Qualitative Comparative Analysis: How inductive use and measurement error lead to problematic inference
Replication materials

Simon Hug*
Département de science politique et relations internationales,
Université de Genève

November 25, 2012

Abstract

An increasing number of analyses in various subfields of political science employ boolean algebra as proposed in Ragin's (1987) qualitative comparative analysis (QCA). This type of analysis is perfectly justifiable if the goal is to test deterministic hypotheses under the assumption of error-free measures of the employed variables. My contention is, however, that only in a very few research areas are our theories sufficiently advanced to yield deterministic hypotheses. Also, given the nature of our objects of study, error-free measures are largely an illusion. Hence, it is unsurprising that many studies employ QCA inductively and gloss over possible measurement errors. In this paper I address these issues and demonstrate the consequences of these problems with simple empirical examples. In an analysis similar to Monte Carlo simulation I show that using boolean algebra in an exploratory fashion without considering possible measurement errors may lead to dramatically misleading inferences. I then suggest remedies that help researchers to circumvent some of these pitfalls.

Keywords: QCA; qualitative methods; measurement error; induction; scope conditions

* Département de science politique et relations internationales, Faculté des sciences économiques et sociales; Université de Genève; 40 Bd du Pont d'Arve; 1211 Genève 4; Switzerland; phone ++41 22 379 83 78; email: simon.hug@unige.ch

Replication Materials

Here I first present a second example of a Monte Carlo simulation, befor presenting the *R*-code employed in the analyses here and in the main text.

Analysis of a second example

The second example used to illustrate the problems of misspecified scope conditions and measurement errors relies on the study by Osa and Corduneanu-Huci (2003).These two authors wish to assess how social movements are mobilized in non-democracies. To explain the presence or absence of mobilization in 24 cases from 15 countries, the authors rely on five explanatory factors, four of which relate to political opportunities. These are the dynamics of repression (DYNA), the access to media and information (ACCES), the presence of influential allies (INFLU), and elite divisions (ELITE). In addition to these four indicators for political opportunities, the authors also employ an indicator for the presence of social networks (SOCIAL).[1] Table 1 reports the values of the binary explanatory and outcome variables for all the 24 cases considered (see table 6 for a list of all cases with the abbreviation used and the full name).

Table 1 about here

Based on this data Osa and Corduneanu-Huci (2003, 615) wish to ". . . generate empirically grounded hypotheses. . . " Employing QCA and eliminating contradictions, the authors find four prime implicants for the occurrence of mobilization and two prime implicants for the absence of mobilization. In the replication of these results I find the following 2 solutions for mobilization, namely $bCE + ABcd + aBDE + AbdE$ and $bCE + ABcd + aBDE + AcdE$.[2] These have as essential prime implicants $bCE + ABcd + aBDE$ which corresponds to the three first prime implicants reported by Osa and Corduneanu-Huci (2003, 617). The fourth set of prime implicants fails to appear for the simple reason that I also included the contradictions in the analysis, which Osa and Corduneanu-Huci (2003) excluded. When also considering the remainders and the contradictions,

[1] The exact coding of these variables as well as the outcome variable are discussed in Osa and Corduneanu-Huci (2003, 611-614).

[2] I adopt here the usual notation, with uppercase letters indicating the presence of a condition, while a lowercase letter denotes the absence of the condition. The letters correspond to the variables as listed in table 1.

the solution yields as key factors $B + E$, namely media access and the presence of social networks. This is again identical to Osa and Corduneanu-Huci's (2003, 623) result.

Starting from this analysis I then moved to Monte Carlo analyses.[3] I generated 1000 new datasets which were created by randomly removing one, and then two, observations. Table 2 shows the results for the Monte Carlo simulations when one observation out of the 24 is dropped. As these results show, in approximately 10 percent of the cases the original solution is altered. More precisely, if the observation BU88 is dropped E appears as unique solution and hence as unique sufficient condition. Dropping the observation EG53, however, leads to two solutions, namely $E+AB$ and $E+Bd$ and consequently the essential prime implicant E. Dropping any of the other 22 observations from the dataset, however, fails to affect the result of the QCA.

Table 2 about here

When two observations are dropped the share of Monte Carlo runs yielding different solutions increases considerably, as table 3 demonstrates. Obviously the results reported in table 3 are related to the results discussed above. Any pair of dropped observations involving BU88 by definition must yield a solution which differs from the original one. In the present analysis 77 cases involved pairs of dropped observations involving BU88 and yielding E as solution. One pair of dropped observations involving BU88 yielded, however, an additional solution. If both BU88 and CH83LCP are dropped, three solutions, namely $B + E$, $D + E$, and $E + ac$ appear, yielding as essential prime implicants E. On the other hand pairs of dropped observations comprising EG53 must also yield a different result. In table 3 79 cases consisted of pairs involving EG53 and all yielded two solutions, namely $E + AB$ and $E + Bd$ and consequently the essential prime implicant E.

Table 3 about here

Table 4 reports the results from the same analysis where in 1000 runs one observation was recoded from 0 to 1, respectively from 1 to 0. As appears in the last column of this table, in approximately three quarters of the cases such potential measurement errors would not affect the result. More precisely if for

[3]Contrary to the strategy employed for the example presented in the main text I randomly selected cases to be dropped or to be recoded. The main substantive results are identical to systematic changes as done above, especially if the number of simulation runs is sufficiently large, which is the case in the analyses that follow.

any of the following observations the outcome is mismeasured, the result does not change: AR77, CH83, CH83LCP, CN89, GR68, HO68, PL60, PL68, PO58, PO62, RO77, RO78, RO87, SA64, SA76, SP52, SP62, and UR73. Changing any of the remaining 6 (out of 24) observations leads in part, however, to dramatically different results. If the outcome of BU88 is changed the solution is E. Changing the outcome CH73 yields $B + D + aE + cE$, while changing BR77 results in $B + D + AE$. More complicated results appear for the remaining three observations: changing PL56 yields 3 solutions ($AB+AE+CE$, $AB+BC+bE$, $AB+bE+CD$), changing IR80 or EG53 yields two solutions also ($ABcd+AbCE+aBDE+AbdE+bCdE$, $ABcd+AbCE+aBDE+AcdE+bCdE$, respectively $E+AB$, $E+Bd$) with essential prime implicants, however ($ABcd+AbCE+aBDE+bCdE$, respectively E).

Table 4 about here

Table 5 reports the same type of analysis, this time, however, in each run the values of the dependent variable for 2 observations were recoded. Now only in approximately half of the runs do we still find the original solution. In addition, the number of other types of solutions has dramatically increased. Instead of six additional solutions we suddenly have 31 additional ones.

Table 5 about here

As discussed above, Braumoeller and Goertz (2000) provide tools which have hardly ever been employed in the literature but which allow the testing of necessary condition hypotheses when faced with measurement error. Using again Osa and Corduneanu-Huci's (2003) data, I illustrate this approach by assuming that the presence of social networks (E: SOCIAL) is proposed as a necessary condition. As table 6 suggests, two counterexamples to this necessary condition are present. With 24 cases the lower-bound of a one-sided 95% confidence interval for this proportion of counterexamples based on a binomial distribution corresponds to 0.015. Following Braumoeller and Goertz's (2000, 848) p_I-test suggests rejecting the necessary condition hypothesis if this lower bound is higher than the error to be expected in the cell corresponding to the counterexamples in a two-by-two table. For the case at hand there is no direct way to assess the reliability of the measures used, but assuming that one case might be on average misclassified (i.e., 1 out of 24) we already have an error rate of more than 0.04. Consequently, allowing for measurement error, we cannot reject the hypothesis that social networks

are a necessary condition for mobilization.

R-code employed for the analyses

Below is the *R*-code used for the Monte Carlo simulations reported in the main text.

R-Code for Monte Carlo simulation with one omitted observation

```
# Welfare state

for (j in 1:(nrow(raginwela))) {data(raginwela)
raginwelar<-raginwela
print(raginwelar[j,])
raginwelarr<-raginwelar[-j,]
qmcc(raginwelarr, outcome="W.", expl.1=TRUE, incl.rem=F, incl.ctr=TRUE,
  details=FALSE, quiet=TRUE)
 j <-j+1
rm(raginwelar,raginwelarr,rr1)   }

# Mobilization

k<-1000
for (j in 1:k) {data(Osa)
Osar<-Osa
rr1 <- as.integer(runif(1,1,25))
Osarr<-Osar[-rr1,]
print(Osar[rr1,])
qmcc(Osarr, outcome="OUT", expl.1=TRUE, incl.rem=TRUE, incl.ctr=TRUE,
 > details=FALSE, quiet=TRUE)
 j <-j+1
rm(Osar,Osarr,rr1)   }
```

R-Code for Monte Carlo simulation with two omitted observations

```
# Welfare state

for (j in 1:(nrow(raginwela))) {
for (k in (j+1):(nrow(raginwela))) {

data(raginwela)
raginwelar<-raginwela
print(raginwelar[j,])
print(raginwelar[k,])
raginwelarr<-raginwelar[-c(j,k),]
qmcc(raginwelarr, outcome="W.", expl.1=TRUE, incl.rem=F, incl.ctr=TRUE,
  details=T, quiet=TRUE)
 k <-k+1
rm(raginwelar,raginwelarr,rr1)   }
 j <-j+1
rm(raginwelar,raginwelarr,rr1)   }

# Mobilization



k<-1000
for (j in 1:k) {data(Osa)
Osar<-Osa
rr1 <- as.integer(runif(1,1,25))
Osarr<-Osar[-rr1,]
print(Osar[rr1,])
rr1 <- as.integer(runif(1,1,24))
Osar<-Osarr[-rr1,]
print(Osarr[rr1,])
qmcc(Osar, outcome="OUT", expl.1=TRUE, incl.rem=TRUE, incl.ctr=TRUE,
```

```
  details=FALSE, quiet=TRUE)
 j <-j+1
rm(Osar,Osarr,rr1)   }
```

R-Code for Monte Carlo simulation with one measurement error

```
# Welfare state


for (j in 1:(nrow(raginwela))) {data(raginwela)
raginwelar<-raginwela
if((raginwela[j,5]==1)) raginwelar[j,5]<-0
if((raginwela[j,5]==0)) raginwelar[j,5]<-1
print(j)
print(raginwelar[j,])
qmcc(raginwelar, outcome="W.", expl.1=TRUE, incl.rem=F, incl.ctr=TRUE,
  details=FALSE, quiet=TRUE)
 j <-j+1
rm(raginwelar,rr)   }

# Mobilization


k<-1000
for (j in 1:k) {data(Osa)
Osar<-Osa
rr <- as.integer(runif(1,1,25))
if((Osar[rr,6]=1)) Osar[rr,6]=0
if((Osar[rr,6]=0)) Osar[rr,6]=1
print(j)
```

```
print(Osar[rr,])
qmcc(Osar, outcome="OUT", expl.1=TRUE, incl.rem=TRUE, incl.ctr=TRUE,
  details=FALSE, quiet=TRUE)
 j <-j+1
rm(Osar,rr)   }
```

R-Code for Monte Carlo simulation with two measurement errors

```
# Welfare state

for (j in 6:(nrow(raginwela))) {
for (k in (j+1):(nrow(raginwela))) {

data(raginwela)
raginwelar<-raginwela
if((raginwela[j,5]==1)) raginwelar[j,5]<-0
if((raginwela[j,5]==0)) raginwelar[j,5]<-1
if((raginwela[k,5]==1)) raginwelar[k,5]<-0
if((raginwela[k,5]==0)) raginwelar[k,5]<-1
print(j)
print(raginwelar[j,])
print(k)
print(raginwelar[k,])
qmcc(raginwelar, outcome="W.", expl.1=TRUE, incl.rem=F, incl.ctr=F,
  details=T, quiet=T, chart=F)
next
 k <-k+1
rm(raginwelar)   }
 j <-j+1
rm(raginwelar)   }
```

```
# Mobilization


k<-1000
for (j in 1:k) {data(Osa)
Osar<-Osa
rr1 <- as.integer(runif(1,1,25))
rr2<-rr1
while(rr2==rr1){rr2 <- as.integer(runif(1,1,25))}
if((Osar[rr1,6]=1)) Osar[rr1,6]=0
if((Osar[rr1,6]=0)) Osar[rr1,6]=1
if((Osar[rr2,6]=1)) Osar[rr2,6]=0
if((Osar[rr2,6]=0)) Osar[rr2,6]=1
print(j)
print(Osar[rr1,])
print(Osar[rr2,])
qmcc(Osar, outcome="OUT", expl.1=TRUE, incl.rem=TRUE, incl.ctr=TRUE,
  details=FALSE, quiet=TRUE)
 j <-j+1
rm(Osar,rr1,rr2)   }
```

Tables

Table 1: Data from Osa and Corduneanu-Huci (2003)

country	A: DYNA	B: ACCES	C: INFLU	D: ELITE	E: SOCIAL	OUTCOME
SP52	1	0	0	0	0	0
EG53	0	1	0	1	0	1
PL56	0	1	0	1	1	1
PO58	1	0	1	1	1	1
PL60	0	0	1	0	0	0
PO62	1	0	1	1	1	1
SP62	0	1	1	1	1	1
SA64	1	0	0	0	0	0
GR68	0	1	1	1	1	1
HO68	0	1	1	1	1	1
PL68	1	0	1	1	1	1
CH73	1	0	1	0	1	1
UR73	1	0	0	0	0	0
AR77	1	1	0	0	1	1
SA76	1	1	0	0	1	1
BR77	0	0	1	0	1	1
RO77	1	0	0	0	1	1
RO78	1	0	0	0	0	0
IR80	0	0	1	1	1	1
CH83	0	1	1	1	1	1
CH83LCP	0	1	0	1	0	0
RO87	1	0	0	0	1	1
BU88	1	1	0	0	0	1
CN89	0	1	1	1	1	1

Table 2: QCA with 1 observation removed (remainder included)

Solution	Essential prime implicants	%	N
B + E		91.40	914
E		4.20	42
E + AB			
E + Bd	E	4.40	44

Table 3: QCA with 2 observations removed (remainder included)

Solution	Essential prime implicants	%	N
B + E		83.90	839
E + AB			
E + Bd	E	7.90	79
E		7.70	77
B + E			
D + E			
E + ac	E	0.30	3

Table 4: QCA with 1 error (remainder included)

Solution	Essential prime implicants	%	N
B + E		73.40	734
E		4.30	43
B + D + AE		3.30	33
B + D + aE + cE		5.30	53
AB + AE + CE			
AB + BC + bE			
AB + bE + CD		4.90	49
ABcd + AbCE + aBDE + AbdE + bCdE			
ABcd + AbCE + aBDE + AcdE + bCdE	ABcd + AbCE + aBDE + bCdE	4.00	40
E + AB			
E + Bd	E	4.80	48

Table 5: QCA with 2 errors (remainder included)

Solution	Essential prime implicants	%	N
B + E		53.20	532
ABcd + abCE + aBDE + AcdE + bCDE		0.30	3
ABcd + AbCE + aBDE + bCdE		0.90	9
ABcd + abCE + aCDE + AcdE + bCDE		0.40	4
ABcd + aBDE + AbdE + bCDE		0.30	3
AbCE + aBDE + AbdE + bCdE + ABcde		1.00	10
AbCE + aBDE + AcdE + bCdE		0.50	5
B + AD + cE + adE		0.30	3
B + AE		0.60	6
B + aE + cE		1.00	10
B + CE		0.30	3
B + D + aE		0.90	9
B + D + AE		6.30	63
B + D + aE + cE		4.40	44
B + D + cE		0.40	4
B + dE		1.10	11
bCE + ABcd + aBDE		0.40	4
bCE + ABcd + aCDE		1.30	13
bCE + aBDE + AbdE + ABcde		1.20	12
bCE + AbdE + aCDE + ABcde		0.80	8
bCE + aCDE + AcdE		0.40	4
D + AE		0.50	5
D + aE + cE		0.50	5
E		7.00	70
AB + AE + CE			
AB + BC + bE			
AB + bE + CD		2.40	24
AB + bE			
Bd + bE	bE	2.20	22
ABcd + AbCE + aBDE + AbdE + bCdE			
ABcd + AbCE + aBDE + AcdE + bCdE	ABcd + AbCE + aBDE + bCdE	3.00	30
ABcd + AbCE + AbdE + aCDE			
ABcd + AbCE + aCDE + AcdE			
ABcd + AbdE + aCDE + bCDE	ABcd + aCDE	0.60	6
ABcd + AbCE + AbdE + bCdE + aBcDE			
ABcd + AbCE + AcdE + bCdE + aBcDE	ABcd + AbCE + bCdE + aBcDE	2.10	21
ABcd + AbCE + AbdE + bCdE + aBCDE			
ABcd + AbCE + AcdE + bCdE + aBCDE	ABcd + AbCE + bCdE + aBCDE	0.40	4
B + D + AC			
B + D + AE	B + D	0.50	5
E + AB			
E + Bd	E	4.80	48

Table 6: List of cases in data from Osa and Corduneanu-Huci (2003)

abbreviation	country and year of event
SP52	Spain 1952
EG53	East Germany
PL56	Poland 1956
PO58	Portugal 1958
PL60	Poland 1968
PO62	Portugal 1962
SP62	Spain 1962
SA64	South Africa 1964
GR68	Greece 1968
HO68	Honduras 1968
PL68	Poland 1968
CH73	Chile 1973
UR73	Uruguay 1973
AR77	Argentina 1977
SA76	South Africa 1976
BR77	Brazil 1977
RO77	Romania 1977
RO78	Romania 1978
IR80	Iran 1980
CH83	Chile 1983
CH83LCP	Chile 1983 (low combative publaciones)
RO87	Romania 1987
BU88	Burma 1988
CN89	China 1989

References

Braumoeller, Bear F. and Gary D. Goertz. 2000. "The Methodology of Necessary Conditions." *American Journal of Political Science* 44(4 October):844–858.

Osa, Maryjane and Cristina Corduneanu-Huci. 2003. "Running Uphill: Political Opportunity in Non-Democracies." *Comparative Sociology* 2(4):605–629.

Ragin, Charles. 1987. *The Comparative Method: Moving Beyond Qualitative and Quantative Strategies*. Berkeley: University of California Press.