

Evaluating Treatment Effects in Longitudinal Panel Data Using ODA from Within Stata

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

This paper demonstrates how ODA can be used to estimate treatment effects in longitudinal panel data using the **oda** Stata package.

Studies in which outcomes of individuals are measured on multiple occasions are commonly analyzed using longitudinal regression models. These models account for the fact that repeated measurements taken on the same individual naturally exhibit *within*-subject dependence (autocorrelation), and accordingly, they provide correct standard errors. Conversely, using an ordinary regression model with repeated measures data violates the assumption that outcomes are conditionally independent given the covariates.

In this paper, we demonstrate how ODA can be used as a non-parametric machine-learning alternative to longitudinal regression models when outcomes for study participants are measured repeatedly. The procedure is implemented using **oda**, a Stata package¹ for implementing ODA from within the Stata environment. This package is a wrapper for the MegaODA software system²⁻⁴, so the MegaODA.exe file must be loaded on the computer for the **oda** package to work.⁵ To download the **oda** package, at the Stata com-

mand line type: “ssc install oda” (without the quotation marks).

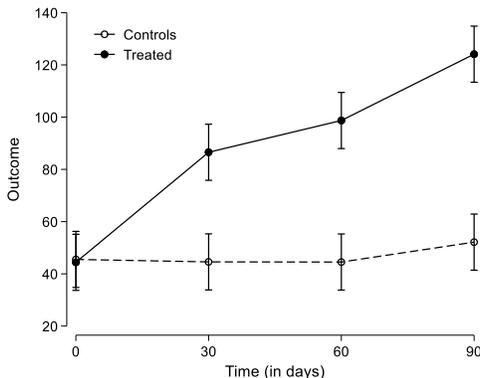
Methods

Data

We generated an artificial dataset for our motivating example to emulate a typical health behavior study.^{6,7} The data generating process produced 200 individuals, half of which were assigned to treatment and half served as controls. An outcome was generated for each subject at 4 time points: 0, 30, 60, and 90 days. For controls, the outcome at all time points were uniformly distributed random variates over the interval 0-100. For the treated subjects, a uniformly distributed random variate over the interval 0-100 was generated for day 0 to ensure that, on average, there would be no difference between treatment and control group at baseline. At day 30, treated subjects had an outcome that was a uniformly distributed random variate over the interval 1.30-1.80 multiplied by their day 0 outcome. Day 60 and day 90 outcomes for the

treated subjects were uniformly distributed random variates over the interval 1.00-1.10 multiplied by their last outcome value. All outcomes were generated with error following a standard normal distribution with a mean of 0 and a standard deviation of 3.

To summarize, the control group was constructed to show no change in outcomes over all measurement periods, whereas the treatment group was expected to show an incremental improvement in the outcome over time. The figure below illustrates the mean values (and confidence intervals) for the two groups at each time point as just described.



Analytic process

The steps involved for evaluating treatment effects using longitudinal data include: (1) performing a *within* group analysis of each study group separately. This analysis informs us as to whether there are time trends of the outcome within each group. Our a priori confirmatory hypothesis is that the level of the outcome will increase over time, more so in the treatment group versus in the control group. An increased trend in the outcome of the control group suggests that there may be external factors influencing all the subjects' outcomes (treatment and control), rather than the intervention, itself. It similarly raises concerns that the effect found in the intervention group

may be due to regression to the mean, and not the intervention.⁸ (2) performing a pooled *between* group analysis to assess the overall treatment effect. This analysis indicates *if* an overall effect of the intervention is observed. Depending on the magnitude of the treatment effect at one or more periods, the overall effect might not be observed. As such, we then perform (3) a *between* group analysis to assess treatment effects at each time-point separately. This analysis indicates *when* an effect of the intervention is observed.

Unlike observational studies where self-selection creates study groups with fundamentally different characteristics,^{9,10} in a randomized trial, we would expect to see no difference between the groups in the outcome at baseline (time = 0), and some effect occurring at a later time. It is also possible that an effect may be transient, that is, the intervention may be effective for a short duration, and then the outcome returns to levels closer to the baseline.

Results

Within-group analyses

The `oda` syntax for performing the within-group analysis for the control group is as follows:

```
oda time y if z==0, pathoda("C:\ODA\")
store("C:\ODA\output") iter(1000) loo
```

This syntax is explained as follows: the time variable (0, 30, 60 and 90 days) is the *class* variable; y is the outcome and serves as the *attribute*; the qualifier "if z==0" means that the analysis is limited to the control group (where z is the treatment variable and 0 represents the control group value); 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 are stored is "C:\ODA\output"; the number of iterations (repetitions) for computing a

permutation *P*-value is 1000; and we request that a leave-one-out-analysis (LOO) be performed (indicated by the “loo” option).

As shown in the **oda** output (Table 1), the ODA model could not identify cut-points of the outcome ordered at the correct time-points. Instead, the cut-points found by the model were ordered at 60, 0, 30, and 90 days.

The effect strength for sensitivity (ESS) is labelled in the output as “Effect Strength PAC”. As expected, ESS was very weak for the control group in both the training and LOO analyses (12% and -6.0% respectively).¹¹

This model was not statistically significant in training analysis ($P < 0.868$) indicating that the outcome could not be discriminated on the basis of time in the control group. Taken together, there did not appear to be any time trend in the control group’s outcomes.

Table 1: Stata **oda** output for within-group analysis of the control group

```
ODA model:
-----
IF Y <= 23.1035225 THEN TIME = 60
IF 23.1035225 < Y <= 47.3962175 THEN TIME = 0
IF 47.3962175 < Y <= 64.058235 THEN TIME = 30
IF 64.058235 < Y THEN TIME = 90

Summary for Class TIME Attribute Y
-----
Performance Index      Train   LOO
-----
Overall Accuracy      34.00% 20.50%
PAC TIME=0           38.00% 20.00%
PAC TIME=30          22.00% 18.00%
PAC TIME=60          30.00%  0.00%
PAC TIME=90          46.00% 44.00%
Effect Strength PAC   12.00% -6.00%
PV TIME=0            35.19% 16.39%
PV TIME=30           36.67% 29.03%
PV TIME=60           27.78%  0.00%
PV TIME=90           37.10% 35.48%
Effect Strength PV    12.24% -6.36%
Effect Strength Total 12.12% -6.18%

Monte Carlo summary (Fisher randomization):
-----
Iterations: 1000
Estimated p: 0.868000

Results of leave-one-out analysis
-----
200 observations
(P-values are computed for binary class variables only)
```

The **oda** syntax for performing the within-group analysis for the treatment group is identical to that of the control group, with the exception of the qualifier now expressed as

“z==1” indicating that the analysis is limited to the treatment group only:

```
oda time y if z==1, pathoda("C:\ODA\")
store("C:\ODA\output") iter(1000) loo
```

As shown in the **oda** output (Table 2), the ODA model identified the cut-points consistent with the correct ordering of time. The effect strength for sensitivity (ESS) was moderate for the treatment group in training (32%) and weak in the LOO analyses (21.33%).¹¹

This model was statistically significant in training analysis ($P < 0.0001$) indicating that the outcome could be discriminated on the basis of time in the treatment group. Taken together, these results suggest that the outcomes appear to statistically trend upward consistent with time in the intervention group.

Table 2: Stata **oda** output for within-group analysis of the intervention group

```
ODA model:
-----
IF Y <= 82.7365875 THEN TIME = 0
IF 82.7365875 < Y <= 97.60672 THEN TIME = 30
IF 97.60672 < Y <= 130.813935 THEN TIME = 60
IF 130.813935 < Y THEN TIME = 90

Summary for Class TIME Attribute Y
-----
Performance Index      Train   LOO
-----
Overall Accuracy      49.00% 41.00%
PAC TIME=0           88.00% 86.00%
PAC TIME=30          18.00% 16.00%
PAC TIME=60          36.00% 14.00%
PAC TIME=90          54.00% 48.00%
Effect Strength PAC   32.00% 21.33%
PV TIME=0            47.31% 46.24%
PV TIME=30           39.13% 23.53%
PV TIME=60           48.65% 24.14%
PV TIME=90           57.45% 54.55%
Effect Strength PV    30.85% 16.15%
Effect Strength Total 31.42% 18.74%

Monte Carlo summary (Fisher randomization):
-----
Iterations: 1000
Estimated p: 0.000000

Results of leave-one-out analysis
-----
200 observations
(P-values are computed for binary class variables only)
```

Overall between-group analysis

The **oda** syntax for performing the overall between-group analysis is as follows:

```
oda z y, pathoda("C:\ODA\") store("C:\ODA\output") iter(1000) loo
```

This syntax is explained as follows: the treatment indicator *z* is specified as the class variable and *y* is the outcome and serves as the *attribute*. All other options are the same as those in the prior models described above.

As shown in the **oda** output (Table 3), ODA identified a model with a relatively strong effect size in both training and LOO (66% and 64%, respectively) with the cut-point at 101.47. Similarly, the effect at this cut-point was statistically significant in both training and LOO ($P < 0.0001$).

Table 3: Stata **oda** output for the overall between-group analysis

```
ODA model:
-----
IF Y <= 101.472036 THEN Z = 0
IF 101.472036 < Y THEN Z = 1

Summary for Class Z Attribute Y
-----
Performance Index      Train   LOO
-----
Overall Accuracy      83.00%  82.00%
PAC Z=0               100.00% 100.00%
PAC Z=1               66.00%  64.00%
Effect Strength PAC    66.00%  64.00%
PV Z=0                74.63%  73.53%
PV Z=1                100.00% 100.00%
Effect Strength PV     74.63%  73.53%
Effect Strength Total  70.31%  68.76%

Monte Carlo summary (Fisher randomization):
-----
Iterations: 1000
Estimated p: 0.000000

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

Fisher's exact test (directional) classification table p = .126E-0012
```

Between-group analyses

The **oda** syntax for performing the between-group analyses at a given time period is as follows:

```
oda z y if time==0, pathoda("C:\ODA\")
store("C:\ ODA\output") iter(1000) loo
```

This syntax is explained as follows: the treatment indicator *z* is specified as the class variable; *y* is the outcome and serves as the *attribute*; and the qualifier “if time==0” means that the analysis is limited to outcomes at 0 days. All other options are the same as those in the prior models described above.

As shown in the **oda** output (Table 4), ODA identified a model with a very weak ESS in both the training and LOO analyses (12% and 10.0% respectively),¹¹ and the model was not statistically significant in either training or LOO ($P=0.87$ and $P=.12$, respectively). These findings are consistent with what is expected at baseline for a randomized trial, that is, the treatment and control group are expected to be comparable on all pre-intervention characteristics.

Table 4: Stata **oda** output for the between-group analysis at day 0 (baseline)

```
ODA model:
-----
IF Y <= 7.5984943 THEN Z = 1
IF 7.5984943 < Y THEN Z = 0

Summary for Class Z Attribute Y
-----
Performance Index      Train   LOO
-----
Overall Accuracy      56.00%  55.00%
PAC Z=0               94.00%  92.00%
PAC Z=1               18.00%  18.00%
Effect Strength PAC    12.00%  10.00%
PV Z=0                53.41%  52.87%
PV Z=1                75.00%  69.23%
Effect Strength PV     28.41%  22.10%
Effect Strength Total  20.20%  16.05%

Monte Carlo summary (Fisher randomization):
-----
Iterations: 1000
Estimated p: 0.870000

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

Fisher's exact test (directional) classification table p = .116801
```

Tables 5, 6, and 7 present results for between-group analyses at time 30, 60, and 90 days. As shown, ESS improves with each successive time-point, and the models are all statistically significant in both training and

LOO. These findings are consistent with the design of the data generating process used to create the artificial data.

Table 5: Stata **oda** output for the between-group analysis at day 30

```
ODA model:
-----
IF Y <= 82.6522715 THEN Z = 0
IF 82.6522715 < Y THEN Z = 1

Summary for Class Z Attribute Y
-----
```

Performance Index	Train	LOO
Overall Accuracy	77.00%	75.00%
PAC Z=0	94.00%	92.00%
PAC Z=1	60.00%	58.00%
Effect Strength PAC	54.00%	50.00%
PV Z=0	70.15%	68.66%
PV Z=1	90.91%	87.88%
Effect Strength PV	61.06%	56.54%
Effect Strength Total	57.53%	53.27%

```

Monte Carlo summary (Fisher randomization):
-----
Iterations: 1000
Estimated p: 0.000000

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

Fisher's exact test (directional) classification table p = .559E-0007
    
```

Table 6: Stata **oda** output for the between-group analysis at day 60

```
ODA model:
-----
IF Y <= 96.419052 THEN Z = 0
IF 96.419052 < Y THEN Z = 1

Summary for Class Z Attribute Y
-----
```

Performance Index	Train	LOO
Overall Accuracy	81.00%	79.00%
PAC Z=0	100.00%	98.00%
PAC Z=1	62.00%	60.00%
Effect Strength PAC	62.00%	58.00%
PV Z=0	72.46%	71.01%
PV Z=1	100.00%	96.77%
Effect Strength PV	72.46%	67.79%
Effect Strength Total	67.23%	62.89%

```

Monte Carlo summary (Fisher randomization):
-----
Iterations: 1000
Estimated p: 0.000000

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

Fisher's exact test (directional) classification table p = .360E-0010
    
```

Table 7: Stata **oda** output for the between-group analysis at day 90

```
ODA model:
-----
IF Y <= 101.472036 THEN Z = 0
IF 101.472036 < Y THEN Z = 1

Summary for Class Z Attribute Y
-----
```

Performance Index	Train	LOO
Overall Accuracy	83.00%	82.00%
PAC Z=0	100.00%	100.00%
PAC Z=1	66.00%	64.00%
Effect Strength PAC	66.00%	64.00%
PV Z=0	74.63%	73.53%
PV Z=1	100.00%	100.00%
Effect Strength PV	74.63%	73.53%
Effect Strength Total	70.31%	68.76%

```

Monte Carlo summary (Fisher randomization):
-----
Iterations: 1000
Estimated p: 0.000000

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

Fisher's exact test (directional) classification table p = .126E-0012
    
```

Discussion

This paper demonstrates how the Stata package **oda** can be used to evaluate treatment effects in studies with longitudinal data. This methodology should be considered an alternative approach to commonly-used longitudinal 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.¹² In contrast to conventional statistical models, ODA has the ability to ascertain where optimal (maximum-accuracy) cut-points are on the outcome variable, which in turn, facilitates the use of measures of predictive accuracy. Moreover, ODA can perform cross-validation using LOO to assess the cross-generalizability of a statistically significant training model to potentially new study participants or non-participants.¹³ 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.¹⁴⁻²⁶

References

- ¹Linden A (2020). ODA: Stata module for conducting Optimal Discriminant Analysis. *Statistical Software Components S458728, Boston College Department of Economics*.
- ²Soltysik RC, Yarnold PR (2013). MegaODA large sample and BIG DATA time trials: Separating the chaff. *Optimal Data Analysis, 2*, 194-197.
- ³Soltysik RC, Yarnold PR (2013). MegaODA large sample and BIG DATA time trials: Harvesting the Wheat. *Optimal Data Analysis, 2*, 202-205.
- ⁴Yarnold PR, Soltysik RC (2013). MegaODA large sample and BIG DATA time trials: Maximum velocity analysis. *Optimal Data Analysis, 2*, 220-221.
- ⁵<https://github.com/njrhodes/ODA>
- ⁶Linden A, Roberts N (2004). Disease management interventions: What's in the black box? *Disease Management, 7*, 275-291.
- ⁷Linden A, Butterworth S, Roberts N (2006). Disease management interventions II: What else is in the black box? *Disease Management, 9*, 73-85.
- ⁸Linden A (2007). Estimating the effect of regression to the mean in health management programs. *Disease Management and Health Outcomes, 15*, 7-12.
- ⁹Linden A, Adams J, Roberts N. (2016) Strengthening the case for disease management effectiveness: un hiding the hidden bias. *Journal of Evaluation in Clinical Practice, 12*, 140-147.
- ¹⁰Linden A, Adams J. Evaluating disease management program effectiveness: an introduction to instrumental variables. *Journal of Evaluation in Clinical Practice* 2006;12(2):148-154.
- ¹¹Yarnold PR, Soltysik RC. *Optimal data analysis: Guidebook with software for Windows*. Washington, D.C.: APA Books, 2005.
- ¹²Yarnold PR, Soltysik RC (2016). *Maximizing Predictive Accuracy*. Chicago, IL: ODA Books. DOI: 10.13140/RG.2.1.1368.3286
- ¹³Linden A, Adams J, Roberts N (2004). The generalizability of disease management program results: getting from here to there. *Managed Care Interface, 17*, 38-45.
- ¹⁴Linden 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.
- ¹⁵Linden A, Adler-Milstein J (2008). Medicare disease management in policy context. *Health Care Finance Review, 29*, 1-11.
- ¹⁶Linden A, Yarnold PR (2018). Identifying causal mechanisms in health care interventions using classification tree analysis. *Journal of Evaluation in Clinical Practice, 24*, 353-361.
- ¹⁷Linden 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.
- ¹⁸Linden 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.
- ¹⁹Linden A, Yarnold PR, Nallomothu BK (2016). Using machine learning to model dose-response relationships. *Journal of Evaluation in Clinical Practice, 22*, 860-867.

²⁰Linden A, Yarnold PR (2016). Using machine learning to assess covariate balance in matching studies. *Journal of Evaluation in Clinical Practice*, 22, 848-854.

²¹Linden A, Yarnold PR (2016). Using data mining techniques to characterize participation in observational studies. *Journal of Evaluation in Clinical Practice*, 22, 839-847.

²²Linden A, Yarnold PR (2017). Using classification tree analysis to generate propensity score weights. *Journal of Evaluation in Clinical Practice*, 23, 703-712.

²³Linden 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.

²⁴Linden 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.

²⁵Linden A, Yarnold PR (2017). Modeling time-to-event (survival) data using classification tree analysis. *Journal of Evaluation in Clinical Practice*, 23, 1299-1308.

²⁶Linden 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.

Author Notes

No conflict of interest was reported.