Sociological Methods & Research
http://smr.sagepub.com/

Assessing the Robustness of Crisp-set and Fuzzy-set QCA Results

Svend-Erik Skaaning

Sociological Methods & Research 2011 40: 391 originally published online 6 April 2011
DOI: 10.1177/0049124111404818

The online version of this article can be found at:
http://smr.sagepub.com/content/40/2/391

Published by:
SAGE
http://www.sagepublications.com

Article

Sociological Methods & Research
40(2) 391–408

DOI: 10.1177/0049124111404818
http://smr.sagepub.com

Assessing the Robustness of Crisp-set and Fuzzy-set QCA Results

Svend-Erik Skaaning[1]

Abstract

Configurational comparative methods constitute promising methodological tools that narrow the gap between variable-oriented and case-oriented research. Their infancy, however, means that the limits and advantages of these techniques are not clear. Tests on the sensitivity of qualitative comparative analysis (QCA) results have been sparse in previous empirical studies, and so has the provision of guidelines for doing this. Therefore this article uses data from a textbook example to discuss and illustrate various robustness checks of results based on the employment of crisp-set QCA and fuzzy-set QCA. In doing so, it focuses on three issues: the calibration of raw data into set-membership values, the frequency of cases linked to the configurations, and the choice of consistency thresholds. The study emphasizes that robustness tests, using systematic procedures, should be regarded as an important, and maybe even indispensable, analytical step in configurational comparative analysis.

Keywords

QCA results, crisp-set, fuzzy-set, configurational comparative methods, robustness tests

As an attempt to bridge the quantitative–qualitative divide, Charles Ragin (1987) published the book *The Comparative Method* in which he introduced an innovative logic of complexity reduction to the social sciences, namely,

[1]Aarhus University, Aarhus, Denmark

Corresponding Author:
Svend-Erik Skaaning, Department of Political Science, Aarhus University, Bartholins Allé 7, DK-8000 Aarhus C, Denmark
Email: Skaaning@ps.au.dk

qualitative comparative analysis (QCA). Since then, QCA as a technique has been refined and supplemented with multivalue and fuzzy-set extensions, and continuous software developments have taken place to facilitate applications. The QCA methods are gaining ground, as shown by the increasing number of scholars using them in publications, and this trend is likely to be fed by a number of new books on why and how to employ QCA (Caramani 2008; Ragin 2008; Rihoux and Ragin 2009; Schneider and Wagemann 2007).

However, as befits new methods, many aspects of their modus operandi still need to be addressed to augment their plausibility. This article takes issue with the question of the robustness of QCA results, that is, the degree to which the solutions are sensitive to (small) changes in the discriminatory choices made by researchers in various stages of the process of systematic complexity reduction once the raw data have been compiled. Thus far, the issue has been underemphasized[1] in technical discussions and empirical applications alike. Contenders—primarily found among scholars from the “variable-oriented” (i.e., statistical) camp—have argued that QCA results are very sensitive to the researcher’s discriminatory power (e.g., Goldthorpe 1997; Lieberson 2004; Seawright 2005). The proponents, on the other hand, have largely rejected this judgment. In doing so, they have emphasized that their prescription of intimate knowledge of theory and cases minimizes the potential problem (Ragin and Rihoux 2004; Schneider and Wagemann 2010). They are basically of the opinion that QCA is best used in a direct dialogue with the cases and thus more compatible with within-case analysis than other forms of cross-case methods. Hence, rather than robustness, they are occupied with the degree to which QCA can contribute to a better understanding of diversity between the cases.

If one takes stock of these arguments, the flagrant lack of theoretical and empirical evidence means that the criticism of QCA rests on a shaky foundation. But also the advocates of QCA methods have to admit that they cannot fully rule out arbitrariness (or bias) in their decisions when employing the method to reduce complexity. Thus, the development of procedures to guide such impact assessment is needed.

In doing this, I primarily make use of the data and the theoretical model from a major research project on the conditions of authoritarianism and democracy in interwar Europe (Berg-Schlosser and Mitchell 2000, 2003). The primary reason is that the data set and the model have been used to illustrate the steps and logic of QCA in a new textbook edited by Benoit Rihoux and Charles Ragin (2009), *Configurational Comparative Methods*. Specifically, they (and I) use QCA to examine the relationship between survival of democratic regimes and socioeconomic development and stability. In this way, my illustrative analyses are based on data and models that have previously been employed by leading experts to illustrate the logic and "best practices" of QCA. Moreover, I also carry analyses of two additional data sets. One of these has been heavily used in other introductions to QCA (Ragin 2000; Schneider and Wagemann 2007; Skaaning and Kvist 2010). It covers advanced industrial democratic countries and provides data on the generosity of welfare states, which represent the outcome, and four conditions. As regards the other supplementary analysis, it is based on my dissertation on civil liberty in Latin American and postcommunist countries (Skaaning 2006). This is based on a "real" and carefully constructed data set and a theoretically guided selection of cases and conditions. To keep a limited focus, the issues of case and model selection are regarded as settled for all the analysis, despite their obvious relevance for the solidity of empirical results and even though this choice goes somewhat against the spirit of the QCA approach.

Attempting to make researchers more conscious about robustness checks when employing configurational comparative methods, the article addresses the topic in the following way: First, different ways of assessing the sensitivity of QCA solutions are presented and the issues of calibration thresholds, case frequency, and consistency levels are selected for further scrutiny. The second part presents the results of the crisp-set (csQCA) and fuzzy-set (fsQCA) applications, emanating from minor adjustments in the calibration of raw data into crisp-set and fuzzy-set values, and the subsequent part focuses on the results based on modifications of frequency and consistency thresholds. The last part provides a general discussion of the relative sensitivity of results based on csQCA and fsQCA, respectively, before the conclusion sums up the suggested guidelines and briefly discusses the preliminary findings on the robustness of QCA results.

Different Ways to Assess the Robustness of QCA Results

Various analytical choices can be made with the intention of warranting the plausibility of findings from the employment of QCA methods. Apart from the cautious compilation and/or construction of data and thorough elaboration of the theoretical framework, a number of procedures exist to support the robustness of empirical results. One of these is methodological triangulation, where a particular research question is subjected to a multimethods test. This means that different methodological approaches and tools are employed in the study. Accordingly, it has been proposed to combine variable-oriented methods, such as statistical analysis, and case-oriented methods, such as process-tracing, in a nested analysis (Lieberman 2005; Rohlfing 2008).

It is obvious that the comparative configurational methods offer yet other alternatives to be considered in this kind of endeavour. Basically, it depends on the causal universe under consideration whether it is more rewarding to see different methodological approaches as complementary or competitive alternatives. Since we cannot determine the character of social phenomena a priori, we have to apply methods based on different assumptions and subsequently evaluate the plausibility of their respective results based on theoretical and substantial insight (Rihoux et al. 2009:172).

Another way to check the robustness of results is to change the case selection, the model specification, and/or the operationalization of the outcome and conditions. These are all well-known procedures for assessing the robustness of results in standard statistical analysis. Concerning case selection, an aggregate-level analysis can be supplemented by an analysis focusing on a more disaggregate level (and vice versa), a wide scope can be replaced by a narrow scope (and vice versa), outliers can be removed to consider their "distortion" of the general picture, and smaller adjustments due to uncertainty about delimitations of the population is yet another option.

As regards the basic theoretical model, one can test whether similar results emerge if it is modified through a reduction, addition, or replacement of explanatory factors or a revision of the hypothesized relationships (sequence, interaction, necessity, sufficiency, etc.). Finally, the effect of different ways to operationalize *explanans* and *explanandum* can be examined, for example, through the replacement of one (proxy) measure by another.

Owing to their functional logic (cf. Ragin 1987, 2008), some extra sensitivity checks apply to the QCA methods. The operation of these methods necessitates a translation of the raw data into set-memberships. Sometimes this procedure is straightforward, whereas other calibrations, despite attempts of theoretical and/or empirical justification, introduce some degree of arbitrariness. Hence, it becomes crucial to explore the similarity of findings across different reasonable breakpoints. Furthermore, the researcher can change the frequency threshold, namely, the number of cases showing a particular combination of conditions that is found to be enough to consider the configuration empirically and/or theoretically relevant. A modification of the absolute or relative criteria, determining the minimal acceptable consistency level differentiating (set-theoretically) consistent configurations from the inconsistent ones, also represents a form of sensitivity control.[2]

To provide focused discussions and illustrations, this study does not consider changes in methods, case selection, theoretical model specification, and operationalization of outcome and conditions. This leaves us with the three types of robustness check to be addressed in relation to csQCA and fsQCA, namely, (1) changes in thresholds related to the calibration of raw data into set-memberships, (2) the frequency of cases linked to the configurations, and (3) the consistency of configurations.

Changing the Calibration Thresholds

The employment of the complexity reducing logic of QCA necessitates a calibration of the data. Along these lines, csQCA requests a dichotomization of the outcome and the explanatory conditions while fsQCA, as a minimum, needs a specification of two qualitative anchors defining full and no membership in the sets defining the conditions and outcome in question (Ragin 2000:161). In addition, the break-even points—separating cases more in particular sets from cases more out of the set—also need a separate specification.

Table 1 provides an overview of the breakpoints originally used in the textbook example to translate the raw scores into crisp and fuzzy values (*i*) and also presents the two alternative breakpoints (*ii* or *iii*). The thresholds used to make the data feasible for the crisp-set analysis correspond to the cross-over points used in the calibration of the original scores into fuzzy values (Ragin 2009:93). All the fuzzy-set values have been assigned to the cases relatively to the raw scores using the qualitative anchors and the direct method of calibration (cf. Ragin 2008:Ch. 5).

The outcome to be explained is the survival (or not) of democratic regimes in interwar Europe. Inspired by Seymour M. Lipset's (1959) seminal study, four conditions—developed, urban, literate, and industrial—are used to measure different aspects of modernization. The measurement of these features follows common practice and is based on data for GNP/cap, percentage of people living in cities, literacy rate, and percentage of people in the industrial labor force, respectively. The last condition, stability, is operationalized as the number of governments (cabinets) during the period.

Also shown in Table 1 are the alternative thresholds for the five conditions that I have used in the assignment of set-membership. While many other breakpoints are possible, they are placed at levels near the original anchors. Only minor changes are made to ensure that similar theoretical justifications could apply to the original as well as the new anchors defining (degree of) set-memberships.

The following crisp-set and fuzzy-set analyses first proceed as standard practice prescribes (Ragin 2009; Rihoux 2009), thereafter comparing the results based on different calibration anchors with each other and the principal findings. As there is much agreement in the literature about whether (or to which degree) democracies survived and to reduce the complexity and number of the replication analyses, alternative thresholds for the outcome are not set and tested.

Table 1. Values Used in the Translation into Set Membership Scores

	Full membership	Cross-over point	No membership
Survival (i)	10	0	–9
Developed (i)	900	550	400
Developed (ii)	1000	600	450
Developed (iii)	800	500	350
Urban (i)	65	50	25
Urban (ii)	70	60	30
Urban (iii)	60	40	20
Literate (i)	90	75	50
Literate (ii)	95	80	60
Literate (iii)	85	70	40
Industrial (i)	40	30	20
Industrial (ii)	45	35	25
Industrial (iii)	35	25	15
Stable (i)	5	9.5	15
Stable (ii)	4	8.5	13
Stable (iii)	6	10.5	17

Note: Original anchors (i), first alternative (ii), and second alternative (iii).

Crisp-set Analysis

The first step in the crisp-set analysis is to construct a truth table. This is done through a regrouping of all identical cases, in terms of their values on the conditions, into a single configuration. If two or more cases belong to the same configuration but have different outcomes, a contradiction is present.[3] The resulting truth table (see Table 2) is contradiction free. The 18 cases have been reduced to 9 observed configurations, out of which 3 show a positive outcome and 6 a negative outcome.

Further reduction of complexity can be achieved by an employment of the minimization rule. It says that if two Boolean expressions (configurations) differ in only one causal condition yet produce the same outcome, then the causal condition distinguishing the two expressions can be considered irrelevant and can be removed to create simpler, combined expressions (Ragin 1987:93). Using this logic of reduction on the binary data, the result obtained for positive outcome is:[4]

$$\text{DEVELOPED*LITERATE*INDUSTRIAL*STABLE}_{\text{(Belgium, Czechoslovakia, Netherlands, United Kingdom + France, Sweden)}} + \text{DEVELOPED*URBAN*LITERATE*STABLE}_{\text{(Finland, Ireland + France, Sweden)}}$$

Table 2. Truth Table Representation of Crisp-set Data

Cases	Survival	Developed	Urban	Literate	Industrial	Stable
Belgium, Czechoslovakia, Netherlands, United Kingdom	1	1	1	1	1	1
France, Sweden	1	1	0	1	1	1
Finland, Ireland	1	1	0	1	0	1
Germany	0	1	1	1	1	0
Austria	0	1	0	1	1	0
Estonia	0	0	0	1	0	1
Hungary, Poland	0	0	0	1	0	0
Italy, Romania	0	0	0	0	0	1
Greece, Portugal, Spain	0	0	0	0	0	0

If also all logical cases, namely, combinations of conditions with no empirical instances, are included in the reduction procedure, the minimized solution for positive outcome becomes very parsimonious, namely: $\text{DEV*STA}_{\text{(Belgium, Czechoslovakia, Netherlands, United Kingdom + France, Sweden + Finland, Ireland)}}$. In ordinary language this result indicates that democracy survived in those countries that were developed and stable.

The next technical move in a standard application of QCA would be to find the "intermediate" solution based on the use of "easy counterfactuals" (cf. Ragin 2008:Ch. 9; Ragin and Rihoux 2004).[5] Thereafter a supplementary test of basically the same model on the negated outcome (not survival) is advisable because of the potential lack of causal symmetry (cf. Ragin and Rihoux 2004). None of these steps are carried out, however, as the aim of this article is to gain methodological rather than substantial knowledge. Instead, I go directly to the robustness checks of the principal results that are similar to those achieved in the textbook.

First, the same model is run with all combinations of the different breakpoints—the original, slightly higher, and slightly lower. This means that a total of 242 (3 × 3 × 3 × 3 × 3 – 1) additional analyses have been run. The solutions provided by the software (TOSMANA) and linked to positive outcome are shown in Table 3. The names of the conditions have been abbreviated. All results refer to analyses including logical cases. C indicates the presence of one or more contradictions. In the case of more than one logical solution, only one is reported. Finally, bold marks the principal result.

Table 3 shows that 78 of the combinations are free from contradictions. All instances involving either the higher calibration point as regards the condition developed or the lower calibration point as regards the condition stable lead to contradictions. A closer look at the results reveals that the changes

Table 3. Results Based on Different Combinations of Breakpoints

			Dev i			Dev ii			Dev iii		
			Urb i	Urb ii	Urb iii	Urb i	Urb ii	Urb iii	Urb i	Urb ii	Urb iii
Lit i	Ind i	Sta i	DEV*STA	DEV*STA	DEV*STA	C	C	C	DEV*LIT*STA	DEV*LIT*STA	DEV*LIT*STA
		Sta ii	DEV*ind + DEV*STA	DEV*ind + DEV*STA	DEV*ind + DEV*STA	C	C	C	DEV*STA + DEV*LIT*ind	DEV*STA + DEV*LIT*ind	DEV*STA + DEV*LIT*ind
		Sta iii	C	C	C	C	C	C	C	C	C
	Ind ii	Sta i	DEV*STA	DEV*STA	DEV*STA	C	C	C	DEV*LIT*STA	DEV*LIT*STA	DEV*LIT*STA
		Sta ii	C	C	C	C	C	C	C	C	C
		Sta iii	C	C	C	C	C	C	C	C	C
	Ind iii	Sta i	DEV*STA	DEV*STA	DEV*STA	C	C	C	DEV*LIT*STA	DEV*LIT*STA	DEV*LIT*STA
		Sta ii	DEV*ind + DEV*STA	DEV* ind + DEV*STA	DEV*ind + DEV*STA	C	C	C	DEV*ind + DEV*STA	DEV*ind + DEV*STA	DEV*ind + DEV*STA
		Sta iii	C	C	C	C	C	C	C	C	C
Lit ii	Ind i	Sta i	DEV*STA	DEV*STA	DEV*STA	C	C	C	DEV*LIT*STA	DEV*LIT*STA	DEV*LIT*STA
		Sta ii	DEV*ind + DEV*STA	DEV*ind + DEV*STA	DEV*ind + DEV*STA	C	C	C	DEV*STA + DEV*LIT*ind	DEV*STA + DEV*LIT*ind	DEV*STA + DEV*LIT*ind
		Sta iii	C	C	C	C	C	C	C	C	C
	Ind ii	Sta i	DEV*STA	DEV*STA	DEV*STA	C	C	C	DEV*LIT*STA	DEV*LIT*STA	DEV*LIT*STA
		Sta ii	C	C	C	C	C	C	C	C	C
		Sta iii	C	C	C	C	C	C	C	C	C
	Ind iii	Sta i	DEV*STA	DEV*STA	DEV*STA	C	C	C	DEV*LIT*STA	DEV*LIT*STA	DEV*LIT*STA
		Sta ii	DEV*ind + DEV*STA	DEV*ind + DEV*STA	DEV*ind + DEV*STA	C	C	C	DEV*ind + DEV*STA	DEV*ind + DEV*STA	DEV*ind + DEV*STA
		Sta iii	C	C	C	C	C	C	C	C	C
Lit iii	Ind i	Sta i	DEV*STA	DEV*STA	DEV*STA	C	C	C	C	C	C
		Sta ii	DEV*ind + DEV*STA	DEV*ind + DEV*STA	DEV*ind + DEV*STA	C	C	C	C	C	C
		Sta iii	C	C	C	C	C	C	C	C	C
	Ind ii	Sta i	DEV*STA	DEV*STA	DEV*STA	C	C	C	C	C	C
		Sta ii	C	C	C	C	C	C	C	C	C
		Sta iii	C	C	C	C	C	C	C	C	C
	Ind iii	Sta i	DEV*STA	DEV*STA	DEV*STA	C	C	C	C	C	C
		Sta ii	DEV*ind + DEV*STA	DEV*ind + DEV*STA	DEV*ind + DEV*STA	C	C	C	DEV*ind + DEV*STA	DEV*ind + DEV*STA	DEV*ind + DEV*STA
		Sta iii	C	C	C	C	C	C	C	C	C

Note: C indicates the presence of one or more contradictions. Bold marks the principal result.

have not had a major effect as the results are identical or, at least, very similar to the original, namely, DEV*STA. In fact, only two alternative results emerge and one of these, DEV*LIT*STA, constitutes a subset of the principal result, whereas the principal result is a subset of the other: DEV*ind + DEV*STA. This finding cannot be generalized on this background. But on the other hand, it shows that small changes in the anchors do not necessarily lead to completely different conclusions.

Fuzzy-set Analysis

Whereas the consistency criterion was not really relevant for the crisp-set analysis previously presented, it plays an important role in the fuzzy-set analysis. When testing the necessity of the conditions with the fuzzy-set method, the membership values in the set of democratic survivors have to be consistently lower than or equal to the membership values associated with the individual conditions. It is advisable to use a demanding threshold for the level of consistency in test of necessity, so it is set to 0.90.

All the conditions—in both their original, negated, and modified (recalibrated) versions—are tested for necessity in relation to positive outcome. The test results in Table 4, linked to the original values, indicate that two of the conditions (LITERATE and STABLE) are necessary for the survival of democracy. However, whereas the explicit connection between LITERATE and the outcome in question tends to be very solid, the conclusion concerning the conditions STABLE and DEVELOPED rests on a more shaky foundation. Then again, very large changes in the consistency levels are not the dominant pattern across different calibration anchors. This means that changes in the consistency criterion to, for example, 0.85 or 0.95, would not have much effect on the conclusions derived from the search for necessary conditions. But is the picture different as regards the conditions' sufficiency?

When Ragin (2009) introduced the truth table algorithm in relation to sufficiency tests in fsQCA, the analytical logic became much more like the one of csQCA. However, this is not to say that similar results are unlikely to appear from the resembling methods as the former implements a more precise and demanding assessment of set theoretic consistency (Ragin 2009:119). Table 5 shows virtually the same truth table that was presented in relation to the crisp-set analysis. The only difference is that the outcome values have been replaced with the consistency levels of the different configurations, namely, the evidence used to determine outcome value in a fuzzy-set analysis.

Table 4. Consistency Levels of Conditions—Tests of Necessity

Developed i	Original	0.83	Developed ii	0.76	Developed iii	0.90
	Negated	0.28		0.35		0.23
Urban i	Original	0.54	Urban ii	0.47	Urban iii	0.61
	Negated	0.58		0.62		0.54
Literate i	Original	0.99	Literate ii	0.97	Literate iii	1.00
	Negated	0.10		0.14		0.08
Industrial i	Original	0.67	Industrial ii	0.53	Industrial iii	0.77
	Negated	0.42		0.55		0.35
Stable i	Original	0.92	Stable ii	0.87	Stable iii	0.96
	Negated	0.22		0.28		0.17

Table 5. Truth Table Representation of Fuzzy-set Data

Consistency	Developed	Urban	Literate	Industrial	Stable
0.90	1	1	1	1	1
0.80	1	0	1	0	1
0.71	1	0	1	1	1
0.53	0	0	1	0	1
0.52	0	0	1	0	0
0.45	1	1	1	1	0
0.39	1	0	1	1	0
0.28	0	0	0	0	1
0.22	0	0	0	0	0

In the textbook example, Ragin (2009) uses a consistency criterion of 0.80. Doing the same,[6] the result of the sufficiency analysis is: DEV*ind + URB*STA.

Concluding the fuzzy-set analysis, 242 additional replication analyses are run to assess the impact of changes in the breakpoints defining set-memberships. The results are shown in Table 6. Once again, all results refer to the analysis for positive outcome using logical cases and a consistency benchmark of 0.80. In the case of more than one logical solution, only one of them is reported.

The findings show that reasonable changes in the translations into fuzzy values do influence the results to a nontrivial degree. On the other hand, most of the formulas are either identical or at least very similar to the baseline result. Moreover, almost all terms show the expected direction as indicated by the dominance of upper-case letters. Once again, the picture is not clear-cut, meaning that proponents and opponents of QCA have room to argue whether the glass is half full or half empty.

Table 6. Sufficiency Test Results Using Alternative Combinations of Calibrations

			Dev i			Dev ii			Dev iii		
			Urb i	Urb ii	Urb iii	Urb i	Urb ii	Urb iii	Urb i	Urb ii	Urb iii
Lit i	Ind i	Sta i	DEV*ind +URB*STA	URB + DEV*ind	DEV*ind + URB*STA	DEV*ind + DEV*URB*STA	URB + DEV*ind	DEV*ind + DEV*URB*STA	URB*STA	URB	URB*STA
		Sta ii	URB*STA	URB	URB*STA	DEV*STA + IND*STA	DEV*STA + IND*STA	DEV*STA	URB*STA	URB	URB*STA
		Sta iii	DEV*ind + URB*STA	URB + DEV*ind	DEV*ind + URB*STA	DEV*ind + DEV*URB*STA	URB + DEV*ind	DEV*ind + DEV*URB*STA	URB*STA	URB	URB*STA
	Ind ii	Sta i	URB*STA	URB	DEV*STA	DEV*STA	DEV*STA + IND*STA	DEV*STA	URB*STA	URB	URB*STA
		Sta ii	DEV*STA	DEV*STA	DEV*STA	DEV*STA + IND*STA	DEV*STA + IND*STA	DEV*STA	URB*STA	URB	URB*STA
		Sta iii	URB*STA	URB	URB*STA	DEV*URB*STA	IND*STA	DEV*URB*STA	URB*STA	URB	URB*STA
	Ind iii	Sta i	DEV*ind + URB*STA	URB + DEV*ind	DEV*ind + URB*STA	DEV*ind + URB*STA	URB + DEV*ind	DEV*STA	DEV*ind + URB*STA	URB + DEV*ind	DEV*ind + URB*STA
		Sta ii	DEV*ind + DEV*STA	DEV*ind + DEV*STA	DEV*ind + DEV*STA	DEV*STA + IND*STA	DEV*STA + IND*STA	DEV*STA	DEV*ind + URB*STA	DEV*ind + DEV*STA	DEV*ind + URB*STA
		Sta iii	DEV*ind + URB*STA	URB + DEV*ind	DEV*ind + URB*STA	DEV*ind + URB*STA	URB + DEV*ind	DEV*ind + DEV*URB*STA	DEV*ind + URB*STA	URB + DEV*ind	DEV*ind + URB*STA
Lit ii	Ind i	Sta i	DEV*ind + URB*STA	URB + DEV*ind	DEV*ind + URB*STA	DEV*ind + URB*STA	URB + DEV*ind	DEV*ind + DEV*URB*STA	URB*STA	URB	URB*STA
		Sta ii	URB*STA + DEV*ind*STA	DEV*STA	DEV*STA	DEV*STA + IND*STA	DEV*STA + IND*STA	DEV*STA	URB*STA	IND*STA	IND*STA
		Sta iii	DEV*ind + URB*STA	URB + DEV*ind	DEV*ind + URB*STA	DEV*ind + URB*STA	URB + DEV*ind	DEV*ind + DEV*URB*STA	URB*STA + DEV*LIT*ind	URB + DEV*LIT*ind	URB*STA + DEV*LIT*ind
	Ind ii	Sta i	DEV*STA	DEV*STA	DEV*STA	DEV*STA + IND*STA	DEV*STA + IND*STA	DEV*STA	URB*STA	URB	URB*STA
		Sta ii	DEV*STA	DEV*STA	DEV*STA	DEV*STA + IND*STA	DEV*STA + IND*STA	DEV*STA	URB*STA	URB	DEV*STA
		Sta iii	URB*STA	URB	URB*STA	URB*STA	URB	DEV*URB*STA	URB*STA	URB	URB*STA
	Ind iii	Sta i	DEV*ind + URB*STA	URB + DEV*ind	DEV*STA	DEV*ind + URB*STA	URB + DEV*Ind	DEV*STA	DEV*ind + URB*STA	URB + DEV*ind	DEV*LIT*STA
		Sta ii	DEV*ind + DEV*STA	DEV*ind + DEV*STA	DEV*ind + DEV*STA	DEV*STA + IND*STA	DEV*STA + IND*STA	DEV*STA	DEV*ind + DEV*STA	DEV*ind + DEV*STA	DEV*ind + DEV*STA
		Sta iii	DEV*ind + URB*STA	URB + DEV*ind	DEV*ind + URB*STA	DEV*ind + URB*STA	URB + DEV*ind	DEV*ind + DEV*URB*STA	DEV*ind + URB*STA	URB + DEV*ind	DEV*ind + URB*STA
Lit iii	Ind i	Sta i	DEV*ind + URB*STA	URB + DEV*ind	DEV*ind + URB*STA	DEV*ind + DEV*URB*STA	URB + DEV*ind	DEV*ind + DEV*URB*STA	URB*STA	URB	URB*STA
		Sta ii	URB*STA	URB	URB*STA	DEV*STA + IND*STA	DEV*STA + IND*STA	DEV*STA	URB*STA	URB	URB*STA
		Sta iii	DEV*ind + URB*STA	URB + DEV*ind	DEV*ind + URB*STA	DEV*Ind + DEV*URB*STA	URB + DEV*ind	DEV*ind + DEV*URB*STA	URB*STA	URB	URB*STA
	Ind ii	Sta i	URB*STA	URB	DEV*STA	DEV*STA	DEV*STA + IND*STA	DEV*STA	URB*STA	URB	IND*STA
		Sta ii	DEV*STA	DEV*STA	DEV*STA	DEV*STA + IND*STA	DEV*STA + IND*STA	DEV*STA	URB*STA	URB	URB*STA
		Sta iii	URB*STA	URB	URB*STA	DEV*URB*STA	URB	DEV*URB*STA	URB*STA	URB	URB*STA
	Ind iii	Sta i	DEV*ind + URB*STA	URB + DEV*ind	DEV*ind + URB*STA	DEV*ind + DEV*URB*STA	URB + DEV*ind	DEV*STA	DEV*ind + URB*STA	URB + DEV*ind	DEV*ind + URB*STA
		Sta ii	DEV*ind + DEV*STA	DEV*ind + DEV*STA	DEV*ind + DEV*STA	DEV*STA + IND*STA	DEV*STA + IND*STA	DEV*STA	DEV*ind + URB*STA	URB + DEV*ind	DEV*ind + URB*STA
		Sta iii	DEV*ind + URB*STA	URB + DEV*ind	DEV*ind + URB*STA	DEV*ind + DEV*URB*STA	URB + DEV*ind	DEV*ind + DEV*URB*STA	DEV*ind + URB*STA	URB + DEV*ind	DEV*ind + URB*STA

Note: C indicates the presence of one or more contradictions. Bold marks the principal result.

Changing the Frequency and Consistency Thresholds

After showing how the results were affected by minor changes in calibration anchors, the attention is now drawn to the two remaining types of robustness tests. For two reasons they are treated under the same, relatively short section: The first sensitivity check included a presentation of the baseline analyses and results, and the possibilities of reasonable changes are fewer as regards the frequency and consistency thresholds. As in the previous section, replications using csQCA are introduced before those based on fsQCA.

Crisp-set Analysis

Beginning with the frequency threshold, the low number of cases (18) means that the most plausible choice is to consider all configurations representing empirical cases as relevant, even though some of them just cover one case. Nevertheless, for the sake of the example, an analysis is run where the frequency criterion is raised to two while the breakpoints are similar to the original analysis. This change of the frequency threshold[7] means that result changes to DEV.[8] The fact that this finding is more parsimonious is not surprising as the more demanding criterion means that more logical cases can be used in the reduction and/or fewer configurations (including fewer cases) have to be covered by the formula.

In the case of contradictory configurations, one way to move on with the analysis is to consider the number of cases showing negative and positive outcomes, respectively, and then assign an outcome to the configuration in question according to the outcome value shown by the majority of cases. Consequently, this procedure is not only related to the issue of frequency but also consistency,[9] and generally it is more relevant for studies including a larger number of cases and/or less reliable data. Given the contradiction free truth table in the example, the result is not amendable by changes in the consistency criterion.

Fuzzy-set Analysis

Also as regards the corresponding fuzzy-set analysis, the relatively low number of cases would prescribe a standard frequency threshold of one case. Nonetheless, if it is changed to criterion of at least two cases, the formula changes to URB + DEV*ind[10] that constitutes a slightly more parsimonious superset of the original solution.

Focusing then on the consistency threshold, the first column in Table 5, representing the consistency levels, shows that three gaps provide a plausible basis for differentiating the consistent configurations from the inconsistent ones. The cut-off values are thus changed from 0.80 to 0.90 and 0.70, respectively. The more demanding threshold means that only one configuration is assigned a positive outcome value, the less demanding breakpoint implies that three configurations fulfill the standard. The result linked to the first alternative is URB*STA, namely, identical to one of the two paths of the baseline result. Yet another solution appears from the other, less obvious alternative, namely, DEV*STA, which is similar to the principal solution of the crisp-set analysis.

On the Relative Sensitivity of Crisp-set and Fuzzy-set Analysis

Several studies have already discussed the (dis)advantages of csQCA versus fsQCA (e.g., Berg-Schlosser and Cronqvist 2005; Ragin 2000; Rihoux 2006; Wagemann and Schneider 2010). But they have merely emphasized the appealing simplicity of csQCA vis-à-vis the greater ability of fsQCA to handle differences in degree, whereas the relative robustness of these methods has hardly received any attention. This situation is surprising since such information would provide important information for selecting which one to employ. Discussions of the relative strength of comparative configurational methods are surely warranted, but what—if anything—does the study's analyses tell us about this issue?

The replications clearly underlined that QCA results are indeed sensitive to the different decisions made by researchers. In many cases, minor adjustments in the thresholds as regards calibration of raw data, case frequency, and consistency levels led to considerable changes in the solution formulas. What is more, if one disregards the high number of contradictions emerging from the alterations, csQCA provided more consistent results. Differently put, csQCA tends to be less sensitive than fsQCA, and QCA results tend to lack robustness.

Notice, however, that drawing such conclusions would be both too hasty and too harsh given the evidence at hand. First, one should *not* disregard the fact that the analyses based on csQCA produce many contradictions. Actually, the mix of contradictions and "full" results is a strong indication of high sensitivity. Second, even though the QCA results are somewhat sensitive to the researcher's choices, the different results tend to approximate each other, so they just rarely support fundamentally dissimilar conclusions. Third, four out of five conditions included in the analyses are often used as indicators of the same phenomenon, namely, modernization. Hence, the model is severely harmed by multicollinearity, meaning that one condition is more likely to "substitute" the position of another in the results than would "normally" be the case. Fourth, the focus on parsimonious results, allowing for the inclusion of all counterfactuals no matter their theoretical plausibility, constitutes a very hard test of sensitivity for the QCA methods. In comparison, solutions that are not based on any logical cases, and intermediate solutions, which are only based on theoretically grounded "easy" counterfactuals (cf. Ragin 2008:Ch. 9), would have shown less diversity. Because they would tend to be more complex, fewer opportunities would be left for conditions to be included and excluded in the results. Finally, one should not base the conclusions on only one replication study.

In support of the last point, I have used the same guidelines to replicate two additional studies. One of them examines the structural causes of civil liberty (Skaaning 2006, 2007) and the other concerns the determinants of generous welfare states (Ragin 2000; Schneider and Wagemann 2007; Skaaning and Kvist 2010). Both of these examples employ csQCA and fsQCA, and their causal conditions are more distinct and show weaker mutual correlations. Interestingly, the results based on these replications (not reported) generally lend support to the opposite conclusions: (1) The QCA results are not very sensitive to small adjustments in the various thresholds, and (2) solutions based on csQCA were more sensitive to changes, supporting that fsQCA applications show a higher degree of robustness.

Elaborating on these findings, the main discrepancies between the applications of csQCA and fsQCA tend to rest on the fact that the latter method is based on a higher standard of set-theoretical consistency (cf. Ragin 2008:138-41; Schneider and Wagemann 2007:226-28). Differently put, the application of csQCA often involves a loss of analytically relevant information, which should make us have comparably more confidence in the results achieved through fsQCA.

Sometimes this difference can make fsQCA results appear less robust because they are usually more influenced by discriminate decisions. At least three thresholds (rather than one) have to be determined when calibrating raw date into set-memberships. Also, frequently—although not necessarily so—the use of csQCA follows a deterministic logic, while the use of fsQCA allows for some degree of deviation from the subset principle, thus requiring an additional choice of consistency level. However, the additional replication studies showed that this picture is not cut into stone. More likely, fsQCA has a higher standard for set-theoretical relationships and a higher degree of result robustness.

Conclusion

Needless to say, QCA techniques are designed to be used in an iterative and reflexive way—relying on extensive theoretical reflection and empirical knowledge—and never follow a "push-button" logic (De Meur, Rihoux, and Yamasaki 2009:158; Ragin 2000:Ch. 11). Still, the robustness of QCA results is a key issue, which deserves greater attention than the few sparse and ill-grounded remarks found in previous discussions. Obviously, researchers always have to justify every analytical step with arguments based on theoretical, methodological, and empirical knowledge. But room for discriminatory choice should certainly face critical discussions of potential bias and/or arbitrariness. Consequently, guidelines are needed to judge whether the results achieved by the use of QCA are robust or not. Systematic sensitivity tests—including both the general and the QCA particular procedures—should thus be part of all serious employment of comparative configurational methods. Following this advice will make the employment of these methods more difficult and time-consuming. Nevertheless, it will be worth the effort if it contributes with a notable increase in the degree to which we trust the findings.

Among the different ways to assess the robustness of our results, I have mentioned method triangulation, modifications in case selection, adjustment to the theoretical models, as well as changes in the operationalization of the outcome and conditions. These are all well-known analytical features that have been used more or less systematically in empirical studies using standard statistical tools. However, the logic of QCA urges us to address yet other aspects of the complexity reducing research process in order to support our results. These considerations specifically concern the calibration of raw data into set-memberships, the frequency of cases linked to the configurations, and the acceptable level of configurational consistency. This article has shown how one can pursue the task with regard to these three issues.

Further research is still needed on the strategies used to assess the robustness of QCA results. First and foremost, however, extensive applications of the guidelines presented in this article should give us some basis for discovering general patterns of sensitivity. This study has shown that the findings derived from an employment of QCA are frequently sensitive to small adjustments. On the other hand, in most instances the modifications did not have much—if any—impact. Replications of two additional studies supported this conclusion and indicated that fsQCA is less sensitive than csQCA. The question now is to which extent these findings can be generalized.

Acknowledgments

The author would like to thank the anonymous reviewers, Jim Mahoney, Charles Ragin, Claudius Wagemann, and the participants in a seminar on comparative methodology at the University of Southern Denmark for their critical comments and suggestions.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

Financial Disclosure/Funding

The author(s) received no financial support for the research and/or authorship of this article.

Notes

1. Recently Axel Marx (2006) has used randomly created data-matrices to test another aspect of robustness, namely, whether qualitative comparative analysis (QCA) can make a distinction between real and random data.
2. A borderline case concerning tests for robustness is represented by comparisons of results based on different uses of logical cases: not to make use of any, a theoretically guided inclusion, and the inclusion of all (relevant) logical cases. However, I consider this procedure of utilizing counterfactual evidence as associated more with the issues of parsimony and inference than robustness.
3. Configurations referring to logical cases, that is, combinations of conditions with no empirical instances, are not presented in the truth table. The terms *remainder*, *logical case*, *counterfactual*, and *simplifying assumption* refer to the same phenomenon and these concepts are used interchangeably.
4. As an extra notification, upper-case letters indicate the presence of an outcome or condition and lower-case letters indicate its absence, and concerning the interpretation of Boolean operators, * means logical AND, whereas + stands for logical OR. The countries covered by the respective expressions are listed in parentheses. A comma between the countries signifies that they belong to the same configuration and a plus that they do not.
5. It bears noting that if the focus had been on intermediate solutions rather than parsimonious solutions, QCA would have produced more identical results because the possibilities for reduction would be more restricted.
6. Besides including logical cases and using a frequency threshold of one case.
7. Once more just focusing on positive outcome and using all relevant logical cases.
8. The software points out LIT*STA to be a (logically substitutable) alternative solution.
9. This aspect of crisp-set is closely related to the consistency criterion of fuzzy-setQCA.

10. The software points out URB + LIT*ind*STA to be a (logically substitutable) alternative solution.

References

Berg-Schlosser, Dirk and Jeremy Mitchell, eds. 2000. *Conditions of Democracy in Europe, 1919-1939: Systematic Case Studies*. New York: Macmillan.

Berg-Schlosser, Dirk and Jeremy Mitchell, eds. 2003. *Conditions of Democracy in Europe, 1919-1939: Comparative Analyses*. New York: Macmillan.

Berg-Schlosser, Dirk and Lasse Cronqvist. 2005. "Macro-quantitative vs. Macro-qualitative Methods in the Social Sciences: An Example From Empirical Democratic Theory Employing New Software." *Historical Social Research* 30:154-75.

Caramani, Daniele. 2008. *Introduction to the Comparative Method With Boolean Algebra*. Thousand Oaks, CA: Sage.

De Meur, Gisèle, Benoit Rihoux, and Sakura Yamasaki. 2009. "Addressing the Critiques of QCA." Pp. 147-66 in *Configurational Comparative Method: Qualitative Comparative Analysis (QCA) and Related Techniques*, edited by B. Rihoux and C. Ragin. Thousand Oaks, CA: Sage.

Goldthorpe, John. 1997. "Current Issues in Comparative Macro-sociology: A Debate on Methodological Issues." *Comparative Social Research* 16:1-16.

Lieberman, Evan S. 2005. "Nested Analysis as a Mixed-method Strategy for Comparative Research." *American Political Science Review* 99:435-52.

Lieberson, Stanley. 2004. "Comments on the Use and Utility of QCA." *Qualitative Methods* 2:13-4.

Lipset, Seymor M. 1959. "Some Social Requisites of Democracy: Economic Development and Political Legitimacy." *American Political Science Review* 53:69-105.

Marx, Axel. 2006. "Towards More Robust Model Specification in QCA: Results From a Methodological Experiment." COMPASS Working Paper 2006-43.

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

Ragin, Charles. 2000. *Fuzzy-set Social Science*. Chicago: University of Chicago Press.

Ragin, Charles. 2008. *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Princeton, NJ: Princeton University Press.

Ragin, Charles. 2009. "Qualitative Comparative Analysis Using Fuzzy Sets (fsQCA)." Pp. 87-122 in *Configurational Comparative Methods*, edited by B. Rihoux and C. Ragin. Thousand Oaks, CA: Sage.

Ragin, Charles and Benoit Rihoux. 2004. "Qualitative Comparative Analysis (QCA): State of the Art and Prospects." *Qualitative Methods* 2:3-13.

Rihoux, Benoit. 2006. "Qualitative Comparative Analysis (QCA) and Related Systematic Comparative Methods: Recent Advances and Remaining Challenges for Social Science Research." *International Sociology* 21:679-706.

Rihoux, Benoit. 2009. "Crisp-set Qualitative Comparative Analysis (csQCA)." Pp. 33-68 in *Configurational Comparative Method: Qualitative Comparative Analysis (QCA) and Related Techniques*, edited by B. Rihoux and C. Ragin. Thousand Oaks, CA: Sage.

Rihoux, Benoit and Charles Ragin, eds. 2009. *Configurational Comparative Method: Qualitative Comparative Analysis (QCA) and Related Techniques*. Thousand Oaks, CA: Sage.

Rihoux, Benoit, Charles Ragin, Sakura Yamasaki, and Damien Bol. 2009. "Conclusions—The Way(s) Ahead." Pp. 167-78 in *Configurational Comparative Method: Qualitative Comparative Analysis (QCA) and Related Techniques*, edited by B. Rihoux and C. Ragin. Thousand Oaks, CA: Sage.

Rohlfing, Ingo. 2008. "What You See and What You Get." *Comparative Political Studies* 41:1492-514.

Schneider, Carsten and Claudius Wagemann. 2007. *Qualitative Comparative Analysis (QCA) und Fuzzy Sets: Ein Lehrbuch für Anwender und jene, die es werden wollen*. Opladen: Verlag Barbara Budrich.

Schneider, Carsten and Claudius Wagemann. 2010. "Standards of Good Practice in Qualitative Comparative Analysis (QCA) and Fuzzy-sets." *Comparative Sociology* 9:397-418.

Seawright, Jason. 2005. "Qualitative Comparative Analysis vis-à-vis Regression." *Studies in Comparative International Development* 40:2-26.

Skaaning, Svend-Erik. 2006. *Democracy Besides Elections: An Inquiry Into the (Dis) Respect for Civil Liberty in Latin American and Post-communist Countries After the Third Wave*. Aarhus, Denmark: Aarhus University.

Skaaning, Svend-Erik. 2007. "Explaining Post-communist Respect for Civil Liberty: A Multi-methods Test." *Journal of Business Research* 60:493-500.

Skaaning, Svend-Erik and Jon Kvist. 2010. "Konfigurationelle komparative metoder"(Configurational Comparative Methods). Pp. 238-63 in *Metoder i Statskundskab*, edited by L. B. Andersen, K.r M. Hansen, and R. Klemmesen. Copenhagen: Hans Reitzels Forlag.

Wagemann, Claudius and Carsten Schneider. 2010. "Qualitative Comparative Analysis (QCA) and Fuzzy-sets: Agenda for a Research Approach and Data Analysis Technique." *Comparative Sociology* 9:376-96.

Bio

Svend-Erik Skaaning is an associate professor at Aarhus University, Denmark. Besides comparative methodology, his research interests include the definition, measurement, and explanation of democracy and human rights. He has recently published on these issues in *Government and Opposition*, *International Political Science Review*, and *Political Research Quarterly*. Currently he is working on a project on the requisites of democracy.