

# 6

## Quantitative Methodologies in Public Policy

*Christian Breunig and John S. Ahlquist*

### Introduction

In 2012, American debt rose to over 16 trillion dollars; Greek government spending reached levels nearly 20 percent lower than 2010; official Spanish unemployment crossed the 25 percent mark; French public spending rose to over 55 percent of the GDP; and income inequality (as measured by the Gini Index) in Italy continued to grow. These estimates illustrate the many quantitative measures, descriptions, and analyses in our daily lives.

The comparative study of public policies also relies on assigning numbers to social, economic, and political phenomena with the goal of examining their relationships.<sup>1</sup> Quantitative analysis provides one avenue for understanding policy variations and their political and institutional causes.

The volume of quantitative work in comparative public policy has grown over the last 50 years. Early studies of public policy were predominantly qualitative in nature (see Simeon, 1976 for an early review) because understanding the entire policy process was seen as the central goal in the public policy literature. From early on, researchers were astutely aware of the complexity of this process. Starting in the 1970s, scholars have broadened their theoretical and empirical inquiries across political entities and time and even across policy domains. This expansion encouraged political scientists to develop and employ methodological tools to summarize, describe, and draw conclusions from large amounts of data. Beginning with the 1990s, we have witnessed a marked sophistication in our methodological tools. Hand in hand with these improvements emerged vigorous discussions about the appropriateness and utility of the newfound tools.

With this brief historical evolution in mind, we introduce three basic goals of quantitative methodology: description, theory testing, and prediction. These three goals are introduced in more detail in the following pages. We illustrate each goal using data and particular quantitative procedures that are frequently employed in the comparative study of the welfare state in advanced democracies. This example is especially valuable here because the welfare state is one of the most researched areas of comparative public policy and one of the most 'quantified' areas of public policy. We specify the relevant data sets when we explore how quantitative methods allow us to describe the welfare state.

Almost all information about the welfare state and its analysis are presented visually. Public policy scholars quantify many political phenomena and collect incredibly detailed data. However, it remains challenging to present this richness in comprehensible and interesting ways. Visualization helps researchers and their consumers to explore, compare, and understand data (see Tukey, 1977; Tufte, 1983). We believe that well-crafted graphical displays allow quantitative scholars to tell their story more convincingly and clearly. This is particularly true for describing outputs from statistical analysis and simulations. Visual displays enable audiences who have no background in the complexities of the model or estimation technique to grasp intuitively the main findings of a research project.

Given the limited scope of this chapter, several themes in methodological debates remain untouched. First, we do not dissect and evaluate differences between qualitative and quantitative methodologies in comparative public policy. Instead, we simply believe that description and inferential leverage are valid and realizable ambitions for both types of research. Second, and related, we do not focus on the many important and difficult decisions that go into an effective research design; we simply assume that the basic components of a successful design, such as comparison and control, are given. Third, this chapter does not touch upon other areas of public policy research that often rely on quantitative methodology. Among them are important fields of public administration research such as evaluation research and cost-benefit analysis. A typical tool of the first area is survey research. Since the 2000s, evaluation research more regularly incorporates clear experimental design. We discuss the role of experiments in the section on inference but do not explicitly refer to evaluation research. The second area, cost-benefit analysis, is often associated with the need to monetize various policy alternatives as well as their potential costs and impacts in order to develop efficient policy choices. It is also outside the realm of this

chapter because comparative public policy research in political science is seldom directed at explaining the effectiveness and efficiency of a particular policy. To put it simply, we focus on quantitative analyses of comparative policy outcomes.

## Measurement and data

Public policy scholars routinely develop concepts that are not directly observable. It is, therefore, a fundamental task to develop procedures that enable us to assign numbers to events or other empirical objects such as institutional structures, control of government, partisan ideology, or the welfare state. This process is called measurement (see, for example, Frankfort-Nachmias and Nachmias, 2008; Jackman, 2008). In this section, we introduce two components of measurement: levels of measurement and assessments of measurement error (that is, validity and reliability). It should be clear that both level and quality of measurement have important consequences for our ability to describe and assess relationships among political phenomena. This section also introduces different types of data structure as well as the data sets employed in the subsequent analysis of welfare states.

A perennial concern of measurement relates to the quality, defined in terms of accuracy and consistency, of the measurement. Measurement error doesn't stem from real differences in the properties of an object but is a consequence of variations in the measurement procedure itself. The two main concerns of measurement are reliability and validity. Reliability describes the extent to which a measurement is consistent under the same procedures. Validity is the idea that researchers are measuring what they intend to measure. In order to validate a measurement, researchers need to supply supporting evidence for its appropriateness. Three types of validity are important here: content validity assesses to what degree the instrument adequately captures all characteristics of a concept; empirical validity captures the degree to which concept and measurement are consistently related. Construct validity indicates to what extent the measurement connects to a broader theoretical framework.

Questions of measurement feature prominently in welfare state studies. In particular, Esping-Andersen's (1990) path-breaking work as well as studies of retrenchment (Pierson, 1996; Korpi and Palme, 2003) highlight how debates about concept formation, operationalization, and measurement can evolve into central concerns of the field. Before Esping-Andersen objected that citizens do not struggle for spending per se (Esping-Andersen, 1990: p. 21), the most common measure of the

welfare state was public social expenditures as a percent of the gross domestic product (GDP). Instead, Esping-Andersen stresses the importance of entitlements for the elderly, sick, and unemployed. His argument recasts the welfare state in terms of social insurance and social services. Similarly, Pierson argues that expenditure measures barely pick up the subtle changes in the welfare state that lead to retrenchment.

The ensuing discussion on the 'dependent variable problem' (see also Green-Pedersen, 2004) highlights the importance of the three types of measurement in quantitative public policy work. First, the retrenchment literature challenges the content validity of the public expenditure measure. In contrast to the narrow focus on total public or social expenditures, programmatic measures that capture the insurance motive of the welfare state were missing or incomplete. Instead of simple government activity on spending, more recent works rely on more concrete individual level benefits (such as unemployment benefit replacement rates) that better capture the decommodifying role of the welfare state. Second, retrenchment scholars contest the empirical validity of the spending measure. They claim that expenditure is a poor instrument for measuring outcomes by showing that restrictions on entitlements are not picked up by the spending variable. They also argue that much of the inclusive evidence with regard to major theoretical expectations, such as partisan ideology or trade openness, at least partly, are due to using government spending as a dependent variable.

For the purpose of this chapter, we simply acknowledge that there is little consensus on how exactly one can capture and operationalize the nature of the welfare state and its scope. We therefore rely on two common measures. First, we use public social expenditures (as percent of GDP) supplied by Busemeyer (2009) as the 'classic' measure of government welfare state effort. Second, we employ the welfare generosity index constructed by Allen and Scruggs as described in Scruggs (2006). The index follows Esping-Andersen and measures the generosity of welfare benefits at the individual level. It combines replacement and coverage rates for three core social insurance programmes: unemployment, sickness, and pensions. The index ranges from 0 (no benefits) to 100 (total replacement of income and coverage). Both measures cover the two competing conceptualizations of the welfare state.

Our data is structured as time-series cross-section (TSCS). This type of data holds repeated observations of a set of units (here countries) over a period of time and is common in comparative public policy. One advantage of pooling data as a TSCS sample is that researchers can learn from both the time as well as the cross-sectional component of the data.

This type of analysis is at the heart of comparative politics and public policy. However, special care needs to be taken with data of this sort because extensive correlation across observations, both longitudinally and in space, can prevail. This is especially true when relying on longer time series for a dozen or two countries. The sample of this study consists of 18 Organization for Economic Co-operation and Development (OECD) member states over 22 years (1980–2002), which results in 396 cases.

TSCS data are essentially a combination of two other types of data structures that public policy scholars regularly encounter. The most basic data is a collection of observations at one point in time. This type of data is called cross-sectional. Typical examples of cross-sectional data are surveys or experiments. Time-series data are measurements of the same unit over time. In addition, many policy scholars are interested in policy diffusion processes. Diffusion scholars are interested in the spatial nature of the data; that is, the units of observation are distributed in space either as regions or point processes (see Chapter 9 in this book). Clearly, some data are combinations of these structures and require special care in accounting for the complex dependencies among observations.

## **Objectives of quantitative analysis in comparative public policy**

Quantitative analysis in political science and public policy concerns three core objectives. First, description enables researchers to illustrate and summarize large amounts of data intuitively and efficiently. Second, causal inference and theory testing allows researchers to state and estimate relationships among variables. It also provides measures of uncertainty for those relationships. The third objective is predicting outcomes based on data and models. At the centre of all three objectives lies the regression model. It is the workhorse for quantitative analysis and is utilized for all three aims.

### **Description**

Both quantitative and qualitative work in comparative public policy engage in describing political phenomena using numbers. Descriptions of single variables usually summarize key properties of the distribution of the interested observations. Typical statistics are measures of central tendencies and dispersion such as means and standard deviations. Similarly, visual techniques such as histograms or box plots can effectively illustrate large amounts of data. For comparative public policy work, line plots that display the development of particular policy outcomes

across time or space are especially important. Description and descriptive statistics summarize data in effective and intuitive ways. Essentially, they are a tool for reducing information in order to gain understanding of the subject at hand.

Figure 6.1 describes the welfare state data. The top row combines all data into a histogram and the bottom row displays country specific developments over time. On average, the countries in our sample spend 21 percent of GDP on public social expenditures. The welfare effort ranges from 9 percent of GDP in 1980 Greece to 36 percent in 1993 Sweden; 50 percent of all observations lie between 17 and 26 percent. The generosity index, which is the product of replacement rate and coverage rate summed over three basic social insurance programmes, is

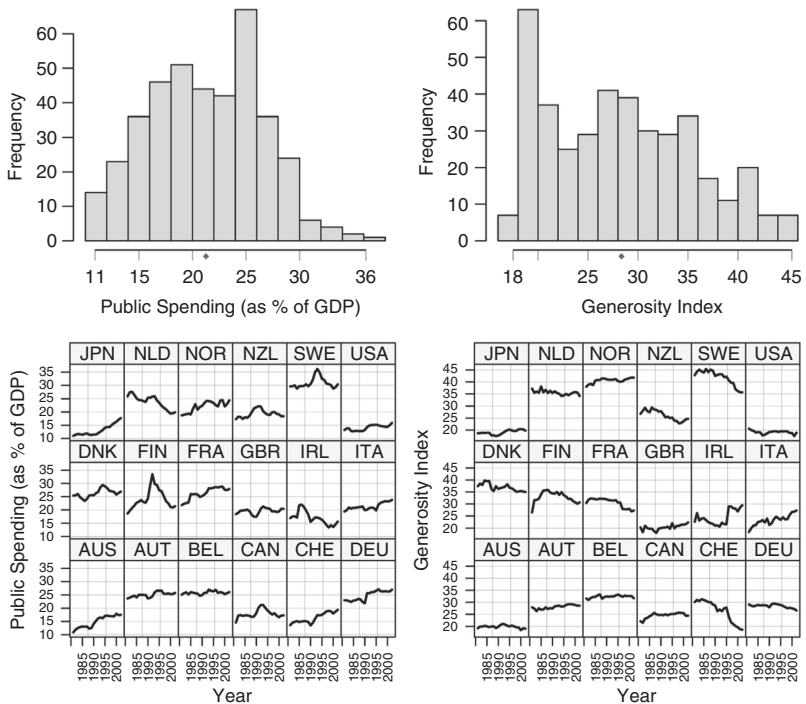


Figure 6.1 Description of the welfare state

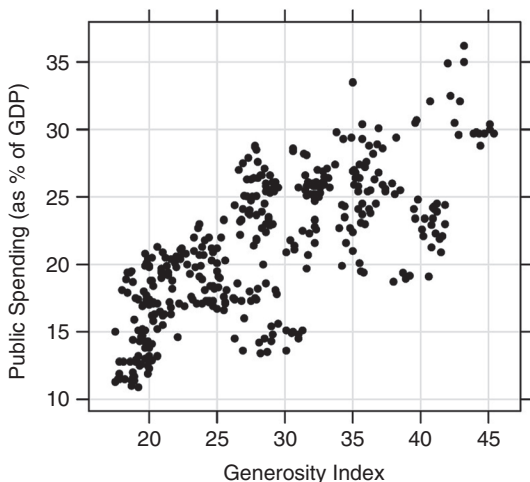
The left side of the plot displays measures of public spending and the right side concentrates on the generosity index. The top row presents histograms. The  $x$ -axis provides information about the range, interquartile range, mean (grey dot), and median of the variable. The bottom row uses line plots to show how the welfare state changes over time in each country.

a bit harder to interpret. The average score is 28, which corresponds to Germany and France's scores in the late 1990s and ranges between 18 (Japan 1980) and 45 (Sweden in the late 1980s).

The bottom row of Figure 6.1 displays the development of the welfare state across countries and time. The line plots strikingly show clear cross-country differences in terms of public spending and generosity. They also reveal substantive variation across time in several countries such as Finland, Italy, and Switzerland.

The side by side comparison of public spending and the generosity index also hint that the two measures of the welfare state do not correlate perfectly. A simple tool for examining the association between two variables is the scatter plot. Figure 6.2 displays both measures. It appears that there is positive relationship between public spending and generosity. Generosity increases jointly with welfare state spending. The plot also suggests that, at high levels of generosity, the spread of public spending is quite large. One reason for this dispersion might be that, despite generous benefits, the state does not necessarily have to make huge spending commitments. This is probably true for periods of economic expansion.

Interesting outcomes in comparative public policy, just like other social phenomena, almost always have more than one cause. In order to



*Figure 6.2* Relationship between two measurements of the welfare state  
The scatter plot displays the correlation between public spending and the generosity index for 18 OECD countries, 1980–2002.

isolate and assess the theoretical expectations that are most relevant to a research agenda, we need to account and control for alternative factors. We use multiple regression in observational studies in order to accomplish this task.

### **Inference and theory testing**

The regression model and its extensions can be employed as a descriptive tool for analysing empirical regularities among variables. The more important concern in comparative public policy is to identify and assess causal relationships. Among the different approaches to causality in the social sciences (see Brady, 2008), the idea of counterfactuals are often employed in comparative public policy. For example, what would have been the environmental standards in Italy had it not been part of the European Union? Would education at the high-school level be better if the government implemented nation-wide curriculum requirements? However, it is impossible to simultaneously observe Italy's environmental standards both with and without the role of the European Union. This obstacle is known as the fundamental problem of causal inference. Since we cannot rerun history, we need to rely on repeated measures and then separate the variable of interest ('treatment') for all other confounding variables. Doing this allows us to estimate a causal effect.

The most effective method for identifying causal effects is a randomized assignment of a treatment to the units. However, random assignments and experimental designs as well as natural experiments are rare or impossible in comparative public policy. Therefore quantitative scholars using observational data need to be cautious about asserting causal relationships. In recent years, matching models have been employed in order to assess causal claims. Matching tools group observations that are as similar as possible along all dimensions except for the variable of interest and then compute differences in the outcome in order to establish causal effects.

The most common approach is to rely on a multiple regression model for theory testing. Typically, a researcher develops some theoretical expectations about relationships among variables and formulates them as distinct and testable hypotheses. Quantitative methods then allow describing whether the theoretical expectations are consistent with the sample of their data. Based on their analysis, researchers are in a position to quantify their uncertainty about the estimated relationships. It is important to point out that these statements of uncertainty – often expressed in terms of statistical significance or standard errors – are derived for statistical assumptions.

The study of the welfare state has been a recurring subject in the scholarly debate about the appropriate use of multiple regression and its challenges. We follow Busemeyer's review on the link between globalization and the welfare state. Two competing hypotheses about the relationship between globalization (that is, economic openness) and the welfare state have been proposed. The competition thesis holds that two mechanisms force governments to decrease public spending as economic globalization increases. First, the exit threat of mobile capital forces governments to lower their taxes and thereby reduce the room for expenditures. Second, international financial markets punish countries that rely on deficit spending. In short, the relationship between economic openness and welfare state spending should be negative. In contrast, the compensation thesis posits that government spending increases as economic openness progresses. Governments are asked to cushion the effects of globalization on domestic labour and employment. Some authors go so far as to argue that increased spending, for example on retraining and education, can in fact contribute to the competitiveness in international markets.

With this theoretical starting point in mind, we would like to elaborate on two 'relatively easy' problems of causal inference in a multiple regression framework. The discussion also highlights some common statistical assumptions. The first concern centres on model specification. Since the social world is complex and potentially many variables might contribute to explaining an interesting empirical regularity in public policy, researchers need to settle on a regression model that they would like to analyse. By selecting what variables to include or exclude in the equation they might run into two issues: inclusion of irrelevant and omission of relevant variables. Both have consequences for the quality of the estimation.

The problem of 'too many' variables is relatively subtle. First, more complicated models sacrifice parsimony and, therefore, become harder to understand and use. Second, more complicated models make it unlikely that all possible specifications can be effectively presented (see Achen, 2002). Third, on a technical level, adding variables can increase our uncertainty about the estimates for the other covariates due to problems of highly correlated regressors. Adding too many variables might lead researchers to miss some systematic relationships.

The generally more severe problem occurs when researchers omit relevant variables. In this situation, our model estimates can be biased, perhaps badly so. In other words, a researcher might over- or underestimate the included variables' effects (depending on the correlation

between the excluded and included variables). Even worse, the downward bias of the standard errors implies that one might overstate the influence of the included variables. Given this problem, researchers in comparative public policy need to think carefully about rival hypothesis and model them. The inability to find a measure for an important concept does not lessen this problem.

For replication and illustration purposes, we follow Busemeyer's model selection. He argues that the compensation thesis contains a static and dynamic aspect. The static argument purports that levels of globalization are associated with spending and that there is a qualitative difference between globalization at its onset in 1980 and in its later state in 2000. This argument is essentially based on Katzenstein's (1985) comparison. Consequently, we estimate the following model, including several factors that might shore up the demand for welfare.

$$\begin{aligned} \text{Welfare}_{it} = & \alpha + \beta_1 \text{Openness}_{it} + \beta_2 \text{Unemployment}_{it} + \beta_3 \text{Income}_{it} \\ & + \beta_4 \text{Age Dependency}_{it} + \beta_5 \text{Deindustrialization}_{it} \\ & + \beta_6 \text{Female Labor}_{it} + \varepsilon \end{aligned} \quad (1)$$

While equation (1) is estimated as a cross-sectional regression in 1980, 1990, and 2000, the dynamic feature of globalization suggests that the relationship is really between changes in openness and changes in spending. This specification attempts to capture the increasing process of economic liberalization. In order to test this logic, we analyse the following first-difference model on the TSCS data.

$$\begin{aligned} \Delta \text{Welfare}_{it} = & \alpha + \beta_1 \text{Welfare}_{it-1} + \beta_2 \Delta \text{Openness}_{it} + \beta_3 \Delta \text{Unemployment}_{it} \\ & + \beta_4 \Delta \text{Income}_{it} + \beta_5 \Delta \text{Age Dependency}_{it} + \beta_6 \Delta \text{Inflation}_{it} \\ & + \beta_7 \Delta \text{Deindustrialization}_{it} + \beta_8 \text{Maastricht}_{it} \\ & + \beta_9 \text{Time}_{it} + \varepsilon \end{aligned} \quad (2)$$

For our example, we evaluate the compensation thesis by controlling for a number of alternative explanations that might affect the demand for more welfare. In particular, we consider the following controls. Unemployment is captured by the commonly used definition of the OECD unemployment rate. Income per capita is measured as current US dollar PPP. Age dependency ratio divides the population share of 0–14 and over 65 years old by the population share of the 15–65 years old. Deindustrialization is computed by subtracting the share of agricultural and industry employment from 100. The female labour force is measured as working women in percentage of population from 15–64 years. Economic openness is measured as the average between exports and imports as a percentage of GDP. Inflation is measured as the annual

changes in consumer price indices. The Maastricht variable indicates EU membership after 1992. All data come from Busemeyer (2009).

The second common problem of multiple regression in comparative public policy is the need to properly account for complicated data structures accompanied by relative small sample sizes. In particular, TSCS data demands that special care be taken when modelling the longitudinal and the spatial dependencies appropriately. In a typical application, researchers observe annual data in a number of advanced democracies. For comparative public policy, it is also often the case that the number of units is smaller than the number of years in the sample. In an influential essay (the most cited article in *American Political Science Review*), Beck and Katz (1995) brought modelling issues of TSCS to a large audience. The relative ease in implementing their recommendations contributed to an explosion of TSCS analyses since the 1990s. In a critical review of political science literature on the issue, Wilson and Butler (2007) conclude that:

It is more than a little ironic that even though B and K's analysis focused on the danger of using estimators without fully understanding their properties, so many in the profession applied the B and K method without paying any attention to the simple textbook issues.

Their review suggests that sufficient methodological knowledge is required in order to grasp the complications of TSCS data (Wooldridge, 2010, is a good start). As a consequence, comparative public policy scholars feel the need to not only follow substantive innovations in the field but also pay attention to methodological developments. For example, TSCS data structures might also be understood as multilevel or spatial data (the 2007 special issue in *Political Analysis* 15 (2) is instructive here) and a Bayesian framework might also be utilized (see Western and Jackman, 1994 for an early example). For our illustrations, we follow Busemeyer and employ panel corrected standard errors in order to abate panel heteroskedasticity and contemporaneous correlation (for example, induced by common international shocks).

As we mentioned before, interpreting regression results from tables is a challenging task for lay readers. Public policy scholars can use simulation and visual aids in order to present their results intuitively and effectively. Following King et al. (2000) (also Williams and Whitten, 2012), we use our estimates to produce counterfactual simulations and then display their conditional first difference and 90 percent confidence bands (Adolph, 2012). In the following graphs, we consider the counterfactual

of a 0.5 standard deviation increase in an independent variable. For example, we ask what would happen to welfare state spending if we increase trade openness from 33 percent of the GDP to 42 percent of the GDP. In each plot, the dots are the point estimates for the counterfactual and the lines are the 90 percent confidence bands. If a confidence band does not cross zero, we are confident that our estimate is different from zero (in other words statistically significant).

Figure 6.3 presents the results from the cross-sectional analyses. For the years 1980, 1990, and 2000, we estimated the determinants of the welfare state. The left plot displays the results for the spending measure and the right plot shows the results for the generosity index. The right graph shows that trade openness increases public social spending in all three cross-sections. For 1980 and 1990, our estimates are clearly different from zero (that is, the confidence band does not touch the vertical zero line). Moreover, the size of the estimated effect is substantial. A 0.5 standard deviation increase (about 9 percent of GDP) leads to an increase of about 1.5 percent of GDP in public social spending, *ceteris paribus*. By 2000, the effect of trade openness on spending is no longer distinguishable from zero. It is also noteworthy that all other covariates in all three cross-sections are not statistically significant. The spending results are mirrored when we measure the welfare state in terms of individual levels of generosity. As the left plot of Figure 6.3 illustrates, a 0.5 standard deviation increase in trade openness leads to a ca. 2 point increase in generosity in 2000 and at least a 3 point increase in 1980 and 1990, holding everything else equal. Roughly speaking, these results translate into the difference between the Danish and the French generosity levels during the mid-1980s. Taken together, we can posit that we cannot reject the compensation hypothesis in the cross-sectional data. Increases in trade openness are associated with higher welfare state efforts in 1980, 1990, and 2000.

Figure 6.4 assesses the dynamic logic of the compensation hypothesis using data from 18 OECD countries for the period between 1980 and 2002. The top plot illustrates the estimated effects on changes in public social spending and the bottom shows the results for changes in the generosity index. Concentrating on spending first, we find that trade openness has a substantive and negative effect on changes in public social spending. A 0.5 standard deviation increase in the change of trade openness (ca. 1 percent of GDP) leads to a 0.7 percent of GDP reduction in public social spending *ceteris paribus*. Since the confidence band does not touch the zero line, the estimated effect is statistically different from zero. This implies that positive changes in trade openness lead to welfare state cutbacks. In addition to trade, changes in unemployment

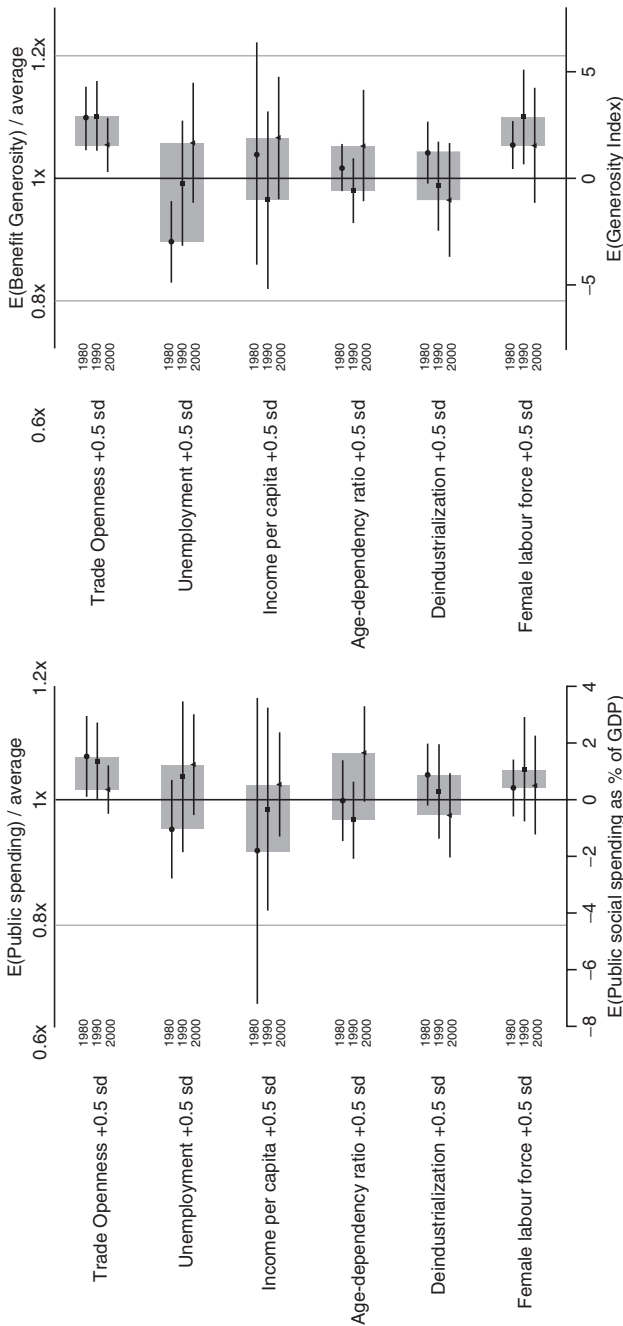


Figure 6.3 Cross-sectional analysis of the welfare state

The plots display the simulated effect for public social spending (left) and the generosity index (right) for each counterfactual listed on the left side of each plot. The dots are the point estimates. Shaded grey boxes show the full range of the point estimates. The lines are the 90% confidence bands. The bottom x-axis displays the expected changes in the welfare state effort.

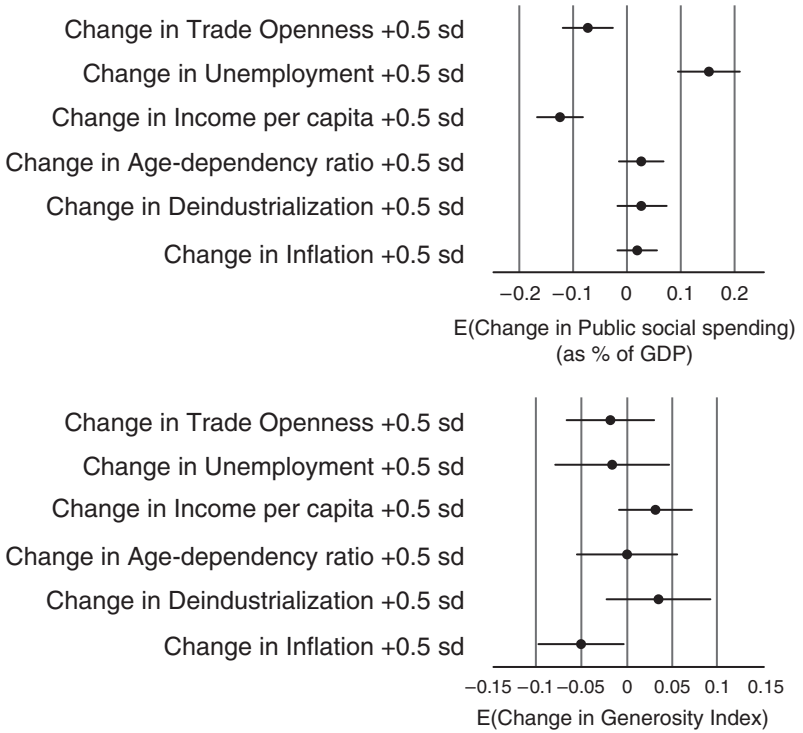


Figure 6.4 Time-series cross-sectional analysis of the welfare state

The plots display the estimated effects for change in public social spending (top) and change in the generosity index (bottom). In each plot, the counterfactual is listed on the left side. The plot displays the point estimate and the confidence band.

and income also display statistically significant effects. While more unemployment leads to spending increases, the growth of personal income reduces welfare state spending. In contrast to the spending measurement, the results for changes in the generosity index (bottom of Figure 6.4) do not display any statistically significant effects. The direction of the trade openness variable is negative. Increases in the change of trade openness lead to reductions in generosity.

Some possible conclusions from the multiple regressions are: first and substantively, the ability of the welfare state to compensate for globalization initially existed, but its effect diminished between 1980 and 2000 as the cross-sectional regressions show. The TSCS analysis indicates that increased openness leads to reductions in the welfare state (measured in spending) across countries and time. Second, the regressions reveal

that careful modelling of theoretical logic can provide researchers with important and varied insights using the same data set. Third, the difference in the results between our two measurements of welfare states in the TSCS analysis indicate that careful measurement and conceptualization remains a crucial task for quantitative public policy research. Fourth, visual displays can make the outputs of statistical and simulation results more easily understandable and comparable. For example, differences in the cross-sectional results are easily captured in Figure 6.3. Given this, visualization and the advancement of their techniques is an important task for comparative public policy scholars.

### **Prediction**

Prediction is the third goal of quantitative analysis. In contrast to predictions about electoral outcomes and wars, this tool hasn't been sufficiently exploited in comparative public policy. In the most common form, researchers rely on regression models in order to develop their predictions. When we speak of prediction we usually mean forecasting, that is, calculating unrealized events into the future. For comparative public policy scholars, forecasting can serve as a helpful tool for illustrating policy consequences of their theoretical models and for making recommendations about policy alternatives.

One other predictive tool might be equally important. Out-of-sample prediction can be employed for assessing how well a model estimated on one set of data is able to predict another set of data not used in the estimation. In addition, out-of-sample prediction (or cross-validation) enables researchers to assess their results in substantively important terms.

The beauty of this method can be illustrated using our welfare data set. We estimate the second model for changes in public social spending. Now we can ask ourselves, 'Which countries do not fit the model well?' In order to answer this question, we leave one country out of the estimation and then compare the model's predictions for that country with the actual observations. The difference between the 'forecast' and the actual observations can be quantified in terms of mean absolute forecast error. This measure has the same units as our original dependent variable – change in public social spending as a percent of GDP.

Figure 6.5 displays the results of this exercise for each country. The larger the mean absolute error, the less well our specification fits a country compared to all the others. For our dynamic TSCS model of changes in public social spending, the model seems to work well for liberal welfare states such as the United States and Japan. On the other hand, Finland and Ireland have the largest 'forecasting errors'. For example,

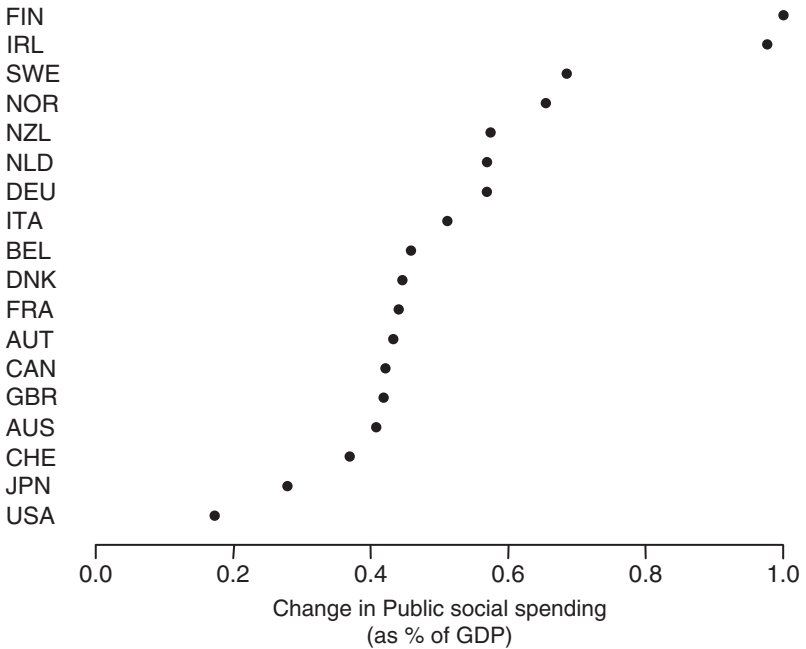


Figure 6.5 Mean absolute error by countries

The results are based on a TSCS regression in the 18 OECD countries, 1980–2002.

the mean forecast error for Finland is over 1 percent point change in spending (as percent of GDP). This suggests that the Finnish and Irish experience of welfare retrenchment since the 1980s is quite different from that of the other countries. A qualitative examination of these two experiences might explore this divergence.

As a note of caution, the link between ‘accurate’ predictions and a correct causal model is not straightforward. While appropriate causal models will result in precise predictions, one can also construct models only with the purpose of improving predictive ability in mind. Likewise, it is possible to remove statistically significant variables from a model without degrading the predictive accuracy of the overall model very much.

### The Interplay of description and inference: Typology formation and classification

We examine the interplay between descriptive and inferential goals of quantitative analysis in the context of typology formation and classification

(Elman, 2005; Ahlquist and Breunig, 2012). One of the essential tasks of social science research is to define, measure, and then classify objects along several dimensions. The classification of objects captures distinct ‘types’ of a phenomenon of interest. However, it is equally plausible that researchers have developed theoretical expectations and then examine data in order to identify typologies ‘in the real world’. This task is clearly inferential.

Traditional cluster analysis is the typical tool for identifying groups in data. The goal of cluster analysis is to minimize differences among objects within groups and maximize difference among groups. Traditional cluster models can serve as an exploratory and descriptive tool. In particular, there is no foundation in statistical theory on which clustering solution in terms of cluster numbers and groupings should be preferred, no measure of uncertainty about the placement of an object, and only a limited number of geometrical arrangements allowed. In short, substantive interpretation of the results is the responsibility of the analyst.

Model-based clustering using mixture models provides researchers with an inferential tool (Banfield and Raftery, 1993; Fraley and Raftery, 1998). It permits evaluation of uncertainty, introduces a statistical yardstick for model selection, and provides guidance on which variable to use for the analysis. So far, mixture models have received little attention in comparative public policy. We use our welfare state data in order to demonstrate the potential payoffs.

Esping-Andersen’s (1990) development of three welfare regimes led to a subsequent effort in identifying welfare state clustering and adding new types – such as the Mediterranean, Latin American, or East Asian welfare state. Using different variables, time periods, and methods, welfare state scholars are still debating on how many worlds of welfare there are, placements of various countries, and changes in regime types over time. The divergent results and findings are at least partly due to the individual researcher’s use of exploratory and descriptive methods. It is important to stress that country clusters are based on clear theoretical arguments. As such, the quantitative analysis should be understood as an inferential task.

Will we find welfare state clusters if we rely on the presented measurements of welfare state and model-based clustering? Figure 6.6 displays the results from model-based clustering using public social spending and the generosity index for 18 OECD countries in 1980, 1990, and 2000. The model with the best fit for the data, (that is, the highest Bayesian Information Criterion or BIC), identifies two components with varying variance and ellipsoidal shape. The middle plot displays the estimated density contours. The typical liberal welfare states such as the United States



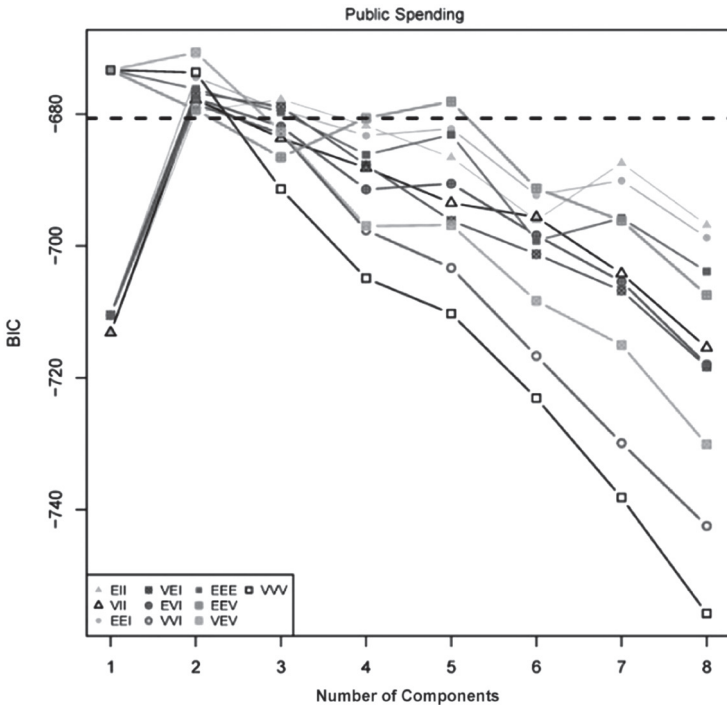


Figure 6.6 Model-based clustering of welfare states

On the top two plots we used country and year labels. The top plot displays the estimated mixture components. The ellipses in the top subfigure are based on the estimated mean and variance parameters for the mixture components. The middle subplot shows the contours of the estimated density surface. The bottom plot guides model selection. It should be the BIC for various models and number of components. The broken line represents a BIC value ten less than the BIC of the best-fitting model.

and Australia are clearly grouped together. The second cluster groups both continental and Nordic countries together. The model suggests that Canada and Ireland moved out of the liberal and into the second cluster by 2000. In addition, there is some uncertainty about the placement of Italy. The probability for being in the second cluster with the other European countries is 0.89 in 1980 and 0.77 in 1990. Similarly, this model indicates that the placement of New Zealand and Great Britain is difficult (low probabilities) in 2000. Finally, the bottom plot indicates the BIC of various models, their shape, and the number of components. While a two cluster solution is the best fitting model, several alternative

models ranging from one to five components are not substantively worse. We would be more confident in the two cluster solution if the BIC of all other models were below the broken line (that is, at least ten units lower than the best fitting model). Given this uncertainty, we would caution analysts from proclaiming clear country grouping based on the presented data.

## Discussion

We highlighted the three objectives of quantitative analysis in comparative public policy. In order to move from the goal of description to inference, analysts are required to be well versed in the substantive, theoretical, and methodological literature. The welfare state literature is a prime example for this connection. As the substantive understanding of welfare states and its theoretical development matured, important discussions about methodological issues such as measurement, causation, and proper modelling of the intricate data structure took centre stage. In short, the dialogue between theory and methodology contributes to a better understanding of the welfare state and comparative public policy more generally.

Second, the development of quantitative methodologies and their usage in comparative public policy enables us to expand our knowledge and to probe new terrains of research. Given the increasing availability and ease of access to quantitative information, analysts need to be mindful of the origin of the data and the incentives of their creators, such as international organizations, national governments, or public and private interests (see, for example, Soroka et al., 2006; Jerven, 2013). At the same time, we need to be cognizant of vast new data sources such as spatial information from satellite images or Internet traffic patterns. Using, analysing, interpreting, and presenting data remains a worthy challenge.

Finally, we like to close by drawing attention to the closing gap between quantitative and qualitative work. Both methodologies serve similar goals. Qualitative works often display and rely on numerical data and quantitative analysis requires appropriate substantive interpretation. For example, one of the frontiers of quantitative work is the analysis of the written or spoken word. The use of qualitative knowledge can also be directly incorporated into quantitative analysis especially in a Bayesian setting.

## Note

- 1 This chapter builds on on Ahlquist, 2010.

## Bibliography

- Achen, Christopher H. (2002) 'Towards a New Political Methodology: Microfoundations and ART', *Annual Review of Political Science*, 5 (1), 423–50.
- Adolph, Christopher (2012) Tile - R Package, [http://faculty.washington.edu/cadolph/software/tile\\_0.4.7.zip](http://faculty.washington.edu/cadolph/software/tile_0.4.7.zip), date accessed 1 June 2012.
- Ahlquist, John S. (2010) 'Quantitative Methodologies', in *The Encyclopedia of Political Science*. New York: CQ Press.
- Ahlquist, John S. and Christian Breunig (2012) 'Model-Based Clustering and Typologies in the Social Sciences', *Political Analysis*, 20 (1), 92–112.
- Banfield, Jeffrey D. and Adrian E. Raftery (1993) 'Model-Based Gaussian and Non-Gaussian Clustering', *Biometrics*, 49 (3), 803–21.
- Beck, Nathaniel and Jonathan N. Katz (1995) 'What to Do (and Not to Do) with Time-Series Cross-Section Data', *The American Political Science Review*, 89 (3), 634–47.
- Brady, Henry E. (2008) 'Causation and Explanation in Social Science', in Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier (eds) *The Oxford Handbook of Political Methodology* (Oxford University Press), 217–71.
- Busemeyer, Marius R. (2009) 'From Myth to Reality: Globalisation and Public Spending in OECD Countries Revisited', *European Journal of Political Research*, 48 (4), 455–82.
- Elman, Colin (2005) 'Explanatory Typologies in Qualitative Studies of International Politics', *International Organization*, 59 (2), 293–326.
- Esping-Andersen, Gösta (1990) *The Three Worlds of Welfare Capitalism* (Cambridge: Polity Press).
- Fraley, Chris and Adrian E. Raftery (1998) 'How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis', *The Computer Journal*, 41 (8), 578–88.
- Frankfort-Nachmias, Chava and David Nachmias (2008) *Research Methods in the Social Sciences* (New York: Macmillan Higher Education).
- Green-Pedersen, Christoffer (2004) 'The Dependent Variable Problem Within the Study of Welfare State Retrenchment: Defining the Problem and Looking for Solutions', *Journal of Comparative Policy Analysis: Research and Practice*, 6 (1), 3–14.
- Jackman, Simon (2008) 'Measurement', in Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier (eds) *The Oxford Handbook of Political Methodology* (Oxford University Press), 119–52.
- Jerven, Morten (2013) *Poor Numbers: How We Are Misled by African Development Statistics and What to Do About It* (Ithaca: Cornell University Press).
- Katzenstein, Peter J. (1985) *Small States in World Markets: Industrial Policy in Europe* (Ithaca: Cornell University Press).
- King, Gary, Michael Tomz, and Jason Wittenberg (2000) 'Making the Most of Statistical Analyses: Improving Interpretation and Presentation', *American Journal of Political Science*, 44 (2), 347–61.
- Korpi, Walter and Joakim Palme (2003) 'New Politics and Class Politics in the Context of Austerity and Globalization: Welfare State Regress in 18 Countries, 1975–95', *American Political Science Review*, 97 (3), 425–46.
- Pierson, Paul (1996) 'The New Politics of the Welfare State', *World Politics*, 48 (2), 143–79.
- Scruggs, Lyle (2006) 'The Generosity of Social Insurance, 1971–2002', *Oxford Review of Economic Policy*, 22 (3), 349–64.

- Simeon, Richard (1976) 'Studying Public Policy', *Canadian Journal of Political Science*, 9 (4), 548–80.
- Soroka, Stuart N., Christopher Wlezien, and Iain McLean (2006) 'Public Expenditure in the UK: How Measures Matter', *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169 (2), 255–71.
- Tufte, Edward (1983) *The Visual Display of Quantitative Information* (Cheshire: Graphics Press).
- Tukey, John (1977) *Exploratory Data Analysis* (Reading: Addison-Wesley).
- Western, Bruce and Simon Jackman (1994) 'Bayesian Inference for Comparative Research', *The American Political Science Review*, 88 (2), 412–23.
- Williams, Laron K. and Guy D. Whitten (2012) 'But Wait, There's More! Maximizing Substantive Inferences from TSCS Models', *The Journal of Politics*, 74 (3), 685–93.
- Wilson, Sven E. and Daniel M. Butler (2007) 'A Lot More to Do: The Sensitivity of Time-Series Cross-Section Analyses to Simple Alternative Specifications', *Political Analysis*, 15 (2), 101–23.
- Wooldridge, Jeffrey M. (2010) *Econometric Analysis of Cross Section and Panel Data*, 2nd edn (Boston: The MIT Press).