p(`*integer*`)` `out(`*varname*`)` `est(`*string*`)` `cv` `bandw(`*real*`)` `bootstrap` `reps(`*integer*`)` `seed(`*integer*`)` `graph` `generate(`*stub*`)`]

covariates is a list of background characteristics one wishes to control for.

2.4 Options

Table 1. Options for `ebct`

	options	Description
*	`treatvar(`*varname*`)`	specify the continuous treatment variable
	`basew(`*varname*`)`	specify base weights q; default is `basew(`$q = 1/n$`)`
	`samplew(`*varname*`)`	specify sampling weights; if units in the sample do not have the same probability of being in the sample, this option can be used to calculate weighted means $\bar{t}^p$ and $\bar{\mathbf{x}}$ for the definition of balancing targets
	`p(`*integer*`)`	specify up to which order p the treatment variable is required to be uncorrelated with covariates; default is `p(1)`; maximum is `p(3)`
	`out(`*varname*`)`	specify the outcome variable for which the DRF or its derivative shall be estimated
	`est(`*string*`)`	specify whether the `drf` or its `derivative` shall be estimated; default is `est(drf)`
	`cv`	uses cross-validation when specified to obtain the bandwidth that minimizes out-of-sample mean squared prediction error instead of the default plugin bandwidth for the local linear regressions
	`bandw(`*real*`)`	sets a user-chosen bandwidth
	`bootstrap`	specify whether the bootstrap shall be used to obtain standard errors (default is no estimation of standard errors)
	`reps(`*integer*`)`	specify how many bootstrap replications shall be used; default is `reps(100)`
	`seed(`*integer*`)`	specify the seed used as a starting point for the resampling process; default `seed(2222)` is used to ensure that results are always reproducible

Continued on next page

options	Description
`graph`	specify whether a graph for the estimated relationship shall be produced; if the `bootstrap` option is specified, the graph includes 95% confidence bands
`generate(`*stub*`)`	if specified, *stub*`_weight`, *stub*`_grid`, and potentially other variables are added to the dataset, depending on the other options chosen; if `out()` is provided, *stub*`_estimates` containing the estimated (derivative of the) DRF is generated; with the `bootstrap` option, *stub*`_stderr`, *stub*`_ci_low`, and *stub*`_ci_high` are also generated, containing standard errors and the lower and upper bounds of the 95% confidence intervals; if any of these variables already exist, they will be overwritten without warning

* `treatvar()` is required.

2.5 Stored results

`ebct` stores the following in `e()`:

Scalars	
`e(N)`	number of observations
`e(bandwidth)`	bandwidth used in the local linear regressions
Macros	
`e(estimand)`	either `drf` or `derivative`
Matrices	
`e(b)`	estimated DRF or its derivative at gridpoints
`e(se)`	estimated standard errors of DRF or its derivative at gridpoints
`e(balance)`	overall balance summary statistics
`e(balance_detail)`	detailed covariate balancing statistics
`e(gridpoints)`	gridpoints used in the estimation
Functions	
`e(sample)`	estimation sample

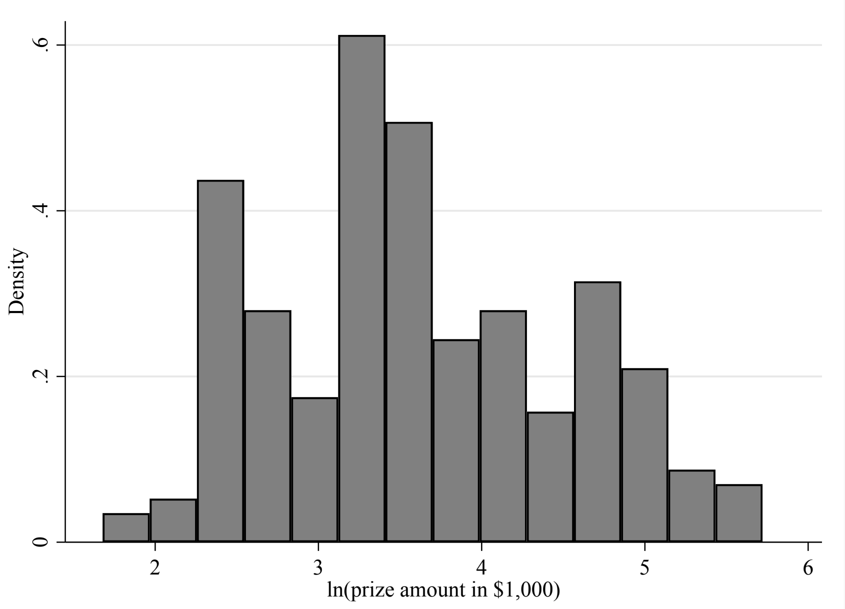
3 Example: Megabucks lottery winners

In this section, the association between the size of lottery winnings and subsequent labor market earnings is reanalyzed using the Hirano and Imbens (2004) survey data on Megabucks lottery winners in Massachusetts from the mid-1980s. Originally, the data were analyzed by Imbens, Rubin, and Sacerdote (2001). The dataset contains information on the prize amount measured in $1,000, labor earnings six years after winning the lottery, and some covariates on age, winning year, working status when winning the lottery, years of high school, years of college, an indicator for being male, the number of tickets bought, and previous earnings in the years one to six prior to winning. While the prize amount is randomly assigned, survey and item nonresponse lead to nonzero correlations of covariates with the treatment variable. The sample consists of $N = 202$ lottery winners with valid information on all covariates, the treatment variable, and the outcome. Following Hirano and Imbens (2004), the treatment variable is used in logarithmic terms. Although not necessary for the entropy balancing scheme to work, this transformation greatly increases efficiency of estimates in this case because it compresses the (relatively sparse) prize distribution at the top.

3.1 Preliminary data analysis

Before estimating balancing weights, we need to inspect the distribution of the treatment variable because inference in areas with low density or regions of support characterized by outliers may yield overly sensitive estimates of causal effects. Hence, these regions should be excluded from the analysis. Moreover, the overlap assumption is also most likely not fulfilled in regions of low density.

```
. use hirano04
. drop if year6==.
(35 observations deleted)
. generate ln_prize=ln(prize)
. label var ln_prize "ln(prize amount in $1,000)"
. summarize ln_prize,d
                 ln(prize amount in $1,000)
-------------------------------------------------------------
      Percentiles      Smallest
 1%      1.67818      .1301507
 5%     2.307573      1.609438
10%     2.420012       1.67818       Obs                 202
25%     2.835211      1.923372       Sum of wgt.         202
50%     3.471362                     Mean           3.591523
                        Largest      Std. dev.      .9505213
75%     4.271681      5.598792
90%      4.92195      5.720607       Variance       .9034907
95%     5.145924      5.778643       Skewness       .1239705
99%     5.720607      6.183716       Kurtosis       3.079105
. keep if inrange(ln_prize,r(p1),r(p99))
(4 observations deleted)
. histogram ln_prize, scheme(s2mono) graphregion(color(white))
(bin=14, start=1.6781801, width=.28874476)
```



Next an ordinary least-squares regression of ln(prize) on covariates is performed to get a sense of the associations of covariates with the treatment variable and the degree of predictiveness of covariates. The regression yields an R^2 of roughly 18% and an overall F statistic of 3.12. The corresponding p-value of 0.0003 shows that covariates significantly predict the size of lottery winnings. Regarding individual covariates, results show that age, male, and earnings four years prior to winning have the largest t statistics in absolute value. These findings underline the need to adjust for confounding before estimating the DRF because these covariates are also most likely to have a substantial impact on posttreatment earnings.

```
. regress ln_prize agew ownhs owncoll male tixbot workthen yearw
> yearm1 yearm2 yearm3 yearm4 yearm5 yearm6

      Source |       SS           df       MS      Number of obs   =       198
-------------+----------------------------------   F(13, 184)      =      3.12
       Model |  27.8778476        13  2.14444982   Prob > F        =    0.0003
    Residual |  126.308846       184  .68646112   R-squared       =    0.1808
-------------+----------------------------------   Adj R-squared   =    0.1229
       Total |  154.186694       197  .782673573   Root MSE        =    .82853

------------------------------------------------------------------------------
    ln_prize | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        agew |    .016769   .0048683     3.44   0.001     .0071641    .0263738
       ownhs |  -.0384305   .0636275    -0.60   0.547    -.1639639    .0871029
     owncoll |   .0296192     .04166     0.71   0.478    -.0525735     .111812
        male |   .4214513   .1362625     3.09   0.002     .1526134    .6902891
      tixbot |  -.0227525   .0188385    -1.21   0.229    -.0599198    .0144147
    workthen |   .0530765   .1686043     0.31   0.753    -.2795698    .3857228
       yearw |   .0254354   .0476658     0.53   0.594    -.0686064    .1194771
      yearm1 |  -.0019863   .0110822    -0.18   0.858    -.0238509    .0198782
      yearm2 |   .0029893   .0177104     0.17   0.866    -.0319521    .0379308
      yearm3 |    .003186   .0174775     0.18   0.856     -.031296     .037668
      yearm4 |   .0292613   .0174843     1.67   0.096    -.0052341    .0637568
      yearm5 |  -.0248633   .0156492    -1.59   0.114    -.0557383    .0060117
      yearm6 |    .000125     .01111     0.01   0.991    -.0217942    .0220443
       _cons |   2.450569   .4805153     5.10   0.000     1.502541    3.398597
------------------------------------------------------------------------------
```

3.2 Estimation of balancing weights

To account for these associations, we use the `ebct` command to estimate covariate balancing weights. Because an additional analysis has shown that none of the higher-order terms of the multivalued covariates significantly improve the fit of the above regression model, only linear terms of all covariates are included in the estimation of balancing weights.

```
. ebct agew ownhs owncoll male tixbot workthen yearw
> yearm1 yearm2 yearm3 yearm4 yearm5 yearm6, treatvar(ln_prize)
Note: Balancing weights are estimated such that T,...,T^P is uncorrelated with
> X, P=1

Estimating balancing weights...

Iteration 0:  f(p) = -2.220e-15
Iteration 1:  f(p) =  .11839583
Iteration 2:  f(p) =  .12816522
Iteration 3:  f(p) =  .12824004
Iteration 4:  f(p) =  .12824004

###########################################
# Summary statistics on balancing quality #
###########################################

Summary statistics from a (weighted) regression of T on X:
```

	before balancing	after balancing
R-squared	0.181	0.000
F-statistic	3.124	0.000
p-value	0.000	1.000

```
###########################################
#     Details on balance and weights      #
###########################################

Balancing statistics from a regression of each element of X on a fourth-order
polynomial in T:
```

	Before Balancing		After Balancing	
	R-square	p-value	R-square	p-value
agew	0.063	0.013	0.001	0.992
ownhs	0.002	0.979	0.005	0.916
owncoll	0.008	0.821	0.002	0.977
male	0.084	0.002	0.041	0.087
tixbot	0.034	0.156	0.027	0.248
workthen	0.006	0.888	0.014	0.595
yearw	0.069	0.008	0.059	0.020
yearm1	0.039	0.102	0.018	0.479
yearm2	0.055	0.027	0.013	0.622
yearm3	0.065	0.011	0.008	0.799
yearm4	0.078	0.003	0.009	0.770
yearm5	0.066	0.010	0.021	0.403
yearm6	0.060	0.017	0.013	0.650

```
Distribution of estimated weights:
```

	value
min	0.000
1st percentile	0.000
5th percentile	0.001
10th percentile	0.002
25th percentile	0.003
Median	0.005
75th percentile	0.006
90th percentile	0.008
95th percentile	0.010
99th percentile	0.013
max	0.015

After estimating balancing weights, `ebct` automatically provides the user with global summary statistics on the balancing quality before and after reweighting. These results are stored in `e(balance)`. The R^2, overall F statistic, and corresponding p-value before balancing stem from a regression of t_i on x_i and are identical to the ones previously presented. In the reweighted sample, balancing indicators are always $R^2 = F = 0$ and p-value $= 1$—implying that covariates are uncorrelated with the treatment variable—unless there are some numerical difficulties when estimating weights. This can be the case if there are holes in the support of the treatment variable, if the distribution of the treatment variable is extremely skewed, or if covariates are almost perfectly collinear. In that case, the treatment variable may need to be transformed or an adjustment to the sample or the specification may be necessary to achieve convergence.

In addition to these global statistics, `ebct` also provides more detailed balancing measures. These are obtained from a linear regression of each covariate on a fourth-order polynomial in the treatment variable. Results from these regressions are stored in `e(balance_detail)`. These results provide evidence on whether there are dependencies between higher orders of the treatment variable and covariates beyond simple correlations. If the balancing indeed works as intended and covariates are truly balanced, regression R^2 and p-values should be fairly close to 0 and 1, respectively. If this is not the case, one should rerun the estimation of balancing weights with $p > 1$ and inspect balance again. In this case, balance has improved considerably through reweighting. The regression R^2 for all covariates has decreased substantially, most are close to zero as intended. Moreover, all covariates except the male indicator and the winning year are no longer statistically associated with the treatment variable on common levels because their p-values > 0.1. Next one may have attempted to increase p until all p-values are above this threshold. In this case, however, the sample is probably too small to justify this step and risk potentially large increases in the variance of estimates.

Next summary statistics on the weight distribution are inspected. This is important to make sure that there are no extreme weights, which would likely substantially reduce the performance of resulting effect estimates (Robins and Wang 2000; Kang and Schafer 2007). Should this be the case, one may opt for an iterative estimation strategy by estimating balancing weights and truncating weights above a certain threshold. For binary treatments, Imbens (2004) suggested a threshold of 4%. However, in the context of continuous treatments, the threshold should probably be lower, for example, 2%. Truncated weights can then be used as new base weights when reestimating balancing weights using `ebct` in the next iteration. Estimated weights from the second iteration should still eradicate the correlation between covariates and the treatment variable but display lower maximum weights. Alternatively, it may be worthwhile to split the sample according to covariates that are highly predictive of the treatment intensity and perform the analysis separately for each subsample. This is likely to reduce problems with extreme weights as well. However, in this application such actions do not seem to be necessary, because the largest weight of roughly 1.5% is relatively small despite the small sample.

3.3 Estimating the DRF and its derivative

Next the `ebct` command is used to estimate the DRF of lottery winnings regarding subsequent labor earnings. To do so, we must include the outcome variable in the option `out()`. To obtain standard errors, we specify the `bootstrap` option. The number of replications is set to 200 using the option `reps()`. The seed for the resampling process is left at the default value set by the `ebct` command. A graph is produced by specifying the `graph` option.

```
. quietly ebct agew ownhs owncoll male tixbot workthen yearw yearm1 yearm2
> yearm3 yearm4 yearm5 yearm6, treatvar(ln_prize)
> out(year6) est(drf) bootstrap reps(200) graph
```

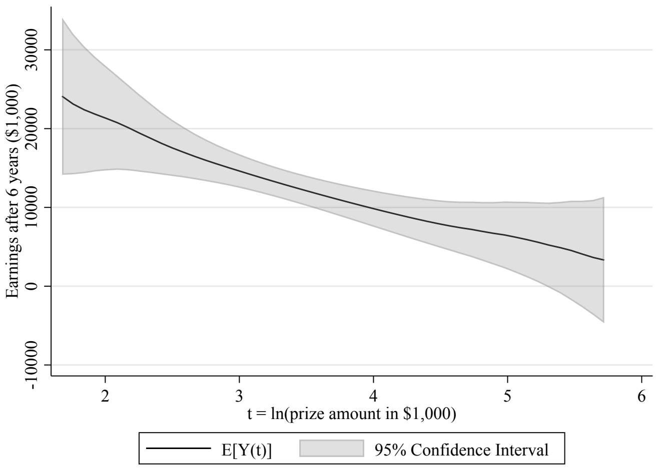


Figure 1. Dose–response function—$E\{y_i(t)\}$

The estimates of the DRF clearly show a downward-sloping relationship between the size of lottery winnings and posttreatment earnings. Confidence bands show that estimates are more noisy in the tails of the ln(prize) distribution. Average posttreatment earnings would have been around \$25,000 had all units received only minimal prize amounts. With an increasing size of prizes, average postwinning earnings decrease substantially. When the prize exceeds about exp(5.3) $\times$ \$1,000 $\approx$ \$200,000, average earnings become statistically indistinguishable from 0 at the 95% level.

Especially when there is no natural comparison level regarding the outcome, the DRF can be somewhat difficult to interpret by itself. Often, the derivative of the DRF has better interpretability because one may infer some "optimal" dose of the treatment (see, for example, Becker, Egger, and von Ehrlich [2012] for such reasoning). To estimate the derivative instead of the DRF, we must augment the `ebct` command by the option `est(derivative)`:

```
. quietly ebct agew ownhs owncoll male tixbot workthen yearw yearm1 yearm2
> yearm3 yearm4 yearm5 yearm6, treatvar(ln_prize)
> out(year6) est(derivative) bootstrap reps(200) graph
```

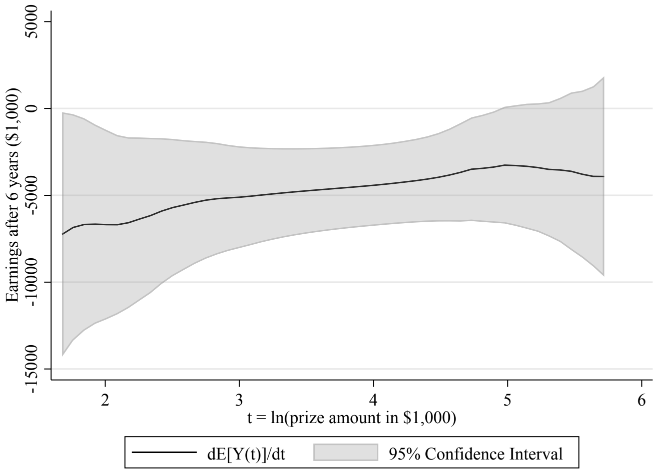
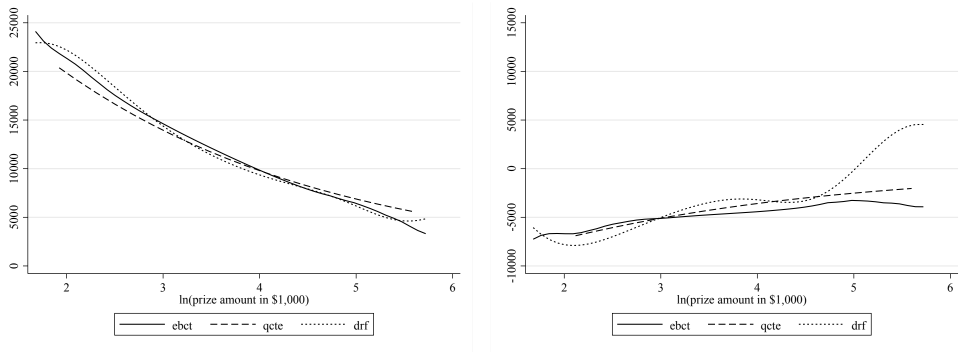


Figure 2. Derivative of the DRF—$dE\{y_i(t)\}/dt$

In line with standard economic theory, the slope of the DRF is negative and decreasing in absolute value. After winners win around exp(5) $\times$ \$1,000 $\approx$ \$150,000, there is no significant marginal effect of prize winnings on subsequent earnings.

3.4 Comparison with other estimation commands

In the following, estimation results for the DRF and its derivative are briefly compared with results from other estimation approaches available in Stata. The commands used for comparison are `drf` (Bia et al. 2014) and `qcte` (Alejo, Galvao, and Montes-Rojas 2020). Both approaches first estimate the GPS parametrically to obtain balancing weights. The approach by Bia et al. (2014) then estimates effects using kernel averaging on the reweighted sample. Alejo, Galvao, and Montes-Rojas (2020) take a parametric approach in the second step, but the command can also estimate quantile treatment effects. See both articles for details. The following graph depicts point estimates for the three approaches in comparison, showing that they all yield very similar estimates of the DRF and its derivative. See the appendix for the code to reproduce this graph.



(a) DRF $E\{y_i(t)\}$ (b) Derivative $dE\{y_i(t)\}/dt$

Figure 3. Command comparison

4 Conclusion

This article presents a new command, `ebct`, for the estimation of DRFs and their derivatives using EBCT and local linear kernel regressions. I illustrated its use by reanalyzing the Megabucks lottery winner dataset and estimating the relationship between lottery winnings and subsequent earnings, showing a downward sloping DRF. The inspection of covariate balance revealed some remaining imbalance after weighting, which ideally should be avoided if one aims to minimize bias in resulting estimates. In other applications with larger samples, one may wish to also render higher orders of the treatment variable uncorrelated with covariates to minimize the risk of such bias. Indeed, realistic simulations have shown that doing so reduces bias and root mean squared error of resulting estimates (Tübbicke 2022). For further guidance, future research should inspect how to choose the power up to which the treatment variable is rendered uncorrelated with covariates optimally, depending on key characteristics of the data at hand.

5 Programs and supplemental materials

To install a snapshot of the corresponding software files as they existed at the time of publication of this article, type

```
. net sj 23-3
. net install st0726      (to install program files, if available)
. net get st0726          (to install ancillary files, if available)
```

6 References

Alejo, J., A. F. Galvao, and G. Montes-Rojas. 2018. Quantile continuous treatment effects. *Econometrics and Statistics* 8: 13–36. https://doi.org/10.1016/j.ecosta.2017.10.004.

———. 2020. A practical generalized propensity-score estimator for quantile continuous treatment effects. *Stata Journal* 20: 276–296. https://doi.org/10.1177/1536867X20930997.

Becker, S. O., P. H. Egger, and M. von Ehrlich. 2012. Too much of a good thing? On the growth effects of the EU's regional policy. *European Economic Review* 56: 648–668. https://doi.org/10.1016/j.euroecorev.2012.03.001.

Bia, M., C. A. Flores, A. Flores-Lagunes, and A. Mattei. 2014. A Stata package for the application of semiparametric estimators of dose–response functions. *Stata Journal* 14: 580–604. https://doi.org/10.1177/1536867X1401400307.

Bia, M., and A. Mattei. 2008. A Stata package for the estimation of the dose–response function through adjustment for the generalized propensity score. *Stata Journal* 8: 354–373. https://doi.org/10.1177/1536867X0800800303.

———. 2012. Assessing the effect of the amount of financial aids to Piedmont firms using the generalized propensity score. *Statistical Methods and Applications* 21: 485–516. https://doi.org/10.1007/s10260-012-0193-4.

Cameron, A. C., and P. K. Trivedi. 2022. *Microeconometrics Using Stata*. 2nd ed. College Station, TX: Stata Press.

Choe, C., A. Flores-Lagunes, and S.-J. Lee. 2015. Do dropouts with longer training exposure benefit from training programs? Korean evidence employing methods for continuous treatments. *Empirical Economics* 48: 849–881. https://doi.org/10.1007/s00181-014-0805-y.

Crump, R. K., V. J. Hotz, G. W. Imbens, and O. A. Mitnik. 2009. Dealing with limited overlap in estimation of average treatment effects. *Biometrika* 96: 187–199. https://doi.org/10.1093/biomet/asn055.

Efron, B., and R. J. Tibshirani. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* 1: 54–75. https://doi.org/10.1214/ss/1177013815.

Fan, J. 1992. Design-adaptive nonparametric regression. *Journal of the American Statistical Association* 87: 998–1004. https://doi.org/10.2307/2290637.

Flores, C. A., A. Flores-Lagunes, A. Gonzalez, and T. C. Neumann. 2012. Estimating the effects of length of exposure to instruction in a training program: The case of Job Corps. *Review of Economics and Statistics* 94: 153–171. https://doi.org/10.1162/REST_a_00177.

Fong, C., C. Hazlett, and K. Imai. 2018. Covariate balancing propensity score for a continuous treatment: Application to the efficacy of political advertisements. *Annals of Applied Statistics* 12: 156–177. https://doi.org/10.1214/17-AOAS1101.

Galdo, J., and A. Chong. 2012. Does the quality of public-sponsored training programs matter? Evidence from bidding processes data. *Labour Economics* 19: 970–986. https://doi.org/10.1016/j.labeco.2012.08.001.

Hainmueller, J. 2012. Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis* 20: 25–46. https://doi.org/10.1093/pan/mpr025.

Hainmueller, J., and Y. Xu. 2013. ebalance: A Stata package for entropy balancing. *Journal of Statistical Software* 54(7): 1–18. https://doi.org/10.18637/jss.v054.i07.

Heckman, J., and R. Robb, Jr. 1985. Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics* 30: 239–267. https://doi.org/10.1016/0304-4076(85)90139-3.

Heckman, J. J., H. Ichimura, and P. E. Todd. 1997. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic Studies* 64: 605–654. https://doi.org/10.2307/2971733.

Hirano, K., and G. W. Imbens. 2004. The propensity score with continuous treatments. In *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, ed. A. Gelman and X.-L. Meng, 73–84. Chichester, U.K.: Wiley. https://doi.org/10.1002/0470090456.ch7.

Imai, K., and D. A. van Dyk. 2004. Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association* 99: 854–866. https://doi.org/10.1198/016214504000001187.

Imbens, G. W. 2000. The role of the propensity score in estimating dose–response functions. *Biometrika* 87: 706–710. https://doi.org/10.1093/biomet/87.3.706.

———. 2004. Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics* 86: 4–29. https://doi.org/10.1162/003465304323023651.

Imbens, G. W., D. B. Rubin, and B. I. Sacerdote. 2001. Estimating the effect of unearned income on labor earnings, savings, and consumption: Evidence from a survey of lottery players. *American Economic Review* 91: 778–794. https://doi.org/10.1257/aer.91.4.778.

Imbens, G. W., and J. M. Wooldridge. 2009. Recent developments in the econometrics of program evaluation. *Journal of Economic Literature* 47: 5–86. https://doi.org/10.1257/jel.47.1.5.

Jann, B. 2021a. ebalfit: Stata module to perform entropy balancing. Statistical Software Components S458975, Department of Economics, Boston College. https://ideas.repec.org/c/boc/bocode/s458975.html.

———. 2021b. Entropy balancing as an estimation command. Working Papers 39, University of Bern Social Sciences. https://ideas.repec.org/p/bss/wpaper/39.html.

Kang, J. D. Y., and J. L. Schafer. 2007. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science* 22: 523–539. https://doi.org/10.1214/07-STS227.

Kluve, J., H. Schneider, A. Uhlendorff, and Z. Zhao. 2012. Evaluating continuous training programmes by using the generalized propensity score. *Journal of the Royal Statistical Society, Series A* 175: 587–617. https://doi.org/10.1111/j.1467-985X.2011.01000.x.

Kreif, N., R. Grieve, I. Díaz, and D. Harrison. 2015. Evaluation of the effect of a continuous treatment: A machine learning approach with an application to treatment for traumatic brain injury. *Health Economics* 24: 1213–1228. https://doi.org/10.1002/hec.3189.

Kullback, S. 1997. *Information Theory and Statistics*. Mineolo, NY: Dover.

Lechner, M. 2001. Identification and estimation of causal effects of multiple treatments under the conditional independence assumption. In *Econometric Evaluation of Labour Market Policies*, ed. M. Lechner and F. Pfeiffer, 43–58. Heidelberg: Physica-Verlag. https://doi.org/10.1007/978-3-642-57615-7_3.

Lechner, M., and A. Strittmatter. 2019. Practical procedures to deal with common support problems in matching estimation. *Econometric Reviews* 38: 193–207. https://doi.org/10.1080/07474938.2017.1318509.

Li, J., and M. W. Fraser. 2015. Evaluating dosage effects in a social-emotional skills training program for children: An application of generalized propensity scores. *Journal of Social Service Research* 4: 345–364. https://doi.org/10.1080/01488376.2014.994797.

MacKinnon, J. G. 2006. Bootstrap methods in econometrics. *Economic Record* 82: S2–S18. https://doi.org/10.1111/j.1475-4932.2006.00328.x.

Manski, C. F. 2013. Identification of treatment response with social interactions. *Econometrics Journal* 16: S1–S23. https://doi.org/10.1111/j.1368-423X.2012.00368.x.

Mitze, T., A. R. Paloyo, and B. Alecke. 2015. Is there a purchase limit on regional growth? A quasi-experimental evaluation of investment grants using matching techniques. *International Regional Science Review* 38: 388–412. https://doi.org/10.1177/0160017613505200.

Racine, J. S. 2016. Local polynomial derivative estimation: Analytic or Taylor? In *Advances in Econometrics, Vol. 36: Essays in Honor of Aman Ullah*, ed. G. González-Rivera, R. C. Hill, and T.-H. Lee, 617–633. Bingley, U.K.: Emerald. https://doi.org/10.1108/S0731-905320160000036027.

Robins, J. M., and N. Wang. 2000. Inference for imputation estimators. *Biometrika* 87: 113–124. https://doi.org/10.1093/biomet/87.1.113.

Roy, A. D. 1951. Some thoughts on the distribution of earnings. *Oxford Economic Papers* 3: 135–146. https://doi.org/10.1093/oxfordjournals.oep.a041827.

Rubin, D. B. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66: 688–701. https://doi.org/10.1037/h0037350.

———. 1980. Comment on Basu, D.—Randomization analysis of experimental data: The Fisher randomization test. *Journal of the American Statistical Association* 75: 591–593. https://doi.org/10.2307/2287653.

Tübbicke, S. 2022. Entropy balancing for continuous treatments. *Journal of Econometric Methods* 11: 71–89. https://doi.org/10.1515/jem-2021-0002.

Wu, X., F. Mealli, M.-A. Kioumourtzoglou, F. Dominici, and D. Braun. Forthcoming. Matching on generalized propensity scores with continuous exposures. *Journal of the American Statistical Association*. https://doi.org/10.1080/01621459.2022.2144737.

About the author

Stefan Tübbicke is a postdoctoral researcher at the Institute of Employment Research of the Federal Employment Agency in Nuremberg, Germany.

Appendix

```
. * ebct
. quietly ebct agew ownhs owncoll male tixbot workthen yearw yearm1 yearm2
> yearm3 yearm4 yearm5 yearm6, treatvar(ln_prize)
> out(year6) est(drf) generate(ebct1)
. quietly ebct agew ownhs owncoll male tixbot workthen yearw yearm1 yearm2
> yearm3 yearm4 yearm5 yearm6, treatvar(ln_prize)
> out(year6) est(derivative) generate(ebct2)
. * Bia et al. (2014)
. quietly drf agew ownhs owncoll male tixbot workthen yearm1 yearm2 yearm3
> yearm4 yearm5 yearm6, outcome(year6) treatment(ln_prize)
> test(L_like) npoints(49) numoverlap(3) method(iwkernel)
> family(gaussian) link(identity) nolog(1) delta(1)
. matrix A = e(b)
. matrix drfBia2014 = A[1,1..50]'
. matrix derivBia2014 = A[1,51..100]'
. quietly svmat drfBia2014
. quietly svmat derivBia2014
```

```
. * Alejo, Galvao, and Montes-Rojas (2020)
. set seed 123
. quietly qcte year6 ln_prize, ynotrans xvar(agew yearw)
> zvar(male ownhs owncoll tixbot workthen yearm1 yearm2
> yearm3 yearm4 yearm5 yearm6) reps(5)
. matrix drfAlejo2020 = r(QDRFplot)
. matrix derivAlejo2020 = r(QCTEplot)
. quietly svmat drfAlejo2020
. quietly svmat derivAlejo2020
. * DRF graph
. twoway (line ebct1_est ebct1_grid) (line drfAlejo20202 drfAlejo20201)
>        (line drfBia2014 ebct1_grid), legend(rows(1))
> legend(label(1 "ebct")) legend(label(2 "qcte"))
> legend(label(3 "drf")) ylabel(0 (5000) 25000)
> scheme(s2mono) graphregion(color(white))
> xtitle("ln(prize amount in $1,000)")
. quietly graph export comparison_drf.pdf, replace
. * Rescale qcte-estimates of derivatives so that delta = 1
. quietly summarize derivAlejo20201 if _n <= 2
. quietly replace derivAlejo20202 = derivAlejo20202/(r(max)-r(min))
. * Derivative graph
. twoway (line ebct2_est ebct2_grid) (line derivAlejo20202 derivAlejo20201)
>        (line derivBia2014 ebct2_grid), legend(rows(1))
> legend(label(1 "ebct")) legend(label(2 "qcte"))
> legend(label(3 "drf")) ylabel(-10000 (5000) 15000)
> scheme(s2mono) graphregion(color(white))
> xtitle("ln(prize amount in $1,000)")
. quietly graph export comparison_derivative.pdf, replace
```