

European Political Science
Accepted for Publication (Dec 2017)

WHAT DID I LEAVE OUT? OMITTED VARIABLES IN REGRESSION AND QUALITATIVE COMPARATIVE ANALYSIS

Claudio M. Radaelli^a and Claudius Wagemann^b

^aDepartment of Politics, Centre for European Governance, University of Exeter, Exeter, UK

Email: C.Radaelli@ex.ac.uk

^bDepartment of Political Science, Goethe University, Frankfurt am Main, Germany

Email: wagemann@soz.uni-frankfurt.de

Abstract

Social scientists often face a fundamental problem: Did I leave something causally important out of my explanation? How do I diagnose this? Where do I look for solutions to this problem? We build bridges between regression models and qualitative comparative analysis (QCA) by comparing diagnostics and solutions to the problem of omitted variables and conditions. We then discuss various approaches and tackle the theoretical issues around causality which must be addressed before attending to technical fixes. In the conclusions we reflect on the bridges built between the two traditions and draw more general lessons about the logic of social science research.

Keywords

causality · econometrics · qualitative comparative analysis · research design regression analysis

Introduction

Social science methods have developed rapidly over the recent years, as have the methods controversies. One of the most sharply discussed additions is certainly Charles C. Ragin's Qualitative Comparative Analysis (QCA) (Ragin 1987; 2000; 2008; Schneider and Wagemann 2012). Since its existence, QCA has triggered critiques about technical matters but also about its very *raison d'être* (Collier 2014; Munck 2016; Paine 2016a; Paine 2016b; Seawright 2005). Yet, this has not led to a decline in the use of the method (Rihoux et al. 2013; Wagemann et al. 2016: 2533).

While, for the skeptics, QCA is 'just too poor' to meet the standards of the classic rules of inference and explanation, for others QCA is 'just another beast' and cannot be related to the mainstream traditions (Schneider and Wagemann 2012). Real conversations and comparisons between QCA and other methodological traditions are therefore still rare (exceptions include Fiss et al. 2013; Grofman and Schneider 2009; Schneider 2016; Seawright 2005; Thiem et al. 2016; Vis 2012). With this contribution, we strengthen the conversation between QCA as a set-theoretic approach and the well-established tradition of quantitative methods within which we focus on regression models.

Instead of raising issues about fundamental differences, we stick to the claim of pluralism in methodology and keep the conversation going at a relatively low level of abstraction, focusing on a specific problem. Practically, we look at a model specification problem that occurs in both regression and QCA, namely the omitted variables problem. Only then do we reflect on what the comparison tells us about the logic of inquiry. We find that regression and QCA are genuinely different, yet they can still be engaged in conversation (in the spirit of Maggetti et al. 2013).

Note that for us, building a bridge does not mean erasing diversity and suggesting a mishmash solution (here we follow Moses and Knutzen 2012: 313). It means generating more awareness around how two different ways of thinking can work in practice, on the shop-floor of the social scientist. Building a bridge simply means that there are two different sides that can be connected, and we are mindful of what connects them and how to go from one side to the other.

What do we look for at the shop-floor level, then? Our central focus is the common problem of having forgotten something important in causal explanation. In regression models, as mentioned, this is called the ‘omitted variables problem’.¹ In one of the most well-known textbooks on research design we are told that we have to systematically look for omitted variables because “if relevant variables are omitted, our ability to estimate causal inferences correctly is limited” (King et al. 1994: 175).

If we leave aside the connotations evoked by the language of ‘omitted variables’, we soon realize that the problem is more general. It doesn’t matter whether a researcher is working with a regression model or QCA: to leave out important explanatory factors seriously flaws causal analysis. In the context of our comparison, the omitted variables problem is still understudied, as opposed to the question of robustness for example, for which a broad literature comparing QCA and econometrics exists (Hug 2013; Schneider and Wagemann 2012: 285ff.; Skaaning 2011) – only Seawright (2005: 16ff.) has devoted a limited number of words to this topic.

¹ In QCA, the term of ‘omitted variables’ is not used, also because it is inappropriate to speak about ‘variables’ (see fn. 2). However, for reasons of readability, we use this expression not only for our discussion on regression, but also when we deal with QCA.

Although the language is different and for very good reasons (variables on one side, conditions and sets on the other)² the basic thrust of the argument is the same. However, and interestingly, the ways in which the problem appears and the solutions are developed differ markedly. These divisions shed light on conceptions of causality, the logic of inquiry, and research design. Yet, we find that the primacy of embedding the causal model in theories is the same.

The regression view of the cathedral³

Let us begin with an example: is regulation an obstacle to growth? Decades of attempts to de-regulate the economy seem to suggest that the proposition that there is a causal relationship between regulation and growth is popular among international organizations and governments (Schrefler and Radaelli 2011). This causal claim has roots in the argument that regulation and de-regulation in goods and labour markets have macro-economic effects (Blanchard and Giavazzi 2003). The specification of a model that can be tested is tricky: is it the total level of regulation that affects the rate of growth, or the quality of rules (Jalilian et al. 2007; Radaelli and De Francesco 2007)? Does the number of rules reach a threshold beyond which its effects on growth are definitively negative? And how do we measure regulation across countries?

² This is more than just a linguistic matter of taste. Set-theoretic methods, such as QCA, do not look at variables but model various degrees of case membership in sets. The difference between the two understandings does not so much result in different values of the variable or the set membership, respectively, but lies in the perspective: variables look at a characteristics, while sets group cases according to a given case property. Sets, therefore, are very closely linked to concepts.

³ We paraphrase a famous article by Calabresi and Melamed (1972) and show how the perspective changes dramatically depending on the analytical lens adopted.

In regression designs, omitted variables can be described as a mistake in model specification: an issue of confounders. A model can be wrongly specified in two ways: one way is a functional form misspecification (e.g. the model describing the relationship between regulation and growth is non-linear but we estimate a linear model), the other way is misspecification of the variables. We focus on the latter. To be sure, functional misspecification generates severe errors, but if the only problem that arises is detecting the functional form, this means that we have the variables and data we need. Therefore, we possess all the ingredients needed to find the functional form that fits the data, along with the possibility of the assistance of the F test for joint exclusion restrictions (Woolridge 2008: 242).

Turning to variables rather than form, the first diagnostic question is whether we suspect the model to be incorrectly specified because we either have omitted relevant variables in the explanation of growth, or if this is due to the inclusion of irrelevant variables (over-specification of the model). In general, irrelevant variables do not affect the un-biasedness of ordinary least squares estimators, although they can have undesirable effects on the variance of the estimators, and thus the efficiency of the estimation. As a result, standard errors are needlessly large (Dougherty 2011: 250f.). Instead, if the model omits a variable that is actually relevant in the 'true' model (under-specification) we generate biased estimators. When there is a single included variable and one omitted variable we can at least infer the direction of the bias (Dougherty 2011). What is more, we also risk finding a flawed or at least not sufficiently good or powerful enough explanation. To use our previous example: even if we establish a correlation between regulation and growth, we could rightly suspect that there are other variables that affect growth.

As banal as it may sound: if we do not insert possible causes into our model, we will not be able to assess their potential contribution to the explanation. As a result, the practice has

evolved to building models from general to specific, that is, to start with a rich set of variables and then to progress to a more specific model by successively imposing restrictions after testing their validity (Dougherty 2011: 460). This is because the cost of including irrelevant variables is perceived to be smaller than the cost of omitting variables. However, this may lead to very large and fundamentally a-theoretical models. We will expand on this later, when dealing with control variables.

In general, the amount of variance explained by the model (regulation and economic growth in our example) and the size of the coefficients indicate that something may be wrong. We can also turn to the broad features of the results obtained via the model that we may have under-fitted, such as the sign of the estimated coefficients in comparison to the prior expectations – this should ring alarm bells. The plot of the residuals may reveal specification errors, hence the good practice of plotting the residuals of the fitted model. More specifically, the main diagnostic devices for whether variables should or should not be included revolve around testing hypotheses for the coefficients, using tests such as the Wald test, the likelihood ratio test and the Lagrange multiplier test. Specification tests and diagnostics vary depending on the type of model used – see Greene (2003, chapter 21) for specification tests that are appropriate for discrete choice models. Nevertheless, if the model is not correct as a result of overlooking an important variable, the test statistics are generally invalidated (Dougherty 2011: 254).

One preliminary and fundamental diagnostic question is whether we suspect a problem of measurement: do we have the correct variable in mind but are unable to measure it for different reasons? If so, we do not really have a problem of confounders at the level of research design (we follow up in the next section with examples). An alternative question to ask is: Is the problem of confounders of a more radical nature, affecting our causal theory? Indeed, variables are often omitted because of poor research design.

Dougherty (2011: 250) explains how this takes us into a catch-22 dilemma: “If we know exactly which explanatory variables ought to be included in the equation when we undertake regression analysis, our task is limited to calculating estimates of their coefficients (...). In practice, however, we can never be sure that we have specified the equation correctly”. In short: if only we had the perfect theory in the first place! But unfortunately, we do not. Thus, issues related to the theory that lead us to specify one model or another cannot be solved with technical recipes. For everything else, however, we can and should mitigate the problem with technical solutions.

Proxies and instruments

Turning our attention to technical solutions, we must first consider proxies and instrumental variables. Let us start with proxies. Imagine we do not have an accurate measure of a variable – but we are confident that the variable (regulation in our example) ought to be in the model of growth. This can be a problem of sheer practical necessity. For example, even in a G7 economy like the UK, only few departments have counted the regulations they manage and enforce. In such circumstances, we can reason that the total number of pages in the official federal-national registers is a good proxy for the number of total regulations. Further, the World Bank has produced cross-country doing business indicators that cover the regulatory regimes of common business operations, such as hiring and firing workers, getting bank credit, protecting equity investors, registering property, enforcing contracts in courts, and starting and closing a business (Djankov et al. 2006). These indicators may be considered a proxy measure for regulation we have in mind. We are not saying that regulation and number of pages in official repositories of laws (or measures of ease of doing business) are the same thing. They *are* not. All we need to say is that there are decent arguments to consider number

of pages or doing business as a proxy of the measure of regulation a macroeconomic model explaining economic growth. This is a sort of plug-in approach, because we plug in or substitute one variable for another.

However, we do not always find proxy variables. The proxies may not exist, or they may not be available in our conceptual toolbox or yet again their operationalization may be contested. In this case, a possible solution is to turn to instrumental variables. Historically, instrumental variables have been used to fix bias from measurement error and to overcome endogeneity problems (Angrist and Krueger 2001: 72). In the last twenty years instrumental variables have been used “to overcome omitted variable problems in estimates of causal relationships” (ibid.). Technically, an instrument is (a) exogenous, thus uncorrelated with the error term of the equation, and (b) correlated with the endogenous explanatory variable. Imagine that, in our model, we want to estimate the causal effect of regulation on growth but there is correlation between regulation and the error term. This may occur because we left out a variable that also correlates with regulation. This omitted factor may be, for example, ‘privatization’. If we find that privatization is correlated with regulation and does not show correlation with the error term we can use data on privatization as instrument and proceed with unbiased estimation.⁴

The question is whether we are turning to instruments exclusively because a variable included in the model is correlated with the error term and therefore we need a fix, or the instrument also makes sense in how we believe causality works – so for example we believe that privatization only influences growth by regulatory channels. The definition of instrumental variables is technical: we justify an instrument if it satisfies properties (a) and

⁴ An instrumental variable has first and foremost a technical definition, contrary to the conceptual approach we followed when dealing with the proxy.

(b) above. This can lead to absurdities in causal analysis. Imagine an instrumental variable that satisfies the properties above but cannot be justified by the theory that has informed our model. We can find a number of variables that technically satisfy the properties and therefore are instruments, but only a subset of these technically derived instruments are also theoretically justified.

Here again, accurate diagnostics is an important step. If we think there is a measurement problem but in order to solve it we create an additional research design problem, the solution is worse than the original state of play. Instruments have been widely used to counter bias from measurement errors in systems of simultaneous equations (Angrist and Krueger 2001: 71). But we should be cautious in their application to omitted variables, considering carefully the risk of the slippery slope: how far do we go in arguing conceptually and theoretically that a variable is an instrument for another variable? What is the conceptual threshold beyond which we no longer consider that ‘regulation explains growth given instrument Z’? If Z is remotely attributable to an alternative instrument W from a conceptual point of view but has ‘nice statistical properties’, do we resist the temptation to use Z as an instrument? In the end, this is a question belonging to research design, and more pertinently to our conceptual thresholds for considering a statistical relationship a causal explanation.

To sum up then, our preliminary conclusions are: if we use proxies, we need a theory informing us about the conceptual boundaries (overlaps, differences, etc.) between the omitted variable and the proxy. If we use instrumental variables, we do not want to end up with a purely technical solution looking in vain for its theoretical grounding.

Control variables

In this section, we expand on the strategy mentioned previously, namely the general-to-specific approach in arriving at the final set of variables. The idea is simple: as we are not sure about the correct model, we start with a fairly general model (Dougherty 2011) and insert a good deal of control variables – using our example, this would mean variables different from regulation, such as taxes, the debt-GDP ratio or the quality of the labour force. This practice is impeccable if we have good theoretical reasons to insert variables that represent alternative rival hypotheses. The questionable practice however is to use control variables as just a way of ending up with bloated specifications of the models, with 10, even 20 control variables.

The classic reasoning for the inclusion of control variables is inherently another catch-22 situation. The assumption is that we would include the correct set of control variables. But this would mean that we already have a robust idea of what the correct model is. We have established rival alternative hypotheses and have come to conclusion that they should be tested. This does not mean that we have to measure all the variables, but that we should know what they are – this way we can answer the question of whether their inclusion increases the bias or not (Clarke 2005: 346, footnote 7). Of course, this level of omniscience does not feature in real-world research, where we have first-order approximations of the true model and we can only decide whether to add a subset of the set of potentially omitted variables.

Thus, unless we know the true and complete specification of the model (and there is no diagnostic, conceptual, theoretical tool for this step) there is no justification for which control variables should be included. This is because including some but not all variables may or may not decrease the bias. This rather depends on the correlations between the included and

excluded variables, the variance of all variables, and the effects of the included and excluded variables (Clarke 2005: 350).

There are also issues with the efficiency of the estimation: the inclusion of control variables does not necessarily decrease the variance of the coefficient(s) of interest. The myth that including some known variables of an un-specified set would be a good strategy is gently debunked with the following words (Clarke 2005: 341): [the unspoken belief seems to be that] “a researcher cannot know all 20 variables that appear in the data-generating process, but if she knows and includes 12 of them, she is better off than if she knows and includes only 10 of them. Unfortunately, there is nothing in the mathematics of regression analysis that supports this conclusion”.

The upshot of this discussion leads authors like Clarke to a fundamental suggestion: let us be cautious with the logic of control variables, and draw more on the logic of research design. Thus, notions of theory and design keep coming back. Either omitted variable bias *is* a problem of a poor research design; or theory helps us to identify proxies; or it is the lack of theory which makes us sceptical about instrumental variables; or we need theory to identify the largest possible subset of control variables which help us detect omitted variable bias.

All these scenarios mean going back to theory in search of a better specification, finding contexts close to those of natural experiments, and controlling for effects that we have not measured by improving on sample stratification. Starr (2005) and Clarke (2005) add the practical suggestion of testing broad theories in focused, controlled circumstances, because under narrow circumstances – what is also known as middle-range theories (Merton 1957) in other approaches – we reduce the number of potential control variables. Most and Starr (1989, see also Starr 2005) take up this reference to research design with the expression of

‘domain specific laws’, i.e., laws that are not universally valid, but are true under certain conditions or within certain domains.

Careful design helps, but Clarke himself acknowledges that there is no amount of theorizing that can provide us with the exact vector of variables to include in a specification. There are local variables on which general theories are silent, yet these variables might determine the effects that we are interested in (Clarke 2009: 62). Results are inevitably vulnerable to omitted variable bias – although sensitivity analysis improves our confidence (details in Clarke, 2009).

Another view of the cathedral

QCA is known for combining sufficient and necessary conditions into configurative formulas. Like in regression, we have to include those conditions into the model which have the potential to be causally relevant.

This becomes even more important as QCA works as a causes-of-effects design. Thus, the identification of a broad (in order not to use the illusionary word ‘complete’) set of at least potential causes beforehand is crucial for the validity of QCA results. Indeed, omitting exactly the condition that would create a difference between otherwise similar cases (and thus create a new configuration) bears the risk of reaching the wrong conclusions.

With regard to the technicalities, a first question is whether there can be an equivalent to an over-specified model in QCA, so whether it is possible that irrelevant (i.e., superfluous) conditions find their way into the explanation. The answer is relatively simple: if in our example of growth a condition like privatization is neither sufficient nor necessary, nor is it an INUS or a SUIN condition, then it will not appear as (part of) the solution in the analysis of necessity or sufficiency. However, in research practice, it seems to be highly unlikely that

an irrelevant condition (as any condition) does not form a part of the solution in one way or another, and if this is the case because it is causally linked to a relevant, but omitted condition, similar to the discussion on instrumental variables above.

Indeed, this reminds us of a situation in regression whereby correlation is not causation.

Theory helps to differentiate causes from correlations. Causal mechanisms assist in this task (Bennett and Checkel 2015: 3f.; Gerring 2014). Indeed, set relations alone do not constitute causal relations (Schneider and Wagemann 2016: 317).

Turning from the inclusion of irrelevant conditions to the omission of relevant ones, we identify three ways to detect omitted conditions. However, unlike with regression, there is no standard literature in QCA which would identify a systematic and generally agreed set of diagnostic rules. Therefore, what we say below is exploration of a novel territory.

Distribution of raw consistency values

Our first suggestion concerns the assessment of sufficiency in a truth table analysis and, connected to this, the frequency of contradictory truth table rows and the distribution of raw consistency values.

Remember that a basic step in QCA is when individual truth table rows are assessed whether they are sufficient for the outcome or not. A truth table row represents a configuration of conditions. If, e.g., the three potentially explanatory factors A, B, and C are used in an analysis, truth table rows can take the form ABC, or AB~C,⁵ or ~A~BC, etc. In a crisp set analysis (csQCA) where the conditions and the outcome are dichotomies, the sufficiency assessment of each of these combinations is straightforward: if, whenever the combination of

⁵ The tilde indicates the negation of a condition.

conditions is observed, the outcome is also observed, then the combination is defined sufficient. The (quite common) situation is that some deviant cases show the same combination of conditions, but not the outcome. Imagine that most cases which share a given combination of explanatory factors for growth have high growth values, while a limited number of cases do not. This situation has been called a ‘contradictory cases’ scenario (Ragin 1987: 113ff.).

‘Contradiction’ means that such a scenario contradicts the conclusion that the configuration of conditions is fully sufficient for the outcome: although there is a great deal of evidence in favour of it, sufficiency cannot be deterministically confirmed. But ‘contradiction’ also means that there are still contradictions in the data which have to be solved in some way. In other words, the conditions of which the configurations (and thus truth table rows) are composed do not differentiate the empirical cases well enough to account for differences in the outcomes, such as different levels of growth. This is a clear indication that conditions which would have been useful, if not indispensable, for the explanation have been omitted. Therefore, one of the earliest recommendations was to search for additional explanatory factors (Ragin 1987: 113 – “omitted causal variables” are even literally mentioned here!). Banal as this may sound, this is of course the gold standard of a research process which we have already discussed for regression: if the set of variables which was chosen in a first moment does not explain an outcome such as growth, then we need more and other variables. In fsQCA, fuzzy sets are used. This means that cases can also have partial membership in sets. Again, truth table rows have to be assessed for their sufficiency. However, sufficiency cannot be defined any more straightforwardly. For perfect sufficiency, every case must confirm the inequality relation $X \leq Y$, with X being the condition and Y the outcome. The consistency measure indicates how deviant the empirical situation is from the ideal situation where all cases fulfil the sufficiency requirement. It does not only take into account the

number of cases which contradict sufficiency, but also how strongly they contradict it and how relevant the cases are. It is an illusion to expect many truth table rows with consistency values of 1 in an fsQCA.

If the researcher is ‘lucky’ (but this is not necessarily a question of luck, but rather of a strong design), then at least some truth table rows show consistency values of 1 and some others at least values close to 1 (and the XY plot looks also fine). Additionally, in an ideal situation, all remaining truth table rows show clearly inferior consistency values so that a final decision about sufficiency assessments is facilitated. This means that a dominance of only high and low raw consistency values indicates that the chosen conditions differentiate well with the different outcome levels.

In consequence, a situation is alarming when raw consistency values of a mid-level are prevalent, e.g. a predominance of values of 0.7x (and hardly any higher – or lower – ones). This means that there are neither clear indications for a decision in favour of sufficiency nor against it. The researcher has chosen conditions which do not differentiate and thus do not potentially explain, while relevant conditions have been left out.

The situation becomes even worse when researchers follow blindly or misinterpret recommendations about the treatment of raw consistency values. Schneider and Wagemann (2012: 279) write that “[i]n general, consistency levels (well) above 0.75 are advisable.” This does not mean that automatically all those rows that show consistency values around the indicated 0.75 can be qualified as sufficient conditions. Such an interpretation does not only bear a superficial reading of the recommendation, but it also overlooks that a *dominant* presence of values of 0.75 (and close to this) indicates serious problems with regards to the conditions not differentiating the outcome well enough. If, therefore, we follow too

superficially any rule of thumb, then we might actually take decisions about sufficiency while there is an ‘omitted variable’ problem in the air.

Although evidence is just anecdotal (systematic studies on this phenomenon are still lacking, but see Wagemann et al. 2016), the suspicion is that this constellation of many mid-level consistency values mainly appears when large numbers of cases are analysed. (Remember that QCA is often presented as an ideal method for mid-sized projects, see Ragin 2000: 25; Schneider and Wagemann 2012: 12). This is not surprising. Large numbers of cases possess more potential for variation in the outcome. An explanation, based only on a small number of conditions might simply not provide the necessary explanatory leverage in order to deal with large numbers of cases, and the rather simplistic explanatory model would not be suitable when accounting for the outcome.

The final XY plot and the parameters of fit

After the discussion of raw consistency values (which are an intermediate step in the analysis), the second suggestion for diagnostics concerns the final result of a QCA. In the analysis of sufficiency, an XY plot is usually produced which locates all cases with regards to their fuzzy values in the explanation (usually a combination of factors) and the outcome (such as growth, in our example). Superficially speaking,⁶ the upper right part of the plot contains those cases for which an explanation has been found, and the lower left part shows the irrelevant cases (Schneider and Rohlfing 2013: 579; Schneider and Wagemann 2012: 308). These two types of cases do not represent a problem of or an indication for an omitted

⁶ The superficiality consists in the fact that our description simplifies actual XY plots a bit. There can also be inconsistent cases in the upper right and lower left corner, namely those below the diagonal. However, for the argument made here, this is not important.

variables phenomenon, and if the XY plot is predominantly filled with these types of cases, the analysis does not suffer from omitted variables.

However, this is different for the other two areas of the XY plot. The lower right part contains the true contradictions. These are those cases for which the result would postulate the presence of the outcome (since the cases exhibit the presumably sufficient condition), but the outcome is absent. Above, we have referred to contradictions within single truth table rows. Cases in the lower right part of the final XY plot are instances of the same phenomenon, but this time not with regard to truth table rows, but at the macro level of the final explanation. If such a contradictory case appears, then there must be an explanatory factor that distinguishes it from other cases which instead confirms the statement of sufficiency and is located in the upper right part of the plot (see also Schneider and Wagemann 2012: 309).

One possibility is then to add this new condition to the explanatory model and to run another QCA, including the condition in question. While the advantage of this is to find a coherent explanation for more (if not all) cases, this can harm the analysis in other respects (e.g., decreasing the coverage value, or leading to more limited diversity). Another possibility is a case-wise comparison of just these two cases, namely the ‘typical case’ and the true logical contradiction (as also Schneider and Rohlfing 2013: 583 suggest). While the second procedure leaves the other cases (and thus the other part of the explanation) untouched, an expansion of the analysis to the other cases often does not even seem advisable, if the condition is not suspected to play any causal role for them.

Finally, there is the upper left part of the plot. Cases which are located here share the characteristic that they have high outcome values, meaning that, in our example, they are all be good examples of ‘growth’, while they have only low values in the explanation. In other

words: they are not explained. This is also expressed through rather low solution coverage values. However, different from the true contradictions, the problem for these ‘deviant cases coverage’ (Schneider and Rohlfing 2013) is not that the conditions do not work as postulated by the research result, but rather that the wrong conditions have been chosen. We need to alter our model in order to cover larger numbers of cases, including more or different conditions which were previously omitted.

Characteristics of the results

A third indication that conditions have been omitted is rather aesthetic. This point is also linked to the final result and is less formal than the other two.

Due to equifinality, QCA usually produces more than one explanatory path of sufficient conditions. Imagine a situation where a considerably low number of cases is analysed, but a high number of paths results which comes very close to or even equals the number of cases.⁷ This is similar to the explanation of every positive outcome case through its very existence, i.e., (nearly) every case of substantial growth is explained by a different combination of conditions. While this might be a correct conclusion given the data, such a result makes the analysis useless, since it does not go beyond a mere case description. This phenomenon is indirectly also linked to the issue of omitted conditions aka variables.⁸ While the problem previously was that the conditions we had chosen did not differentiate enough, in this scenario they do not provide any possibility to identify commonalities among the cases. This is certainly also a problem of case selection. But it also indicates that there might be a

⁷ Since the extreme situation is already that every case has its ‘own’ explanation, there cannot be more explanatory paths than cases, unless we allow for values of ‘unique coverage’ being zero.

⁸ Another source for this phenomenon might be the generally low number of cases.

condition which we have not included in our model, but which would have enabled us to minimise previous results in a way that results in a more parsimonious and more useful explanation. It is even plausible to imagine that such a new condition would substitute more than one of the conditions which we had used before.

The problem just described also becomes visible if the single explanatory paths contain many, if not all conditions, i.e., if the paths describe very long terms. Also in this situation, there have not been enough communalities for the researcher to arrive at more parsimonious solutions, and there might be yet another condition which has been overlooked and which might substitute some of the other conditions.

Will we ever know? Discussion and conclusions

Building bridges

On the one hand, reflecting on shop-floor problems (that is, the concrete problems of researchers at work) assists us in building bridges. On the other, our contribution also raises the attention towards subtle yet fundamental differences. Let us first look at the bridges.

A first bridge consists of the role of theory. True, there are technical proposals on how to discover omitted variable problems, and there are ways to remedy this. And yet...apart from instrumental variables which are defined because of their formal characteristics, all other technical solutions reach a limit when it comes to the question *which* variables and conditions are proxies, the correct control variables, or additional QCA conditions. Even in the case of the instrumental variables, the fact that they are defined through their formal characteristics raises some uneasiness. Thus, *both* approaches teach us the importance of theoretical reasoning when identifying explanatory models.

This brings us to a second point: when can we consider a result satisfactory? Imagine that a QCA analysis results in a nice XY plot which does not show any true contradictions or unexplained cases, and which is characterised by high solution consistency and coverage values. In such a scenario, it is a common understanding that the outcome is considered ‘explained’. However, this does not liberate us from the doubt that there could yet be any better explanatory factors that – either additionally, or substituting some of them – would improve the explanation (what more, do we look at them if they improve ‘slightly’ and what does ‘slightly’ mean?).

If such factors are then substantively very similar to those we have already used, then this is hardly a problem of having omitted something; there is simply a better version of what we have already done. The puzzle however starts, if we see that the inclusion of a new variable (or condition) A, for which we remove another variable B from the model, improves the research result, but A and B do *not* show any substantial overlaps.

There is a lesson about humility here: we can never be sure to have found and included all relevant variables. Even if we arrive at an (illusionary) perfect explanation, we do not know if other variables or conditions that we have not included would have also worked out and would have provided us with yet another explanation for the same phenomenon. Above all for y-centered research, this means that our results can always only be preliminary, until somebody else arrives with stronger results.

This second ‘bridge’ is about different expectations. Bluntly stated, QCA wants to explain more than regression. While an explained variance of, say, 25% is not per se so bad for regression analysis, a QCA with a solution coverage of 0.25 would be quickly rejected by most reviewers.

These different expectations are rooted in the linear *versus* configurative logic; the symmetric *versus* asymmetric patterns; the probabilistic *versus* the close-to-deterministic logic; and the effects-of-causes *versus* the causes-of-effects perspective. With regard to the latter point, we have argued that omitting explanatory factors is important for both scenarios, but a bit more in the former than in the latter. What is more, QCA is (still) mainly used for small and mid-sized N situations, and regression for large N research. Scenarios with less cases (i.e., less empirical material) can be expected to be more coherent, to show more regularities and commonalities, to be influenced by less coincidental ‘noise’, and to require more determinism, since cases are more likely to be known and to be narrated in detail.⁹ In general, omissions seem a more relevant issue for methods such as QCA than for others. In consequence, if we consider theory fundamental for dealing with omissions, theory plays a notably strong role in a method such as QCA.

One more difference shines through. Conventionally, set-theoretic scholars argue that their explanations are not about the net effect of individual variables of an outcome. It is rather the configuration of conditions which is considered (Rihoux and Ragin 2009; Schneider and Wagemann 2012: 78), different from statistical approaches where the individual variable (omitted or not) matters a lot. We do not know much about which conditions are more important than others in QCA – for an exception, see Goertz (2006).

And to conclude...

This article has compared two research traditions through the lens of a common problem encountered in identifying the ‘cathedral’ of causality. Our contribution is novel because

⁹ One may ask at this point if it is then justified to lower expectations for large N QCAs and whether a low-expectation large N QCA is really desirable.

there has been almost nothing on omitted conditions in the QCA view of the cathedral. And it is the first time that regression models and QCA have been compared on this dimension of empirical research.

Our lens allows us to shed light also on the nature of explanation. A conversation across the two views is possible, provided that both acknowledge the existence of different ontologies. Then the conversation can usefully proceed in distinguishing issues of omitted variables and conditions that pertain to theory and design, and those that belong to measurement or other technical issues. This distinction is fundamental. And there is hierarchy: no measurement fix can substitute for poor thinking at the level of design. Our comparison has also shown that regression analysis has its shared techniques and ways to fix the problem – there is debate, but there are also common beliefs on what can be done practically. If anything, the problems belong to the abuse of certain ‘fixes’ as evidenced by our examination of instrumental variables.

By contrast, the discussion of omitted variables in QCA has just begun, partly because the ontological presuppositions are unique to this method. But partly this is the state of play we would expect in a relatively new field. This is where future research is mostly needed.

The reality of researchers at work is rather shaped by a continuous move between research ideas and evidence (Ragin 1994: 76). Usually, a researcher starts from an observation or a doubt, reformulates this in one or several hypotheses, and in doing so s/he draws on guiding assumptions, hunches, suspicions, and puzzles. Already the very choice of the research object, i.e. the phenomenon to be explained, suggests some possibilities of how the phenomenon has come about or why the phenomenon is the way it is. In other words: apart from very rare instances of purely hypothesis-generating research (which might occur when the field is really new), a first step in every causally oriented research process is to think

about which variables or conditions to choose. In conclusion, the best check for whether researchers have included the correct variables is still the use of theory, previous knowledge, the relevant literature, a mindful attitude, and plausible imagination.

Acknowledgements

We would like to thank the reviewers. We also thank Fabrizio Gilardi, Jonathan C. Kamkhaji and Gabriel Katz and the participants to the ‘cake for comments’ seminar series 2016-2017 at the University of Exeter. We are grateful to Francesca Farmer for checking style and grammar. Research for this project was funded by the ERC project 694632 Protego – Procedural Tools for Effective Governance and by the Goethe University of Frankfurt.

References

- Angrist, J. D. and Krueger, A. B. (2001) ‘Instrumental variables and the search for identification: From supply and demand to natural experiments’, *Journal of Economic Perspectives* 15(4): 69-85.
- Bennett, A. and Checkel, J.T. (2015) ‘Process Tracing: From Philosophical Roots to Best Practices’, in A. Bennett and J.T. Checkel (eds.) *Process Tracing: From Metaphor to Analytical Tool*, Cambridge: Cambridge University Press, pp. 3-38.
- Blanchard, O. and F. Giavazzi (2003) ‘Macroeconomic effects of regulation and deregulation in goods and labor markets’ *Quarterly Journal of Economics*, 118(3): 879-907.
- Blatter, J., Langer, P.C. and Wagemann, C. (2017) *Qualitative Methoden in der Politikwissenschaft*, Wiesbaden: Springer.
- Calabresi, G. and Melamed, D. (1972) ‘Property rules, liability rules and inalienability: One view of the cathedral’, *Harvard Law Review*, 85(6):1089-1128.

- Clarke, K.A. (2005) 'The phantom menace: Omitted variable bias in econometric research', *Conflict Management and Peace Science* 22(4): 341-352.
- Clarke, K.A. (2009) 'Return of the phantom menace: Omitted variable bias in political research', *Conflict Management and Peace Science* 26(1): 46-66.
- Collier, D. (2014) 'Comment: QCA should set aside the algorithms', *Sociological Methodology* 44(1): 122-126.
- Djankov, S., C. McLiesh, and R.M. Ramalho (2006) 'Regulation and growth', *Economic Letters*, 92: 395-401.
- Dougherty, C. (2011) *Introduction to Econometrics, Fourth Edition*, Oxford: Oxford University Press.
- Fiss, P., Sharapov, D., and Cronqvist, L. (2013) 'Opposites attract? Opportunities and challenges for integrating large-N QCA and econometric analysis', *Political Research Quarterly* 66(1): 191-197.
- Gerring, J. (2014) 'Causal mechanisms: Yes, but...', *Comparative Political Studies* 43(11): 1499-1526.
- Goertz, G. (2006) 'Assessing the trivialness, relevance, and relative importance of necessary and sufficient conditions in social science', *Studies in Comparative International Development* 41(2): 88-109.
- Greene, W.H. (2003) *Econometric Analysis, Fifth edition*, Upper Saddle River, N.J.: Pearson Education.
- Grofman, B. and Schneider, C.Q. (2009) 'An introduction to crisp set QCA, with a comparison to binary logistic regression', *Political Research Quarterly* 62(4): 662-672.

Hug, S. (2013) 'Qualitative comparative analysis: How inductive use and measurement error lead to problematic inference', *Political Analysis* 21(2): 252-265.

Jalilian, H., C. Kirkpatrick, D. Parker (2007) 'The impact of regulation on economic growth in developing countries: A cross-country analysis', *World Development*, January: 87-103.

King, G., Keohane, R.O. and Verba, S. (1994) *Designing Social Inquiry: Scientific Inference In Qualitative Research*, Princeton: Princeton University Press.

Leightner, J.E. and Inoue, T. (2012) 'Solving the omitted variables problem of regression analysis using the relative vertical position of observations', *Advances in Decision Sciences*. doi:10.1155/2012/728980.

Maggetti, M., Gilardi, F. and Radaelli, C.M. (2013) *Designing Research in the Social Sciences*, Los Angeles et al.: Sage.

Merton, R.K. (1957) 'On sociological theories of the middle range', in R.K. Merton (ed.) *On theoretical sociology. Five essays, old and new*, New York: The Free Press, pp. 39-72.

Moses, J. and Knutsen, T. (2012) *Ways of Knowing: Competing Methodologies in Social and Political Research*, Basingstoke: Palgrave Macmillan.

Most, B.A. and Starr, H. (1989) *Inquiry Logic and International Politics*, Columbia: University of South Carolina Press.

Munck, G.L. (2016) 'Assessing set-theoretic comparative methods: A tool for qualitative comparativists?', *Comparative Political Studies* 49(6): 775-780.

Paine, J. (2016a) 'Set-theoretic comparative methods: Less distinctive than claimed', *Comparative Political Studies* 49(6): 703-741.

Paine, J. (2016b) 'Still searching for the value added: Persistent concerns about set-theoretic comparative methods', *Comparative Political Studies* 49(6): 793-800.

- Radaelli, C.M. and F. De Francesco (2007) *Regulatory Quality in Europe: Concepts, Measures and Policy Processes*, Manchester: MUP.
- Ragin, C.C. (1987) *The Comparative Method*, Berkeley: The University of California Press.
- Ragin, C.C. (1994) *Constructing Social Research: The Unity and Diversity of Method*, Thousand Oaks: Pine Forge Press.
- Ragin, C.C. (2000) *Fuzzy-Set Social Science*, Chicago: Chicago University Press.
- Ragin, C.C. (2008) *Redesigning Social Inquiry: Fuzzy Sets and Beyond*, Chicago: University of Chicago Press.
- Rihoux, B., Álamos-Concha, P., Bol, D., Marx, A., and Rezsöhazy, I. (2013) 'From niche to mainstream? A comprehensive mapping of QCA applications in journal articles from 1984 to 2011', *Political Research Quarterly* 66(1): 175-184.
- Rihoux, B., and C.C. Ragin, eds. (2009) *Configurational Comparative Methods: Qualitative Comparative Analysis and Related Techniques*, Thousand Oaks: SAGE.
- Schneider, C.Q. (2016) 'Real differences and overlooked similarities. Set-methods in comparative perspective', *Comparative Political Studies* 49(6): 781-792.
- Schneider, C.Q. and Rohlfing, I. (2013) 'Combining QCA and process tracing in set-theoretic multi-method research', *Sociological Methods & Research* 42(4): 559-597.
- Schneider, C.Q. and Wagemann, C. (2012) *Set-Theoretic Methods for the Social Sciences*, Cambridge: Cambridge University Press.
- Schneider, C.Q. and Wagemann, C. (2016) 'Assessing ESA on what it is designed for: A reply to Cooper and Glaesser', *Field Methods* 28(3): 316-321.
- Schrefler, L. and C.M. Radaelli (2011) 'Deregulation', in B. Badie, D. Berg-Schlosser and L. Morlino (eds.) *International Encyclopedia of Political Science*, London, Sage.

- Seawright, J. (2005) 'Qualitative comparative analysis vis-à-vis regression', *Studies in Comparative and International Development* 40(1): 3-26.
- Skaaning, S.E. (2011) 'Assessing the robustness of crisp-set and fuzzy-set QCA results', *Sociological Methods & Research* 40(2): 391-408.
- Starr, H. (2005) 'Cumulation from proper specification: Theory, logic, research design, and 'nice' laws', *Conflict Management and Peace Science* 22(4): 353-363.
- Thiem, A., Baumgartner, M., and Bol, D. (2016) 'Still lost in translation! A correction of three misunderstandings between configurational comparativists and regression analysis', *Comparative Political Studies* 49(6): 742-774.
- Vis, B. (2012) 'The comparative advantages of fsQCA and regression analysis for moderately large-N analyses', *Sociological Methods & Research* 41(1): 168-198.
- Wagemann, C. (2017) 'Qualitative Comparative Analysis (QCA) and Set Theory', *Oxford Research Encyclopedia of Politics*, Oxford, Oxford University Press.
- Wagemann, C., Buche, J. and Siewert, M.B. (2016) 'QCA and business research: Work in progress or a consolidated agenda', *Journal of Business Research* 69(7): 2531-2540.
- Woolridge, J.M. (2008) *Introductory Econometrics*, Andover: Cengage Learning.

About the Authors

Claudio M. Radaelli is professor of political science at the University of Exeter where he directs the Centre for European Governance. In 2016 he was awarded the European Research Council grant Protego, Procedural Tools for Effective Governance. He is co-editor, with Claire Dunlop and Philipp Trein, of *Learning in Public Policy: Analysis, Modes and Outcomes*, Palgrave, 2018.

Claudius Wagemann is professor of political science at the Goethe University Frankfurt. He has co-authored, with Carsten Q. Schneider, *Set-Theoretic Methods for the Social Sciences*, Cambridge University Press, 2012, and, more recently, with Joachim Blatter and Phil C. Langer, *Qualitative Methoden in der Politikwissenschaft*, Springer, 2018.