political science, such as policy studies and historical institutionalism, provide a new and self-conscious exploration of epistemology and methodology with an overall increase in rigorous qualitative approaches.

analy-
orary
op the
phy.

from David Marsh and Gerry Stoker, eds.,

Theory and Methods in Political Science,

3rd edition (Basingstoke: Palgrave Macmillan, 2010)

and-
eds)

uali-
tative research

- Grbich (2007) provides an account of key ... nd a practical guide to a range of research approaches.

Chapter 13

Quantitative Methods

PETER JOHN

In spite of many valiant attempts to integrate quantitative and qualitative methods, the divide between the two remains. Many researchers still tend to use one approach, but not the other. Not only is the divide personal, it often sorts out researchers into topics of study. As a result, many academics assume that quantitative investigation only concerns elections, voting systems, party manifestoes and political attitudes rather than having a more general application. The division becomes manifest in the descriptors researchers apply to themselves and to others: quantitative researchers are known as political scientists; the rest often have the labels of students of politics, area specialists, biographers and public policy specialists. Not only do different topics, skills and networks help create the divide; it is sustained by apparently clashing conceptions of the purpose and practice of social science. Some qualitative researchers think that quantitative work is underpinned by a crude version of positivism. Instead, qualitative work describes complex realities, acknowledges that researchers cannot separate their values from the political world, engages with and seeks to understand the beliefs and aspirations of those who are being researched and rejects the idea that there are universal rules of human behaviour. In this context, a review of quantitative methods cannot just be a description of the different techniques on offer. Such an account would reinforce the idea that quantitative research is a set of techniques rather than a practice. Instead, this chapter aims to persuade sceptics of the depth and subtlety of quantitative work. For much of the debate about quantitative and qualitative research is shallow and rests on stereotypes of the research process.

The argument develops that of King, Keohane and Verba in their influential book, *Designing Social Inquiry* (1994). Writing with the tools of quantitative analysis in mind, they argue that both fields apply a 'unified logic of inference with differences mainly of style and specific technique' (1994: 1). They recommend qualitative inferences could be improved by the adoption of a few straightforward procedures. Whilst the book should be compulsory reading for every research student, experienced researchers often feel uncomfortable with the clean and tidy nature of their programme, which seems to squeeze out the messy, problem-solving and practical way in which most qualitative researchers actually do the job. Often, investigators respect their hunches; they discover bits of new data accidentally, they carry out detective work and follow up leads. Sometimes they start with the wrong research question and, after many blind alleys, come to a moment of revelation. It is sometimes quite late in the project that the student or even experienced academic knows how to frame the research problem and can test alternative hypotheses. This chapter claims that quantitative researchers also engage in unplanned and unpredictable data analysis; they solve problems incrementally and follow their intuitions just like their qualitative counterparts. They discuss their strategies with their colleagues and seek the advice of others in the research community. The message is that all researchers should design their projects to be capable of testing hypotheses, but they should also use their practical knowledge to carry out exciting and imaginative pieces of work.

Quantitative researchers sometimes help their critics because convention requires them not to report the interpretive aspects to their craft. They report complex statistical analysis as though they had run their data through a black box, making knowledge of the technique a necessary prerequisite to understanding the article. This chapter aims to demystify both the theory and presentation of quantitative research. The idea is not to knock down quantitative work in politics, but to show that much of its practice coheres with the rich traditions of social science. The chapter also reports recently established conventions and rules of scholarly journals that encourage or require political scientists to present as much information as possible about how they gather their data, choose their models and ensure that others can replicate their results (King, 1995). Moreover, in spite of rapid advances in statistical techniques, the use of programming and new software, the leading political methodologists argue that researchers should make further efforts to present their data more effectively so the ordinary reader can understand how the research shows the relationships between the variables of interest (King *et al.*, 2000).

The collection and management of data

Quantitative work rests on the observation and measurement of repeated incidences of a political phenomenon, such as voting for a political party, an allocation of resources by a government agency or citizen attitudes toward taxation and public spending. By observing variables over a large number of cases, it is possible to make inferences about a class of political

behaviour, such as who votes for a political party, which area gets resources from governments and what is the distribution of attitudes to public spending in the adult population. With large numbers, social scientists can confidently make generalizations about the empirical world. Statistical theory shows that the larger the number of cases (or the greater number in proportion to the whole population), the surer data analysts can be that what they observe is not a random occurrence. Moreover, political scientists often want to analyze whole populations, such as the voting record of all Members of Parliament or all electoral systems in the world, which involves large numbers.

Some qualitative researchers are often suspicious about the way in which their quantitative colleagues generate observations, particularly when what they measure is attitudinal or behavioural, such as opinions drawn from large-scale surveys using standardized questions. These measures appear to ignore social and political contexts (Kirk and Miller, 1986). Even official statistics that government departments produce may reflect political decisions about what kinds of data to collect. In the end, official information is what politicians and bureaucrats wish to make public. Some techniques, such as content analysis (the classification and counting of data drawn from the texts of media or political debates), appear to strip out the context of the language and render it meaningless or sterile. Quantitative researchers appear to be blind to the relationship between the observer and observed that makes each act of collecting data unique. Critics claim that quantitative researchers ignore the complexity of the world in their quest to turn politics into a series of repeated and identical experiences or events (Ragin, 2000).

But the practice of quantitative research does not live up this stereotype. Quantitative researchers are acutely aware that complex social realities may not always be captured by repeated observations. In certain situations, quantification is not appropriate as what is being measured could be made either meaningless or biased by ignoring the social construction of the data. There is usually a long discussion as to whether measures are valid or not. For example, research that depends on standardized questions may not be replicated across countries because of differences in culture and language. In the qualitative prelude to most surveys and in pilots, questions are bandied about, interviewers evaluate interviews and respondents fill in an additional questionnaire about their experience of completing questions. For example, survey researchers have frequent discussions about the effect of question wording and order on the responses to their questions. Quantitative researchers also pay a lot of attention to reliability (that data are produced independently of the activity of measurement) and seek to improve it where possible. Content analysis researchers, who seek to extract key terms from documents like newspapers, use inter-coder reliability scores to find out whether two different researchers coded an item in the same way (Krippendorff, 1980: 129–154). Such problems do not just occur in surveys and the analysis of texts. Statisticians who use data from government departments frequently investigate how the data is collected. They consider the possible biases in the results and think of ways to correct for them. There is even discussion about the extent to which research instruments, such as survey questions, reflect biases within social science, such as in favour of class-based explanations in voting behaviour (Catt, 1996: 67–9 , 92).

Quantitative researchers spend much time and effort thinking about their data. Choosing data or sampling appears an easy task but it contains many hidden pitfalls. The sample must allow the investigator to make inferences, but often it is not clear what constitutes the population. If the topic of study is about change over time, which years should the researcher choose to analyze? Surveys pose many dilemmas, such as how to define a household. Surveys may need to be re-weighted because of the stratification of the sample (Skinner *et al.,* 1989). There are also choices about how to measure variables. No perfect set of data exists; for example, response rates to surveys may be low and archives may contain missing years. Although the electronic storage of data gives the impression of permanence, and has massively improved from the expansion of the internet, files sometimes become corrupted and data get lost.

The collection and manipulation of data invite errors. Interviewers, research assistants or survey companies sometimes input responses to questionnaires incorrectly; researchers accidentally delete cases and variables; the transfer of files between software packages and across the internet can create dirt in the data; and researchers can even accidentally work on the wrong or old dataset because they did not label it correctly. They may even forget how they created their variables because they did not note what they did at each stage or failed to save the command file correctly, when using a statistical software package such as a 'Stata do' file. One of the problems is that the speed and efficiency of modern computers encourage researchers to think that their data are clean. But most political scientists learn to be careful after making silly errors when their concentration lapses. As mistakes are so easy to make, researchers spend a large amount of their time carefully collecting data, checking and re-checking how they or the research assistant entered the information. Even with this culture of paranoia, mistakes still occur in published work, sometimes in the best quality journals (see, for example, the correction of Garrett (1998) by King *et al.,* 2000: 356).

Data collection and management requires attention to practical issues and to theory about what is the best data for the study. No solution is ideal, but researchers pick up practical knowledge from their colleagues

and friends about how to solve these problems and learn about the pitfalls of particular choices. A few words in an internet search engine can produce all sorts of useful information, some of it posted by political methodologists. Debates also occur in footnotes and appendices, in listserv discussions and in emails between colleagues; they become part of the common stock of knowledge that members of the research community acquire. These critical activities show that quantitative researchers do the same things as their qualitative colleagues: they use a variety of strategies to find the best data to answer their research questions.

The power of description

One of the advantages of descriptive measures is that they allow the observer to split the observations and to examine the proportions, such as the percentage of a group who support a political party. Judgements about these proportions form an essential part of the interpretation of data. In journalism and other forms of commentary, there are debates about whether a percentage is too big or too small, and descriptive political science is no exception. Consider an imaginary statistic showing that five per cent of the electorate believes in repatriation. Commentators can either interpret it as evidence of alarming racism or of tolerance of the bulk of the population. To resolve this dilemma, social scientists would place the statistic in its proper context, taking into account arguments about what defines a liberal society and existing empirical knowledge. The interpretation of the 5 per cent would differ with the additional information that, for example, 10 per cent of the population believed in it twenty years previously.

Summary statistics are useful to understand the data, such as measures of central points so that researchers can know the average or typical point for a variable. The most common is the mean or average, but there is also the median (middle observation) and mode (the most frequent value). Equally important are measures of dispersion. Observers find it useful to know whether the observations converge on the average value or are widely distributed. For example, if the interest is in response times of fire brigades in different locations, researchers and residents may be interested in finding out which area has the lowest average response time. But they should also be interested in the dispersion around the average as residents would like to know how likely the fire engines will arrive close to the mean time. As with central points, there are a number of measures, such as the inter-quartile range (the distance between the upper and lower quartiles) and the standard deviation (the square root of the variance). When deciding which measure to use, researchers need to think carefully about their data and decide whether they are nominal (with categories that are just different to each other, such as male or female), ordinal (with measures that involve ranking), and ratio/interval (with values that have equal intervals between categories). Investigators may wish to look at the shape of the distribution, such as whether it is unimodal (spread evenly around one point) or bimodal or multimodal (having a number of peaks), which can inform much about what the data reveals. Alternatively the data may be skewed or symmetrical, leptotkurtotic (bunched around the mean) or normal. The normal is particularly interesting because it shows the distribution is random.

When technical terms appear, like the ones in the paragraph above, qualitative researchers and students start to think that quantitative topics are not for them, but often merely formalize what people do in everyday life. Imagine a person walking into a room full of people. The person would immediately size up the gathering by asking how many people are there, how many are of a certain type or how many old or young people there are. When coming to these judgements, people make approximate proportions, averages and distributions. Descriptive statistics standardize these common-sense ideas (or common-sense ideas make sense of the statistics). Moreover, such statistics appear regularly in newspapers and in qualitative research.

Paradoxically, quantitative researchers do not use descriptive statistics enough, often only reporting them as the prelude to applying sophisticated tests and models. But much can be gained by their careful and imaginative use. To obtain the best results, quantitative researchers must first immerse themselves in their data and explore the myriad of possible ways of cutting and representing the information. Familiarity with descriptive measures assists an understanding of the topic and can help researchers interpret the output from more complex statistical models. Much can be gained by representing descriptive data pictorially in the form of bar charts, pies and plots – most software packages easily provide these. In short, quantitative researchers should be as intimate with their research materials as their qualitative colleagues. As Jenkins writes, 'The statistician should fall in love with his data' (cited by Franzosi, 1994: 45).

Tables and inferential statistics

Social scientists often want to infer or deduce models of causation that they wish to test. Such models often hypothesize a strong relationship between two variables (either positive or negative). Social scientists assume that the values of one variable cause or influence variation in another. The explaining terms are called independent variables and what

is being explained is known as the dependent variable. For example, consider a project on what causes people to volunteer, which is an important topic in the literature on social capital (see, for example Verba *et al.*, 1995). Theory – in the form of the socio-economic status (SES) model of political behaviour – may suggest that those from wealthy families are more likely to join organizations. This suggests finding the variables of wealth and social capital to see if the former leads to the latter.

One of the simplest ways to find out if one variable determines or is associated with another is tables or cross-tabulations. Tables show how the values or categories of one variable are expressed as the categories of another. Researchers frequently use tables in survey research. If the volunteering project had been carried out in the days before computers, researchers would have sorted all the cards containing the records of the interviews into the piles of wealthy volunteers, non-wealthy volunteers, wealthy non-volunteers and non-wealthy non-volunteers. Then they would have counted the numbers of cards of each category, worked out their percentages as a proportion of each variable and represented the results in a two by two table. A table is usually titled in the following way: 'Table N: dependent variable by independent variable' or in the example, 'Table N: volunteering by wealth'. If the tables are set up to display column percentages, with totals of 100 per cent at the bottom of the table, the dependent variable (volunteering) is shown in the rows with the columns displaying the independent variable (wealth). The researcher can compare the proportion of the independent variable taken up by the dependent variable, the numbers of wealthy and non-wealthy who volunteer – the eye can look across the table to compare the proportion of volunteers in the wealthy and non-wealthy groups. But it is just as respectable for the independent variables to be the column and to compare row per cents. In the end, the analyst learns to read the table by comparing the amount of influence the independent variable has on the response or dependent term.

Now that the records of surveys can be stored as data matrixes in software packages, such as Stata 11.0 (StataCorp, 2009) or in free open source software, such as R (R Development Core Team, 2008), researchers can create such a table in seconds. But their construction is surprisingly tricky. Often variables need to be re-coded, such as by transforming the individual ages of respondents into bands of age groups. Working out which measure to use requires knowledge of the data and attention to theory to select the appropriate units. There is an art in creating a table that is attractive to look at and is formatted professionally, which is surprisingly difficult to do. Researchers should not paste across the output from a software package like SPSS, as the result is awkward to look at and is often hard to understand. Report or paper writers need to spend time ensuring all the required information is present, such as a clear labelling of the variables and the totals of each column or row. Rounding up the percentages to a number or to one decimal place helps too.

Researchers who use tables from surveys also need to run tests to show that the associations could not have happened just by chance. Because surveys are samples from larger populations, associations in the data could appear because the data have an unusual selection of people. Statisticians conventionally argue that researchers should have ninety-five per cent confidence that the association is not random. The humped shape of the normal distribution indicates that the mean value of the variable in the sample is going to be close to the population mean whereas the chance that it is far from the mean is much less. The ninety-five per cent confidence level is convenient because it is just under two standard deviations (typical deviations) from the mean or average level and also is the point at which the normal distribution becomes flatter. Survey researchers calculate the probability and most computer packages routinely produce a figure. If the figure was 0.04, for example, researchers would believe that the association had not occurred by chance. But the ease with which computers run these tests sometimes makes researchers forget to examine the strength of the associations, which show how much one variable affects another. In large samples, such as those in excess of 4,000 respondents, it is easy to find statistically significant but meaningless or imperceptible relationships.

One common objection to testing hypotheses from using correlations presented in tables and regressions is that they do not establish causation but only show associations. Unlike natural scientists, the claim is that political ones rarely carry out experiments, so they have no way of knowing whether the relationships they observe in their data are accidental, spurious or causal. Theory comes to the aid of the social scientist because a relationship between two variables needs to be logical and consistent as well as following from existing empirical studies. The association between wealth and volunteering is not a correlation found by what is called dredging the data, but derives from sociological theory that argues that as some people have more resources and advantages than others, so they are more able to engage with public life. The relationship is logical in the sense that social background can affect political participation. Logically, it would not be possible for volunteering to affect social background (at least in one generation). Such research can only test whether background affects voluntary activity or not, but not the other way round. It is plausible because investigators compare the SES with other models, such as the rational choice model of participation or models that emphasize contextual factors, such as friendship networks or the neighbourhood. As always, theory, rather than the computer or technique, should specify the direction of the causal arrow, but as long as the independent variable is genuinely thought to be prior to the dependent variables and where each independent variable is independent from each other, it is possible to make an inference.

When researchers appraise hypotheses they are not satisfied with just observing relationships in the data. To support their case they would look for other relationships to make a set of plausible arguments. They might be interested in change over time; they could run multivariate models as indicated below. Just like detectives on a case, quantitative researchers gradually piece together the evidence. At all times they are aware of academic communities of reviewers and conference participants who are likely to be sceptical about the results. They think of the likely criticisms and devise strategies about how to convince the sceptics. Rarely do quantitative researchers claim that an association in the data proves causation, but that correlation has importance only when applied by theory and used alongside other evidence.

Multivariate analysis

The social and political worlds are multi-causal, which makes it hard to identify one specific relationship. For example, there may be no relationship between wealth and volunteering because wealthy volunteers tend to go to schools which encourage voluntary activity. The causal relationship between schooling and volunteering makes the correlation between wealth and volunteering spurious because wealthy people go to a certain type of school which also produces volunteering as well as good examination results. So it would be entirely possible for poor people to have as high a level of volunteering as rich people if they had been to a school that encouraged it. So how is it possible to know how much each one influences the response variable? Sometimes it is possible to overcome this problem by using multiple regression to examine all the determinants. This technique allows a test as to whether other factors, rather than wealth, affect voluntary activity. Researchers do not aim to show that *X* causes *Y*, but that *X* causes *Y* alongside or allowing for or controlling for *Z* or *W*. Analysts become more confident of testing hypotheses because they have allowed for all the possible causes of behaviour or attitudes. They can run one model against another and carry out robust tests of each one. However, multivariate analysis carries more risks than descriptive statistics because the regression models that social scientists commonly use make restrictive assumptions about the data.

The most common multivariate model is ordinary least squares (OLS). The intuitive idea is that a plot of the points between two interval variables, *X* and *Y*, may contain a relationship. If the points are not randomly distributed, it may be possible to plot a line that minimizes the distance between it and the data points. It is then fairly easy to see the relationship in data by moving the eye along the bunch of data points in the scatter plot of X and Y. This line would have a gradient or slope that indicates the constant relationship between the two variables. Rather than eyeballing the data, OLS uses a formula to estimate the slope of the line from the mean or average value of the independent variable and from the data points. In addition, OLS estimates the distances between the regression line and the data points – what are called the residuals or errors. OLS calculates the overall explanatory power of the model offers which is a statistic, the r-square, which falls between 0 and 1. The same mathematics governs models with more than one independent term. This neat extension allows the estimation of the effects of each of the independent terms upon the dependent variable. OLS allows researchers to test hypotheses in the knowledge that they are controlling for all the known hypothesized effects and that these are independent of each other.

Because OLS assumes the data is a sample from the population of possible data points, everything that the model does not explain is random. For each variable there is a standard error or measure of spread that indicates the probability that the relationship between the independent and dependent variable is there by chance or not. Political scientists have been happy to run hypothesis tests based on the 95 per cent confidence level. If the probability is equal to or greater than 95 per cent, researchers accept that an independent variable has an effect on the dependent one; if it is less then they reject the hypothesis that there is a statistically significant relationship. The procedure easily tests models that derive from social science theory. This procedure appears to correspond to the scientific method because investigators allow the variable of interest to pass or fail a test.

The other advantage of OLS is that it is a standard: it is very comprehensible across the profession, indeed most of social science. Most political scientists know how to read an OLS table or output as they can look at a column of coefficients and see if an effect is big or not (either by comparing standardized measures or thinking about the units of measurement). They will also know that they can divide the coefficient by the standard error to create a t-statistic, which they know must exceed 1.96 to meet the standard five per cent probability test. Most tables display stars next to the coefficients, which people often glance at when looking for the statistically significant relationships. The eye is naturally drawn to a star and can conclude that one star is good at 0.05 probability, two stars are better at 0.01, and three stars may even deserve the popping open of some champagne at 0.001! They can also look at the r-square statistic to see if

it is big or not. This knowledge can allow non-technical researchers to know a little about what is going on rather than skipping the middle sections of quantitative papers.

For the bulk of the post-war period the OLS model held sway, particularly as it was taught as the central component of most political science methods courses. Most well-trained political scientists understand its output. In spite of its ease of comprehension, OLS has disadvantages. It is worth recalling that the model depends on ten assumptions that are frequently breached in most contexts. For example, one assumption is that the variables are constant and linear over time and space (Wood, 2000). Also, in many research situations the number of cases is too small, for example with studies that use the developed or OECD countries as the units of analysis.

In recent years, political scientists have moved away from OLS, partly because they know more about the properties of the variables they wish to explain. Some variables may be dichotomous, so requiring a logit or a probit model; other variables may be ordered rather than interval, which would require an ordered logit or probit model; other variables may be censored, with a cut-off point at one or both ends of the distribution, requiring a censored or tobit model; and other data may be count or event data, like wars, requiring a poisson model. Just as OLS was standard fare for a previous generation of political scientists, these different estimators are now part of a familiar menu of choices for today's, easily implemented by commands in most software packages. Most regression output has the same format to OLS, so it is possible to read them in the same way, transposing the R-coefficient to another measure of fit and searching out for the ubiquitous stars. Another change is that statistical theory and its applications have advanced massively in recent years. This means it is possible to estimate relationships with different statistical assumptions (see Box 13.1 on non-parametric estimation).

Although the multiple regression model is the workhorse of empirical political science, it is worth knowing that the structure of causal relationships may be more complex than it often implies. For example, the existence of marginal Westminster seats causes governments to direct public resources to them (Ward and John, 1999), but the receipt of those resources will affect which areas are going to be marginal seats in the following election. Over time, how can a researcher know what level of resources it takes to win marginal seats? This is the problem of endogeneity or selection. It can be partly overcome by more sophisticated use of statistics, such as two-stage models, or more recently, selection models (Heckman, 1979). These models depend on restrictive assumptions, such as finding a perfect variable with which to instrument the data, something that rarely occurs. Structural equation models (SEM) (Schumacker and

Box 13.1 Non-parametric models

Monte Carlo simulation allows the investigator to estimate a variable and to make inferences to the population. It needs vast amounts of computer memory to generate data from an artificially created population that resembles the process being investigated. Then the researcher estimates a statistical model from this population and assesses its performance. Political scientists use bootstrapping models that are similar to Monte Carlo simulation and relax the restrictive assumptions of the OLS model (Mooney and Duval, 1993; Mooney, 1996; Mooney and Krause, 1997), arguing that the OLS model only developed because of the limitations of computational capacity and now the microchip revolution makes other forms of estimation possible. Bootstrapped estimators are available on statistical packages, such as STATA, and articles in journals now appear with reports of both OLS and bootstrapped estimates.

Lomax, 1996; Maruyama, 1998) available in software packages, such as LISREL, MPLUS and AMOS, can estimate more complicated sets of relationships when there are many measures of the same underlying concept. But the analyst should be careful: more complex statistics cannot cover up the difficulty of specifying a causal relationship. Sometimes it is better to be modest in describing the data rather than make too many claims about the direction of causation.

Testing and reporting models

When non-specialists read quantitative articles they may come away with the impression that political scientists only tested a model that derived from theory. But even with a small number of independent variables, there are many choices about which ones to exclude or include in the final model. These choices should be driven by theory, but sometimes theory provides arguments and counter-arguments for a number of models. For example, researchers could include all or some of the independent variables in the final model irrespective of whether they reach the 95 per cent confidence level or not. Alternatively, they could include only those variables that reach the required significance level. Moreover, the number of choices increases if researchers include interaction effects. These are terms created by multiplying two variables to indicate a joint impact on the dependent variable and they may be included along with the original independent terms (Friedrich, 1982). In many situations, it does not matter which model to run as all of them show the same kinds of relationships and levels of probability. But competing models can show the hypothesized variables to be

sometimes significant and sometimes not. Moreover, the profusion of new techniques of estimation means that researchers face many choices over the estimator. Then there are different ways in which the data may be presented, such as whether to have clustered or robust standard errors, which can affect the statistical significance of a variable. Or it can be a multilevel model to take account of the different levels in data, so the individual's behaviour or values is affected by his or her individual chacteristics, but also by the community context in which he or she is located or nested.

Researchers may be tempted to present the model that shows the hypothesized variable to be outside the ninety-five per cent confidence level. With the speed of current computers and the easy manipulation of software packages, modellers can engage in the much despised practice of significance hunting, which involves running many hundreds of equations until the preferred one emerges. Because journal editors cannot require researchers to report every model they run, it is hard to detect this practice. Gerber and Malhotra (2006) show that reported papers in political science journals cluster just over the 0.05 probability level at the same time as they show a gap on the non-significant side of this cut, something that would not be expected in the real world. Basically, political scientists select and present results that meet the 0.05 arbitrary cut-off point and reject models that do not.

The incentive to present the most favourable model exists because few journals publish papers containing negative results. Most journal editors and reviewers find these papers to be less interesting and less publishable than those that reach positive conclusions; alternatively, there is self-selection at work whereby researchers only send off papers to journals when they have positive results. The alternative explanation is that political scientists choose to carry out and research councils usually fund research projects that are likely to yield new findings. In the natural sciences the bias has been studied and is called the file drawer problem (Rotton *et al.*, 1995; Csada *et al.*, 1996; Bradley and Gupta, 1997).

Qualitative researchers may become suspicious that advanced statistics creates a screen behind which the modeller seems to cook the results. When practice breaches the stereotype of the pure model of scientific investigation, the effect is something of a fall from grace. Rather than a devious manipulation of data, the art of building models involves the assessment of different possibilities or pathways, each of which is trailed with theory. Researchers think about what is going on in their models and go back and forth between theory and the results they produce. Along the way is much dialogue – often internal, but also with colleagues along the department corridor or across the internet. Such conversations show that quantitative research is above all problem-centred. Problems and solutions are continually traded amongst the research community to overcome the many pitfalls. And if the entrant to the profession does not know anyone, the many specialist courses on the theory and practice of quantitative political science can fill the gap. These courses are as much about finding out about the hidden knowledge and getting tips on shortcuts as they are about formal tuition and instruction on statistical theory. A folklore of practices emerges and complex networks link together researchers. Researchers engage with their data. They neither test pure models, nor do they dredge for significant results; but they carefully consider each one and come up with plausible explanations of the routes they have chosen.

The dialogue becomes hidden by the time investigators submit their articles to learned journals. But after the publication of research, the discursive aspect to the production of knowledge starts again. Researchers discuss results of papers with varying degrees of scepticism or respect that draws upon their knowledge about their data and about the people involved. Members of the research community often detect cooked models because they cannot understand how researchers arrived at their results. When researchers find they cannot replicate the results of a paper, this knowledge gradually diffuses to affect the reputation of the investigator. The informal control may become formal when one academic questions the findings of another by publishing a comment, to which there may be replies and rejoinders.

A more recent approach has been to call for greater transparency and accountability in the research process. Transparency may be aided by the advance publication of an analysis plan, which is a document where the researcher commits to a scheme of data analysis in advance of getting the data. The researcher emails this plan to a nominated and independent third party. This is designed as a self-denying ordinance to incentivize researchers to keep to what they promised rather than to select good results from their data. It is very hard to keep to these plans, however, as new ideas emerge about how to analyze the data. But they could form a good record about what the researchers intended and how they have departed from their original plans. This information could be posted on the researcher's website, along with the original data and the programme code or Stata do file, so that those who are interested can see how the researchers got their results and allow them to play with the data to generate their own, which would be much less cumbersome than downloading it from a data archive. When authors submit their papers to academic journals, they should also send their original analysis plans, data and command files. They should also compose a note of how they implemented their plans so the journal's reviewers can re-run the data to see if minor changes to the commands change the results or whether there are alternative specifications.

King (1995) has campaigned for a standard of replication, whereby any person may repeat other scholars' work using the same dataset and coding of the variables. Such a standard has now been adopted by the main US journals. The ability to replicate not only guards against the false presentation of data, which is in any case rare, but it ensures that researchers carefully check their data for mistakes. Replication encourages researchers to consider the steps toward the presentations of their final results and to check the health of their models, such as for breaches of the assumptions of OLS. It is good to teach replication workshops to demystify data analysis. I got smiles even out of unconfident students when I showed them that changing the month when the Falklands war was assumed to start alters the Sanders *et al.* (1987) findings that the war did not help prime minister Thatcher win the UK's General Election of 1983. The choice is whether to use the month when hostilities started as opposed to the one where troops were sent to the islands (see Clarke *et al.*, 2000).

As a result, standards of reporting in political science have improved and most articles in good journals convey at least some of the vast range of diagnostic statistics, rather than just r-squares and probability values. This caution is wise as King (1986) shows that the r-square statistic can be misleading, making 'macho' comparisons of its size rather meaningless. For example, the r-square can increase by including more variables in the model rather than because of any real improvement in explanation. Similarly, stepwise regression has now fallen into disuse. Stepwise is a facility on some of the more popular software programmes, such as SPSS, which allows the researchers to select their variables by automatically discounting non-significant terms or including significant ones in each equation.

The current wave of reforms could go further as there is a range of tests that researchers can apply to the interior of their regression models (Franzosi, 1994). For example, it is common that one case in a model can cause a variable to be significant, and researchers need to find out why this is (sometimes it is caused by a data entry error). There are tests of the contribution each case makes to the final model, which help the researcher to understand what is going on inside the model. Moreover, political scientists could consider abandoning some of the shibboleths of their art. The most sacred is the 0.05 and 0.01 significance tests (or 95 and 99 per cent confidence levels) that can lead researchers to reject or accept a hypothesis because the probability exceeds or does not reach the required level only by a small margin. But there is no theoretical reason why these rules should exist. Tanenbaum and Scarbrough (1998:15) remind us they derived from the period before computers automatically calculated the probability values and researchers had to look up the values in printed tables, which had limited space so they summarized cut-off points. Now no one uses tables, so researchers should be forbidden from adding asterisks to the variables in models they publish to indicate that a variable has passed a significance test. They should only report the standard errors and the probability levels and discuss them in the text. Such a practice would not be so satisfying for the researcher, but it would lessen the file drawer problem and lead to a more balanced and nuanced discussion of the research. Psychologists have already conceived of life beyond significance tests (Harlow and Mulaik, 1997) and a discussion has begun in political science (Gill, 1999). Along with analysis plans and the replication standard, such a practice would reduce the incentive for researchers to fiddle their results; but at the same time it would not interfere with the inspirational and creative aspects of quantitative work.

Recent developments

Political methodology is a fast-moving field, which is responding to new statistical theories, new applications (such as new R packages), and developments from econometrics. It is not possible to be just to these in the space here. They are listed with some references or weblinks for enthusiasts to follow up.

- *Visualization of regression findings*: The graphing of predicted values or probabilities from a regression can show in a simple and attractive way the influence of a variable on another controlling for other factors in the regression. Although they were always possible to implement, new software makes this much easier, such as Gary King's 'Clarify' programme, which uses simulation (gking.harvard. edu/clarify/docs/clarify.html).
- *New forms of content analysis*: Software and analytic developments have produced a cottage industry of different programs and methods to collect text-based data. Particularly influential has been *Wordscores*, developed by Ken Benoit, which is suitable for coding left-right scores (wordscores.com/). One that does not have so many assumptions built in is Yoshikoder www.yoshikoder.org/.
- *Advances in panel data analysis:* Panel data is where there are repeated observations of a cross-section, which are particularly useful in political science which wishes to compare changes across countries and other large units. The main impetus and source of innovation is from economics (for example, Arellano, 2003).
- *Increasing use of Bayesian statistics*: Bayesian models use an updating model of human behaviour, which generates a more flexible approach to estimation, which acknowledges the bounded nature of human behaviour (Gill, 2007).

- *Field experimental methods*: There is a growing band of political scientists who have been carrying out real world experiments on topics such as voting behaviour, deliberation, political participation, collective action, and media impacts (see Chapter 15).
- *Natural experiments*: This is where a feature of the political world resembles an experiment, such as an accidental division of the population in two differently treated groups, which have made an appearance in political science (Dunning, 2008).
- *Renewed interest in quasi-experiments*: These seek to look at different features of the data to gain leverage or use a technique called matching to select cases that are very similar to each other bar that one has the causal variable of interest (see review by Sekhon, 2008).
- *Greater attention to interaction models*: With the expansion of visualization of the results of data analysis (see above), there has been greater attention to what interaction models actually show us and improved visualization of the relationships they show (Kam and Franzese, 2007)
- *Spatial models:* Politics varies across space and what happens in one place may affect what happens elsewhere. Spatial econometrics seeks to model these processes. Their application to political science has been modest so far (Darmofal, 2006).

Conclusions

This chapter shows the complexity and subtlety of quantitative work. Far from being mindless number crunchers testing unrealistic models, researchers who use large numbers of observations are acutely aware of the context and character of their data and the assumptions that underlie statistical models. Whether through descriptive statistics, tabulations, OLS or more advanced statistical models, quantitative researchers immerse themselves as much in their data as their qualitative counterparts. Imagination and intuition have their rightful place in the craft of quantitative analysis. Moreover, a highly critical research community exists to appraise and scrutinise the methods that investigators deploy.

In the spirit of a subtle defence, this chapter criticizes some of the practice of quantitative work, such as the tendency to present results too cleanly and to hide much of the messiness of data analysis. More improvements can still be made. A cultural shift would acknowledge the importance of exploratory data analysis. As Tanenbaum and Scarbrough (1998) argue, the revolution in the speed of computers and the ease of using software packages help researchers and students utilize the benefits of exploratory data analysis as they can flexibly handle and present data.

However, the space in journals is a constraint on the possibilities for elaboration. It is also tedious to read articles that recount how the researchers did the research with tales of blind alleys and mistakes (though the internet can help here). But much has already been achieved through the campaign for a replication standard and the new culture of resistance against cookbook data analysis. Rapid advances in statistical techniques, made possible by the speed of today's computers, have transformed the field. Quantitative researchers today now seek to be both more advanced in their methods and more comprehensible to a non-technical audience.

Further reading

- A textbook for beginners in statistics is Wonnacott and Wonnacott (1990).
- *How to Lie With Statistics* (Huff 1991) provides an accessible approach to the topic.
- For introductions to quantitative methods in political science, see Miller's (1995) chapter in the first edition of *Theory and Methods in Political Science*; the classic book by Tufte (1974); and introductions and reviews (for example, Pennings *et al.*, 2006; Burnham *et al.*, 2008 (Chs 5 and 6) and Jackson, 1996).
- More advanced readers could read the volume edited by Scarbrough and Tanenbaum (1998) and the quantitative sections of *The Oxford Handbook on Political Methodology* (2008), edited by Box-Steffensmeier, Brady and Collier.
- Econometrics books are essential once you have got beyond the basics: for example, Gujarati (2003), then Greene (2007).

Chapter 14

The Comparative Method

JONATHAN HOPKIN

The role of comparison in political science is widely misunderstood, probably because of the entrenched use of the term 'comparative politics' to describe research into 'foreign' countries (in the United States, empirical political scientists work in either 'American politics' or 'comparative politics'). Apart from the obvious paradox that a US scholar working on American politics thus becomes a comparativist once she crosses the Atlantic, this definition also misleadingly restricts the domain of comparative political analysis. In fact, comparison of some form is present wherever political scientists make claims about causality, whether they are studying one country, two countries, 192 countries, or indeed cases from some other unit of analysis. This chapter will present an introductory picture of the uses of the comparative method, describe its logic and some of its techniques, assess its strengths and limitations, and discuss the problems involved in designing comparative research.

Theory and the comparative method

Comparison and the comparative method are used implicitly or explicitly across political science and the social sciences in general. Comparison serves several purposes in political analysis. Observation of the ways in which political problems are addressed in different contexts provides valuable opportunities for policy learning and exposure to new ideas and perspectives. Comparison across several cases (usually countries) enables the researcher to assess whether a particular political phenomenon is simply a local issue or a broader trend. But perhaps the principal function of comparison in political science is that of developing, testing and refining theories about causal relationships, and all political research – even purely descriptive narratives – involves causal claims of some kind. The comparative method is 'one of the primary means for establishing social scientific generalizations' (Ragin *et al.*, 1996: 749).

Ironically, a lot of research in the disciplinary subfield of 'comparative