Article

Comparative Political Studies
2016, Vol. 49(6) 793–800

DOI: 10.1177/0010414015626456
cps.sagepub.com

Still Searching for the Value-Added: Persistent Concerns About Set-Theoretic Comparative Methods

Jack Paine[1]

Abstract

What novel leverage for understanding the social world do set-theoretic comparative methods (STCM) offer? I have argued that although important methodological ideas underlie this method, many STCM techniques converge with existing quantitative tools of statistical modeling. Unfortunately, STCM scholars have often obscured this crucial point by instead emphasizing stark differences from quantitative tools. Here, I further develop two key arguments about these convergences by addressing the astute commentaries offered by Thiem et al. and Schneider: (a) Regarding necessary and sufficient conditions, STCM's procedures for incorporating cases from a 2 × 2 table may yield erroneous conclusions that can easily be avoided by using more conventional techniques. (b) Regarding causal complexity, STCM and statistical interaction terms often provide the same information. These arguments demonstrate that STCM scholars have yet to establish distinctive advantages of their methods over statistical modeling. Furthermore, alternative qualitative tools offer considerably more promise than does STCM.

Keywords

comparative method, necessary and sufficient conditions, qualitative comparative analysis (QCA), qualitative methods, set theory

[1]University of Rochester, NY, USA

Corresponding Author:
Jack Paine, Wallis Institute of Political Economy, University of Rochester, 107 Harkness Hall, Rochester, NY 14627-0158, USA.
Email: jackpaine@rochester.edu

What novel leverage for understanding the social world do set-theoretic comparative methods (STCM) offer? I have argued (Paine, 2016) that although important methodological ideas underlie this method, many STCM techniques converge with the existing quantitative tools of statistical modeling. Unfortunately, STCM scholars have often obscured this crucial point by instead emphasizing stark differences between their methods and quantitative tools. Here, I further develop two key arguments about these convergences by addressing the astute commentaries offered by Thiem, Baumgartner, and Bol (2016) and Schneider (2016): (a) Regarding necessary and sufficient conditions, STCM's procedures for incorporating cases from a 2 × 2 table may yield erroneous conclusions that can easily be avoided by using techniques based upon conventional statistical ideas about association. (b) Regarding causal complexity, STCM and statistical interaction terms often provide the same information. Consistent with my conclusions and those of Munck (2016), these arguments demonstrate that STCM scholars have yet to establish distinctive advantages of their methods over statistical modeling. Furthermore, alternative qualitative tools offer considerably more promise than does STCM for inferring causal relationships.

Inferring Necessity and Sufficiency: Relative Frequency Is Crucial

The focus on necessary and sufficient causal conditions is a fundamental feature of STCM. The point of departure in this discussion is the standard STCM argument that scholars do not need to incorporate data from all four cells in a 2 × 2 table.[1] Schneider's (2016) proposal to focus on all four cells is a valuable step forward, yet the standard rejection of using data from all four cells is so common in the STCM literature that my critique of this position hardly involves a "straw man," as he puts it (p. 5).[2]

Recognizing that all four cells matter raises crucial issues about how to combine information from the cells to make valid inferences. Because STCM encourages researchers to ignore certain cells, scholars in this tradition have generally overlooked these issues. Furthermore, even contributions such as Schneider and Wagemann (2012) that have proposed STCM procedures for productively incorporating all four cells face important shortcomings. When the distribution of $X = 0$ and $X = 1$ cases is skewed, set-theoretic concepts—which do not account for relative frequency—do not adequately capture notions of necessity or sufficiency. They likewise can generate easily avoidable false negatives or false positives. Therefore, core STCM questions such as "is X a superset of Y?" (Schneider, 2016, p. 5) may provide misleading inferences about necessity or sufficiency. It is crucial to instead use metrics that adjust for relative frequency, as I have sought to do, in contrast to Schneider's rejection of my proposed consistency measure (Paine, 2016).

How STCM Can Produce False Negatives

Schneider (2016) alters the baseball example from my article by asking, "In 1927, was a Babe Ruth at-bat a necessary condition for a home run?" Contrary to Schneider's argument, my proposed consistency measure (Paine, 2016, p. 12, Equation 6) provides a more sensible conclusion than standard STCM metrics. In 1927, Babe Ruth homered in 60 of his 691 times at-bat (8.7%), whereas all other hitters homered in 862 of their collective 94,906 at-bats (0.9%). Equation 6 from my article accounts for the relative rarity of Babe Ruth at-bats and calculates a high necessary condition consistency score, 0.91. Similarly, using a standard quantitative measure that adjusts for relative frequency, a Babe Ruth at-bat raised the odds of a home run by a factor of almost 10—an enormous amount. Inferring relatively strong support for the necessary condition hypothesis is not, as alleged by Schneider (2016), an "absurd claim" (p. 9). Home runs occurred infrequently when Babe Ruth was not batting, which corresponds with STCM's conceptualization of an "almost" necessary condition.

In contrast, by not accounting for the low frequency of Babe Ruth at-bats relative to the rest of the league, the STCM necessary condition metric produces a basically meaningless low consistency score of 0.09 (see Schneider, 2016, Table 2). More troubling, and moving beyond the focus on necessity and sufficiency, STCM misses a substantively important relationship: Babe Ruth hit home runs with far greater relative frequency than the rest of the league. Overall, an example that Schneider uses to demonstrate the superiority of the STCM metric instead produces the opposite conclusion. His approach yields a false negative.

How STCM Can Produce False Positives

Throughout the history of major league baseball, is a non-Babe Ruth at-bat necessary for a home run? Table 1 provides data to address this question posed in my article (Paine, 2016).

The set of home runs hit by individuals other than Babe Ruth is almost a perfect subset (in percentage terms, 99.7%) of all home runs hit in baseball history—consistent, by the STCM metric, with the hypothesis that an at-bat by a player other than Babe Ruth is almost a necessary condition for a home run. Although Schneider correctly notes that using Schneider and Wagemann's (2012) secondary triviality metric would guard against this mistaken inference, Schneider's conclusion overlooks two important conceptual concerns with his

Table 1. Home Runs by Batter, 1871 to 2015.

	Home run	
Batter	No	Yes
Not Babe Ruth	15,835,542	282,336
Babe Ruth	9,908	714

Source. Data from Sports Reference (2015).

procedure. First, it is troubling that a "consistency" measure can suggest that data are almost completely *consistent* with a deterministic hypothesis—because of skewed case distribution—when they are obviously not.

The second concern is that Schneider and Wagemann (2012) bracket two very different ideas under the label "trivial." The first is standard: one value of *Y* occurs rarely. For example, although oxygen may be a necessary condition for a social revolution, the presence of oxygen carries almost no predictive power for when social revolutions occur because social revolutions happen infrequently. The second idea is non-standard: one value of *X* occurs rarely. The two Babe Ruth examples above represent the second idea because there was only one Babe Ruth. Crucially, in the Babe Ruth examples, the presence or absence of *X* provides *strong* predictive power for the outcome. Therefore, it is not clear why *X* should be considered a trivial necessary condition. STCM scholars have introduced confusion in their causal analysis by conflating disparate concepts under the same label. Alternatively, if one rejects Schneider and Wagemann's (2012) unsatisfying second type of triviality, then the problem remains of false positives caused by skewed case distribution.

Inferring Causal Complexity: Regression Interaction Terms Provide Valuable Insight

The focus on causal complexity—in the sense of multiple conjunctions of causal conditions—is another STCM centerpiece. Thiem et al. (2016) and Schneider (2016) expand on existing arguments by claiming irreconcilable methodological differences, arguing that regression interaction terms simply fail to address casual complexity as they conceptualize it.[3] I counter that both methods can be used to study causally complex relationships. I also question whether distinctive STCM approaches to limited diversity and multichotomous variables truly advance political methodology.

My article provides an extended example demonstrating how, in a binary setting, modeling data with regression interaction terms can provide the same information as an STCM truth table (Paine, 2016, Tables 6 and 7). Hence, Schneider's (2016) contention that I failed to show how regression distinguishes between necessity and sufficiency, or to explain the outcome versus non-outcome, is based on a misunderstanding of my argument. Because a fully saturated regression model can recover all four cells in a complex 2 × 2 table, a scholar can use whatever metric they want to (a) compute necessity or sufficiency scores, (b) analyze the outcome versus non-outcome, or (c) carry out any other procedure that one could perform with STCM tools.

Thiem et al. (2016) are correct that STCM's conjunctions and statistical interactions are different concepts. However, their argument does not preclude the possibility that regression interaction terms can provide insight into causal complexity. To provide a simple example, suppose there are two posited causal conditions, $A_i \in \{0,1\}$ and $B_i \in \{0,1\}$, and an outcome $Y_i \in \{0,1\}$. To assess the working hypothesis that $A_i = 1$ is a necessary condition for $Y_i = 1$, a scholar should estimate the regression model in Equation 1:

$$Y_i = \beta_0 + \beta_A \cdot A_i + \beta_B \cdot \beta_i + \beta_{AB} \cdot A_i \cdot B_i + \varepsilon_i \qquad (1)$$

If $Y_i = 1$ indeed *never* occurs empirically when $A_i = 0$, then estimating Equation 1 with ordinary least squares will produce $\hat{\beta}_0 = \hat{\beta}_B = 0$. These coefficient estimates—whatever the frequency of $Y_i = 1$ when $A_i = 1$ (which determines $\hat{\beta}_A$ and $\hat{\beta}_{AB}$)—imply $E[Y_i \mid A_i = 0] = 0$. Furthermore, fitting the model with heteroskedastic-robust standard errors will estimate $\text{Var}(\varepsilon_i \mid A_i = 0) = 0$. These estimates correspond perfectly with a necessary condition hypothesis.[4]

A similar exercise could be performed for any necessary or sufficient condition hypothesis with any number of binary variables. Despite true differences between conjunctions and interactions, they provide the *same* insights into the social world in a binary setting.

In two other regards, the results yielded by STCM are distinct from that of regression, but STCM's methodological value-added is questionable. First, Schneider (2016) mentions limited diversity. One useful consideration is that quantitative scholars frequently, albeit implicitly, focus on limited diversity when evaluating multicollinearity in their data—a topic covered in any introductory econometrics textbook. If there are few cases with certain values on a higher order term, then including both the higher and lower order terms in the same regression model will introduce considerable multicollinearity (or, in some cases, perfect multicollinearity, making it impossible to estimate the higher order terms). STCM scholars have commendably devoted considerable attention to limited diversity. However, it may often be preferable to avoid estimating coefficients based on data characterized by limited diversity, rather than imposing very strong and often unverifiable assumptions about "logical remainders," that is, the combinations of conditions not observed empirically.

Finally, whereas STCM is less distinctive than claimed with binary variables, the value-added of its distinctiveness with multichotomous variables remains to be established. Although regression does not perfectly mimic STCM when using fuzzy, multi-valued, or generalized sets (Schneider, 2016), this observation does not necessarily support using STCM. Brady (2013) expresses concern that necessity and sufficiency become less meaningful concepts when one moves beyond binary variables. In addition, concerns arise about aggregating fuzzy sets to evaluate necessity and sufficiency (Dunning, 2013) and using unreliable calibration procedures (Krogslund, Choi, & Poertner, 2015). Finally, the strong STCM claim to be an alternative to statistical methods is called into question by the crucial importance of using conventional statistics to assess whether a particular STCM finding arose due to chance (Braumoeller, 2015).

Conclusion: Qualitative Methodologists Should Turn From STCM to Other Tools

These arguments support the core contention that STCM is less distinctive than claimed vis-à-vis quantitative methods. Furthermore, many novel STCM techniques produce problematic inferences, which—in addition to the points emphasized above—include the low standards for inferring deterministic causation. Paine (2016) and Schneider (2016) agree on this latter point.

Furthermore, Munck (2016) is correct in arguing that these critiques of STCM do not reject qualitative methods more broadly. The field of qualitative methodology would be better served by focusing on tools such as process tracing that do possess distinctive advantages relative to quantitative methods—specifically, using knowledge about the causal process to draw causal inferences. This focus is far preferable to perpetuating claims about STCM's distinctiveness that yield an unproductive methodological divide.

Declaration of Conflicting Interests

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

Funding

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

Notes

1. Throughout, this comment presumes there is a convincing reason that empirical counterexamples are not sufficient to eliminate a deterministic hypothesis, a central but not altogether convincing STCM tenet discussed in Paine (2016).
2. Paine (2016, p. 7) provides citations.
3. Paine (2016, p. 14) provides additional citations.
4. Paine (2016, pp. 14-20, Appendix B) elaborates upon the flexibility of regression for modeling causal complexity. In addition, my definition of necessity differs from that in Clark, Gilligan, and Golder (2006)—whom Thiem et al. (2016) critique—because I conceptualize necessary conditions as *limiting* cases of probabilistic relationships (see Paine, 2016, Equations 13 and 14) rather than following Clark et al. and devaluing necessary conditions by equating them with any probabilistic relationship.

References

Brady, H. E. (2013). Do two research cultures imply two scientific paradigms? *Comparative Political Studies*, *46*, 252-265.

Braumoeller, B. F. (2015). Guarding against false positives in qualitative comparative analysis. *Political Analysis*, *23*, 471-487.

Clark, W. R., Gilligan, M. J., & Golder, M. (2006). A simple multivariate test for asymmetric hypotheses. *Political Analysis*, *14*, 311-331.

Dunning, T. (2013, August 29-September 1). *Contributions of fuzzy-set/qualitative comparative analysis: Some questions and misgivings*. Presented at the 109th annual convention of the American Political Science Association, Chicago, IL.

Krogslund, C., Choi, D., & Poertner, M. (2015). Fuzzy sets on shaky ground: Parameter sensitivity and confirmation bias in fsQCA. *Political Analysis*, *23*, 21-41.

Munck, G. L. (2016). Assessing set-theoretic comparative methods: A tool for qualitative comparativists? *Comparative Political Studies*, *49*, 775-780.

Paine, J. (2016). Set-theoretic comparative methods: Less distinctive than claimed. *Comparative Political Studies*, *49*, 703-741.

Schneider, C. Q. (2016). Real differences and overlooked similarities. *Comparative Political Studies, 49*, 781-792.

Schneider, C. Q., & Wagemann, C. (2012). *Set-theoretic methods for the social sciences: A guide to qualitative comparative analysis*. Cambridge, UK: Cambridge University Press.

Sports Reference. (2015). *Baseball statistics and history*. Available from http://www.baseball-reference.com

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

Author Biography

Jack Paine is a postdoctoral associate at the Wallis Institute of Political Economy at the University of Rochester. His primary research agenda uses game theoretic models to understand how authoritarian survival considerations affect prospects for civil war onset, focusing particularly on oil wealth.