Optimal Data Analysis
Vol. 9 (June 12, 2020), 178-182

2155-0182/10/$3.00

Implementing ODA from Within Stata: A Reanalysis of the National Supported Work Experiment

Ariel Linden, Dr.P.H. and Paul R. Yarnold, Ph.D.
Linden Consulting Group, LLC Optimal Data Analysis LLC

Data from the National Supported Work (NSW) randomized experiment have been used frequently over the past 30 years to demonstrate implementation of various non-experimental methods for drawing causal inferences about treatment effects. In this paper we reanalyze these data using the new Stata package for implementing ODA.

Studies in which participants are randomized to treatment are considered the gold standard for assessing causal inference because randomization putatively ensures that the study groups do not differ systematically in their characteristics, and consequently, treatment effects are assumed to be unbiased.[1] If randomization is infeasible, investigators rely on statistical techniques which model treatment assignment in order to control for threats to validity which may compromise causal interpretation of the results.[2-6]

In this paper we reanalyze data from the National Supported Work (NSW) experiment which was originally discussed by LaLonde[7] in the context of economic evaluation approaches, but has since then been utilized frequently to demonstrate the implementation of various non-experimental techniques, such as propensity scoring methods, for assessing causal inference.

Herein we apply the new Stata package called **oda**,[8] that implements ODA from within the Stata environment, to these data to assess whether results are consistent with findings reported by Dehejia and Wahba.[9]

The **oda** package is a wrapper for the MegaODA software[10], and the megaODA.exe file must be loaded on the computer for the **oda** package to work (ODA software is available at https://odajournal.com/resources/). To download the **oda** package, at the Stata command line type: “ssc install oda” without quotation marks.

Methods

Data

The NSW was a US federally- and privately funded program that aimed to provide work experience for individuals who had faced economic and social problems prior to enrollment in the program. Candidates for the experiment were selected on the basis of eligibility criteria, and then were either randomly assigned to, or

2155-0182/10/$3.00

excluded from, the training program. We use the same subset of NSW data used by Dehejia and Wahba[9], joining the 185 treated units from the NSW experiment to comparison units from the 15,992 individuals in the Current Population Survey (CPS). Data were retrieved from: http://users.nber.org/~rdehejia/nswdata2.html.

Available variables included were age, education, black, hispanic, no degree, married, real earnings in 1974, 1975 and 1978 (all adjusted to 1982 US dollars), and indicators for unemployed status in 1974 and 1975. The outcome (primary model attribute) was real earnings in 1978, and the treatment (class) variable indicates whether individuals participated in the NSW intervention or were untreated from the CPS data.

Analysis

Dehejia and Wahba[9] estimated a propensity score in which the binary treatment indicator was regressed on age, age^2, age^3, education, education^2, married, no degree, black, Hispanic, earnings in 1974 and 1975, unemployed in 1974 and 1975, and an interaction of education and earnings in 1974. They then used various matching algorithms and compared the outcomes across the methods. Here we replicate their propensity score estimation and the 1:1 matching without replacement method.

```
* Estimate the propensity score
    logit treat age c.age#c.age
    c.age#c.age#c.age educ c.educ#c.educ
    married nodegree black hispan re74 re75
    u74 u75 c.educ#c.re74

* save predictions
    predict pscore

* use psmatch2 to perform 1:1 matching
    psmatch2 treat, outcome(re78)
    pscore(pscore) neighbor(1) noreplace

* Evaluate treatment effects using regression
  with robust standard errors and frequency
  weights from the psmatch2 output
      regress re78 treat [fw =_weight],
      vce(robust)
```

The regression produced a non-statistically significant estimated treatment effect of $1282 (95% CI: -204, 2768, *P* = 0.091).

We now evaluate these data using **oda** with the following syntax (see the help file for **oda** for a complete description of the syntax options):

```
oda treat re78 if _weight !=. , pathoda("C:\
ODA\") store("C:\ODA\ output") iter(10000)
loo, seed(1234)
```

The above syntax is explained as follows: The variable “treat” is the *class* variable; the outcome variable “re78” (earnings in 1987) is the *attribute*; the [if] statement indicates that the sample should be limited to observations with a non-missing weight (i.e., matched treated and control observations), the directory path where the megaODA.exe file is located on my computer is "C:\ODA\"; the directory path where the output and other files generated during the analysis should be stored is "C:\ODA\output"; the number of iterations (repetitions) for computing a permutation *P*-value is 10,000; leave-one-out (LOO) analysis should be performed, and the seed should be set to 1234 to ensure replication of the permutation results.

The **oda** package produces an extract of the total output produced by the ODA software (the complete output is stored in the specified directory with the extension “.out”).

As shown in the **oda** output below, the ODA model can be interpreted as follows: “if real earnings in 1978 <= $1237.291, then predict that the treatment group is 0 (controls). If the earnings are > $1237.291, then predict that the treatment group is 1 (treatment).”

2155-0182/10/$3.00

The effect strength for sensitivity (ESS) is labelled in the output as "Effect Strength PAC". In the training data the ESS is 16.76% and in the LOO analysis it is 15.68% (a very weak effect).[11] The permutation *P*-value for the training sample was 0.010 and for the LOO analysis was 0.005. In contrast to the regression which found no statistical difference between treatment groups (i.e., no treatment effect), ODA found a statistically significant treatment effect in these data, although the model found by ODA had a difficult time in discriminating between treatment groups.

```
ODA model:
----------
IF RE78 <= 1237.291 THEN TREAT = 0
IF 1237.291 < RE78 THEN TREAT = 1


Summary for Class TREAT  Attribute RE78
---------------------------------------

Performance Index           Train    LOO
-----------------          ------  ------
Overall Accuracy            58.38%  57.84%
PAC TREAT=0                 49.19%  49.19%
PAC TREAT=1                 67.57%  66.49%
Effect Strength PAC         16.76%  15.68%
PV TREAT=0                  60.26%  59.48%
PV TREAT=1                  57.08%  56.68%
Effect Strength PV          17.34%  16.16%
Effect Strength Total       17.05%  15.92%


Monte Carlo summary (Fisher randomization):
-------------------------------------------
Iterations:  10000
Estimated p: 0.010400


Results of leave-one-out analysis
---------------------------------
370 observations

Fisher's exact test (directional) classification table  p = .001530
```

Discussion

In this paper, we demonstrated how the new Stata package **oda** can be used in conjunction with a matching algorithm to evaluate treatment effects in observational data. ODA should be considered the preferred approach over commonly-used parametric models because ODA avoids the assumptions required of parametric models, is insensitive to skewed data or outliers, and has the ability to handle any variable metric including categorical, Likert-type integer, and real number measurement scales.[10] Moreover, in contrast to regression models, ODA also has the distinct ability to ascertain where the optimal (maximum-accuracy) cutpoints are on the outcome variable, which in turn, facilitates the use of measures of predictive accuracy. Moreover, ODA can perform cross-validation using LOO which allows for assessing the cross-generalizability of the model to potentially new study participants or non-participants.

Finally, the findings continue to support our recommendation to employ the ODA and CTA frameworks to evaluate the efficacy of health-improvement interventions and policy initiatives.[12-30]

References

[1]Linden A, Yarnold PR (2017). Minimizing imbalances on patient characteristics between treatment groups in randomized trials using classification tree analysis. *Journal of Evaluation in Clinical Practice*, *23*, 1309-1315. DOI: 10.1111/jep.12792

[2]Linden A, Adams J (2006). Evaluating disease management program effectiveness: an introduction to instrumental variables. *Journal of Evaluation in Clinical Practice*, *12*, 148-154.

[3]Linden A, Adams JL (2010). Using propensity score-based weighting in the evaluation of health management programme effectiveness. *Journal of Evaluation in Clinical Practice*, *16*, 175-179.

[4]Linden A, Adams JL (2010). Evaluating health management programmes over time. Application of propensity score-based weighting to longitudinal data. *Journal of Evaluation in Clinical Practice*, *16*, 180-185.

[5]Linden A (2014). Combining propensity score based stratification and weighting to improve causal inference in the evaluation of health care interventions. *Journal of Evaluation in Clinical Practice*, *20*, 1065-1071.

2155-0182/10/$3.00

[6]Linden A, Uysal SD, Ryan A, Adams JL (2016). Estimating causal effects for multivalued treatments: A comparison of approaches. *Statistics in Medicine*, *35*, 534-552.

[7]LaLonde R (1986). Evaluating the econometric evaluations of training programs. *American Economic Review*, *76*, 604–620.

[8]Linden A (2020). ODA: Stata module for conducting Optimal Discriminant Analysis. *Statistical Software Components S458728, Boston College Department of Economics*.

[9]Dehejia RH, Wahba S (2002). Propensity score matching methods for nonexperimental causal studies. *Review of Economics and Statistics*, *84*, 151-61.

[10]Yarnold PR, Soltysik RC (2016). *Maximizing Predictive Accuracy*. Chicago, IL: ODA Books. DOI: 10.13140/RG.2.1.1368.3286

[11]Yarnold PR, Soltysik RC. *Optimal data analysis: Guidebook with software for Windows*. Washington, D.C.: APA Books, 2005.

[12]Linden A, Adams J, Roberts N (October, 2003). *Evaluation methods in disease management: determining program effectiveness*. Position Paper for the Disease Management Association of America (DMAA).

[13]Linden A, Roberts N (2005). A Users guide to the disease management literature: recommendations for reporting and assessing program outcomes. *American Journal of Managed Care*, 11, 81-90.

[14]Linden A, Adler-Milstein J (2008). Medicare disease management in policy context. *Health Care Finance Review*, 29, 1-11.

[15]Linden A, Yarnold PR (2018). Identifying causal mechanisms in health care interventions using classification tree analysis. *Journal of Evaluation in Clinical Practice*, *24*, 353-361.

[16]Linden A, Yarnold PR (2018). Estimating causal effects for survival (time-to-event) outcomes by combining classification tree analysis and propensity score weighting. *Journal of Evaluation in Clinical Practice*, *24*, 380-387.

[17]Linden A, Yarnold PR (2016). Combining machine learning and propensity score weighting to estimate causal effects in multivalued treatments. *Journal of Evaluation in Clinical Practice*, *22*, 875-885.

[18]Linden A, Yarnold PR (2017). Using classification tree analysis to generate propensity score weights. *Journal of Evaluation in Clinical Practice*, *23*, 703-712.

[19]Yarnold PR, Linden A (2017). Computing propensity score weights for CTA models involving perfectly predicted endpoints. *Optimal Data Analysis*, *6*, 43-46.

[20]Linden A, Yarnold PR (2016). Using data mining techniques to characterize participation in observational studies. *Journal of Evaluation in Clinical Practice*, *22*, 839-847.

[21]Linden A, Yarnold PR (2016). Using machine learning to assess covariate balance in matching studies. *Journal of Evaluation in Clinical Practice*, *22*, 848-854.

[22]Linden A, Yarnold PR (2016). Using machine learning to identify structural breaks in single-group interrupted time series designs. *Journal of Evaluation in Clinical Practice*, *22*, 855-859.

[23]Linden A, Yarnold PR, Nallomothu BK (2016). Using machine learning to model dose-response relationships. *Journal of Evaluation in Clinical Practice*, *22*, 860-867.

2155-0182/10/$3.00

[24]Yarnold PR, Linden A (2016). Using machine learning to model dose-response relationships via ODA: Eliminating response variable baseline variation by ipsative standardization. *Optimal Data Analysis*, *5*, 41-52.

[25]Linden A, Yarnold PR (2016). Combining machine learning and matching techniques to improve causal inference in program evaluation. *Journal of Evaluation in Clinical Practice*, *22*, 868-874.

[26]Linden A, Yarnold PR (2017). Modeling time-to-event (survival) data using classification tree analysis. *Journal of Evaluation in Clinical Practice*, *23*, 1299-1308.

[27]Linden A, Yarnold PR (2018). The Australian gun buy-back program and the rate of suicide by firearm. *Optimal Data Analysis*, *7*, 28-35.

[28]Linden A, Yarnold PR (2018). Using machine learning to evaluate treatment effects in multiple-group interrupted time series analysis. *Journal of Evaluation in Clinical Practice*, *24*, 740-744.

[29]Linden A, Yarnold PR (2018). Using ODA in the evaluation of randomized controlled trials. *Optimal Data Analysis*, *7*, 46-49.

[30]Linden A, Yarnold PR (2018). Using ODA in the evaluation of randomized controlled trials: Application to survival outcomes. *Optimal Data Analysis*, *7*, 50-53.

Author Notes

No conflict of interest was reported.