The real dependent variable problem: the limitations of quantitative analysis in comparative policy studies

Abstract *Comparative public policy relies heavily on processing quantitative data, typically done by looking for the relationship between variables or by grouping empirical data into categories. In methodological terms, comparative data commonly deal with nations as policy units, and observations have to be understood as part of an interconnected tissue. In statistical terms, there are problems in the identification of appropriate data, violations of the assumptions are rife, and there are just not enough nations to be able to make comparisons sensibly in these terms.*

The 'dependent variable' problem relates to the definition, operationalisation and measurement of key variables, but the problems of comparison go deeper than that. The real dependent variable problem is whether it makes sense to look for a dependent variable at all. The evidence is always equivocal; a methodology that attempts to bracket off disparate influences cannot be valid; and attempting to apply general principles across different circumstances and conditions is inconsistent with what we know about policy development. Quantitative methods offer ways to sort and systematize information, but they do not provide a basis for generalisation.

Comparative public policy has come to be heavily dependent on quantitative analysis, and in particular on two paradigmatic approaches. One is the comparison and analysis of country-based indicators, taken as the basis of a search for relationships and predictors of outcomes. This approach is widespread in econometrics. It was pioneered in social policy by Harold Wilensky (e.g. Wilensky, 1975) and was subject to extensive criticism in the 1990s, but it is ubiquitous, featuring in many analyses. From national figures, researchers search for correlates of variables as a way of explaining either what determines issues of interest - inequality, welfare spending, family structure and the like - or what effect these issues have. So, for example, we might be told that

- spending on welfare affects family stability (Halla et al, 2015)
- policies on employment, debt and support for long-term unemployment affect household indebtedness (Angle, Heitzmann, 2015)
- countries where there are more women have less corruption (Zhao, Xu, 2015, p 420]
- the economy shrinks when government grows too large (Mitchell, 2005)
- the effectiveness of population policy depends on the age of the first child. (Pierrakos et al, 2014]

The other main approach is based on normative models, offering comparisons between régimes or systems. The empirical basis of this approach generally depends on a summary of

the cumulative effect of a range of subordinate variables. There are problems, obviously enough, in dealing with aggregates, and one of the principal criticisms of régime analysis has been its disregard for the detail (Mabbett, Bolderson, 1999); but ultimately the process is built, like the first, on the relevance and appropriateness of the subordinate variables.

Both these approaches rest, in my view, on shaky foundations. The data they rely on are often patchy and loosely defined, but that is not the central problem - it may still be possible to gain insights from a disparate range of pointers. The difficulty lies in the use of the empirical methods that have come to dominate the field: complex, sophisticated quantitative methods that rely on a series of delicate assumptions. They are subject in practice to serial compromises, where researchers gloss over deficiencies and apply methods of questionable validity. It is not at all clear that we can usefully select appropriate variables to test, find relevant data, run a procedure to determine what influences a dependent variable and come to a general conclusion about what the association really means. There are persistent problems of definition, operationalisation and measurement. The issue has been generally discussed as "the dependent variable problem", (Clasen, Siegel, 2007), but for most of the critics the issue is about which dependent variables are appropriate, and how the variables can best be identified and refined, not whether makes sense to go through this process at all. This paper focuses on the theoretical and methodological problems of the approach.

Methodological problems

The first issue in this kind of work concerns the question of what it is possible to do with national data. One of the key reasons for using national data is practicality - data from different countries are usually collected to reflect national units. Dealing with data at a national level has been called "methodological nationalism" (Yeates, Holden, 2009, p 77), and for some purposes - for example, considering the impact of poverty or global inequality (Milanovic, 2009) - it can mislead. Consider two apparently similar sorts of comparative study. One relates fertility figures to infant mortality figures; the other relates family structure to family policy. The first test makes sense. Using fertility figures at national level is only a convenience; even if there is a national average figure, the nature of the relationship being tested is one which should be expected to hold at different levels of aggregation. That implies that using time-series data or regional data would offer comparable observations in greater depth, and that for practical purposes these can be treated as independent of each other (though that assumption needs to be put to the test). In the second case, however, disaggregation is not possible - 'family policy', if it means anything, is generally exercised at a national level. The point of comparison in comparative policy analysis is usually to examine differential impacts of different policies. When it comes to policy variables, it is usually appropriate to treat each national government as a policy unit - for the purposes of data analysis, a single subject. That is how we end up with studies that treat "Belgium" or "Germany" as equivalent subjects, regardless of the internal differences of policy, their complex federal structures or their very different populations.

Even if the information is capable of disaggregation, it may not be valid to do it. If the relationship under examination is of general application, regardless of the context, then in principle variables can be combined across nations. When a medical study reviews the progress of two hundred subjects in fifty hospitals, it combines the data and treats it like one

large sample. The fact that we wouldn't presume to do this with most cross-national samples in politics, sociology and social policy speaks volumes: it's a recognition that the social, political and institutional context shapes and conditions the data. That may be because the data are defined in terms that are relative to the national context (unemployment rates, crime rates, inequality, growth, gender relationships) or because they mean different things in different places (social assistance, assent to ideologies, democratic participation or access to justice). Wilensky's work on welfare states focused on public expenditure, as have many others since. Castles is critical: "different countries employ quite different mixes of policy instruments to achieve similar policy goals" (Castles, 1994, 349), and using a summary indicator like expenditure disguises that.

The best-known attempt to resolve this issue has been to treat national data as part of a cluster or constellation of policies - a welfare régime, as exemplified by the work of Gøsta Esping- Andersen. (Esping-Andersen, 1990) The identification of categories might seem to overcome one of the issues raised by the dependent variable problem, which is the question of what should be compared: but it is vulnerable to others. The field is littered with "black swans" - exceptions to the generalisations that regime analysis depends on. (Castles, 2010) Australia is not 'liberal', but 'radically redistributive'. (Cass, Freeland 1994) France is not corporatist, and (I would argue) not the kind of hybrid between Bismarck and Beveridge that many French commentators believe, but has a distinctive model in its own right. Several writers have tried to expand the number of models to take into account the bewildering number of exceptions. The attempt is doomed to failure: there are simply too many variations to be fitted into the boxes available. Even within régimes, country-level or system-based data is a poor predictor of the operation of welfare in subordinate spheres. The most obvious problem with classifying systems is that what is true for one field, such as unemployment provision, does not tell us what will happen in another field, such as pensions. (See Castles, 2008) Being "corporatist" or "social democratic" does not tell us what governments will actually do.

It is all too common for published studies to cut corners - to under-specify concepts and terms, to miss out steps in the reasoning, to rely on the computer to provide the answers. Martin Ravallion, formerly the Chief Economist of the World Bank, recommends that to identify predictors of poverty we might use: "a regression model in which a potentially large set of variables is identified as potential explanatory factors ... and one lets them "fight it out" statistically to determine how much each variable matters, controlling for the other variables." (Ravallion, 2016, p.249) That account at least has the virtue of honesty. The common pattern of cross-national data analyses is that there will be lots of factors, lots of potential variables and lots of associations that might be identified. Deductive research depends on the generation and testing of hypotheses, but if the objective is not to prove a particular case - a different kind of scientific solecism - the hypothesis will be refined to fit the data after the work has started. Students who are learning to manipulate this data have to understand the techniques that are applied before they can find out what sorts of hypothesis can be tested. In other words, it becomes an exercise in data mining. If the researchers have the data, and can see that an association looks promising, they will look for it; they will find the data to improve the results; if there are no results, they will not publish them. And the basic problem with data mining is that if you sift and sort enough dross, you're sure to come across some nuggets eventually. Part of the problem is that associations arise by chance, and

the bigger the data set, the more likely that is to be true. The other part of the problem is the tendency to assume that whatever is true ought to be explicable. So when we come up against specific, unexpected results, such as the high correlation of military expenditure with welfare (Wilensky, 1975) or the relationship of social expenditure to a country's distance from Vienna (Barnes, Srivenkatamarana, 1982), we feel we have to invest the findings with meaning.

Statistical problems

Any of the problems mentioned up to this point might be considered statistical, in the sense that they concern the rationale for the analysis of data. But there are also specific problems in developing research designs that analyse data by multivariate processes such as regression analysis, and hope to achieve methodologically valid results. The first problem relates to the character of the data. Various organisations, such as the Luxembourg Income Study, have sought to iron out definitional differences between the variables, so that like is more or less compared with like. The problem runs deeper than definitions. For example, one of Esping-Andersen's key concepts is decommodification, but the difficulties of attributing specific benefits to expenditure in decommodified settings threatens the integrity of the basic classification. In the UK, health care is decommodified: people do not pay for the health care they receive, and more important they do not pay for health insurance either. That means that every UK citizen has, implicitly, the equivalent of health insurance - which has a monetary value, but is not counted in any national statistics. The effect of omitting it is to distort figures on social security, public expenditure and inequality. There is an similar issue relating to the value of housing. Some people have subsidised housing; some have benefits to pay for housing; some have to pay while others do not. In several developing countries, and even in some transitional economies, squatting is widespread; in others, poor people have to pay rents. To make the figures comparable, the World Bank has been considering whether it should impute rents to people who do not actually pay any. (Balcazar et al, 2012)

A second problem relates to the mathematical assumptions concealed in the statistical method. Multivariate analysis is a powerful tool, but it depends in principle on conformity to certain standards. Valid analysis depends procedurally on cleaning up the data, identifying and compensating for multicollinearity, avoiding outliers and dealing plausibly with missing values. (see e.g. Tabachnick, Fidell, 2013) One of the intrinsic weaknesses of the method is the assumption that the data are parametric or can be transformed into something like parametric data. The central limit theorem tells us that the means of independent variables will take on a normal distribution as the numbers increase; the lognormal central limit theorem, less a theorem than a practical insight, that positive numbers will have a positive skew, but that can be compensated for through transformation. Many policy variables are not like this at all. Some are dichotomous. Some are quantified only by looking at monetary value, which produces a very different kind of distribution. The country variables are neither probabilistic nor truly independent: countries lie within spheres of influence, they are influenced by common economic factors, and they imitate each other. When it comes to social variables, they are just as unlikely to fit the bill: there are reasons why sociologists focus on issues like gender, race and class, and a propensity to assume the properties of the normal distribution as samples grow is not usually what they have in mind when they do.

The third problem relates to the number of cases. The pseudo-normality of parametric methods generally relies on getting a largish sample to work with. Ragin writes: "There are very few comparativists who conduct studies of 10 or 20 countries, but many who study 1or 2 or 75 (i.e. enough to permit the use of conventional quantitative methods.)" (Ragin, 1998, p 106.) Maybe that ought to be true, but it isn't. There are not enough countries in comparable circumstances to do this sort of thing convincingly or well. It is commonplace to find work that is based on the OECD, or the European Union, or some selection of countries within them. This kind of work is now widespread in economics, and often forms part of the basic training of comparative policy studies. Recent examples from an influential journal look for associations between

- cohabitation and formal legal status, in 12 countries (Gassen, Perelli-Harris, 2015)
- job creation for low educated workers, in 19 countries (Abrassart, 2015)
- carers' well-being and policies to support them, in 18 countries. (Verbakel, 2014)

The selection of countries for *The Spirit Level*, the authors explain, began with a list of the 50 richest countries in the world. (Wilkinson, Pickett, 2009) Then they exclude countries with populations below three million, on the basis that very small populations could distort the global results. That left them with 23 countries. They show links between income inequality and a range of social problems: illiteracy, obesity, murder and higher prisoner populations. (Wilkinson, Pickett, pp 106, 91, 135 and 148). In several cases, there were data only for some of these countries: so, for example, the link between income inequality and children's experience of conflict is based on 18 countries, social equality and mobility is based on 8 countries, preparedness to recycle rubbish is based on 11 countries. (Wilkinson, Pickett, 2009, pp 139, 160 and 232.) If *The Spirit Level* is persuasive, it is partly because there are good theoretical reasons to believe that inequality, itself a form of disadvantage, produces further disadvantage, and partly because indicators that point repeatedly in the same direction offer some degree of corroboration. However, the quantitative evidence the authors present only reinforces that viewpoint; it cannot be of itself decisive.

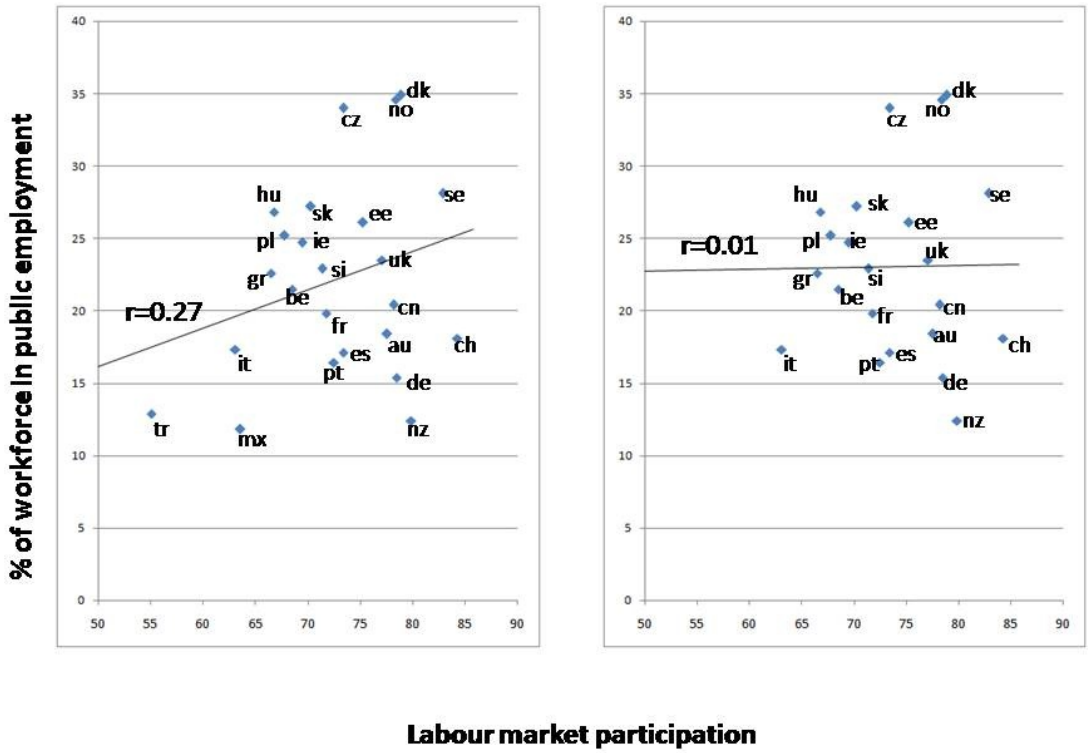
The evidence base is necessarily thin. The OECD currently tracks 33 countries. Now, in most cases where we are dealing with observations about 33 subjects or less, we would usually understand that we have to be very cautious about generalisations.
It may seem more plausible within the limitations of available data to take advantage of complex data sets within a handful of countries, and to process multiple points of data within those countries - for example, using data from a time-series. Typically what happens is that each country is represented by a number of variables and data points. It might look as though we are dealing with 33 countries, the argument runs, but here are 330 subjects for analysis.

There is an obvious problem here: for comparative purposes, observations within a country are not genuinely independent of each other. Where data are hierarchically structured, clustered or grouped, rather than genuinely independent, it is usually appropriate to develop a multilevel research design (Hox, 2010). The use of that kind of analysis is now widespread in comparative social policy, but, Bryan and Jenkins warn, "we believe that many researchers do not appreciate the problems that can arise when the number of countries in a multi-country data set is small." (Bryan, Jenkins, 2015, 2) They calculate that at least 25 countries are

needed for linear models, and 30 for logit models (2015, p 17). Their explanation is technical, but the conceptual obstacles are just as great. Let us imagine that we are talking not about 30 countries or less, but a similar number of human beings, and we decided to put together data from different time periods and different data points. Shifting the subjects from countries to human beings should make it immediately obvious why this is methodologically problematic. In an individual human being, everything is interconnected; the subordinate data about each subject depends on the makeup of that individual. We cannot say for certain, when we are considering information about a country, whether that applies or not. If we are talking about examining inter-relationships within a single country, or in a group of countries, it may well be true that there are multiple, independent observations to consider. We can sensibly examine the associations between social, economic and political phenomena, where there are many subjects and many types of data. At the same time, it might equally be true that we are considering each country as a policy unit, focusing on for example legislation or national policy, and in that case, it is unlikely to be the case that inferential statistics can be used legitimately. It is difficult, then, to offer any general rule as to numbers.

Even if the techniques chosen are valid, the small number of countries that can be referred to makes any statistical process subject to selection bias. We usually can't tell what the effects of missing values might be. Here is an illustration. The graphs in Figure 1.1, drawn from OECD data, are based on the same information. The graph on the left seems to show a positive relationship between labour market participation and the proportion of public employment. That sounds initially plausible: countries which create public employment may well have higher rates of employment overall. Unfortunately, this relationship does not pass a simple test. The graph on the right removes the data for two countries - Turkey and Mexico - and the apparent relationship disappears with them.

Figure 1.1: The apparent relationship depends on the data that are selected



This seems to be a case of spurious correlation; the data for two relatively untypical countries, more recently introduced to the OECD, cannot possibly demonstrate an underlying causal relationship in the data for the other members. At the same time, the data for several other OECD members is missing (it was not listed on the OECD website at the time the graph was drawn): it is not evident that these two countries are outliers, or that their data is a distortion.

Small numbers and missing values are not fatal to statistical processes - only to some of them. There are a couple of non-parametric tests which work well for relatively small numbers of subjects, such as Fisher's exact test - though what that is mainly telling us that the distribution of results has not arisen by chance, and in comparative studies that can usually be taken for granted. The advocates of cluster analysis would argue that the process of grouping and classifying cases descriptively avoids many of the pitfalls of inferential analysis, (Uprichard, 2009) though that has to depend on the specific techniques applied. However, the problem here goes deeper than an inadequate information base. Even in the most scrupulous research reports, it is not often possible for a reader to be sure that researchers have chosen the right tests, and eliminated the sources of bias. Wherever there are omissions, there is unlikely to be enough information given to know whether or not the omission is important. It is in the nature of missing values that no-one can quite know what the missing data might show - it is hardly possible to know what the implications might be. For example, in any multivariate analysis, the residuals should be checked and free of

remaining biases - the randomness of unconnected residuals should mean that they fall out as normal - but that would mean that researchers are being asked routinely to report factors which do not add much by way of insight, and it is still relatively unusual to see the process mentioned in published papers in this field. Every quantitative relationship identified in a multivariate analysis which does not cover all the ground has to be treated as contingent and subject to reservation or revision - and that is true of virtually everything done with these methods.

These arguments are not fatal to conventional statistical analysis, but a degree of humility is in order. The main role of quantitative methods, Castles suggests, is to provide "a preliminary sorting process, informing us about combinations of variables which fit together to produce possible accounts". (Castles, 1998, p 20) Ragin attempts to formalise that sorting process: Qualitative Comparative Analysis reviews constellations of interacting factors, systematizing the permutations of selected variables. (Ragin, 1987) That has its appeal, because it treats the variables as a constellation rather than leaving them in isolation; it has proved to be a particularly effective tool in showing those who are wedded to linear models some of the limitations of their method. (e.g. Ragin, 1994a) However, the approach also has its weaknesses. It relies on the theoretical assumption that the variables, taken together, identify a generative process. This can only be done with a limited number of pre-selected variables: if the range of variables is too varied, there is a combinatorial explosion, and most of the conceivable categories will be empty. In any case, the method is still vulnerable to the methodological objections considered in the previous section. When we compare cases, as Ragin himself argues, we need to accept that variables cannot be considered in isolation, that causes can be heterogeneous and outcomes can be complex. He sums this up in a well-chosen phrase: "Most of these features of case-oriented, qualitative research are anti-statistical." (Ragin, 1994b, p 306.)

The dependent variable problem revisited

The 'dependent variable problem' is a broad term, used to embrace a wide range of studies which have been concerned with what is being examined, and what relationships are being identified - issues like spending, decommodification, retrenchment and so forth. (Green-Pedersen, 2004) Pierson is concerned that any analysis is likely to be reductive:

> "The problem lies partly in the concept of the welfare state itself, partly in data limitations, and partly in limitations of current theorizing about welfare state change. ... As the concept of the welfare state or welfare regime 'stretches', it becomes inevitable that quite distinct processes and outcomes will be joined together under the umbrella of a single master variable. ... In a context where actors have complex motives, and the dependent variable is so heterogeneous, attempts to reduce change to a single dimension will be counterproductive." (Pierson, 2001, pp 420-1)

The dependent variable problem, Clasen and Siegel argue, needs to be understood more broadly still: the term encompasses the conceptualisation, operationalisation and measurement of issues that will make it possible to assess the impact of welfare states. (Clasen, Siegel, 2007) Hudson and Kuhner add, to that list, consideration of the methodological techniques used to analyse the variables. (Hudson, Kuhner, 2010) This is a

field where there is very little agreement about criteria, methodology or interpretation.

Some of the problems of using inferential statistics can be side-stepped. Byrne argues that, while it is true that modelling inevitably involves some simplification, the study of complex systems is likely to be based on consideration of all cases of interest, rather than a probabilistic sample. Rather than working forward from generative influences towards outputs, he argues, it makes much more sense to work backwards - to go from what we know about effects, interpreting those effects in the context of a specific, complex policy system. Byrne recognises explicitly, however, that the nature of complex systems means that it is not possible to generalise such findings across a range of cases. Even if it was possible to agree on the underlying concepts, the criteria for selection, the definition of terms, the process of data collection and the basis of the evidence, it would not necessarily follow that the conclusion was generalisable, because none of those conditions could establish whether the agreed findings would be applicable in other circumstances. There would be reason to doubt the outcomes of a comparative analysis even if there were no methodological reservations.

There are three critical limitations in any process which searches for the relationship between variables. The first is evidential: the nature of the available evidence cannot justify generalisations about generative relationships. Associations might indicate a relationship between variables, but they can also point to relationships between those variables and a common source; and multivariate analysis is often predicated on the supposition that there is a common, underlying factor that helps to explain the constellation of factors that occur together. In comparative policy studies, none of that may be true. The policies of different countries are influenced by geographical proximity, culture, language, policy transfer and diffusion; the determination of the pattern of influence on a country may be shared, or its experience may be unique, and there is no way of being sure which is true. There may well be an association between welfare spending and the marriage rate (Halla et al, 2013), or the length and frequency of breastfeeding and the willingness of government to legislate for breastfeeding breaks (Heymann et al, 2013) - but how can we say which is the cause, and which is effect? Wherever there is a complex, multi-faceted set of phenomena, it is almost impossible to distinguish generative mechanisms, or to demonstrate genuinely effective processes.

The second key problem is methodological. We have come to rely on quantitative analysis to winnow out the linking strands in a complex set of data. It can be done, but in a complex environment it cannot be done with any certainty that the procedure is capable of replication in different contexts. Pawson and Tilley make a devastating critique of the process of policy evaluation. The standard technique is to identify a variable of interest and to seek to determine how that variable changes under different conditions. This is done by isolating, or bracketing off, the influences - a range of independent variables - from other potential influences. A control trial works by examining the effect of changes in those influences, and taking other factors to be equal. The fundamental problem with this process, they argue, is that we cannot understand the impact of policy without considering its interplay with other factors - the same factors which are being bracketed off. (Pawson, Tilley, 1997, p 52) Their critique applies even more powerfully in the context of comparative policy analysis. In comparative policy, we are typically dealing not with a specific, isolated variables, but a skein of interconnected issues. The effect of policies in different countries cannot be

understood without appreciating the setting in which they are being introduced, the economic and social conditions which apply, and the way that those policies relate to others in the same place. Byrne argues that each of these sets should be treated as a 'case', that each case is effectively a complex system in its own right, and that what matters in a case is not the relationship of isolated variables, but the way that everything is composed. (Byrne, 2009)

This points to the third, outstanding problem, which is theoretical. As comparative studies have developed, it has become more obvious that no two systems are truly alike. Howlett and Cashore point to a 'new orthodoxy' in comparative analysis of policy development:

> "First, there is widespread acceptance that any analysis of policy development must be historical in nature and cover periods of years or even decades or more. Second, it has generally been agreed that political institutions and their embedded policy subsystems act as the primary mechanisms of policy reproduction. Third, paradigmatic change, a process in which there is a fundamental realignment of most aspects of policy development, is generally understood to occur only when the policy institutions themselves are transformed." (Howlett, Cashore, 2009)

Implementation, understood as the process of delivering policy in practice, is inherently dependent on context. Studies of policy outcomes fare little better. Even where there is direct policy transfer, the 'same' policies - structural adjustment, Conditional Cash Transfers, Essential Health Packages, or whatever - are designed and implemented differently, and they mean different things in different contexts. 'What works' in one place is difficult enough to assess; what works in several nations is even less determinate.

There are no absolute rules to be complied with: studies might be thought of as better or worse, according to the extent to which they take into account of the many reservations. At the same time, many of those reservations are serious and fundamental. No formal method of analysis can manage the depth, range and variety of the social, institutional, economic and political settings where policy is made. Statistical methods depend on reduction, and reduction drains the value of much that is done in comparative social policy. It can be helpful to set out differences in a systematic way, but if the key influences are unique and distinctive, the process of implementation has to be understood in the place where it is operated and the outcomes follow from policy development and implementation, it makes little sense to examine the issues by generalising across conditions. Comparative studies work best, not when they look for common elements, but when they set the distinct cases side by side, using each example to drawn insights from the other about those differences. And that is why, ultimately, the broad generalisations that come from quantitative analysis are worth far less than qualitative, in-depth considerations of the political, economic and social context of a country.

References

Abrassart, A., 2015, Low skill jobs or jobs for low skill workers? *Journal of European Social Policy* 25(2) pp.225 –241
Angle, S., Heitzmann, K.,2015, Over-indebtedness in Europe, *Journal of European Social Policy* 25(3) pp.331 –351

Balcazar, C., Ceriani, L., Olivieri, S., Ranzani, M., 2012, *Rent imputation for welfare measurement* (Washington:World Bank Group).
Barnes, J., Srivenkatamarana, T., 1982, Ideology and the welfare state, *Social Service Review* 56(2), pp. 230-246.
Bryan, M., Jenkins, S., 2015, Multilevel modelling of country effects: a cautionary tale, European Sociological Review, doi: 10.1093/esr/jcv059
Byrne, D., 2009, Complex realist and configurational approaches to cases, in: Ragin, C., Byrne D. (Eds), *The Sage handbook of case-based methods* (London: Sage).
Byrne, D., 2011, What is an effect? Coming at causality backwards, in Williams, NM., Vogt, W. (eds) The Sage handbook of innovation in social research methods, London: Sage.
Cass, B., Freeland, J., 1994 Social Security and Full Employment in Australia, in: Hills, J., Ditch, J., Glennerster., H., (Eds.) *Beveridge and social security* (Oxford: Clarendon Press).
Castles, F., 1994, Is expenditure enough? On the nature of the dependent variable in comparative public policy analysis, *The Journal of Commonwealth & Comparative Politics* 32(3), pp. 349-363.
Castles, F., 1998, *Comparative public policy* (Cheltenham: Elgar).
Castles, F., 2008, What welfare states do, *Journal of Social Policy*, 38(1) pp. 45–62
Castles, F., 2010, Black swans and elephants on the move, *Journal of European Social Policy*, 20(2) pp. 91-101
Clasen, J., Siegel, N., 2007, Comparative welfare state analysis and the 'dependent variable problem', in: Clasen, J., Siegel, N., (Eds.) *Investigating welfare state change* (Cheltenham: Edward Elgar).
Esping-Andersen, G., 1990, *Three worlds of welfare capitalism* (Brighton: Polity).
Gassen, N., Perelli-Harris, B., 2015, The increase in cohabitation and the role of union status in family policies, *Journal of European Social Policy* 25(4) pp. 431-449
Green-Pedersen, C., 2004, The Dependent Variable Problem within the Study of Welfare State Retrenchment, *Journal of Comparative Policy Analysis* 6(1), pp 3-14
Halla, M., Lackner, J., Scharler, J., 2013, *Does the welfare state destroy the family?*, (Vienna: Forschungsinstitut zur Zukunft der Arbeit), http://ftp.iza.org/dp7210.pdf
Heymann, J., Raub, A., Earle, A., 2013, Breastfeeding policy: a globally comparative analysis, *Bulletin of the World Health Organization*,91 pp 398-406.
Howlett, M., Cashore, B., , 2009, The Dependent Variable Problem in the Study of Policy Change, *Journal of Comparative Policy Analysis*, 11(1), pp 33-46
Hox, J., 2010, Multilevel analysis, London: Routledge.
Hudson, J., Kuhner, S., 2010, Beyond the dependent variable problem, *Social Policy and Society* 9(2) pp. 167-79.
Mabbett, D., Bolderson, H., 1999, Theories and methods in comparative social policy, in J Clasen (ed) Comparative *social policy: concepts, theories and methods*, Oxford: Blackwell.
Milanovic, B., 2009, Globalisation and inequality, ch 2.5 in: Yeates, N., Holden C.(Eds), *The global social policy reader* (Bristol: Policy Press).
Mitchell, D., 2005, *The impact of government spending on economic growth,* (Washington DC: Heritage Foundation).
Pawson, R., Tilley, N., 1997, *Realistic evaluation*, (London: Sage).
Pierrakos, G., Balourdos, D., Soulis, S., Sarris, M., Pateras, , J., Skolarikos, P., Farfaras, A., 2014, Comparative analysis and evaluation of the effectiveness of demographic policies in EU countries (2009-2010*). World Health and Population*. 15(1) pp. 31-43
Pierson, P., 2001, *The new politics of the welfare state*, (Oxford: Oxford University Press).

Ragin, C., 1987, *The comparative method*, (Berkeley: University of California Press).
Ragin, C., 1994a, A qualitative comparative analysis of pensions systems, ch 13 in: Janoski, T., Hicks, A, *The comparative political economy of the welfare state*, (Cambridge: Cambridge University Press).
Ragin, C., 1994b, Qualitative comparative analysis, ch 12 of Janoski, T., Hicks, A., *The comparative political economy of the welfare state,* (Cambridge: Cambridge University Press).
Ragin, C., 1998, The logic of Qualitative Comparative Analysis*, International Review of Social History* 43 Supplement pp 105-24.
Ravallion, M., 2016, *The economics of poverty*, Oxford: Oxford University Press.
Tabachnick, B., Fidell, L., 2013, *Using multivariate statistics*, (Boston: Pearson).
Verbakel, E., 2014, Informal caregiving and well-being in Europe, *Journal of European Social Policy* 24(5) pp. 424-441
Wilensky, H., 1975, *The welfare state and equality*, (Berkeley: University of California Press).
Wilkinson, R., Pickett, K., 2009, *The spirit level*, (London: Allen Lane).
Uprichard, E., 2009, Introducing cluster analysis, in Byrne,D., Ragin C. (Eds) 2009, *The Sage handbook of case-based methods* (London: Sage).
Yeates, N., Holden, C., 2009, *The Global Social Policy Reader,* (Bristol: Policy Press).
Zhao, X., Xu,H., 2015, E-Government and corruption*, International Journal of Public Administration* 38 pp. 410-421