PROPENSITY SCORES USING MORE THAN 2 GROUPS: WHAT ARE MY OPTIONS AS A PRACTITIONER?

Antonio Olmos
Priyalatha Govindasamy
Research Methods & Statistics
University of Denver

American Evaluation Association Conference,
Chicago, Ill, November 2015

Paper overview

- Propensity scores with more than two groups
- Quick review of the packages
- Methods
- Results
- Suggestions for practitioners

Propensity scores with more than 2 groups

- Recent development, despite the fact that Imbens (2000) demonstrated the extension from 2 to > 2 groups
- Very few implementations that can be used to conduct PSA (in R)
  - Trimatch (Bryer, 2013)
  - Twang (McCaffrey et.al., 2012)
- Very different implementations, with very different goals in mind
- As a practitioner, what are my options?

TriMatch (Bryer, 2013)

- PS matching for 3 groups using a measure of distance in a multidimensional space
  - PS estimation using 3 separate logistic regressions (Tr1-C, Tr2-C, Tr1-Tr2)
  - Estimation of 3 distances (one for each pair). The triplets with the smallest distances (next slide) are kept
    - Since distances are estimated across all groups, matches across all 3 groups are possible

TriMatch (cont)

- Options to generate different number of matches across all three groups
  - MaximumTreat: matching without replacement with some limitations; Caliper includes all units within a specified caliper (multiple duplicates); one to N user sets how many times each T1 and T2 can be reused
    - Number of non-used cases or duplicates in an analysis, dependent on these options
- Balance is assessed by the package through plots
- Either RM ANOVA or Friedman's Rank Sum test to estimate the outcome.
  - Follow up tests using paired-t-tests

twang (McCaffrey et. al. 2012)

- Creates multiple propensity scores (one for every pair) using Generalized Boosted Models (GBM)
  - Iterations fitting many simple regression trees combined to create an overall piecewise constant function
- Originally designed for 2 groups, but extended in 2013 to more than 2 groups
  - The package applies propensity scores through weights to the group participants

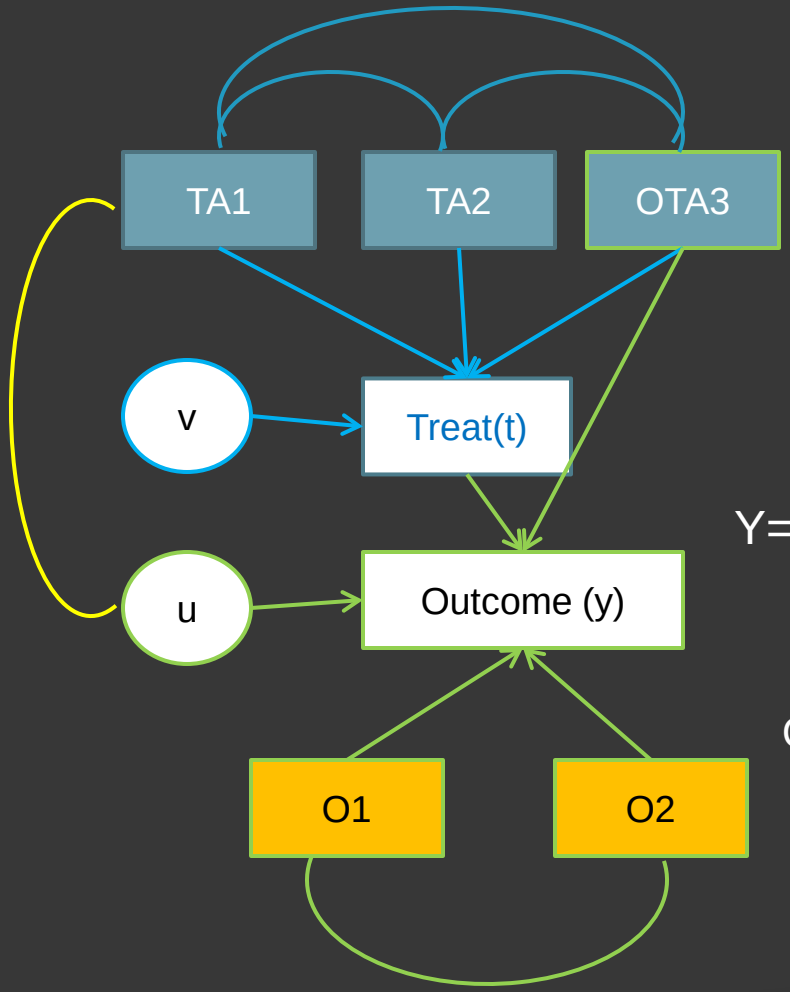
twang (cont)

- Given that the algorithm can overfit the data, some stopping rules
  - Based on summary statistics (absolute standardized mean difference) or Kolmogorov–Smirnov (KS)
  - Parameters such as number of trees, the stopping method (means, KS), and the estimation approach (ATE, ATT) can be defined
  - Balance is assessed by the package through plots, pairwise standardized differences
  - All cases are included
- Outcome model estimated through weighted regression

Research questions

- Questions tested with both twang (weighted regression) and TriMatch (matching):
  - Effect of distributional characteristics (normal, positive skewness) of the variables included in the selection model
  - Effect of the Degree of correlation (no, small, large) between the variables in the selection model and between the variables in the outcome model
  - Effect of the selection bias defined as the correlation between one of the variables in the selection model and the error term for the outcome model (low, high)

Simulation model



Every model was simulated 100 times
Sample size: 500

Models generated for the simulation

D1=0.7(TA1)+0.4(TA2)+0.1(OTA3)+v

D2=0.8(TA1)+0.5(TA2)+0.15(OTA3)+v

Y=0.4(O1)+0.7(O2)+0.3(OTA3)+5(D1)+10(D2)+u

r(TA1, u) = 0.3, 0.6

OTA_3 is part of both Treatment (T) and Outcome (Y)

Simulation conditions

- Selection bias. 2 conditions:
  - r(u, TA1) = 0.3, 0.6
- Distributional characteristics. 3 conditions:
  - Normal, Positive Skewness, Mixed (Normal, Positive Skewness)
- Correlation betweeen IV's (selection model). 3 conditions:
  - No, small(0.3), large(0.6)
- Initial sample size: 500 cases

Results

Preliminary results

- twang:
  - Need a reference group (*control*; *treatment*). If you want a different reference, you need to reorganize the file
    - Group 3 in the file is the reference group
  - The analysis assumes independence between groups
  - Twang can estimate either ATE or ATT
  - Runs a weighted regression model (dummy variables)

Preliminary results

- TriMatch
  - Assumes dependency among the matched observations
    - RM-ANOVA, Friedman test for non-parametric
  - Matches all groups in every combination
    - Post-hoc analyses using matched-test
  - Cases may be lost due to matching (ways to increase/decrease sample size)
    - Observations can be used more than once is some instances
  - Cannot account for covariates

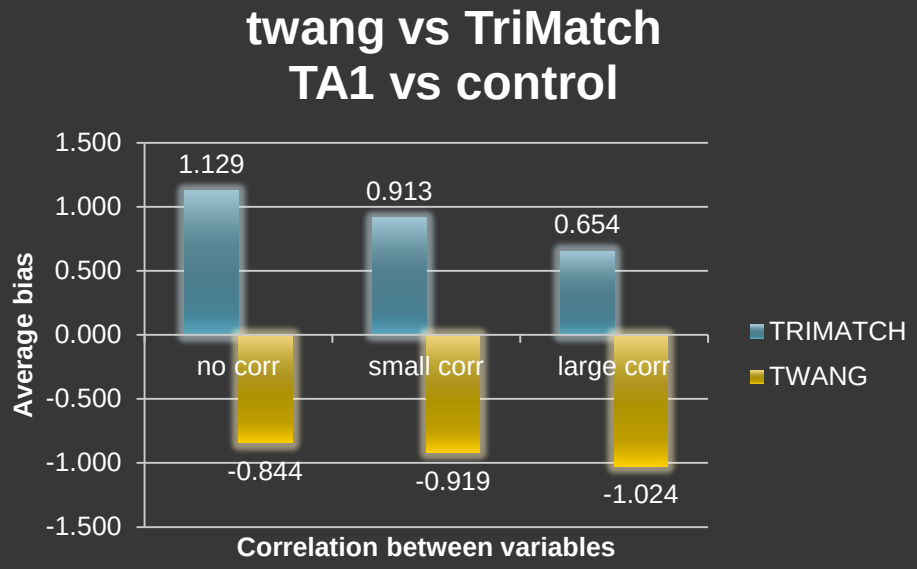
Results: twang vs. TriMatch General

- Comparison TA1 vs control: twang bias (M=-0.929) always larger than TriMatch (M=0.899)
  - KW(1) = 26.27, p < 0.01*
- Comparison TA2 vs. control: twang bias was always larger than for TriMatch (M=0.555 vs. M=0.141)
  - KW(1) = 26.27, p < 0.01*
- 2-way interactions (not statistically tested)

* Kruskal Wallis, based on ranks

twang vs. TriMatch

Correlation among IV's in the selection model (no, small, large)

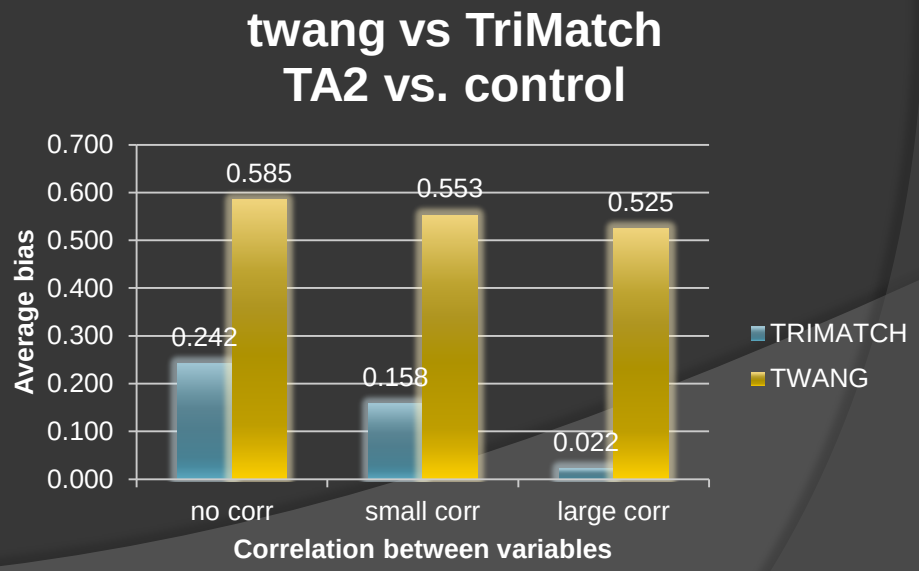


TA1 vs control

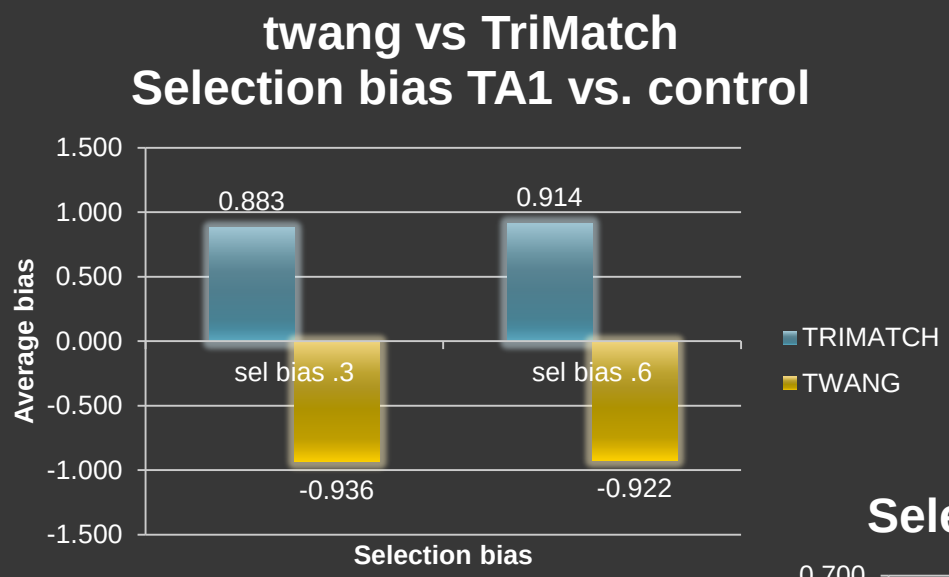
- Larger bias for twang
- Bias for TriMatch got smaller for larger correlations among variables

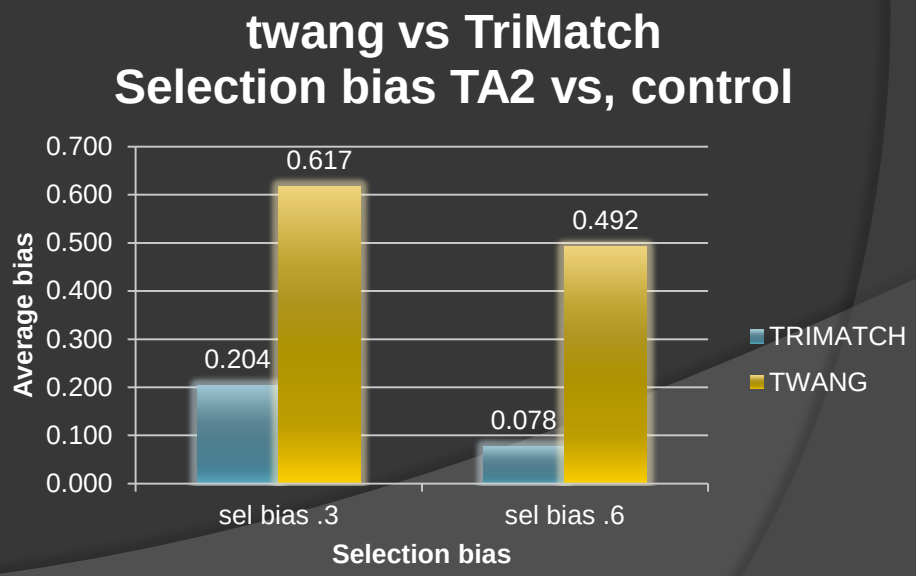
TA2 vs. control

- Larger bias for twang
- Bias for TriMatch got smaller for larger correlations among variables



twang vs. TriMatch
Selection bias r(u, TA1) = 0.3, 0.6

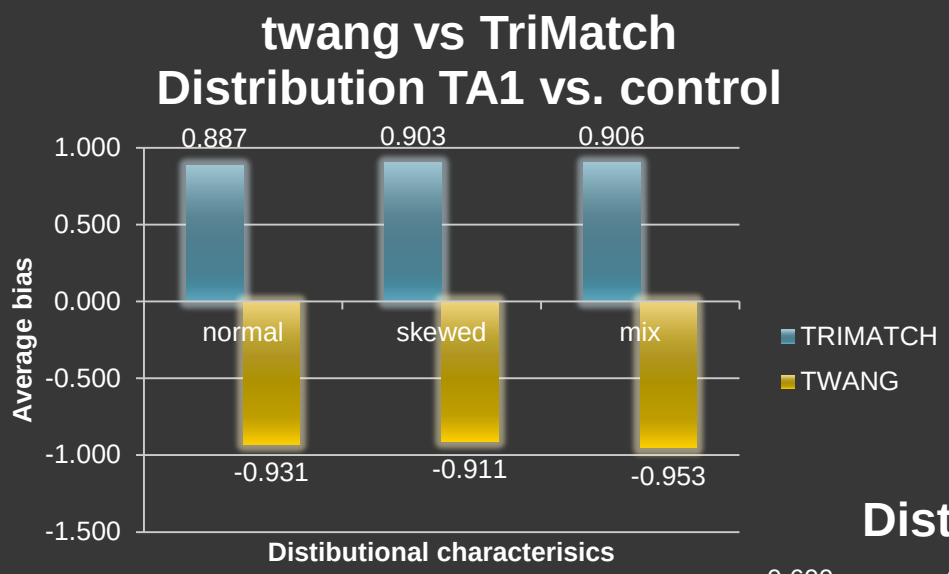


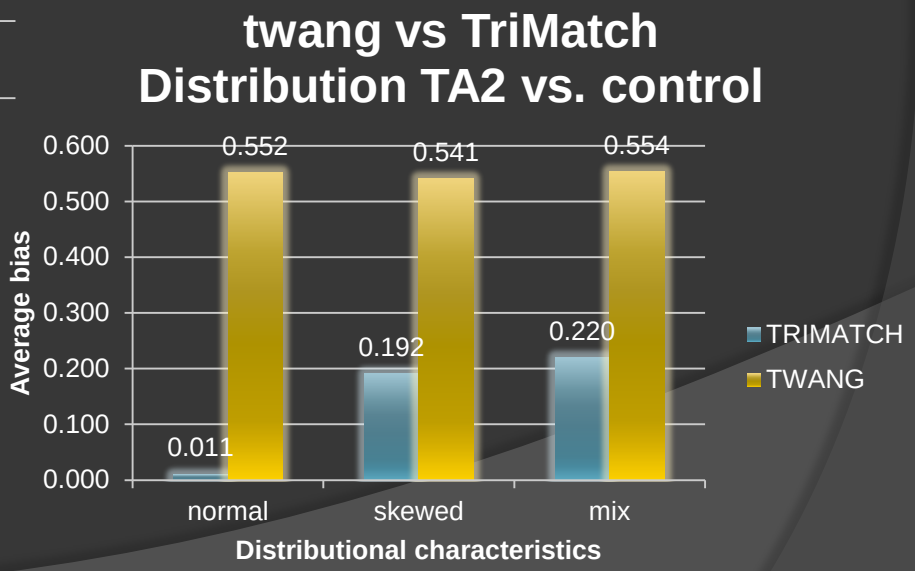


TA2 vs control

- Larger bias for twang
- Bias for TriMatch got smaller for larger selection bias

twang vs. TriMatch
Distributional characteristics (normal, positive skew, mix)





TA2 vs control

- Larger bias for twang
- Bias for TriMatch got larger for skewed/mix

Conclusions

- Under the present conditions, TriMatch seems a better choice than twang
  - Overall, less bias than twang
  - More likely to provide an accurate estimate when IV's in the selection model are correlated
  - Better, if distributions are normal
- Only main effects were tested for significance

As a practitioner, what are my choices?

- TriMatch, if you have max three groups, and you don't have covariates
- At this point, your only choice if you have more than 3 groups or if you have covariates is limited to twang
  - Be aware that twang is very sensitive to correlation between IV's in the selection model, larger selection bias, distributional characteristics of the variables in your selection model

Future work

- Multiple options within each package that were not tested
  - TriMatch: type of match
  - twang: number of trees, stopping method, estimation approach
- Sample size (small, medium, large)
- Effect of hidden bias

Thanks

Antonio Olmos

Antonio.olmos@du.edu

Priyalatha Govindasamy

Priyalatha.govindasamy@du.edu