

CS112: Final Assignment

Xavier Ruiz, Nazar Yaremko

Minerva Schools at KGI

To: The Editor, Journal of Peace

From: Xavier Ruiz & Nazar Yaremko

Date: 13 December 2019

Re: Re-evaluation of the relationship between violence, kinship networks, and political resilience.

Executive Summary

In an effort to increase the validity and interpretability of the results of a study on the effects of kinship networks on political participation post-victimization (Dorff, 2017) we have improved and replicated said study. To do so, we have conducted a matching procedure to balance the covariate distribution between victims and non-victims, as well as performed a sensitivity test that informs us of the hidden bias necessary to skew our results. Our matching results were highly successful (leximin p-value of 0.53) and the replication of the marginal effect of kinship from the original paper yielded extremely similar results as the original. This corroborates the original assessment on the need to strengthen kinship ties for victimized individuals or those living in violence environments as an effort to increase their participation in the political activity that will improve their conditions. Its direct importance on peacebuilding efforts relates to the identification of the most at-risk communities and the immediate improvement of those suffering a post-violent period.

Introduction

In the following memo, we re-examine Cassy Dorff's conclusions on the relationship between victimization and the strength of kinship networks surrounding the victimized individual on political participation. After careful review of the original paper (Dorff, 2017) we have identified significant differences between the victimized and not-victimized groups that may render a causal inference invalid. To find more robust results we have introduced a matching procedure preceding the causal evaluation, which achieved an extremely high balance between the group's covariates (leximin p-value = 0.5317) and introduced a sensitivity test to account for likely endogeneity. We hope the re-evaluation of the mentioned study will help better guide recovery action and prevention of at-risk communities by establishing a sturdier case for choosing a course of action.

The remainder of the memo follows our process of re-evaluation as follows:

1. Data preprocessing prior to matching attempt
2. Matching procedure using genetic matching
3. Sensitivity testing
4. Replication of marginal effect plot for kinship vs political participation
5. Proposals for further analysis

1. Data Pre-processing

To prepare the dataset for our matching procedure there were two main alterations we performed on the data. The first consisted of translating all relevant string variables to numeric values. Victimization (*victim*), perceived neighborhood security (*neighsec*), and attendance to political party meetings (*party*) were all translated into binary outputs since their outcomes consisted of only two possible options. The second main alteration was filling in all NAs in the data with the mean for the specific covariate. Note that we do not believe this to significantly skew our results considering the low number of missing data points in the original dataset (1.01%).

2. Matching Process

First, we established the variables on which we would hope to match as all those relevant to the original study's logistic model:

*“kinship”, “tenure”, “age”, “women”, “edu”, “neighsec”, “party”, “vote”, “rural”, “I(kinship*victim)”*

Matching was performed via genetic matching, which consistently led to high match balance between the groups (leximin p-value of .20 - .55). Three of the outcomes with a leximin p-value higher than 0.40 were recorded, as shown below.

Table 1.
Match treatment effect estimates post-matching for balances above a leximin p-value of 0.40.

Balance Found	Estimate	SE	P-value
0.434	0.08435	0.038269	0.027515
0.45045	0.076544	0.036106	0.034008
0.53173	0.087934	0.036917	0.01722

These results help set an initial approximation to the matches' estimated effect. It seems like a balance between the groups increases, the estimate holds almost constant at around ~0.08, with a standard error of ~0.03 and an estimate p-value of ~0.02. Note that not only are all estimates extremely similar, but all of the recorded p-values yield significant results (lower than our significance level threshold of 0.05). This justifies our use of a sensitivity test to corroborate the robustness of this initial inference.

Balance itself proved extremely high for all variables. Following the last example of the table, we continue our interpretation of the matching results (examples of covariate balance post-matching shown below, for more detailed matching results, see Appendix B), noting that seven of the nine relevant covariates balance was checked on showed a leximin p-value higher than .75, out of which four were higher than .9. This is extremely relevant, as before the matching process only two of the nine covariates surpassed the usual balance threshold of a 0.10 leximin p-value. In other words, the balance was reached successfully.

***** (V1) kinship *****		
	Before Matching	After Matching
mean treatment.....	0.63109	0.63109
mean control.....	0.56158	0.62577
std mean diff.....	37.336	2.855
var ratio (Tr/Co)....	0.84201	1.2194
T-test p-value.....	3.4231e-07	0.53173
KS Bootstrap p-value..	< 2.22e-16	0.556

***** (V2) tenure *****		
	Before Matching	After Matching
mean treatment.....	4.1563	4.1563
mean control.....	4.2155	4.2122
std mean diff.....	-5.1775	-4.885
var ratio (Tr/Co)....	1.09	1.2044
T-test p-value.....	0.45788	0.54849
KS Bootstrap p-value..	0.684	0.76

***** (V3) age *****		
	Before Matching	After Matching
mean treatment.....	37.989	37.989
mean control.....	41.763	37.832
std mean diff.....	-26.619	1.1125
var ratio (Tr/Co)....	0.75338	1.0893
T-test p-value.....	0.00032786	0.63693
KS Bootstrap p-value..	< 2.22e-16	0.818

***** (V4) women *****		
	Before Matching	After Matching
mean treatment.....	0.48029	0.48029
mean control.....	0.53537	0.48746
std mean diff.....	-11.005	-1.4322
var ratio (Tr/Co)....	1.0057	0.99907
T-test p-value.....	0.1189	0.65495

Note that with the addition of further interaction terms to the matching process it is likely that the balance could increase even higher, but increasing the number of matches to increase sample size is not advised as it greatly diminished both the balance and the p-value of the estimate found.

3. Sensitivity Test

We perform a sensitivity test using the rbounds library from R (see Appendix C for relevant code for sections 2 and 3). This process is extremely important as Dorff's paper itself was born from the consideration of an additional confounder previously unconsidered, kinship strength, thus the consideration of other covariates we may still be missing could also be significant in finding relevant results. As can be seen in the table below, the critical gamma value needed for our estimate to lose significance would fall between 1.3 and 1.4, as that is the bound where we pass the significance level threshold. This can be interpreted as a somewhat high sensitivity to hidden bias, meaning that the introduction of several small confounders could be enough to render our results insignificant. While this does not mean we cannot continue with our analysis, it does encourage the consideration of additional possible confounders that may be at play to reduce the possibility of endogeneity in our analysis.

Table 2. Sensitivity analysis results.

Gamma	1.0	1.1	1.2	1.3	1.4	1.5
Lower Bound	0.0025	0.0006	0.0001	0.0000	0.0000	0.0000
Upper Bound	0.0025	0.0084	0.0221	0.0479	0.0889	0.1461

4. Replication of Dorff, 2017, figure 6

Using our post-matching dataset we re-plot Dorff's marginal effect figure for kinship vs its effect on participation in political party meetings (see Fig 1).

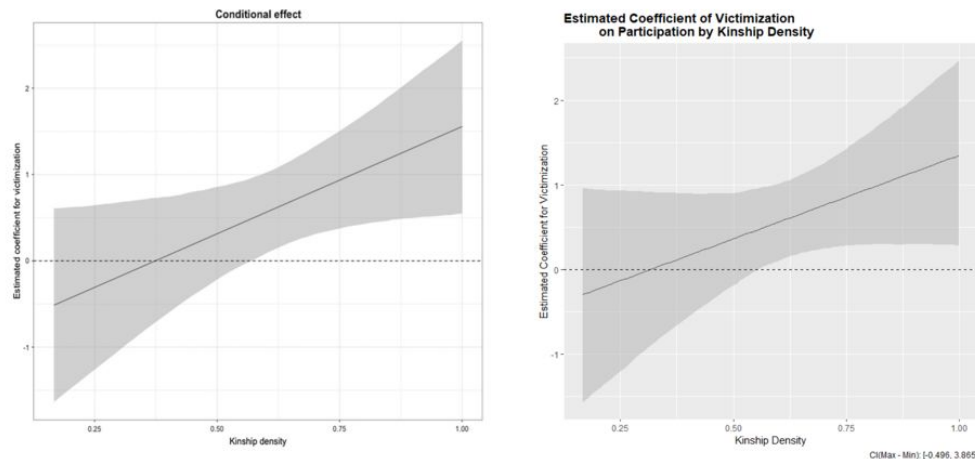


Figure 1. Marginal effect plot replication. On the left, the original marginal effect plot for unbalanced control and treatment groups. On the right, the replicated figure built after adding a matching mechanisms beforehand.

Both figures above are quite similar, showing a similar positive trend across similar bounds. The main difference between the replicated figure and the original is the smaller slope of the trendline, suggesting a slightly lesser effect on kinship on participation due to victimization. The confidence intervals for more extreme cases of kinship have also increased, and the expected outcomes for very low kinship values are now higher than those originally expected.

5. Further analysis proposals

Moving forward we would advise following the same methodology we did for the evaluation of kinship's effect on all three other outcomes of political activity discussed by Dorff and shown in the dataset (participation in neighborhood associations, NGOs, and unions). To facilitate this step, you will find additional code to perform the analysis below the relevant code for this memo also attached in the gist of Appendix A. This additional analysis can help corroborate the positive effect on active participation that stronger kinship networks may lead to.

An alternative test one must consider would be establishing a new outcome variable that somehow encompasses participation trends overall. This can be an outcome that measures the total number of different events the individual participated in by adding the four outcomes discussed in the original paper, a binary outcome noting whether a specific combination of outcomes happened (e.g. if we believe NGOs and unions together have an increased effect on political improvement, we might measure participation for both simultaneously), or a completely new outcome independent of those discussed in this analysis. This broader consideration may help identify bigger trends arising from stronger kinship networks that may have greater applicability in policy decisions when seeking to encourage peace.

Conclusions

After our replication analysis, we have mostly verified Dorff's initial conclusions, now supported by balanced covariate distribution between groups and sensitivity analysis. We have found that as kinship networks are stronger, then their effect on a victimized individual's participation in political party meetings positively increases as well. Turning towards possible applications based on these results we would once again find ourselves in agreement with Dorff's initial assessment. We strongly advise shifting recovery efforts on violent communities to be more collective focused. It is clear that for an increased interest in the betterment of the community through political means one needs to support the kinship networks that support the individuals affected.

Additionally, this analysis has implications on the prevention and identification of individuals and communities most likely to suffer more strongly from the effects of victimizations. Those places with looser family ties (e.g. those with higher divorce rates, emancipated individuals, and/or cases of family violence) are less bound to be active in helping out after victimization has occurred, falling into a negative reinforcing loop that would be tough to get out of. To address this, we would advise encouraging strong kinship ties as a preventive measure if violence could be likely in the future.

References

Dorff, C. (2017). Violence, kinship networks, and political resilience: Evidence from Mexico. *Journal of Peace Research*, 54(4), 558-573. doi:10.1177/0022343317691329

Appendix A: Code

For all relevant code used in this calculation, as well as the necessary code for the evaluation of the three other outcomes in the original dataset, please refer to the following gist:

<https://gist.github.com/xrj-22/b7670fb4faaf73d856b4ad247d3f1bdb>

Appendix B: Detailed Matching Results

```
***** (V1) kinship *****
                                Before Matching      After Matching
mean treatment.....          0.63109                0.63109
mean control.....            0.56158                0.62577
std mean diff.....           37.336                  2.855

var ratio (Tr/Co).....         0.84201                1.2194
T-test p-value.....           3.4231e-07              0.53173
KS Bootstrap p-value.. < 2.22e-16                    0.556
```

```
***** (V2) tenure *****
                                Before Matching      After Matching
mean treatment.....           4.1563                  4.1563
mean control.....             4.2155                  4.2122
std mean diff.....            -5.1775                 -4.885

var ratio (Tr/Co).....          1.09                  1.2044
T-test p-value.....            0.45788                 0.54849
KS Bootstrap p-value..          0.684                   0.76
```

```
***** (V3) age *****
                                Before Matching      After Matching
mean treatment.....           37.989                  37.989
mean control.....             41.763                  37.832
std mean diff.....            -26.619                  1.1125

var ratio (Tr/Co).....          0.75338                 1.0893
T-test p-value.....            0.00032786              0.63693
KS Bootstrap p-value.. < 2.22e-16                    0.818
```


***** (V4) women *****

	Before Matching	After Matching
mean treatment.....	0.48029	0.48029
mean control.....	0.53537	0.48746
std mean diff.....	-11.005	-1.4322
var ratio (Tr/Co).....	1.0057	0.99907
T-test p-value.....	0.1189	0.65495

***** (V6) neighsec *****

	Before Matching	After Matching
mean treatment.....	0.53558	0.53558
mean control.....	0.63423	0.53763
std mean diff.....	-19.882	-0.41404
var ratio (Tr/Co).....	1.0783	0.98684
T-test p-value.....	0.0045724	0.79446
KS Bootstrap p-value..	0.006	0.86

***** (V7) party *****

	Before Matching	After Matching
mean treatment.....	0.10677	0.10677
mean control.....	0.11484	0.10677
std mean diff.....	-2.6451	0
var ratio (Tr/Co).....	0.93594	1
T-test p-value.....	0.71032	1
KS Bootstrap p-value..	0.842	1

***** (V8) vote *****

	Before Matching	After Matching
mean treatment.....	3.2913	3.2913
mean control.....	3.4277	3.2965
std mean diff.....	-14.891	-0.56435
var ratio (Tr/Co).....	1.1768	1.0106
T-test p-value.....	0.031454	0.5476
KS Bootstrap p-value..	0.04	0.992

```
***** (v9) rural *****
```

	Before Matching	After Matching
mean treatment.....	0.19713	0.19713
mean control.....	0.35368	0.19713
std mean diff.....	-39.278	0
var ratio (Tr/Co).....	0.69391	1
T-test p-value.....	2.0388e-07	1

```
Before Matching Minimum p.value: < 2.22e-16
Variable Name(s): kinship age Number(s): 1 3

After Matching Minimum p.value: 0.53173
Variable Name(s): kinship Number(s): 1
```

Appendix C: Genetic Matching/Sensitivity Test

```
#STEP 2 - Matching Process
#Defined the predictors and relevant outcomes for genetic matching
attach(foo)
X <- cbind(kinship, tenure, age, women, edu, neighsec, party,
vote, rural, I(kinship*victim))
Y <- foo$att_party.bin
detach(foo)

#Performed genetic matching with victimization as treatment on the predictors above
genout <- GenMatch(Tr=Tr, X=X, M=1, pop.size=200, max=200,
max.generations=10, wait.generations=25)

#Matched units based on the weight matrix generated above and evaluated the effect estimate
mout <- Match(Y=Y, Tr=Tr, X=X, Weight.matrix = genout)
summary(mout)

#Checked for balance between control and tmt groups post-matching process
mb2 <- MatchBalance(victim ~ kinship + tenure + age
+ women + edu + neighsec + party
+ vote + rural, data=foo, match.out=mout)

#STEP 3 - Sensitivity Analysis
#Performed a sensitivity test using the rbounds library
psens(mout, Gamma=1.5, GammaInc=.1)
```