Optimizing Consistency and Coverage in Configurational Causal Modeling

Michael Baumgartner* and Mathias Ambühl†

Abstract

Consistency and coverage are two core parameters of model fit used by configurational comparative methods (CCMs) of causal inference. Among causal models that perform equally well in other respects (e.g. robustness or compliance with background theories), those with higher consistency and coverage are typically considered preferable. Finding the optimally obtainable consistency and coverage scores for data δ, so far, is a matter of repeatedly applying CCMs to δ while varying threshold settings. This paper introduces a procedure called *ConCovOpt* that calculates, prior to actual CCM analyses, the consistency and coverage scores that can optimally be obtained by models inferred from δ. Moreover, we show how models reaching optimal scores can be methodically built in case of crisp-set and multi-value data. ConCovOpt is a tool, not for blindly maximizing model fit, but for exploring the space of viable models at optimal fit scores—which, as we demonstrate by various data examples, may have substantive modeling implications.

1 Introduction

Over the past three decades, different variants of *configurational comparative methods* (CCMs) have gradually been added to the toolkit for causal data analysis in many disciplines—ranging from social and political science to business administration and management, environmental and evaluation science, on to public health and psychology. CCMs differ from other techniques as regression analytical methods (RAMs) (e.g. Gelman and Hill 2007) or Bayes-nets methods (BNMs) (e.g. Spirtes et al. 2000) in a number of respects. Most importantly, while RAMs and BNMs study statistical and probabilistic properties of causal

*University of Bergen, Norway, michael.baumgartner@uib.no

†ConsultAG, Switzerland, mathias.ambuehl@consultag.ch

structures as characterized by probabilistic theories of causation (e.g. Suppes 1970), CCMs scrutinize Boolean properties as described by regularity theories of causation (e.g. Mackie 1974). Furthermore, unlike RAMs and BNMs, CCMs do not search for causal dependencies among *variables* but among concrete *values of variables*, and they do not quantify net effects and effect sizes but place a Boolean ordering on sets of causes by grouping their elements conjunctively, disjunctively, and sequentially (for more on these differences see Thiem et al. 2016).

The most well-known CCM is *Qualitative Comparative Analysis* (QCA) (Ragin 1987; Cronqvist and Berg-Schlosser 2009; Ragin 2008; Thiem 2014). *Coincidence Analysis* (CNA) is a more recent addition to the family of CCMs (Baumgartner 2009; Baumgartner and Ambühl 2018). There are various differences between QCA and CNA—in the underlying methodological principles, in the implemented algorithms, or in the search target—, but also important commonalities. Both methods process configurational data featuring crisp-set, fuzzy-set, or multi-value variables (Thiem 2014), which are called *factors* in CCM jargon. They both exploit relations of sufficiency and necessity for causal inference and output models accounting for the values taken by endogenous factors in terms of redundancy-free Boolean functions of exogenous factor values. And they share two of their core parameters of model fit, which will constitute the topic of this paper: *consistency* and *coverage* (Ragin 2006). Informally, consistency reflects the degree to which the behavior of an outcome obeys a corresponding sufficiency or necessity relationship or a whole model, whereas coverage reflects the degree to which a sufficiency or necessity relationship or a whole model accounts for the behavior of the corresponding outcome. What counts as acceptable scores on these fit parameters is defined in threshold settings determined by the analyst prior to the application of QCA or CNA, based on method-specific conventions.

Among causal models that perform equally well with respect to other fit criteria, for example, robustness and compliance with case knowledge or background theories, the ones with higher consistency and coverage—that is, with a higher *product* of consistency and coverage—are considered preferable in both QCA and CNA. This raises the question of how to systematically find the models with optimal products of consistency and coverage. Currently, neither the procedural protocols of QCA nor of CNA have answers to that question on offer. Rather, optimizing consistency and coverage scores is a matter of repeatedly applying QCA and CNA to the data while varying relevant thresholds and comparing the fit scores of resulting models. Such a trial-and-error approach is neither guaranteed to recover all consistency and coverage optima, of which there may be many, nor is it efficient, as it may require a multitude of data re-analyses. Variations in the thresholds may—depending on the data—induce substantive changes in the composition of the issued models as well as in their consistency and coverage scores. These changes may not be proportional to the threshold variations. That is, higher thresholds are not guaranteed to produce models with higher (i.e. optimal) products of consistency and coverage. In consequence, the range of acceptable threshold settings may have to be searched in its entirety in fine-grained steps.

This paper will show that it is possible to exhaustively identify optimal consistency and coverage scores of CCM models inferrable from data δ independently of actually applying CCMs to δ. We introduce an explicit procedure, called *ConCovOpt*, that calculates all consistency and coverage optima for δ prior to CCM analyses. ConCovOpt will be complemented by a second procedure, called *DNFbuild*, that purposefully builds models reaching optimal scores for crisp-set and multi-value data. For these data types, models are hence guaranteed to exist at the consistency and coverage optima. ConCovOpt can also be applied to fuzzy-set data, in which case optimas amount to upper bounds that cannot possibly be outperformed by actual models, but there is no guarantee that models *de facto* exist at those bounds. The upper bounds, thus, constrain the interval of threshold settings within which optimal actual models must be searched.

ConCovOpt is a tool, not for blindly maximizing model fit, but for systematically exploring the space of viable models at optimal consistency and coverage scores. Sometimes, models with optimal scores will turn out to be the best models overall, while sometimes optimizing consistency and coverage is only possible at the price of overfitting or of compromising on robustness or compliance with background theories. Choosing the best model(s) among all viable models, which may be numerous (Baumgartner and Thiem 2017), is a delicate task that requires balancing various criteria. Consistency and coverage are only two of those criteria. But making an informed choice presupposes that the whole model space is brought to the analyst's attention. ConCovOpt contributes to that objective.

The article is organized as follows. Section 2 reviews some conceptual preliminaries. In section 3, the procedure is presented using a simple crisp-set data example. Section 4 applies it to large-*n* data. DNFbuild is introduced in section 5 on the basis of multi-value data; and a fuzzy-set application is discussed in section 6. Finally, section 7 puts ConCovOpt into proper methodological perspective. We implemented ConCovOpt and DNFbuild as R functions, which are extensively used in the R script that is available in the supplementary material and that allows for replicating all calculations of this paper.

2 Conceptual preliminaries

We begin by introducing the conceptual (and notational) preliminaries of our ensuing discussion. As indicated above, CCMs study Boolean dependence relations between factors taking on specific values. Factors represent categorical properties that partition sets of units of observation (cases) either into two sets, in case of binary properties, or into more than two (but finitely many) sets, in case of multi-value properties. Factors representing binary properties can be *crisp-set* (cs) or *fuzzy-set* (fs); the former (typically) take on 0 and 1 as possible values, whereas the latter can take on any (continuous) values from the unit interval $[0, 1]$. Factors representing multi-value properties are called *multi-value (mv)* factors; they can take on any of an open (but finite) number of non-negative integers as possible values. Values of a cs or fs factor X are often interpreted as membership scores in the set of cases exhibiting the property represented by X, while the values of an mv factor X designate the particular way in which the property represented by X is exemplified.

As the explicit 'Factor=value' notation yields convoluted syntactic expressions with increasing model complexity, we subsequently use—whenever possible—a shorthand notation that is conventional in Boolean algebra (and CCM modeling; but see Thiem 2014 for an alternative notation): membership in a set is expressed by italicized upper case and non-membership by lower case Roman letters. Hence, in case of cs and fs factors, we normally write 'X' for X=1 and 'x' for X=0. It must be emphasized that, while this notation significantly simplifies the syntax of CCM models, it introduces a risk of misinterpretation, for it yields that the factor X and its taking on the value 1 are both expressed by 'X'. Disambiguation must hence be facilitated by the concrete context in which 'X' appears. Therefore, whenever we do not explicitly characterize italicized Roman letters as 'factors', we use them in terms of the shorthand notation. In case of mv factors and in case of explicit definitions, value assignments are always written out, using the 'Factor=value' notation; that is, we write 'Y=v' for factor Y taking the value v.

CCM models may feature all the standard Boolean operations: negation $\neg X$ ('not X'), conjunction $X{*}Y$ ('X and Y'), disjunction $X + Y$ ('X or Y'), implication $X \rightarrow Y$ ('If X, then Y), and equivalence $X \leftrightarrow Y$ ('X if, and only if, Y'). In case of cs and mv factors, Boolean operations are given a rendering in classical logic (which we do not reiterate here; see e.g. Lemmon 1965, ch. 1, for a canonical introduction). In case of fs factors, these operations are rendered in fuzzy logic: negation $\neg X$ amounts to $1 - X$, conjunction $X{*}Y$ to $\min(X, Y)$, disjunction $X+Y$ to $\max(X, Y)$, an implication $X \rightarrow Y$ is taken to express that the membership score in X is smaller or equal to Y ($X \leq Y$), and an equivalence $X \leftrightarrow Y$ that the membership scores in X and Y are equal ($X = Y$).

Based on the implication operator, the notions of *sufficiency* and *necessity* are defined, which are the two dependence relations exploited by CCMs: X is sufficient for Y if, and only if (iff), $X \rightarrow Y$ ('if X is given, then Y is given'), and X is necessary for Y iff $Y \rightarrow X$ ('if Y is given, then X is given'). CCM models have the form $\Phi \leftrightarrow Y$, where Y is an endogenous factor value and Φ stands for an expression $X_1 * \ldots * X_i + \ldots + X_m * \ldots * X_n$ in *disjunctive normal form (DNF)* such that all factors in that DNF are different (and logically, conceptually, and metaphysically independent) from Y. An expression is in DNF iff it is a disjunction of one or more conjunctions of one or more literals (i.e. factors or their negations; see Lemmon 1965, 190). All in all, thus, CCM models explain Y in terms of a necessary disjunction of sufficient conditions of Y.

Sufficiency and necessity relations amount to mere patterns of co-occurrence. As such, they carry no causal connotations whatsoever, and, hence, most of these relations do not reflect causation. Still, some of them do. Regularity theories of causation (Mackie 1974; Graßhoff and May 2001; Baumgartner 2008) are designed to filter out those sufficiency and necessity relations that do track causation. According to regularity theories, an expression of the form $\Phi \leftrightarrow Y$ tracks causation only if Φ is *redundancy-free*, meaning that no conjuncts or disjuncts can be removed from Φ without violating the truth of $\Phi \leftrightarrow Y$. There are important differences between QCA and CNA in regard to how rigorously Φ needs to be freed of redundancies before it is amenable to a causal interpretation. In QCA (as standardly practiced), complete redundancy elimination as implemented in so-called *parsimonious* models is not always required for causal interpretability—partial redundancy elimination as in *intermediate* or *conservative* models may suffice as well. In CNA, by contrast, complete redundancy elimination is crucial for causal interpretability. These differences shall be bracketed in the following. However, to facilitate comparability between models output by QCA and CNA, we will (mainly) focus on parsimonious QCA models.

Since CCM-processed data δ tend to feature various deficiencies (e.g. fragmentation, noise, etc.), expressions of type $\Phi \leftrightarrow Y$ that adhere to the strict standards of the equivalence operation ('$\leftrightarrow$') often cannot be inferred from δ. To relax these standards, that is, to approximate the ideal (deterministic) sufficiency and necessity relations expressed in $\Phi \leftrightarrow Y$, Ragin (2006) introduced the consistency and coverage measures into the QCA protocol, which have subsequently also been imported into CNA (Baumgartner and Ambühl 2018). As the implication operator underlying the notions of sufficiency and necessity is defined differently in classical and in fuzzy logic, the two measures are defined differently for crisp-set and multi-value data, which both have a classical footing, and for fuzzy-set data. *Cs-consistency*

(con^{cs}) and *cs-coverage* (cov^{cs}) of $X \rightarrow Y$ are defined as follows, where '$|\Gamma|$' represents the cardinality of the set of cases instantiating Γ:

$$con^{cs}(X \rightarrow Y) = \frac{|X{*}Y|}{|X|} \qquad cov^{cs}(X \rightarrow Y) = \frac{|X{*}Y|}{|Y|} \tag{1}$$

Fs-consistency (con^{fs}) and *fs-coverage* (cov^{fs}) of $X \rightarrow Y$ are defined as follows, where n is the number of cases in the data:

$$con^{fs}(X \rightarrow Y) = \frac{\sum_{i=1}^{n} \min(X_i, Y_i)}{\sum_{i=1}^{n} X_i} \qquad cov^{fs}(X \rightarrow Y) = \frac{\sum_{i=1}^{n} \min(X_i, Y_i)}{\sum_{i=1}^{n} Y_i} \tag{2}$$

Whenever the values of X and Y are restricted to 1 and 0 in the crisp-set measures, con^{cs} and cov^{cs} are equivalent to con^{fs} and cov^{fs}, but for binary factors with values other than 1 and 0 and for multi-value factors the cs and fs variants of consistency and coverage diverge. Nonetheless, we will in the following not explicitly distinguish between the cs and fs measures because the particular contexts of our discussion will make it sufficiently clear which of them are at issue.

Consistency and coverage take values from the unit interval, with 1 representing perfect consistency and coverage. What counts as acceptable scores on these measures is defined in threshold settings determined by the analyst prior to the application of QCA or CNA. The implementation of these thresholds differs in important ways in the two methods. In QCA, a consistency threshold is imposed only on conjunctions of all exogenous factors (so-called *minterms*) in the course of the generation of *truth tables*, which are intermediate calculative devices for QCA. The final models issued may or may not meet the chosen threshold. It is recommended, by convention, not to set the consistency threshold below 0.75 or 0.8 (Ragin 2008, 46; Schneider and Wagemann 2012, 129). Moreover, it is common QCA practice not to impose coverage thresholds at all. In CNA, by contrast, thresholds for both consistency and coverage are used as authoritative model building constraints. The thresholds define what counts as sufficient and necessary conditions, to the effect that models not meeting the thresholds cannot be built. By convention, consistency and coverage thresholds should not be set below 0.75 (Baumgartner and Ambühl 2018).

Despite these differences, in both QCA and CNA models with higher consistency and coverage are preferred over models with lower scores on these measures, provided they fare equally well in other respects (e.g. robustness). The following section introduces our procedure, ConCovOpt, calculating consistency and coverage optima.

3 The optimization procedure

The goal of ConCovOpt is to identify both optimal and maximal consistency and coverage scores—*con-cov optima* and *con-cov maxima*, for short—, the distinction being that an optimum optimizes at least one of consistency and coverage, whereas a maximum optimizes their product. The procedure's input is a set of configurational data δ and a set of outcomes $\mathbf{O}$ in δ. By suitably aggregating δ for the modeling of every $Y \in \mathbf{O}$, ConCovOpt first identifies output values for Boolean functions, so-called *rep-assignments*, which reproduce the behavior of Y as closely as possible, and, by calculating consistency and coverage scores for these rep-assignments, it then infers all con-cov optima and maxima that CCM models of Y can possibly reach.

We introduce the procedure using the very simple crisp-set data example in Table 1a drawn from Giugni and Yamasaki (2009, 476), who investigate the policy impact of different social movements between 1975 and 1995. The exogenous factors are high protest activity (P), public opinion favorable to the movement (O), and powerful institutional allies (A), with values 0 and 1 representing 'no' and 'yes' for all factors. The endogenous factor C takes the value 1 whenever a movement manages to significantly change the countries policy, and 0 otherwise. The authors analyze the data for various western countries separately; Table 1a features the data for the United States.

We begin by searching for con-cov optima for outcome C (i.e. C=1) in Table 1a. The notion of a con-cov optimum shall be defined as follows:

Con-cov optimum. An ordered pair $\langle con, cov \rangle$ of consistency and coverage scores is a con-cov optimum for outcome Y=v in data δ iff, prior to applying a CCM, it can be excluded that a CCM model of Y=v inferred from δ scores better on one element of the pair and at least as well on the other, whereas it cannot be excluded that such a model reaches $\langle con, cov \rangle$.

If models of Y (i.e. Y=1) inferred from data δ can be modeled with perfect consistency and coverage, $\langle 1, 1 \rangle$ is the only con-cov-optimum for Y in δ. Outcome C in Table 1a, however, does not have a con-cov optimum of $\langle 1, 1 \rangle$. The reason is that the cases P87, P92, N80 feature the same configuration of the exogenous factors—*viz.* the configuration $p{*}o{*}a$—while C is given in P87 and P92 and c is given in N80. The configurations $p{*}o{*}a{*}C$ and $p{*}o{*}a{*}c$ constitute what we will call an *imperfect pair* for C in Table 1a.[1]

[1] In the QCA literature, such a pair is often labelled *contradictory* (e.g. Rihoux and De Meur 2009, 46-49). We find this terminology misleading because a contradiction cannot possibly be realized. But, as Table 1a vividly demonstrates, it is very well possible for $p{*}o{*}a$ to be realized in combination with both C and c. It is far from contradictory for a configuration to be combined with both the presence and the absence of an outcome.

	P	O	A	C
E75	1	0	1	0
E87	1	1	0	1
P81	1	0	0	0
P90	1	0	0	0
N77	1	1	1	0
E80	0	1	0	0
E92	0	1	0	0
P75	0	0	1	1
P87	0	0	0	1
P92	0	0	0	1
N80	0	0	0	0

(a)

conf.	P	O	A	C	n	exo-groups	$\varphi(C{=}1)$	φ_1	φ_2
σ_1	1	0	1	0	1	$\{\sigma_1\}$	0	0	0
σ_2	1	1	0	1	1	$\{\sigma_2\}$	1	1	1
σ_3	1	0	0	0	2	$\{\sigma_3\}$	0	0	0
σ_4	1	1	1	0	1	$\{\sigma_4\}$	0	0	0
σ_5	0	1	0	0	2	$\{\sigma_5\}$	0	0	0
σ_6	0	0	1	1	1	$\{\sigma_6\}$	1	1	1
σ_7	0	0	0	1	2	$\{\sigma_7, \sigma_8\}$	0, 1	0	1
σ_8	0	0	0	0	1				

(b)

Table 1: Subtable (a) is the data matrix for the United States from Giugni and Yamasaki (2009, 476) with C being the outcome. Subtable (b) is a configuration table with configuration labels in column 'conf.', case frequencies in 'n', imperfect pairs with gray shading, as well as added exo-groups, rep-list $\varphi(C{=}1)$, and rep-assignments φ_1 and φ_2.

Imperfect pair. An imperfect pair for $Y{=}v$ in data δ is an (unordered) pair of configurations $\{\sigma_i, \sigma_j\}$ in δ such that $Y{=}v$ is instantiated in one element of the pair and not instantiated in the other element, while all other factors in δ take constant values in both σ_i and σ_j.

In crisp-set and multi-value data δ, there is a tight logical connection between imperfect pairs for an outcome Y and $\langle 1, 1\rangle$ being the con-cov optimum for Y in δ: Y has a con-cov optimum of $\langle 1, 1\rangle$ in δ iff there does not exist an imperfect pair for Y in δ.[2] The reason is that a CCM model with perfect consistency and coverage expresses Y as a strict Boolean function of the other factors in δ; and such a function (as any other function) issues exactly one output for every input. If there does not exist an imperfect pair for Y, this principle is straightforwardly satisfied, as every input (i.e. every configuration of factors other than Y) can be mapped either on Y or on y. But if there exists an imperfect pair, there exists an input to which no determinate output can be assigned, meaning Y cannot be expressed as a strict Boolean function of the other factors in δ. On average, the more imperfect pairs Y has in δ the lower the con-cov optima for Y in δ.[3]

The existence of imperfect pairs indicates that there are varying causes of Y in the uncontrolled causal background. The variation of Y in the imperfect pair must have some cause

[2] This principle does not hold for fuzzy-set data (see section 6).

[3] Not only the amount of imperfect pairs affects the con-cov optima but also the number of cases instantiating the different elements of imperfect pairs.

or other; but that cause cannot be among the other factors in δ because they are constant in the imperfect pair. Since varying latent causes are a source of confounding, it is standardly recommended to try to resolve imperfect pairs prior to a CCM analysis; and there are various approaches on offer for how to do this (e.g. Rihoux and De Meur 2009). Of course, as suppressing the variation of latent causes, especially in observational studies, is very difficult (in particular, when these latent causes are unknown), these approaches may be incapable of improving the data quality. For the purposes of this paper, we will hence assume that the quality of all our example data sets has been improved as far as possible, meaning that the remaining imperfect pairs cannot be resolved.

A first step towards determining con-cov optima for an outcome Y is to identify imperfect pairs for Y in δ. To do this in a methodical manner, we re-organize the data such that the instantiated configurations are rendered more transparent by aggregating all cases in δ instantiating the same configuration in a single row of what we will call a *configuration table*. A configuration table CT of data δ merges multiple rows of δ in which all factors have identical values into one row, such that each row of CT corresponds to one determinate configuration of the factors in δ.[4] The configurations in CT are labeled and the number of cases instantiating each configuration is stored in an additional frequency column. The first three (line-separated) columns of Table 1b amount to a configuration table of our example data in Table 1a.

A configuration table then allows for splitting the configurations into groups in which all factors other than a scrutinized outcome Y, that is, all factors that are exogenous with respect to Y, take constant values. We shall speak of *exo-groups*, for short.

Exo-group. An exo-group of an outcome $Y{=}v$ in a configuration table CT is a group of configurations in CT with constant values in all factors in CT other than Y.

The imperfect pairs for Y (i.e. $Y{=}1$) in data δ can then be directly read off the list of Y's exo-groups: exo-groups with more than one element such that Y is instantiated in one element and not instantiated in another element correspond to imperfect pairs. To illustrate with our example, the exo-groups of C are listed in the fourth (line-separated) column of Table 1b. For instance, there is only one configuration in Table 1b featuring $P{*}o{*}A$, *viz.* σ_1, meaning that $\{\sigma_1\}$ is a singleton exo-group of C. By contrast, there are two configurations featuring $p{*}o{*}a$, *viz.* σ_7 and σ_8, which accordingly constitute an exo-group of C with two elements

[4]Note that a configuration table is *not* a QCA truth table. The former type of table merely aggregates cases to configurations, whereas the latter type, additionally, determines for each configuration whether it is sufficient for an outcome (according to a chosen consistency threshold).

$\{\sigma_7, \sigma_8\}$. As C is instantiated in one element of that group and not instantiated in the other, $\{\sigma_7, \sigma_8\}$ amounts to an imperfect pair for C; and as C has no other exo-groups with more than one element, it is C's only imperfect pair.

In order for a CCM model, which, to recall, has the form $\Phi \leftrightarrow Y$, to have highest possible consistency and coverage, its redundancy-free DNF Φ must reproduce the instantiation behavior of the outcome Y as closely as possible. The notion of *reproducing the behavior* of an outcome as closely as possible will be of crucial importance for ConCovOpt. It must be understood somewhat differently for crisp-set and multi-value data, on the one hand, and fuzzy-set data, on the other. In case of *cs* and *mv* data, we say that Φ reproduces the behavior of an outcome as closely as possible iff Φ returns the value 1 for every exo-group in which the outcome is constantly instantiated, 0 for every exo-group in which it is constantly non-instantiated, and either 0 or 1 for every exo-group with varying instantiation behavior on the outcome. Applied to our example, this means that a Φ—whichever concrete DNF this may be—reproduces the behavior of C as closely as possible iff Φ returns 1 for exo-groups $\{\sigma_2\}$, $\{\sigma_6\}$; 0 for $\{\sigma_1\}$, $\{\sigma_3\}$, $\{\sigma_4\}$, $\{\sigma_5\}$; and either 0 or 1 for $\{\sigma_7, \sigma_8\}$. Taken together, these value assignments yield what we will call the *rep-list* (reproduction list) $\boldsymbol{\varphi(C{=}1)}$ for outcome C in Table 1b.

Rep-list. A rep-list $\boldsymbol{\varphi(Y{=}\nu)}$ for an outcome $Y{=}\nu$ assigns all the values reproducing the behavior of $Y{=}\nu$ as closely as possible to every exo-group of $Y{=}\nu$.

Moreover, to an assignment that returns a value from a rep-list for every exo-group we will refer as a *rep-assignment* (reproduction assignment).

Rep-assignment. A rep-assignment φ for an outcome $Y{=}\nu$ assigns exactly one value from a rep-list $\boldsymbol{\varphi(Y{=}\nu)}$ to every exo-group of $Y{=}\nu$.

Whatever concrete factor values CCMs may ultimately incorporate in DNFs accounting for outcome Y, it is clear—prior to applications of CCMs—that a DNF not returning a rep-assignment does not reach a con-cov optimum for Y. At the same time, as will be shown in section 4, some rep-assignments yield non-optimal consistency or coverage scores. That is, returning a rep-assignment for Y is necessary but not sufficient for a DNF to reach a con-cov optimum for Y. In order to identify those rep-assignments that actually yield con-cov optima, all possible rep-assignments must be built from $\boldsymbol{\varphi(Y{=}\nu)}$ and the consistency and coverage scores they induce must be individually tested for optimality.

The number of rep-assignments that can be built from a rep-list $\boldsymbol{\varphi(Y{=}\nu)}$ is equal to the number of combinatorially possible value distributions drawn from $\boldsymbol{\varphi(Y{=}\nu)}$, that is, to $\prod_{i=1}^{n} |\boldsymbol{\varphi(Y{=}\nu)}_i|$, where n is the number of exo-groups and $|\boldsymbol{\varphi(Y{=}\nu)}_i|$ the cardinality of the set of possible values assigned to exo-group i. In our example, the complete set of rep-assignments is easily built, as there is only one exo-group with more than one value in the rep-list $\boldsymbol{\varphi(C{=}1)}$. Hence, outcome C has a total of two rep-assignments, φ_1 and φ_2, which are featured in the last two columns of Table 1b. φ_1 and φ_2 coincide except for the fact that they contain the values 0 and 1, respectively, for exo-group $\{\sigma_7, \sigma_8\}$. φ_1 induces perfect consistency but does not cover the instance of C in σ_7, whereas φ_2 covers the instance of C in σ_7 but violates perfect consistency in σ_8.

It only remains to be determined which of all rep-assignments actually reach con-cov optima. To this end, consistency and coverage scores are calculated for all rep-assignments. In case of crisp-set and multi-value data, this can be done by plugging the values of a rep-assignment φ_i and the corresponding instantiation behavior of outcome $Y{=}v$ into the definitions con^{cs} and cov^{cs} in (1). In our example, this means that columns 'φ_1' and 'φ_2' of Table 1b yield the X-values of con^{cs} and cov^{cs}, column 'C' the Y-values, and column 'n' the case frequencies. We get the following consistency and coverage scores:

$$\varphi_1 : \quad con = 1 \;; \quad cov = 0.5 \tag{3}$$

$$\varphi_2 : \quad con = 0.8 \;; \quad cov = 1 \tag{4}$$

Due to the imperfect pair in exo-group $\{\sigma_7, \sigma_8\}$, it is impossible, in principle, for a CCM model of C inferred from the data in Table 1a to score better on consistency and coverage. φ_1 outperforms φ_2 in consistency and φ_2 outperforms φ_1 in coverage. As neither of the two scores better than the other on one measure and at least as well on the other, it turns out that they are both con-cov optima. φ_1 optimizes consistency, φ_2 optimizes coverage. At the same time, φ_2 clearly outperforms φ_1 in the product of consistency and coverage. That is, φ_2 has the better overall model fit; it reaches a *con-cov maximum*:

Con-cov maximum: An ordered pair $\langle con, cov \rangle$ of consistency and coverage scores is a con-cov maximum for outcome $Y{=}v$ in data δ iff $\langle con, cov \rangle$ is a con-cov optimum for $Y{=}v$ in δ with highest product of consistency and coverage (con-cov product).[5]

In sum, without having applied CCMs to Guigni and Yamasaki's (2009) data, we have been able to identify optimal and maximal consistency and coverage scores for it. Before we

[5]Maximizing $con \times cov$, of course, is only one option among many. Maximizing $min(con, cov)$ or $max(con, cov)$ would yield different con-cov maxima that might be of interest in certain modeling contexts.

search for actual CCM models for our example, let us assemble the different steps calculating con-cov optima and maxima in explicit procedural form. To this end, one generalization is still needed. For simplicity, all data analyzed in this paper comprise a single outcome only, but, of course, configurational data may feature *multiple outcomes*. If that is the case, exo-groups, rep-lists, and rep-assignments must be formed and consistency and coverage scores calculated for each outcome separately. For generality, we thus let the input of ConCovOpt be data δ along with a set of outcomes $\mathbf{O}$ in δ. If no prior knowledge is available as to which values of which factors in δ are possible outcomes, ConCovOpt can simply be run by setting $\mathbf{O}$ equal to all values of all factors in δ.

Procedure 1 (**ConCovOpt**) consistency and coverage optimization

Input: configurational data δ with a set of outcomes $\mathbf{O}$ in δ.
Output: all con-cov optima and maxima for all outcomes in $\mathbf{O}$.

(1) Aggregate δ in a configuration table CT.

(2) For every outcome $Y{=}v$ in $\mathbf{O}$, split the configurations in CT into exo-groups of $Y{=}v$.

  (3) Build the rep-list $\varphi(Y{=}\nu)$.
  (4) Build all rep-assignments φ_1 to φ_m from $\varphi(Y{=}\nu)$.
  (5) Calculate the consistency and coverage scores of φ_1 to φ_m.
  (6) Eliminate all rep-assignments with scores that do not reach a con-cov optimum.
  ⇒ The scores of the remaining rep-assignments correspond to all *con-cov optima* for $Y{=}v$ in δ, the optima with highest con-cov product are the *con-cov maxima*.

Let us now apply CCMs to our example data in order to find actual models returning φ_1 and φ_2. At a consistency threshold anywhere between 1 and 0.75, QCA produces the following parsimonious (QCA-PS) and conservative models (QCA-CS),[6] the latter of which is also the model published by Giugni and Yamasaki (2009, 479):

$$\text{QCA-PS:} \quad p{*}A \;+\; P{*}O{*}a \;\leftrightarrow\; C \qquad con = 1\;;\;\; cov = 0.5 \tag{5}$$

$$\text{QCA-CS:} \quad p{*}o{*}A \;+\; P{*}O{*}a \;\leftrightarrow\; C \qquad con = 1\;;\;\; cov = 0.5 \tag{6}$$

[6]The relevant replication code is available as an R script in the supplementary material. To build QCA models, we use the **QCA** R package (Dusa 2019). It should be emphasized, though, that the **QCA** package, in our view, has a number of shortcomings, especially when it comes to finding the complete space of viable models. For the latter purpose, the **QCApro** R package (Thiem 2018) is preferable. We nonetheless use the **QCA** package here because, while its shortcomings do not distort our example analyses, it is the only QCA software providing a new algorithm called *CCubes* that allows for setting coverage thresholds, which, as we shall see promptly, is a crucial asset when it comes to finding optimal models.

In both (5) and (6), not only the solution consistency but also the consistency of all sufficient conditions (i.e. disjuncts) is above the minimum of 0.75. Our previous calculations now attest that the QCA models reach a con-cov optimum, as they return rep-assignment φ_1. At the same time, we can now also say that (5) and (6) do not reach a con-cov maximum. Rep-assignment φ_2 shows that it is possible to significantly improve on the overall model fit. However, it turns out that, in the $[1, 0.75]$ interval of acceptable consistency thresholds, standard QCA does not find a better scoring model—for two main reasons. On the one hand, QCA builds models *from the top down*, that is, by first searching for complete minterms satisfying a chosen consistency threshold and then eliminating redundancies. The minterm $p{*}o{*}a$ of the exo-group $\{\sigma_7, \sigma_8\}$, however, does not meet the minimum consistency threshold of 0.75 and is therefore not further considered by QCA (in QCA jargon: it is coded '0')—although, as we shall see below, a proper part of that minterm in fact meets the 0.75 threshold. On the other hand, standard QCA does not allow for setting a coverage threshold and, hence, cannot be 'asked' to build models with specific target scores on coverage.

CNA, by contrast, accepts separate thresholds for consistency of sufficient conditions and of whole models as well as for coverage of whole models. It can thus be 'asked' to build models at any target scores. Moreover, it builds models *from the bottom up* by first testing single factor values for compliance with chosen thresholds and by then gradually adding further factor values until threshold compliance is established.[7] Dusa (2018) has recently presented a promising new algorithm for QCA called *CCubes* that also builds models from the bottom up and accepts the same types of thresholds as CNA.[8] At a threshold setting of $\langle 1, 0.5 \rangle$, CNA and CCubes return the same model as QCA-PS, *viz.* (5), but at $\langle 0.75, 1 \rangle$ they find the following model realizing φ_2:

$$\text{CNA / CCubes:} \quad p{*}o \,+\, P{*}O{*}a \;\leftrightarrow\; C \qquad con = 0.8 \;;\;\; cov = 1 \tag{7}$$

Like the QCA models, (7) and all of its disjuncts satisfy the consistency minimum of 0.75, but unlike the QCA models, the coverage of (7) also meets a minimum of 0.75. Subject to QCA conventions, both (5) and (7) are viable model candidates, whereas according to CNA conventions, only (7) has acceptable model fit. This is not the place to resolve

[7]For more on the difference between a top-down and bottom-up search see Baumgartner and Ambühl (2018).

[8]Dusa (2018) describes CCubes as a minimization algorithm that is optimized for speed but produces the exact same models as traditional QCA algorithms (e.g. Quine-McCluskey optimization). In fact, however, as our example shows, CCubes accepts much more fine-grained fit parameters and its bottom-up approach may generate models that differ from standard QCA models. CCubes is not equivalent to standard QCA algorithms; its models often significantly outperform common QCA models in model fit.

that tension in the CCM acceptability conventions (nor to substantively interpret the resulting models). What matters for our current purposes is that computing con-cov optima and maxima by means of ConCovOpt prior to actually conducting CCM analyses has (at least) three important payoffs. First, it allows us to determine how close actually obtained models come to optimal and maximal model fit. Second, it renders transparent whether the obtained models exhaust the space of con-cov optima or whether further models should be searched at different threshold settings. Third, without having to try out a whole range of threshold settings, CCMs can be run by directly constraining them towards optimal thresholds.

4 Large-*n* crisp-set data

The data in Table 1a are very simple and, although the resulting optimal models differ significantly in overall fit, they have a considerable overlap in causally relevant factor values. Thus, they induce only marginally different causal conclusions. To show that optimizing consistency and coverage scores by means of ConCovOpt can also make a substantive difference in causal conclusions, we now turn to a more intricate, large-*n* data example. Britt et al. (2000) investigate the determinants leading to the parental decision to terminate a pregnancy after a prenatal diagnosis of trisomy 21. Four exogenous factors are examined: existing children (C; 0 := 'none', 1 := '1 or more'), maternal age in years (M; 0 := '37 and under', 1 := '38 and above'), prior voluntary abortions (A; 0 := 'none', 1 := '1 or more'), and gestational age in weeks (G; 0 := '16 and under', 1 := '17 and over'). The endogenous factor is termination (T; 0 := 'continue', 1 := 'terminate'). The cases are 142 pregnant women receiving a trisomy 21 diagnosis at Wayne State University Clinic from September 1989 through October 1998. The complete data are provided in our replication script, the corresponding configuration table is reproduced in Table 2.

As is frequently the case in large-*n* configurational data, there are numerous imperfect pairs, highlighted with gray shading in Table 2. Instead of first calculating con-cov optima and maxima and only afterwards looking at concrete CCM models, we proceed in reverse order for this example. We begin by presenting the model Britt et al. (2000, 412) offer for their data. They choose a consistency threshold of 0.875. At this threshold, QCA builds two parsimonious models (only the second of which is mentioned by the authors):[9]

$$\text{QCA:} \quad c{*}G + C{*}g + m{*}A + c{*}M{*}a \leftrightarrow T \qquad con = 0.98\,;\ \ cov = 0.70 \tag{8}$$

$$c{*}G + C{*}g + m{*}A + M{*}a{*}g \leftrightarrow T \qquad con = 0.98\,;\ \ cov = 0.70 \tag{9}$$

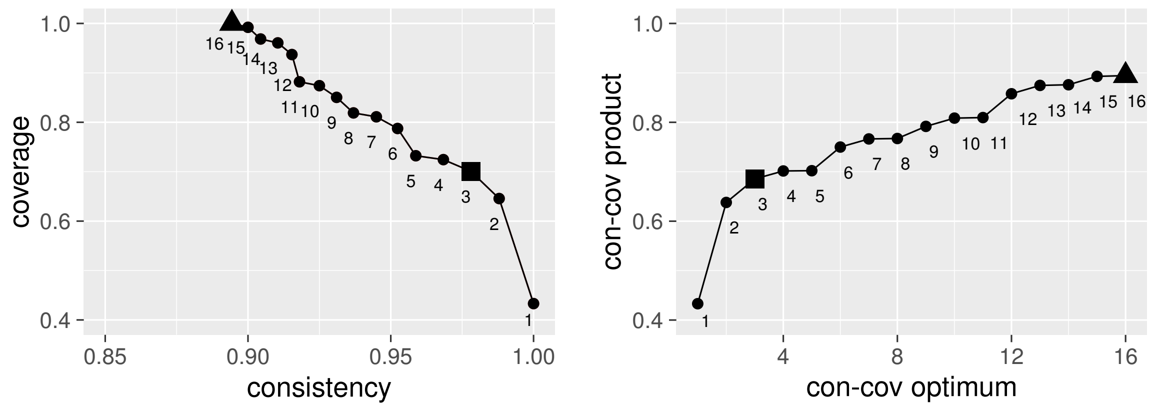
[9]The conservative models are the same at a consistency threshold of 0.875.

conf.	C	M	A	G	T	n	exo-groups	$\varphi(T{=}1)$	φ_{pub}	φ_{max}
σ_1	1	1	0	0	1	27	$\{\sigma_1, \sigma_2\}$	0, 1	1	1
σ_2	1	1	0	0	0	1				
σ_3	1	1	1	0	1	7	$\{\sigma_3, \sigma_4\}$	0, 1	1	1
σ_4	1	1	1	0	0	1				
σ_5	1	0	0	0	1	11	$\{\sigma_5\}$	1	1	1
σ_6	1	0	1	1	1	9	$\{\sigma_6\}$	1	1	1
σ_7	0	0	1	0	1	1	$\{\sigma_7\}$	1	1	1
σ_8	1	1	1	1	1	3	$\{\sigma_8, \sigma_9\}$	0, 1	0	1
σ_9	1	1	1	1	0	1				
σ_{10}	0	1	0	1	1	1	$\{\sigma_{10}\}$	1	1	1
σ_{11}	0	0	0	1	1	8	$\{\sigma_{11}\}$	1	1	1
σ_{12}	1	0	0	1	1	19	$\{\sigma_{12}, \sigma_{13}\}$	0, 1	0	1
σ_{13}	1	0	0	1	0	6				
σ_{14}	1	0	1	0	1	7	$\{\sigma_{14}\}$	1	1	1
σ_{15}	1	1	0	1	1	11	$\{\sigma_{15}, \sigma_{16}\}$	0, 1	0	1
σ_{16}	1	1	0	1	0	3				
σ_{17}	0	0	0	0	1	4	$\{\sigma_{17}, \sigma_{18}\}$	0, 1	0	1
σ_{18}	0	0	0	0	0	2				
σ_{19}	0	1	1	1	1	1	$\{\sigma_{19}\}$	1	1	1
σ_{20}	0	1	1	0	1	1	$\{\sigma_{20}, \sigma_{21}\}$	0, 1	0	1
σ_{21}	0	1	1	0	0	1				
σ_{22}	0	0	1	1	1	6	$\{\sigma_{22}\}$	1	1	1
σ_{23}	0	1	0	0	1	11	$\{\sigma_{23}\}$	1	1	1

Table 2: Configuration table for the data in Britt et al. (2000, 412) with (highlighted) imperfect pairs, exo-groups, rep-list $\varphi(T{=}1)$, and the rep-assignments φ_{pub} and φ_{max}.

As the consistency scores of the whole models and of the individual disjuncts are above the 0.75 minimum, (8) and (9) are acceptable by QCA standards. By contrast, according to CNA standards, their coverage is slightly too low. Is it thus possible to improve the overall model fit for Britt et al.'s data?

A handful of QCA re-runs with only slightly varied consistency thresholds will show that the results are highly volatile, yielding many different models with different overall fit. Identifying a con-cov maximum for these data calls for a systematic approach. We thus apply ConCovOpt with $\mathbf{O} = \{T{=}1\}$. The result of step (1) is the configuration table in Table 2. Step (2) yields 9 singleton exo-groups, which entail determinate values for the rep-list in step (3). By contrast, in all 7 non-singleton exo-groups the outcome T varies, meaning that DNFs reproducing the behavior of T as closely as possible can return either 0 or 1 for those groups. In total, the resulting rep-list $\varphi(T{=}1)$ induces $2^7 = 128$ rep-assignments $\varphi_1, \ldots, \varphi_{128}$ in step (4). In step (5), consistency and coverage scores are calculated for every φ_i of these



(a) (b)

Figure 1: Plot (a) shows the 16 con-cov optima for outcome T in the data of Britt et al. (2000), plot (b) the con-cov products of each optimum. ▲ is the con-cov maximum, ■ the published model.

128 assignments, by plugging the values of φ_i and T into the definitions for con^{cs} and cov^{cs}. Finally, 16 con-cov optima remain after the elimination of all non-optimal scores in step (6). They are plotted Figure 1a.

In light of these results, we can now say that the published model (9), which scores $\langle 0.98, 0.70\rangle$, indeed reaches a con-cov optimum (φ_3; ■ in Figure 1). However, with its con-cov product of 0.69 it is quite far away from a con-cov maximum. The con-cov products of all 16 optima are plotted in Figure 1b. The con-cov maximum for outcome T is the optimum φ_{16} $\langle 0.89, 1\rangle$, which reaches a con-cov product of 0.89 (▲). For transparency, we add the rep-assignment φ_{pub} (i.e. φ_3) realized by the published model as well as the assignment φ_{max} (i.e. φ_{16}) yielding the con-cov maximum to the last two columns of Table 2.

The value distribution of φ_{max} is striking: it assigns 1 to *every* exo-group, resulting in a (vacuous) *tautology*. In other words, a best fitting CCM model for the data of Britt et al. (2000) entails that a pregnancy is terminated in case of a trisomy 21 diagnosis whatever the values of the exogenous factors, that is, independently of the pregnant woman's existing children, prior abortions, age, and gestational age. To render this more concrete, let us now search for actual models at the con-cov maximum of $\langle 0.89, 1\rangle$. While standard QCA algorithms do not find a model reaching the con-cov maximum, CNA and CCubes have no problem finding a model realizing φ_{max}. At a threshold setting of $\langle 0.89, 1\rangle$, the following tautologous model is returned:

$$\text{CNA / CCubes:} \quad M + m \leftrightarrow T \qquad con = 0.89 \; ; \; cov = 1 \tag{10}$$

Of course, any other tautologous model, as $C + c \leftrightarrow T$ or $A + a \leftrightarrow T$, reaches the same overall fit scores, but (10) is the only model such that the two minimally sufficient conditions M and m individually have a consistency of 0.89 as well, which is why it is the only model issued by CNA and CCubes. (10) is acceptable both by QCA and CNA standards, and it also meets the consistency threshold imposed by Britt et al. (2000). It is the best fitting model for their data, and, for principled reasons, its fit scores cannot be outperformed. But of course, it is *not a causally interpretable* model. Causes are difference-makers of their effects, yet a tautology does not make a difference to anything.

This finding casts doubts on all causal conclusions Britt et al. (2000) have drawn from their data. 127 of all 141 women receiving a trisomy 21 diagnosis in their sample choose to terminate, making termination the canonical response to the diagnosis. No non-tautologous function of the examined exogenous factors can account for the outcome better than the tautologous model that always entails termination. The data contain too little variation on the outcome T to conclude anything about its causes; in particular, there is no evidence that T is caused by any of the factors C, M, A, or G. A causal interpretation of the published model (9) is unwarranted. This shows that a systematic search for con-cov maxima (prior to a CCM analysis) may have implications that go way beyond minor model adjustments or improvements. The optimization of consistency and coverage scores rendered possible by ConCovOpt may thoroughly change the conclusions drawn from a study.

5 Multi-value data

ConCovOpt is straightforwardly applicable to multi-value data. Although the factors in multi-value data can take more than two values, CCM models for a multi-value outcome $Y{=}v$ have the same logical form as crisp-set models: they account for $Y{=}v$ in terms of redundancy-free DNFs featuring factors exogenous to Y. And DNFs, irrespective of whether they feature *cs* or *mv* factors, are disjunctions of conjunctions that are true or false, that is, they only return the values 1 or 0. Hence, for an optimal multi-value DNF to reproduce the instantiation behavior of an outcome $Y{=}v$ as closely as possible the exact same conditions must be satisfied as in the crisp-set case (see p. 10). It follows that rep-lists and rep-assignments can be built and evaluated for consistency and coverage in exactly the same way for *mv* data as for *cs* data.

conf.	E	M	I	S	n	exo-groups	$\varphi(S{=}3)$	φ_1	φ_2	φ_3	φ_4
σ_1	0	0	2	3	1	$\{\sigma_1\}$	1	1	1	1	1
σ_2	1	0	1	2	2	$\{\sigma_2\}$	0	0	0	0	0
σ_3	1	1	0	2	2	$\{\sigma_3, \sigma_4\}$	0, 1	0	0	1	1
σ_4	1	1	0	3	3						
σ_5	1	0	0	3	1	$\{\sigma_5\}$	1	1	1	1	1
σ_6	1	1	1	3	1	$\{\sigma_6\}$	1	1	1	1	1
σ_7	1	1	2	2	1	$\{\sigma_7, \sigma_8\}$	0, 1	0	1	0	1
σ_8	1	1	2	3	2						
σ_9	0	0	1	2	4	$\{\sigma_9\}$	0	0	0	0	0
σ_{10}	0	0	0	3	1	$\{\sigma_{10}\}$	1	1	1	1	1

Table 3: Configuration table for the data in Verweij and Gerrits (2015), with re-coded factor S, exo-groups, rep-list, and all resulting rep-assignments.

To illustrate, we apply ConCovOpt to the *mv* data collected by Verweij and Gerrits (2015), who investigate the impact of different management strategies in response to unplanned events occurring in the context of the implementation of a large infrastructure project in Maastricht. Their data comprise 18 unplanned events (between 2009 and 2011) measuring the following exogenous factors: nature of the event (E; 0 := 'physical/remote', 1 := 'social/project/public'), nature of the management response (M; 0 := 'internal', 1 := 'external'), nature of the interaction between public and private managers (I; 0 := 'autonomous public', 1 := 'autonomous private', 2 := 'cooperation'). The endogenous factor is the satisfactoriness of the management response (S). To emphasize the data's *mv* nature, we replace the original values of S (i.e. 0 and 1) by two (arbitrarily chosen) different ones; that is, we will say that S takes the value 3 if satisfactoriness is high and the value 2 if it is not high.

The input to ConCovOpt, hence, is Verweij and Gerrits' complete data with $\mathbf{O} = \{S{=}3\}$ (the data are reproduced in our replication script). In step (1), ConCovOpt aggregates that data in a configuration table, which is contained in the first three (line-separated) columns of Table 3. As there is only one outcome, step (2) then forms only one set of exo-groups. They are listed in column 4. There are two non-singleton exo-groups, $\{\sigma_3, \sigma_4\}$ and $\{\sigma_7, \sigma_8\}$, each inducing two possible values in the rep-list in step (3). In that constellation, steps (4) generates $2^2 = 4$ rep-assignments φ_1 to φ_4. For transparency, we list them all in Table 3. In step (5), the value distributions of φ_1 to φ_4 and the instantiation behavior of outcome $S{=}3$ are plugged into con^{cs} and cov^{cs}. Step (6) then checks the resulting consistency and coverage scores for optimality, which check, in this case, is positive for all scores, meaning that none of them are eliminated. Overall, thus, ConCovOpt identifies the following four con-cov optima for outcome $S{=}3$, the last of which, φ_4, has the highest con-cov product and, thus, is a con-cov maximum:

$$\varphi_1 : \; con = 1 \; ; \quad cov = 0.44 \tag{11}$$

$$\varphi_2 : \; con = 0.86 \; ; \quad cov = 0.67 \tag{12}$$

$$\varphi_3 : \; con = 0.78 \; ; \quad cov = 0.78 \tag{13}$$

$$\varphi_4 : \; con = 0.75 \; ; \quad cov = 1 \tag{14}$$

Verweij and Gerrits (2015, 21) present a conservative solution, (15), that scores $\langle 1, 0.44\rangle$ and, thus, realizes the optimum φ_1. Independently of where in the interval $[1, 0.75]$ the consistency threshold is placed, QCA finds two parsimonious solutions, (16) and (17), which likewise realize φ_1 by scoring $\langle 1, 0.44\rangle$.

$$\text{QCA-CS:} \quad M{=}0 * I{=}0 \; + \; E{=}0 * M{=}0 * I{=}2 \; + \; E{=}1 * M{=}1 * I{=}1 \;\leftrightarrow\; S{=}3 \tag{15}$$

$$\text{QCA-PS:} \quad M{=}0 * I{=}0 \; + \; E{=}0 * I{=}2 \; + \; M{=}1 * I{=}1 \;\leftrightarrow\; S{=}3 \tag{16}$$

$$M{=}0 * I{=}0 \; + \; M{=}0 * I{=}2 \; + \; M{=}1 * I{=}1 \;\leftrightarrow\; S{=}3 \tag{17}$$

That is, in the range of acceptable threshold settings for consistency, standard QCA only finds models realizing one of four con-cov optima. Instead of now applying CNA and CCubes to find DNFs returning φ_2 to φ_4, we next introduce a procedure for building CCM models realizing *any* con-cov optimum for crisp-set and multi-value data.

A con-cov optimum for an outcome $Y{=}v$ in δ is realized by a DNF composed of values of factors that are exogenous to Y in δ. Such a DNF outputs either 1 or 0 for every exo-group. Any two DNFs that return the same output for all exo-groups have the same consistency and coverage scores. For convenience, let us call the set of exo-groups to which a rep-assignment φ_i assigns the value 1 the *positive group* of φ_i. One particularly interesting DNF returning φ_i then is what we label φ_i's *canonical DNF*—in reference to canonical normal forms of logical expressions (Lemmon 1965, 198).

Canonical DNF. The canonical DNF returning a rep-assignment φ_i (in a configuration table) for outcome $Y{=}v$ is the disjunction of the configurations of all factors exogenous to Y in φ_i's positive group.

Just as any Boolean function is guaranteed to have unique canonical normal forms, so is every rep-assignment guaranteed to have exactly one canonical DNF, which moreover is easily built. For example, disjunctively concatenating the configurations of factors E, M and

I in φ_4's positive group in Table 3 yields the canonical DNF returning φ_4; it is a DNF that scores $\langle 0.75, 1\rangle$ in accounting for $S{=}3$:

$$\begin{aligned} E{=}0 * M{=}0 * I{=}2 \;+\; E{=}1 * M{=}1 * I{=}0 \;+\; E{=}1 * M{=}0 * I{=}0 \;+ \\ E{=}1 * M{=}1 * I{=}1 \;+\; E{=}1 * M{=}1 * I{=}2 \;+\; E{=}0 * M{=}0 * I{=}0 \end{aligned} \tag{18}$$

In the very same way, the canonical DNFs of the other con-cov optima for outcome $S{=}3$ can be construed. For example, as the exo-group $\{\sigma_7, \sigma_8\}$ is not in φ_3's positive group, removing the configuration of the exogenous factors in $\{\sigma_7, \sigma_8\}$, *viz.* $E{=}1 * M{=}1 * I{=}2$, from (18) is all it takes to build the canonical DNF returning φ_3.

Of course, as canonical DNFs are built by disjunctively combining all configurations of all exogenous factors in an optimum's positive group, they will normally not account for corresponding outcomes in a redundancy-free manner, and consequently, not be causally interpretable. For instance, if we remove $E{=}0$ from the first and the last disjuncts in (18), we are still left with a DNF that returns exactly the same output for all exo-groups in Table 3; that is, we are still left with a DNF returning φ_4. The same holds if we continue to eliminate $E{=}1$ from all disjuncts. By contrast, if $I{=}2$ is eliminated from the first disjunct of (18), we are left with a DNF that scores $\langle 0.56, 1\rangle$ in accounting for $S{=}3$ and, thus, no longer returns φ_4. The reason is that a DNF featuring $E{=}0 * M{=}0$ as a separate disjunct (instead of $E{=}0 * M{=}0 * I{=}2$) outputs 1 for exo-group $\{\sigma_9\}$, whereas φ_4 assigns 0 to that group. In sum, while some factor values can be removed from (18) such that the remaining DNF still returns φ_4, other factor values are indispensable for (18) returning φ_4.

These considerations suggest that (18) can be turned into a *redundancy-free DNF* returning φ_4 by systematically removing factor values and checking whether the resulting DNF still outputs the same as (18) for all exo-groups of Table 3. All factor values for which this check is positive are redundant; all factor values for which the check is negative are indispensable (non-redundant). If all redundancies are removed from (18) along these lines, we are left with this redundancy-free DNF: $I{=}2 + I{=}0 + M{=}1$. Accounting for $S{=}3$ on its basis yields the following model, which reaches the con-cov maximum for Verweij and Gerrits' data:

$$I{=}2 \;+\; I{=}0 \;+\; M{=}1 \;\leftrightarrow\; S{=}3 \qquad con = 0.75\;;\;\; cov = 1 \tag{19}$$

(19) has a peculiarity. Although the model as a whole meets the consistency minimum of 0.75, two of its component disjuncts do not. $M{=}1$ and $I{=}0$ are sufficient for $S{=}3$ with consistencies of 0.67 and 0.71 only. It follows that in order to find (19) with CNA and CCubes, the consistency threshold must be lowered to 0.66 (see the replication script for

details). Whether models should be considered acceptable that meet the 0.75 consistency minimum even though some of their components do not is a question that has not been explicitly discussed in the literature—and this is not the place to fill that gap. But if the fact that (19) cannot be built at the consistency threshold of 0.75 is taken to render that model unacceptable, our previous discussion has shown that there are other con-cov optima on offer for Verweij and Gerrits' data that score better than the published model. For example, by first building the canonical DNF returning φ_2 and then systematically eliminating redundancies, we find the following redundancy-free DNF, which additionally satisfies the requirement that all disjuncts have consistencies of 0.75 or higher and which, hence, is returned by both CNA and CCubes at that consistency threshold:

$$I{=}2 \;+\; M{=}0 * I{=}0 \;+\; M{=}1 * I{=}1 \;\leftrightarrow\; S{=}3 \qquad con = 0.86\,;\;\; cov = 0.67 \qquad (20)$$

(20) comes close to the two parsimonious QCA solutions (16) and (17). The only difference is that $I{=}2$ is conjunctively combined with $E{=}0$ in (16) and with $M{=}0$ in (17), which creates a model ambiguity, whereas $I{=}2$ is a stand-alone disjunct in (20), which is non-ambiguous. Hence, replacing QCA's parsimonious solutions by redundancy-free DNFs returning φ_2 or φ_4—systematically built via their canonical DNFs—not only increases the con-cov product from 0.44 to 0.58 and 0.75, respectively, but also resolves a model ambiguity—two clear advantages of (20) and (19) over (16) and (17).[10]

We end this section by assembling the steps building DNFs returning rep-assignments in procedural form, which we label *DNFbuild*, for short.

Procedure 2 (DNFbuild) build redundancy-free DNFs returning rep-assignments

---

Input: *cs* or *mv* configuration table CT and rep-assignment φ_i for outcome $Y{=}v$ in CT
Output: redundancy-free DNF(s) returning φ_i

(1) Build the canonical DNF_{cano} returning φ_i by disjunctively concatenating the configurations of the factors exogenous to $Y{=}v$ in φ_i's positive group.

(2) Eliminate all factor values from DNF_{cano} for which it holds that the result of the elimination, *viz.* DNF_{elim}, still returns φ_i (i.e. produces the same output as DNF_{cano} for all exo-groups in CT).

⇒ If no further factor values can be eliminated from DNF_{elim}, it is a redundancy-free DNF returning φ_i.

---

[10] At the same time it is worth noting that (20) and (19) do not contradict (16) and (17); rather, the former are mere submodels of the latter (Baumgartner and Ambühl 2018).

In case of *cs* and *mv* configuration tables, all φ_i's have a canonical DNF, which is efficiently built by step (1) of DNFbuild. Step (2), by contrast, involves some intricacies. The reason is that different orders in which factor values are eliminated from DNF_{cano} may result in different redundancy-free DNFs. Generating all of these DNFs can be done more or less efficiently, but the running time of all algorithms that solve this problem grows exponentially with the number of factors and the number of values these factors can take. In the replication script, we use an algorithm called *ereduce* to generate all redundancy-free forms of DNF_{cano}.[11] In the end, what matters for our current purposes is not the concrete implementation of step (2) but the fact that DNFbuild is guaranteed to find a redundancy-free DNF Φ_{rf} of every con-cov optimum $\langle h, k\rangle$. The equivalence $\Phi_{rf} \leftrightarrow Y{=}v$ then amounts a CCM model accounting for $Y{=}v$ with consistency h and coverage k. That is, in case of *cs* and *mv* data, there exists a CCM model for every con-cov optimum.

6 Fuzzy-set data

This section, first, applies ConCovOpt to fuzzy-set data, and second, shows that, contrary to crisp-set and multi-value data, there is no guarantee that actual CCM models exist for every con-cov optimum in fuzzy-set data. As concrete background for this discussion, we choose the study by Basurto (2013) who analyzes the autonomy (i.e. the ability to exercise self-governance) among local institutions for biodiversity conservation in Costa Rica. The study aims to identify causally relevant factors for, on the one hand, the emergence of autonomy between 1986 and 1998 and, on the other, the endurance of that autonomy between 1998 and 2006. Basurto investigates three groups of potentially causally relevant factors: local, national and international ones. In what follows, we focus on the local influence factors of high local communal involvement through direct employment (E), high local direct spending (S), and co-management with local or regional stakeholders (C), and we concentrate on the outcome of endurance of high local autonomy (A), with 0 representing 'no' and 1 'yes' for all factors (Basurto 2013, 577). The data cover 16 Costa Rican biodiversity conservation programs; the factors are calibrated on a membership scale with increments of 0.2.

As input to ConCovOpt we, thus, use Basurto's data (which can be consulted in the replication script) with $\mathbf{O} = \{A{=}1\}$. In step (1), ConCovOpt builds the configuration table in Table 4. In this example, each case corresponds to exactly one configuration—which is not uncommon for *fs* data. Since there, again, is only one outcome in $\mathbf{O}$, step (2) then builds

[11] In a nutshell, *ereduce* searches for minimal hitting sets in DNF_{cano} that prevent DNF_{cano} from being false in the data from which it is inferred.

conf.	E	S	C	A	n	exo-groups	$\varphi(A{=}1)$	φ_{max}	φ_1
σ_1	1.0	1.0	1.0	1.0	1	$\{\sigma_2\}$	1.0	1.0	1.0
σ_2	1.0	0.6	1.0	0.4	1				
σ_3	1.0	0.6	1.0	1.0	1	$\{\sigma_2, \sigma_3, \sigma_4\}$	0.4, 0.6, 1.0	0.6	0.6
σ_4	1.0	0.6	1.0	0.6	1				
σ_5	1.0	0.8	1.0	0.8	1	$\{\sigma_5\}$	0.8	0.8	0.8
σ_6	1.0	0.4	1.0	0.4	1	$\{\sigma_6\}$	0.4	0.4	0.4
σ_7	0.4	1.0	1.0	1.0	1	$\{\sigma_7\}$	1.0	1.0	1.0
σ_8	1.0	1.0	0.0	1.0	1	$\{\sigma_8\}$	1.0	1.0	1.0
σ_9	0.4	0.4	0.2	0.4	1	$\{\sigma_9, \sigma_{10}\}$	0.2, 0.4	0.4	0.4
σ_{10}	0.4	0.4	0.2	0.2	1				
σ_{11}	0.4	0.4	0.6	0.4	1	$\{\sigma_{11}\}$	0.4	0.4	0.4
σ_{12}	0.4	0.4	0.4	0.2	1	$\{\sigma_{12}, \sigma_{13}\}$	0.4, 0.6	0.6	0.4
σ_{13}	0.4	0.4	0.4	0.6	1				
σ_{14}	0.2	0.4	0.2	0.4	1	$\{\sigma_{14}\}$	0.4	0.4	0.4
σ_{15}	0.4	0.4	0.0	0.6	1	$\{\sigma_{15}, \sigma_{16}\}$	0.4, 0.6	0.6	0.4
σ_{16}	0.4	0.4	0.0	0.4	1				

Table 4: Configuration table for the data in Basurto (2013) with exo-groups, rep-list, and rep-assignments φ_{max} and φ_1.

one set of exo-groups, which are listed in column 4 of Table 4.

The main peculiarity of fs data is that configurations and outcomes (or generally, factors) are not merely instantiated or not instantiated in each case of the data but instantiated with different set membership scores (i.e. to different degrees) in many different cases. One consequence is that there are often not only imperfect *pairs* but imperfect *n-tuples*, where n is the number of different membership scores an outcome has in an exo-group. An example is exo-group $\{\sigma_2, \sigma_3, \sigma_4\}$ in Table 4. Since A has three different membership scores in that group, *viz.* A=0.4, A=1.0, and A=0.6, it corresponds to an imperfect triple. That the other 3 non-singleton exo-groups in Table 4 have only two elements is mere happenstance; fs exo-groups can have any number of members. Still, just as CCM models for other data types, models for fs data have highest consistency and coverage if they reproduce the behavior of the outcome, meaning the value distribution of the outcome factor, as closely as possible. That is, a DNF Φ of an fs model $\Phi \leftrightarrow Y$ scores the higher on consistency and coverage, the closer the value of Φ comes to the value of Y for every exo-group.

Due to the instantiation by degrees in fs data, however, the notion of reproducing the behavior of an outcome as closely as possible cannot be spelled out exactly as for cs and mv data (cf. p. 10). While a rep-assignment for the latter data types must reproduce the behavior of the outcome by only assigning one of two values, 0 or 1, a rep-assignment φ for fs data may assign many more values, *viz.* the minimum of the conjuncts' scores in a conjunction and the maximum of the disjuncts' scores in a disjunction. The values φ can assign to a particular exo-group are constrained by the membership scores of the exogenous factors in that group. More specifically, as φ is the output of a disjunction of conjunctions of positive or negative factor values, it can only assign the membership scores of the positive or of the negative factors in that group, but no other values. To illustrate, take again the exo-group $\{\sigma_2, \sigma_3, \sigma_4\}$, in which E=1.0, S=0.6, and C=1.0. The conjunction $E{*}S{*}C$ issues 0.6 for that group, *viz.* $min(1.0, 0.6, 1.0)$; correspondingly, $e{*}c$ and $E{*}C$ return 0.0 and 1.0, respectively, or $s{*}C$ outputs 0.4. But any value other than 0.0 (negation of E and C), 0.4 (negation of S), 1.0 (value of E and C) or 0.6 (value of S) cannot be assigned to that exo-group by φ, meaning that those are φ's *possible values* for that group. It follows that a rep-assignment φ *reproduces the behavior* of an outcome as closely as possible if, and only if, it holds for every exo-group x that φ assigns one of its possible values to x that comes as close as possible to the outcome's membership scores in x. In the case of exo-group $\{\sigma_2, \sigma_3, \sigma_4\}$ that means that φ reproduces the behavior of A as closely as possible iff it assigns one of 0.4, 0.6, or 1.0 to that exo-group (but not 0.0); or for exo-group $\{\sigma_{12}, \sigma_{13}\}$ φ must assign either 0.4, which is equally close to A=0.2 and A=0.6 in that group, or 0.6, which is closest to A=0.6. Sometimes the closest possible value of φ exactly matches an outcome's score, sometimes it is far away from it. Sometimes only one of the possible values is closest to the outcome's score(s), sometimes multiple values are closest.

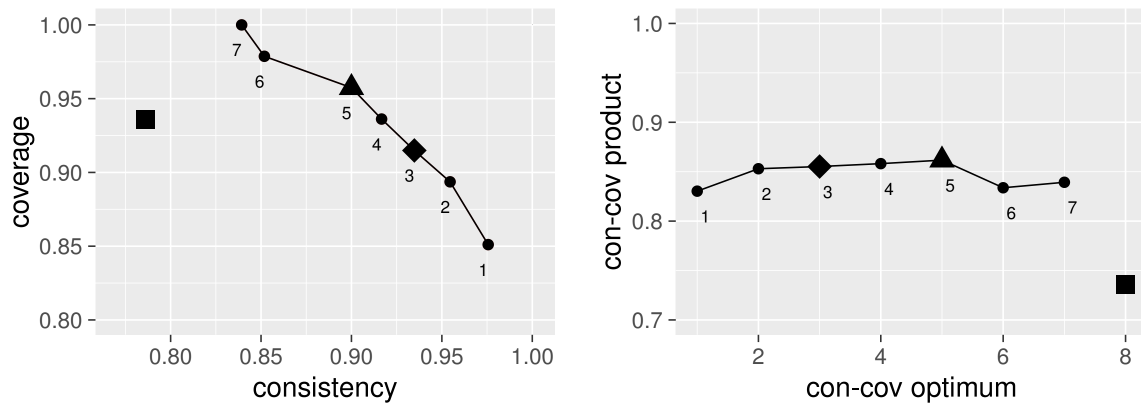
Based on this notion of reproducing the behavior of an outcome, step (3) builds the rep-list $\boldsymbol{\varphi(A{=}1)}$ in Table 4 for Basurto's data. From this list, 24 rep-assignments $\varphi_1, \ldots, \varphi_{24}$ for outcome A are built in step (4). Just as in case of *cs* and *mv* data, the values of the rep-assignments and of the outcome are then plugged into the definitions for consistency and coverage in step (5)—this time, of course, using the fuzzy-set definitions con^{fs} and cov^{fs} in (2). After eliminating all rep-assignments inducing non-optimal scores, 7 con-cov optima remain, which are plotted in Figure 2a with their con-cov products in Figure 2b.

Basurto (2013) chooses a consistency threshold of 0.79 and produces an intermediate solution, (21), which coincides with QCA's conservative solution. At that consistency threshold, QCA issues the two parsimonious solutions (22) and (23).

$$\text{QCA-IS/CS:} \quad C + E{*}S \leftrightarrow A \qquad con = 0.79\ ;\ \ cov = 0.94 \tag{21}$$

$$\text{QCA-PS:} \quad C + E \leftrightarrow A \qquad con = 0.79\ ;\ \ cov = 0.94 \tag{22}$$

$$C + S \leftrightarrow A \qquad con = 0.79\ ;\ \ cov = 0.96 \tag{23}$$



(a) (b)

Figure 2: Plot (a) shows the 7 con-cov optima for outcome A in the data of Basurto (2013) and the con-cov score of the published model (■), plot (b) the con-cov products of each optimum and of the published model. ▲ is the con-cov maximum and ◆ the best optimum realizable by a CCM model.

Contrary to our previous examples, these QCA models fall significantly short of a con-cov optimum, let alone a con-cov maximum. Moreover, the models are not robust under variations of the consistency threshold. But thanks to ConCovOpt, we now have concrete threshold settings at which to search for models with optimal fit. The con-cov maximum of $\langle 0.9, 0.957\rangle$ (▲ in Figure 2) is reached by the rep-assignment φ_{max} in Table 4. It turns out, however, that neither QCA nor CNA nor CCubes find a model at the threshold setting $\langle 0.9, 0.957\rangle$, not even—as is possible in CNA and CCubes—if the consistency of individual disjuncts is allowed to fall short of 0.9. There simply does not exist a CCM model for outcome A at the con-cov maximum. This is not some idiosyncrasy of Basurto's data. Con-cov optima for fs data frequently do not have actual CCM models realizing them.

In a cs and mv configuration table, the configuration Γ_i of exogenous factors in an exo-group $\{\sigma_i\}$ of some outcome Y is guaranteed not to be instantiated in any other exo-groups of Y. Hence, if Y is given in $\{\sigma_i\}$, Γ_i—properly freed of redundancies—can safely be included as a disjunct in a model of Y without affecting the model's consistency in any exo-groups other than $\{\sigma_i\}$. It is therefore possible to modularly build CCM models for every rep-assignment φ_i along the lines of DNFbuild (cf. section 5). The DNFbuild approach, however, does not work for fuzzy-set data, because if Y is an fs outcome, the configuration Γ_i in exo-group $\{\sigma_i\}$ may likewise be instantiated with some non-zero membership score in many other exo-groups. In consequence, including Γ_i as a disjunct in a CCM model of Y is likely to affect the model's consistency in exo-groups other than $\{\sigma_i\}$. It may hence happen that a DNF only returns an optimal value for exo-group $\{\sigma_i\}$ provided it contains factor values that, at the same time, return a non-optimal value for another exo-group $\{\sigma_j\}$, meaning that no DNF returns optimal values for both $\{\sigma_i\}$ and $\{\sigma_j\}$.

To make this (abstract problem) concrete, compare the exo-groups $\{\sigma_9, \sigma_{10}\}$ and $\{\sigma_{15}, \sigma_{16}\}$ in Table 4. The rep-assignment φ_{max} assigns 0.4 to $\{\sigma_9, \sigma_{10}\}$ and 0.6 to $\{\sigma_{15}, \sigma_{16}\}$. Given the membership scores E=0.4, S=0.4 and C=0.2 in $\{\sigma_9, \sigma_{10}\}$, a DNF only returns 0.4 for $\{\sigma_9, \sigma_{10}\}$ if it includes E or S or $E*S$ (but not C) as a disjunct. But E, S and $E*S$ also return 0.4 for group $\{\sigma_{15}, \sigma_{16}\}$, while φ_{max} assigns 0.6 to that group. Given the membership scores E=0.4, S=0.4 and C=0.0 in $\{\sigma_{15}, \sigma_{16}\}$, a DNF would have to include e or s or $e*s$ in order to return 0.6 for $\{\sigma_{15}, \sigma_{16}\}$. But if those factor values are included, 0.6 is issued for $\{\sigma_9, \sigma_{10}\}$, which again is not the value assigned by φ_{max}. In sum, no DNF outputs 0.4 for $\{\sigma_9, \sigma_{10}\}$ and 0.6 for $\{\sigma_{15}, \sigma_{16}\}$, meaning that no actual CCM model will realize φ_{max}.

This problem is resolved if we do not require 0.6 to be issued for $\{\sigma_{15}, \sigma_{16}\}$ but 0.4, which indeed happens to be assigned by another rep-assignment, φ_1 (cf. Table 4), yielding the con-cov optimum $\langle 0.93, 0.91 \rangle$. Moreover, it turns out that there exists an actual DNF also reproducing all other values of φ_1. At a threshold setting of $\langle 0.93, 0.91 \rangle$, both CNA and CCubes find the following model:[12]

$$\text{CNA / CCubes:} \quad S \leftrightarrow A \qquad con = 0.93 \; ; \; cov = 0.91 \tag{24}$$

Overall, even though we found that DNFbuild does not work for fuzzy-set data and that no CCM model exists realizing the con-cov maximum for Basurto's data, first calculating con-cov optima by means of ConCovOpt and then purposefully searching for CCM models at the optimal thresholds resulted in model (24) with considerably better fit than Basurto's published model (21). Moreover, that improvement of model fit substantively alters the causal conclusions to be drawn from Basurto's (2013) study. A causal interpretation of (24) suggests that the endurance of autonomy only depends on local direct spending. Contrary to Basurto's findings, there is no evidence in his data that the other exogenous factors might be difference-makers of autonomy endurance as well.

[12]At a consistency threshold of 0.93, (24) is not found by standard QCA, but if that threshold is lowered to 0.9, (24) is also returned by QCA.

7 Discussion

We end this paper by putting ConCovOpt into proper methodological perspective. Most importantly, ConCovOpt is not intended as a tool for a “simple hunt for high values of consistency and coverage” (Schneider and Wagemann 2012, 148). As emphasized in the introduction, there are other parameters of model fit, for example, robustness/sensitivity and compliance with established background theories, the former of which, in particular, has attracted a lot of attention in recent years (Skaaning 2011; Braumoeller 2015; Thiem, Spöhel, and Dusa 2016; Cooper and Glaesser 2016). That is, high consistency and coverage are merely one asset of a model among others, and models with higher consistency and coverage should not be unequivocally preferred if they are outperformed by rival models in other respects.

What is more, the problem of model overfitting is still underinvestigated in configurational causal modeling, and there is ample evidence that CCMs have a strong tendency to overfit if they are induced to do so by overly high consistency and coverage thresholds (see e.g. Braumoeller 2015; Rohlfing 2015). One clear indication that overfitting might be taking place is that the complexity of resulting models increases disproportionally to their increase in model fit. That is a phenomenon that regularly occurs if CCMs are ‘forced’ to build models reaching con-cov maxima—an example is provided in the replication script. What would hence be needed is a tool analogous to, say, the Akaike Information Criterion (AIC) in statistical modeling that strikes a balance between model fit and simplicity. As long as such a tool is not available for CCMs, a general preference of models reaching con-cov maxima would be blind and hazardous.

Still, we have discussed various examples in this paper for which ConCovOpt has helped to significantly increase the model fit without an increase in model complexity, thus steering clear of overfitting dangers. Moreover, systematically scanning the model space at optimal consistency and coverage scores has led to the resolution of model ambiguities in case of Verweij and Gerrits’ (2015) as well as Basurto’s (2013) data; and it has even called into question the causal interpretability of a whole data set, *viz.* in case of the study by Britt et al. (2000). All of this shows that rendering con-cov optima and maxima transparent may importantly affect the causal conclusions drawn from configurational data.

Hence, ConCovOpt is intended as a tool for exploring the space of viable CCM models at optimal consistency and coverage scores in a systematic and exhaustive manner. In recent years, the awareness in the CCM literature has grown that the space of viable models for analyzed data may be larger than anticipated (cf. e.g. Baumgartner and Thiem 2017).

When the data quality is very high, meaning when there are no imperfect pairs (or n-tuples), uncovering the whole model space is a matter of applying the relevant CCM once, at one determinate threshold setting. But in the presence of imperfect pairs, multiple CCM runs at various threshold settings are needed. Slight changes in thresholds may greatly change the resulting models and there may be no systematicity in how threshold changes affect model changes. ConCovOpt efficiently uncovers all con-cov optima and maxima prior to applications of CCMs. This information makes it possible to apply CCMs by directly constraining them towards optimal thresholds, without having to go through a whole array of threshold settings. Also, it can be determined how close actually obtained models come to con-cov optima and whether the obtained models exhaust the space of con-cov optima or whether further models should be searched at different threshold settings.

In sum, models reaching con-cov optima will sometimes turn out to be the best models overall, sometimes not, and sometimes they will be of methodological interest even without being causally interpreted. But in one way or another, transparency on the model space at consistency and coverage optima is univocally valuable for configurational causal modeling.

References

Basurto, X. (2013). Linking multi-level governance to local common-pool resource theory using fuzzy-set Qualitative Comparative Analysis: Insights from twenty years of biodiversity conservation in Costa Rica. *Global Environmental Change 23*, 573–587.

Baumgartner, M. (2008). Regularity theories reassessed. *Philosophia 36*, 327–354.

Baumgartner, M. (2009). Inferring causal complexity. *Sociological Methods & Research 38*, 71–101.

Baumgartner, M. and M. Ambühl (2018). Causal modeling with multi-value and fuzzy-set coincidence analysis. *Political Science Research and Methods*. doi: 10.1017/psrm.2018.45.

Baumgartner, M. and A. Thiem (2017). Model ambiguities in configurational comparative research. *Sociological Methods & Research 46*(4), 954–987.

Braumoeller, B. F. (2015). Guarding against false positives in Qualitative Comparative Analysis. *Political Analysis 23*(4), 471–487.

Britt, D. W., S. T. Risinger, V. Miller, M. K. Mans, E. L. Krivchenia, and M. I. Evans (2000). Determinants of parental decisions after the prenatal diagnosis of down syndrome: Bringing in context. *American Journal of Medical Genetics 93*(5), 410–416.

Cooper, B. and J. Glaesser (2016). Exploring the robustness of set theoretic findings from a large n fsQCA: an illustration from the sociology of education. *International Journal of Social Research Methodology 19*(4), 445–459.

Cronqvist, L. and D. Berg-Schlosser (2009). Multi-value QCA (mvQCA). In B. Rihoux and C. C. Ragin (Eds.), *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*, pp. 69–86. London: Sage Publications.

Dusa, A. (2018). Consistency Cubes: a fast, efficient method for exact Boolean minimization. *The R Journal 10*(2), 357–370.

Dusa, A. (2019). ***QCA****: A Package for Qualitative Comparative Analysis.* R Package Version 3.4.0. https://cran.r-project.org/package=QCA.

Gelman, A. and J. Hill (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.

Giugni, M. and S. Yamasaki (2009). The policy impact of social movements: A replication through Qualitative Comparative Analysis. *Mobilization: An International Quarterly 14*(4), 467–484.

Graßhoff, G. and M. May (2001). Causal regularities. In W. Spohn, M. Ledwig, and M. Esfeld (Eds.), *Current Issues in Causation*, pp. 85–114. Paderborn: Mentis.

Lemmon, E. J. (1965). *Beginning Logic*. London: Chapman & Hall.

Mackie, J. L. (1974). *The Cement of the Universe. A Study of Causation*. Oxford: Clarendon Press.

Ragin, C. C. (1987). *The Comparative Method*. Berkeley: University of California Press.

Ragin, C. C. (2006). Set relations in social research: Evaluating their consistency and coverage. *Political Analysis 14*, 291–310.

Ragin, C. C. (2008). *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago: University of Chicago Press.

Rihoux, B. and G. De Meur (2009). Crisp-set Qualitative Comparative Analysis (csQCA). In B. Rihoux and C. C. Ragin (Eds.), *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*, pp. 33–68. Thousand Oaks: Sage.

Rohlfing, I. (2015). Mind the gap: A review of simulation designs for Qualitative Comparative Analysis. *Research & Politics*. doi: 10.1177/2053168015623562.

Schneider, C. Q. and C. Wagemann (2012). *Set-Theoretic Methods: A User's Guide for Qualitative Comparative Analysis (QCA) and Fuzzy-Sets in the Social Sciences*. Cambridge: Cambridge University Press.

Skaaning, S.-E. (2011). Assessing the robustness of crisp-set and fuzzy-set QCA results. *Sociological Methods & Research 40*(2), 391–408.

Spirtes, P., C. Glymour, and R. Scheines (2000). *Causation, Prediction, and Search* (2 ed.). Cambridge: MIT Press.

Suppes, P. (1970). *A Probabilistic Theory of Causality*. Amsterdam: North Holland.

Thiem, A. (2014). Unifying configurational comparative methods: Generalized-set Qualitative Comparative Analysis. *Sociological Methods & Research 43*(2), 313–337.

Thiem, A. (2018). ***QCApro**: Advanced Functionality for Performing and Evaluating Qualitative Comparative Analysis*. R Package Version 1.1-2.

Thiem, A., M. Baumgartner, and D. Bol (2016). Still lost in translation! A correction of three misunderstandings between configurational comparativists and regressional analysts. *Comparative Political Studies 49*, 742–774.

Thiem, A., R. Spöhel, and A. Dusa (2016). Enhancing sensitivity diagnostics for Qualitative Comparative Analysis: A combinatorial approach. *Political Analysis 24*(1), 104–120.

Verweij, S. and L. M. Gerrits (2015). How satisfaction is achieved in the implementation phase of large transportation infrastructure projects: A Qualitative Comparative Analysis into the A2 tunnel project. *Public Works Management & Policy 20*(1), 5–28.