Optimal Data Analysis
Vol. 9 (March 4, 2020), 45-50

2155-0182/10/$3.00

Implementing ODA from Within Stata: Evaluating Treatment Effects in Multiple-Group Interrupted Time Series Analysis (*Invited*)

Ariel Linden, Dr.P.H.
Linden Consulting Group, LLC

In this paper, I describe how to evaluate treatment effects in interrupted time-series studies in which the treated unit is contrasted with one or more comparable control groups, using the new Stata package for implementing ODA.

Prior papers[1-5] introduced the new Stata package called **oda**[6] for implementing ODA from within the Stata environment. This package is a wrapper for the MegaODA software[7], and therefore the MegaODA.exe file must be loaded on the computer for the **oda** package to work (MegaODA software is available at https://odajournal.com/resources/). To download the **oda** package, at the Stata command line type: "ssc install oda" (without the quotation marks).

In this paper, I demonstrate how the **oda** package can be used to evaluate treatment effects in interrupted time-series studies in which the treated unit is contrasted with one or more comparable control groups.

Interrupted time series analysis (ITSA) is an increasingly popular study design for evaluating the effectiveness of large-scale interventions and policy changes,[8-10] in which a single aggregate entity (e.g., hospital, county, or state) is the treatment unit, and accordingly, the outcome of interest is reported serially over time at the aggregate level (e.g., morbidity or mortality rates). The design is called an *interrupted time series* because the intervention is expected to "interrupt" the level and/or trend of the outcome variable subsequent to its introduction.[11]

ITSA is considered a fairly strong quasi-experimental design, primarily through its control over *regression to the mean*.[12-14] However, recent studies have demonstrated that the single-group ITSA can either fail to identify the effects of external factors on the time series—resulting in a false causal attribution, or conversely, confuse the causal interpretation when a directionally correct change in the time series also occurs prior to the intervention.[15,16] These issues appear to persist even when the study design includes multiple cross-overs from the treatment to non-treatment condition (or vice versa).[17]

2155-0182/10/$3.00

In order to reduce confounding and strengthen causal interpretation in ITSA studies, the treated unit should be compared with a control group that is balanced on observed characteristics (including, at a minimum, the baseline level and trend of the outcome).[16-24] Once a comparable control group is identified, either ordinary or generalized least-squares methods that can handle serially correlated errors are typically used to estimate treatment effects. However, linear models require strong statistical assumptions that, when violated, may bias the treatment effect estimates. As such, ODA is a preferred evaluation approach since it provides the most robust treatment effect estimates while minimizing the number of statistical assumptions that must be satisfied in the process.

The implementation of ODA in multiple-group ITSA consists of two parts: assessing comparative balance between the treated unit and control group on the pre-intervention time series of the outcome, and estimating a treatment effect (i.e., difference between the treated unit and control group on the post-intervention time series of the outcome). An integral component of both analyses includes cross-validation to assess the generalizability of the results to other points in the time series (past and future), or to similar series (e.g., other entities planning to implement a similar intervention) assuming they are comparable on other characteristics.

In **oda** we assess balance on the baseline time series and treatment effects by specifying the treatment indicator as the *class* variable and the outcome variable as the *attribute*. The only difference is that we limit the analysis to the pre-intervention period for assessing balance, and we limit the analysis to the post-intervention period for evaluating treatment effects.

Methods

Data

This paper uses data from a prior evaluation of the effects of California's voter-initiative Proposition 99 which was passed in 1988 to reduce smoking rates by raising the cigarette excise tax by 25 cents per pack and using the revenue to fund anti-smoking campaigns and other related anti-smoking activities throughout the state.

The per-capita cigarette sales (in packs) serves as the aggregate outcome variable under study, measured at the state level from 1970 until 2000 (with 1989 representing the first year of the intervention). Eleven states were discarded from the dataset because of their adoption of some other large-scale tobacco control program at some point during California's intervention period under study (1989-2000), leaving 38 states to serve as potential controls.

Analytic process

A comprehensive discussion on use of ODA for evaluating treatment effects in multiple group ITSA is available elsewhere.[24] Discussion here is limited to its implementation using the **oda** package.

Matching

The Stata package ***itsamatch***[20] was employed to identify the states that most closely matched California on level and trend of per capita cigarette sales (and lagged per capita cigarette sales) in the preintervention period (1970-1988). This approach is analogous to assessing equality (proportionality) of survival functions across groups in survival analysis.[25] The cutpoint significance level was set to $P > 0.40$ to identify balanced matches at the highest threshold possible for which controls could still be found. Montana emerged as the single best match to California.

Assessing balance

For assessing balance of the outcome time-series before the intervention was initiated, **oda** can be implemented with the following syntax (see the help file for **oda** for a complete description of the syntax options):

2155-0182/10/$3.00

```
oda treat cigsales if year <1989,
pathoda("C:\ODA\") store("C:\ODA\output")
iter(25000) loo seed(1234)
```

The above syntax is explained as follows: The binary treatment variable (1 if California, 0 if Montana) is the *class* variable; annual cigarette sales is the *attribute*; the [if] statement indicates that the sample should be restricted to the years prior to 1989; 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 25,000; leave-one-out analysis is additionally requested; and the seed is set to 1234 to allow us to replicate permutation results (any integer value can be used as seed).

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").

```
ODA model:
----------
IF CIGSALE <= 122.7 THEN TREAT = 0
IF 122.7 < CIGSALE THEN TREAT = 1


Summary for Class TREAT  Attribute CIGSALE
------------------------------------------

Performance Index           Train    LOO
-----------------          ------  ------
Overall Accuracy            60.53%  60.53%
PAC TREAT=0                 78.95%  78.95%
PAC TREAT=1                 42.11%  42.11%
Effect Strength PAC         21.05%  21.05%
PV TREAT=0                  57.69%  57.69%
PV TREAT=1                  66.67%  66.67%
Effect Strength PV          24.36%  24.36%
Effect Strength Total       22.71%  22.71%


Monte Carlo summary (Fisher randomization):
-------------------------------------------
Iterations:  25000
Estimated p: 0.804960


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

Fisher's exact test (directional) classification table  p = .147648
```

As shown in the **oda** output, the ODA model is interpreted as follows: "if the cigarette sales <= 122.7, then predict that the treatment group is 0 (Montana). If cigarette sales are > 122.7, then predict that the treatment group is 1 (California)." The effect strength for sensitivity (ESS) is labelled in the output as "Effect Strength PAC" which yielded a relatively weak ESS = 21.05%[26] for both the training sample and LOO analysis. This model was not statistically significant in training ($P < 0.805$) or LOO ($P < 0.148$) analysis, indicating that California and Montana could not be discriminated on the basis of preintervention annual cigarette sales (i.e., the units were balanced).

Evaluating treatment effects

Evaluating treatment effects for multi-group ITSA can be performed in **oda** with the same syntax as before with the exception being that here the [if] statement indicates that the time period is limited to the intervention period:

```
oda treat cigsales if year >1988,
pathoda("C:\ODA\") store("C:\ODA\output")
iter(25000) loo seed(1234)
```

```
ODA model:
----------
IF CIGSALE <= 82.5 THEN TREAT = 1
IF 82.5 < CIGSALE THEN TREAT = 0


Summary for Class TREAT  Attribute CIGSALE
------------------------------------------

Performance Index           Train    LOO
-----------------          ------  ------
Overall Accuracy            95.83%  87.50%
PAC TREAT=0                 91.67%  83.33%
PAC TREAT=1                100.00%  91.67%
Effect Strength PAC         91.67%  75.00%
PV TREAT=0                 100.00%  90.91%
PV TREAT=1                  92.31%  84.62%
Effect Strength PV          92.31%  75.52%
Effect Strength Total       91.99%  75.26%


Monte Carlo summary (Fisher randomization):
-------------------------------------------
Iterations:  25000
Estimated p: 0.000000


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

Fisher's exact test (directional) classification table  p = .000322
```

As shown in the **oda** output, the ODA model is interpreted as follows: "if the cigarette sales <= 82.5, then predict that the treatment group is 1 (California). If cigarette sales are > 82.5, then predict that the treatment groups is 1 (California)." The ESS for the training sample is a very strong 91.7%, $P < 0.0001$. Thus, the two units are thus nearly perfectly discriminated on

2155-0182/10/$3.00

the basis of annual cigarette sales after the intervention began in 1989, with annual sales of 82.5 packs per capita or lower in California. In the LOO analysis, the ESS was strong (75.0%) and statistically significant ($P < 0.0003$), suggesting that the lower annual cigarette sales noted for California will cross-generalize strongly to future years or to other entities with similar characteristics that implement a similar intervention.

Discussion

This paper demonstrates how the new Stata package **oda** can be used to evaluate treatment effects in multiple group interrupted time series studies. ODA should be considered the preferred approach over commonly-used parametric models because ODA avoids the assumptions required of these 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.[7] Moreover, in contrast to conventional statistical models, ODA also has the distinct ability to ascertain where 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 assesses the cross-generalizability of the model to potentially new study participants or non-participants.[27]

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.[28-44]

References

[1]Linden A (2020). Implementing ODA from Within Stata: An Application to Data From a Randomized Controlled Trial (*Invited*). *Optimal Data Analysis,* 9, 9-13.

[2]Linden A (2020). Implementing ODA from Within Stata: Implementing ODA from Within Stata: An Application to Estimating Treatment Effects using Observational Data (*Invited*). *Optimal Data Analysis,* 9, 14-20.

[3]Linden A (2020). Implementing ODA from Within Stata: An Application to Dose-Response Relationships (*Invited*). *Optimal Data Analysis,* 9, 26-32.

[4]Linden A (2020). Implementing ODA from Within Stata: Assessing Covariate Balance in Observational Studies (*Invited*). *Optimal Data Analysis,* 9, 33-38.

[5]Linden A (2020). Implementing ODA from Within Stata: Evaluating Treatment Effects for Survival (Time-to-Event) Outcomes (*Invited*) (*Invited*). *Optimal Data Analysis,* 9, 39-44.

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

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

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

[9]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).

[10]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.

2155-0182/10/$3.00

[11]Shadish WR, Cook TD, Campbell DT (2002). *Experimental and Quasi- Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.

[12]Linden A (2007). Estimating the effect of regression to the mean in health management programs. *Disease Management and Health Outcomes,* 15, 7-12.

[13]Linden A (2013). Assessing regression to the mean effects in health care initiatives. *BMC Medical Research Methodology,* 13, 1-7.

[14]Yarnold PR, Bryant FB, Soltysik RC (2013). Maximizing the accuracy of multiple regression models via UniODA: regression *away from* the mean. *Optimal Data Analysis,* 2, 19-25.

[15]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.

[16]Linden A (2017). Challenges to validity in single-group interrupted time series analysis. *Journal of Evaluation in Clinical Practice,* 23, 413-418.

[17]Linden A (2017). Persistent threats to validity in single-group interrupted time series analysis with a crossover design. *Journal of Evaluation in Clinical Practice*, 23, 419-425.

[18]Linden A (2015). Conducting interrupted time-series analysis for single- and multiple-group comparisons. *Stata Journal*, 15, 480-500.

[19]Linden A, Adams JL (2011). Applying a propensity-score based weighting model to interrupted time series data: improving causal inference in program evaluation. *Journal of Evaluation in Clinical Practice*, 17, 1231-1238.

[20]Linden A (2018). A matching framework to improve causal inference in interrupted time series analysis. *Journal of Evaluation in Clinical Practice,* 24, 408-415.

[21]Linden A (2018). Combining synthetic controls and interrupted time series analysis to improve causal inference in program evaluation. *Journal of Evaluation in Clinical Practice,* 24, 447-453.

[22]Linden A (2018). Using permutation tests to enhance causal inference in interrupted time series analysis. *Journal of Evaluation in Clinical Practice*, 24, 496–501.

[23]Linden A (2018). Using group-based trajectory modelling to enhance causal inference in interrupted time series analysis. *Journal of Evaluation in Clinical Practice,* 24, 502–507.

[24]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.

[25]Linden A, Adams J, Roberts N (2004). Evaluating disease management program effectiveness: an introduction to survival analysis. *Disease Management*, 7, 180-190.

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

[27]Linden A, Adams J, Roberts N (2004). The generalizability of disease management program results: getting from here to there. *Managed Care Interface*, 17, 38-45.

[28]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

[29]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.

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

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

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

[33]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.

[34]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.

[35]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.

[36]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.

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

[38]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.

[39]Yarnold PR, Linden A (2016). Theoretical aspects of the D statistic. *Optimal Data Analysis*, 22, 171-174.

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

[41]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.

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

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

[44]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.