Memo Detailing Revisions: RJournal 2021-167

"rbw: An R package for Constructing Residual Balancing Weights"

We thank the reviewers for the helpful comments on the paper and the R package accompanying it. In response, we have made the following changes to the manuscript:

- Added a footnote to the introduction, which acknowledges the existence of doubly robust methods for estimating causal effects of continuous exposures and points to an R package for implementing them.
- Simplified the point treatment example to avoid extraneous details.
- Moved the code used to create estimates from alternative methods (CBPS and IPW) to the supplementary materials.
- Added results from the `ipw` package to Table 1.
- Removed bulleted lists of the variables included in our data sets.
- Included comparisons of computational performance between RBW and CBPS for the example involving a time-varying treatment.
- Added as a limitation of the method and the package the fact that RBW relies on the "sandwich" variance estimator instead of an alternative tailored to addressing the estimation uncertainty introduced by the RBW weights.
- Acknowledged the possibility of expanding RBW to survival data.

Additionally, we have upgraded the package in the following ways:

- Used the functions from the `usethis` package to add several automatic unit tests to `rbw`.
- Renamed the function `rbwATE()` to `rbwPoint()`.
- Included a comprehensive set of argument checks to all the main functions.
- Changed the functions `rbwPoint()` and `rbwMed()` to allow the user to pass a character vector to the `baseline_x` argument.
- Included a `CITATION` file for the package.

We are confident that these revisions have improved the paper and the package considerably. Next, we provide detailed responses to the reviewer comments. The first section addresses changes to the article, and the second, changes to the package. We reproduce direct quotes in bold font.

Article

Response to Reviewer 1

Causal inference is a hot topic in statistics and data science. Compared to when treatment is binary, few methodologies address problems where the analyst is trying to assess the causal effect of a continuous variable. Therefore I do think this article is important to the R community.

We appreciate that you have found our work relevant to the R community. Below we respond to your suggestions and describe the changes they have motivated. We believe that these revisions have greatly improved the paper.

I recommend the authors comment on other approaches for estimating the causal effect of continuous exposures (i.e., G-computation, doubly-robust methods) and point to R packages for these.

We have added a footnote to the introduction to acknowledge doubly robust methods for estimating causal effects with continuous exposures (p. 2):

> Other methods for estimating causal effects of continuous exposures include doubly robust estimators, which model both the treatment and outcome processes and give consistent estimates as long as one of these models is correctly specified. For example, Kennedy et al. (2017) introduce an approach that does not require any parametric assumptions about the effect curve. Instead, it uses flexible data-adaptive methods both to estimate the treatment and outcome models and to subsequently fit the dose-response curve. Readers can install the R package that implements this method from GitHub (Kennedy 2021).

The first example demonstrates how to use `rbwATE()` with an application to the effects of political advertisements on campaign contributions. This example is too complex. Examples should be minimal in the sense that they demonstrate only enough of the functionality that the analyst understands how to use the software properly; anything extra just creates mental clutter. A more helpful example would avoid the complexity of the treatment variable transformation.

Thank you for your comment. We agree that the original example had too many extraneous details that diverted it from its primary goal of illustrating how the function works. Instead of developing a new example from scratch, we have decided to adapt the existing one. We have opted for this approach because we believe the current point treatment situation — i.e., the effects of political advertisements on campaign contributions — aligns well with the other two examples in the paper. As shown in the introduction and conclusion, our work speaks to a social science audience throughout. We believe that showing RBW's applicability across a range of contexts in the political science literature can help us draw attention to our method's usefulness in many situations encountered in the social sciences more broadly.

That said, we have made considerable changes to the first example:

- Recall that CBPS assumes that the treatment variable is normally distributed. Instead of Fong et al.'s (2018) Box-Cox transformation, we now use a log transformation of our treatment variable, the total number of political advertisements in a zip code. Q-Q plots show no sizable differences between the two approaches in achieving normality.
- Our MSM no longer includes a quadratic term for the transformed treatment. By doing so, we avoid the need to consider the covariance between the coefficients of the linear and quadratic terms when constructing the standard error for our estimated causal effect.
- We have removed the square terms of all continuous covariates from the treatment model for CBPS and IPW. As a result, both the treatment model for these methods and the vector of baseline confounders passed to `rbwPoint()` now include only the linear terms of our covariates.

The authors also don't provide an interpretation of the MSM effect estimate and assume the analyst understands how to interpret such an MSM.

We have added an interpretation of the estimated causal effects in the paragraph describing the Table 1 findings (p. 10):

> All methods yield relatively similar point estimates and standard errors. In particular, they indicate that going from zero to 1,000 political advertisements increases campaign contributions by around four thousand dollars, on average, although the point estimates from CBPS and IPW are slightly larger than those produced by RBW and npCBPS.

Much of this example is spent demonstrating how to use *CBPS* to estimate the same effect; estimating the effect using *CBPS* doesn't add to the analysts understanding of how to use the `rbw` package and is therefore unnecessary. While the results in Table 1 are illustrative (especially the timing results) I recommend moving the code used to create the estimates from `CBPS` and `npCBPS` to the supplementary materials.

Thank you for pointing this issue out. Indeed, the original manuscript spent much of the point treatment example describing how to estimate CBPS weights. In response to your comment, we have moved the code detailing the construction of the CBPS and IPW weights to the supplementary material.

I would also recommend adding similar results from the `ipw` package to this table.

We have added similar results from the `ipw` package to Table 1. As mentioned above, readers can find the code for constructing the IPW weights in the supplementary material.

Each example includes lists of the covariates used therein as bullet points. This seems unnecessary. A sentence stating what covariates are included as confounders would suffice.

We have removed bulleted lists of the variables included in our data sets — we now use only complete sentences to state the variables used in our models.

Response to Reviewer 2

This paper describes the R package `rbw`, which implements the method of residual balancing weights (RBW) for estimating marginal structural models (MSMs) based on the methodological paper of (Zhou and Wodtke 2020). This package solves a relevant problem for researchers working with longitudinal data. Making it relevant for researchers in the R community that work with observational data. The approach is an extension of the entropy balancing procedure developed by (Hainmueller 2012) (see Zhou and Wodtke 2020) to the longitudinal setting and it expands the available toolkit that researchers have to estimate causal effects in the context of observational data.

We appreciate that you have found our work relevant to the R community. Below we respond to your suggestions and describe the changes they have motivated. We believe that these revisions have greatly improved the paper.

The authors provide three motivations examples to showcase the performance of the package: (1) the estimation of the average treatment effect in a cross-sectional observational study, (2) a mediation analysis, and (3) the estimation of effects in the context of time-varying treatments in longitudinal data. For the first example, the author compares the estimate of different approaches and the computation time (in seconds). RBW is able to produce results in a fraction of the time of its competitors. However the same table is not produced for the second and third motivating examples. I would encourage the authors to add the comparative performance of RBW in other settings beyond the ones presented in their first example, specially metrics regarding computational performance.

Thank you for this valuable suggestion. Unfortunately, we are limited in the extent to which we can compare RBW to alternative methods in the last two examples. This is because these methods have not been extended to causal mediation analysis and only partially to contexts involving time-varying treatments. Indeed, CBPS does not work with continuous treatments in longitudinal settings, and its nonparametric version can only handle point treatment situations. We explain this in the article (p. 15):

> We conclude by comparing RBW's computational performance to CBPS for a dichotomized version of the treatment representing whether more than 10% of the candidate's advertising was negative for each week. We use this dichotomized variable because CBPS has not been extended to work with continuous treatments in longitudinal settings. Additionally, we do not compare RBW to npCBPS because the latter's use is restricted to point treatment situations. Lastly, though we can construct IPW weights almost immediately, Zhou and Wodtke (2020) show that IPW yields considerably larger effect estimates than RBW and CBPS for this particular case (likely due to the method's susceptibility to model misspecification), so we do not report the IPW results.

The following excerpt contrasts the computation times (p. 16):

> Whereas RBW takes less than one second to construct the weights, CBPS takes much longer. Hence, the longitudinal setting presents the same pattern we saw above for the point treatment situation: RBW has considerable gains in computational performance over alternative methods of constructing weights for MSMs. Since our focus here is computational performance, we omit the effect estimates for the dichotomized treatment. As shown in Zhou and Wodtke (2020), the

> results are broadly consistent with those based on the continuous treatment, with RBW and CBPS yielding similar point estimates.

Furthermore, to my understanding, the RWB method does not provide an estimator for the variance covariance matrix. Thus, users need to implement a robust estimation of the variance covariance matrix in order to get appropriate coverage of the confidence intervals. The use of the robust standard errors (SE) has been justified solely via simulations. I would encourage the authors to add this as a limitation of the package (and the method).

We have acknowledged this point in the paragraph describing RBW's limitations (p. 17):

> Finally, we have relied on the "sandwich" variance estimator to compute standard errors. Though Zhou and Wodtke (2020) demonstrate by simulation studies that this estimator is likely conservative for RBW, the method does not provide a variance estimator tailored to addressing the estimation uncertainty of the RBW weights.

Finally, I wonder if this same approach could also be used to account for censoring and if it could be used in the context of survival data. If so, I would also encourage the authors to add comments about that in the article, since this is also a relevant topic to public health research.

We appreciate your remark. Though an in-depth exploration of our method's application to survival data is beyond the scope of this paper, the final paragraph now indicates our plan to expand RBW in this direction (p. 17):

> We will continue to upgrade the package by expanding RBW's ranges of applicability — specifically, to censored data, to contexts involving repeated outcome measures including survival data (Hernán, Brumback, and Robins 2002, 2000), and to cases where, as discussed above, exact balance is infeasible and approximate balance must be pursued.

Package

Response to Reviewer 1

The package does not have any automatic unit tests. Without automatic testing it is impossible to know that any future changes to the software don't break things or that the current code is doing what it's supposed to. While 100% test coverage is not necessary, some level of automatic testing is needed to demonstrate that the software is robust.

Thank you for stressing the importance of automated unit testing for making the package robust. We have used the functions from the `usethis` package to add several automatic unit tests to `rbw`. In particular, we have written tests that:

(1) check whether the functions throw an error message when the user omits the data;
(2) check whether the functions throw an error message when the user omits the treatment;
(3) check whether the functions throw an error message when the base weights are not the same length as the data;
(4) check whether the user supplies the confounders and models necessary for running the models; and
(5) check whether the output for each function has the expected structure — i.e., a list of four objects.

Core functions don't perform complete argument checks. Providing internal argument checks is an easy way to improve the experience of the analyst when using an R package. As an example, the documentation for `rbwPoint()` states the argument, `baseline_x`, is "[a]n expression for a set of baseline confounders stored in data." This is rather ambiguous as to if the analyst should pass a character vector of these column names or something else. The former would be standard in R. Doing so, however, produces this uninformative error: `Error in terms.formula(object, data = data) : invalid model formula in ExtractVars`. Including an argument check with an informative error message would avoid this issue altogether (packages such as checkmate [https://CRAN.Rproject.org/package=checkmate] can help with this).

On this note, it is confusing that the treatment argument accepts a character string while the `baseline_x` argument doesn't.

Thank you for your remarks. We have now included a comprehensive set of argument checks. For example, all functions now check if the variables required for running the models are included in the data and whether the user has specified them as symbols or character strings. Moreover, we have changed the functions `rbwPoint()` and `rbwMed()` to allow the user to pass a character vector to the `baseline_x` argument.

The software doesn't include a `CITATION` file. I recommend populating a `CITATION` file (https://devguide.ropensci.org/building.html#citation-file) that refers to both the manual and the paper describing the methods.

We have used the function `usethis::use_citation()` to create a boilerplate `CITATION` file, which we later populated with the appropriate references: the package's reference manual on CRAN and Zhou and Wodtke's (2020) paper introducing RBW. We will add a reference to the current article in due course.

The naming convention for `rbwPoint()` is inconsistent with `rbwMed()` and `rbwPanel()`. "Med" and "Panel" refer to the type of data required for those functions while "ATE" refers to a statistical estimand. Renaming `rbwPoint()` to something like `rbwPoint()` would be more consistent.

Thank you for catching this inconsistency in the function names. We have accepted your suggestion and renamed `rbwATE()` to `rbwPoint()`.

The API across `rbwATE()`, `rbwMed()`, and `rbwPanel()` is inconsistent. `rbwATE()` fits the models required to compute the response residuals using an internal call to `lm()`. `rbwMed()` and `rbwPanel()` require the analyst to fit these models themselves. This makes `rbwMed()` and `rbwPanel()` much more complex to use. If the authors wish to use this form of the API, why not just have the user pass a list of the residuals instead of the models? Doing so would allow models outside of those inheriting from `lm`.

We appreciate your thorough examination of the functions' inner workings. However, we are afraid that having the user pass a list of the residuals instead of the models to `rbwMed()` and `rbwPanel()` is impracticable considering the balancing conditions involved in constructing the weights. To see why, let A_t represent a time-varying treatment at time t and L_t be a vector of time-varying confounders at time t. Recall from the paper that RBW invokes a model for the conditional mean of L_2 given L_1 and A_1 and balances the residuals from this model across levels of A_2 and levels of (L_1, A_1) (the observed past). Analogously, in the context of causal mediation analysis, we have models for the conditional means of post-treatment confounders given the treatment and baseline confounders, and the method balances the residuals from this model across levels of the mediator and levels of the treatment and baseline confounders. As Zhou and Wodtke (2020) point out, when the GLMs are Gaussian, binomial, or Poisson regressions with canonical links, the response residuals are orthogonal to the regressors in the original *unweighted* sample. However, we need to include the regressors in the set of functions $H(\bar{L}_{t-1}, \bar{A})$ constituting the balancing conditions described in Equation 6 (p. 3 of the paper) to ensure orthogonality in the *weighted* sample. If users were to pass the residuals only, `rbwMed()` and `rbwPanel()` would not have access to the regressors and, therefore, would not be able to balance the residuals against them in the weighted sample using an internal call to `eb2()`. By contrast, `rbwPoint()` computes the response residuals for each baseline confounder by centering it around its sample mean. That is, the required models are a set of regressions without predictors, with the baseline confounders as the response variables, so we decide to include an internal call to `lm()` in `rbwPoint()` instead of asking users to pass a list of constant-only models.

Response to Reviewer 2

I would encourage the authors to flesh-out more the built-in documentation.

Yes, we have elaborated more on the logic and implementation of each of the four functions in the built-in documentation.

I would also suggest that the package should provide by default robust SE, since a naive implementation of this package may lead to incorrect inferences.

Your remark that a naive implementation of MSMs may lead to incorrect inferences is entirely relevant. However, like `CBPS` and `ipw`, `rbw` constructs the weights and lets users estimate the actual MSMs. We believe that limiting the package's scope makes for added flexibility in that users can use the same set of weights to try out different modeling specifications and alternative methods for estimating standard errors (e.g., "sandwich," bootstrap, jackknife, etc.).

Bibliography

Fong, Christian, Chad Hazlett, and Kosuke Imai. 2018. "Covariate Balancing Propensity Score for a Continuous Treatment: Application to the Efficacy of Political Advertisements." *The Annals of Applied Statistics* 12 (1). https://doi.org/10.1214/17-AOAS1101.

Hainmueller, Jens. 2012. "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies." *Political Analysis* 20 (1): 25–46. https://doi.org/10.1093/pan/mpr025.

Hernán, Miguel A., Babette A. Brumback, and James M. Robins. 2000. "Marginal Structural Models to Estimate the Causal Effect of Zidovudine on the Survival of HIV-Positive Men." *Epidemiology (Cambridge, Mass.)* 11 (5): 561–70. https://doi.org/10.1097/00001648-200009000-00012.

———. 2002. "Estimating the Causal Effect of Zidovudine on Cd4 Count with a Marginal Structural Model for Repeated Measures." *Statistics in Medicine* 21 (12): 1689–709. https://doi.org/10.1002/sim.1144.

Kennedy, Edward H. 2021. *Npal: Nonparametric Causal Inference Methods*. https://github.com/ehkennedy/npcausal/blob/master/npcausal.pdf.

Kennedy, Edward H., Zongming Ma, Matthew D. McHugh, and Dylan S. Small. 2017. "Non-Parametric Methods for Doubly Robust Estimation of Continuous Treatment Effects." *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 79 (4): 1229–45. https://doi.org/10.1111/rssb.12212.

Zhou, Xiang, and Geoffrey T. Wodtke. 2020. "Residual Balancing: A Method of Constructing Weights for Marginal Structural Models." *Political Analysis* 28 (4): 487–506. https://doi.org/10.1017/pan.2020.2.