2

Global Comparative Methods

Paul Pennings, Hans Keman and Jan Kleinnijenhuis

INTRODUCTION

Comparative research is generally defined either at the level of systems (often national), or as a process by assessing politics over time (often yearly). Both descriptions are generally considered to differentiate the comparative approach from other approaches within social science such as rational choice. In line with this definition of comparative research, this chapter focuses on comparative studies in which the nation-state is the unit of analysis and countries are researched at one point in time or over time.

Both time and space are important dimensions in most research designs. Depending on the units of variation and the causal relationship under review, inter-temporal and/or cross-sectional variation will define the type of cases that are needed to organize the comparative data. This chapter discusses the strengths and weaknesses of two types of comparative research designs whose logic is closely related to the dimensions of time and/or space: synchronic cross-sectional comparisons and diachronic (pooled) cross-time comparisons. We make an important restriction by focusing on quantitative research, in particular time series analysis and the use of pooled cross-section time series data sets.

Global comparative methods enable us to compare many countries by using abstract concepts that can travel in order to discover universal factors that account for the phenomenon to be explained. These types of analyses are often characterized by a trade-off between the level of abstraction and the scope (or number) of countries so that they have per definition both strengths and weaknesses attached to them.

The focus on quantitative analysis might give the impression that theory is not considered to be important for comparative research. But the contrary is true. Theory, considered as a set of plausible research answers to a research question, always precedes comparative research. Often it consists of a number of causal relations that are to be confirmed by means of empirical evidence, which refute or confirm the tenability of the proposed relations. Without theory or by using flawed theory, quantitative comparative research becomes meaningless and cannot lead to valid results and insights.

The restriction to quantitative methods does not imply that these provide better means for a comparative study of social reality than qualitative methods. Neither of these approaches is better suited for comparative research by themselves: all depends on how they are used given the research question and the research design. The discussion will focus on methodological problems associated with this form of analysis and makes explicit which problems and pitfalls need to be taken into account in order to arrive at valid results.

The discussion is divided into two main parts. The first part focuses on the measurement of democracy and democratization on a global scale using a synchronic large-N design. This theme is among the most prominent examples of global research. The second part is explanatory and examines factors that account for democratization, such as economic development. In this part the emphasis is diachronic analysis within a large-N design by taking time into account.

CONCEPTS AND MEASUREMENT OF DEMOCRACY AND DEMOCRATIZATION: SYNCHRONIC ANALYSIS WITHIN A LARGE-N DESIGN

Doing research on many countries confronts us with the problem of conceptualization. A classic example from comparative politics is the research on democracy and democratization. The concept of democracy may seem unproblematic, but it is highly complex and multi-dimensional which means it is not self-evident how to apply this concept within a large-N design. Below, various ways of conceptualizing, measuring and transforming democracy into a valid and reliable cross-national variable are explored. We adopt Keman's (2002) view on democracy which distinguishes two dimensions

- pluralism – representing the possibilities of organizing as a group on the societal level free from the state;

- polyarchy – indicating the positive conditions for the population to participate in national decision-making.

The combination of both variables presents the degree of 'democraticness' in a society from a comparative perspective. Keman's study is based on 172 countries in the world (40% non-democracies; 10% old democracies; 50% recent (established after 1945) or new (after 1988) democracies). The starting point is the well-known conceptualization of democracy by Dahl as polyarchal democracy, being a political system with the six institutions listed below (Dahl, 1984; 1998).

1. Universal suffrage and the right to run for public office.
2. Free and fairly conducted elections.
3. Availability and observance of the right to free speech and protection to do so.
4. The existence and free access to alternative (and often competing) information (not controlled by government).
5. The undisputed right to form and to join relatively autonomous organizations, in particular political parties (and crucially: parties in opposition).
6. The responsiveness of government (and parties) to voters and accountability of government (and parties) to election outcomes and parliament.

It is this combined set of institutions that distinguishes polyarchic regimes from other regime types. The coming about of these institutions can then be seen as the process toward democratization. The persistance of the whole set is the hallmark of an established democracy (see also Keman, 2002; Schmidt, 2000: 393–5).

Among many comparativists Tatu Vanhanen can count as a prime example who has attempted to describe and analyze the process of democratization (Vanhanen, 1990; 1997; 2003). His index of polyarchy is based on two measures representing 'participation' and 'competition' that together form an Index of Democratization (ID) The degree of legal competition (in a democracy there will be at least two equal groups which are free to compete for power) is operationalized as 100 minus the percentage of the votes won by the

largest party (a high score indicates a high degree of competition). The degree of participation is operationalized as the number of voters as a percentage of the total population (a high score indicates a high degree of participation). From his analysis it appears that on average the countries score higher today than in the 1980s (1980 = 8.96; 1990 = 13.9) on the Index of Democratization. Indeed, the world has changed towards more democratization and now contains a growing number of countries that have taken the road to greater polyarchy.

Coppedge and Reinicke (1990) have developed a scale that examines the available institutions that promote a pluralist organization of society. In addition to examining the requirements for free and fair elections, they have developed indicators to measure the degree of freedom of organization, of speech and information, and of access to government sources of information. This operationalization is quite close to Dahl's idea of polyarchy. Coppedge and Reinicke measure the extent to which groups in society can organize themselves and are capable of conducting a viable opposition. As Schmidt (2000: 402) rightly observes, this kind of operationalization tends to ignore the formal institutions (i.e., Rule of Law) that restrict the powers of government and the state. To some extent this defect has been solved by Jaggers and Gurr (1995), who within the research programme 'Polity III', have collected data across most nation-states on:

- those institutions that facilitate and promote political choice by citizens;
- the availability of basic civil and political rights for all citizens; and
- the existence of constitutional requirements that limit the executive powers.

Jaggers and Gurr have developed a scale that enables them not only to differentiate between 'autocracy' and 'democracy', but also the level of democracy available. What do these cross-national variables tell us about the level of democratization?

First of all, it appears that the dissimilar conceptualizations and operationalizations lead to different results. The number of non-democratic countries is proportionally twice as high according to Coppedge and Reinicke than that found by Jagger and Gurr (the difference is 30 cases). Yet, Keman has found that the differences are less if one controls the results for regime types such as the ones developed by Alvarez et al. (1996): Presidentialism, parliamentarism, dictatorships and autocracies. It should be noted that on the level of individual cases the differences are – again – not large, but certain cases appear to be odd or even out of place (partly due to fact that the data used are more often than not supplied by public authorities or derived from constitutional documents).

Contrary to the indicators and scales discussed here, there is also research that focuses explicitly on the execution of individual rights not interfered with by the state (and its agencies). An example is the Freedom House index of political and civil rights (Freedom House, 2007) which has been established since 1972. This scale runs from 1 to 7, where a low value implies actual availability and observation for these rights. Taken together these two scales provide information on the extent to which a nation is not only formally democratic, but can also be considered as truly liberal democratic in practice and therefore as close as can be to Dahl's polyarchy. Studies that apply this scale show that the prevalence and observance of political and civil rights do make a difference. What is striking is the marked difference between parliamentarism and presidentialism in this respect. The latter regime type consistently shows a worse record in observing civil and political rights, notwithstanding its rule of law (Riggs, 1998).

Are these scales satisfactory as truly comparative variables? According to Bollen and Paxton (2000) this is not the case, mainly due to the (ab)use of 'subjective' measures (such as, for instance, those of Coppedge and Reinicke and of the Freedom House). Subjective measures often contain

judge-specific errors that may produce reliable (whilst measured consistently and dependably), but not valid ratings of democracy due to bias that comes from the inclusion of extraneous factors. A combination by means of grouped panel rating on the basis of judges with diverse orientations or experiences could reduce this bias (Bollen and Paxton, 2000: 79). In Keman's approach objective measures are combined with subjective ones. To this end he collected a number of scales and indexes that have been developed both subjectively and objectively and grouped these variables as being productive for creating pluralistic conditions or promoting polyarchic institutions (see Bollen, 1993; Bollen and Paxton, 2000; Keman 2002; Schmidt, 2000). By ex ante dividing the measures into more pluralistic and polyarchic the *validity* of the variables in use is improved. A statistical procedure to combine variables on pluralism and polyarchy is factor analysis that can be used to merge several variables into one or two that indicate the extent of democracy and degree of democraticness across the world. Of the 127 nations that have positive scores on both dimensions – pluralism and polyarchy – about one-third (N = 43) of the countries included can be considered – according to this operationalization – as genuinely democratic (i.e., the score is > 1.0). This is a relatively high number of countries. The 'older' and the 'richer' the countries are the stronger their democraticness appears to be. In addition, the parliamentary types of democracy score consistently higher than any other type of regime, including presidentialism. Finally, Latin-American countries do fare better than postcommunist ones. This supports the idea that 'ageing' is an important factor in developing higher levels of democraticness.

Doing causal analysis in large-N designs: Synchronic designs

In this section we shall employ these three indexes of democracy to (re)consider a number of associations with the other variables that can be seen as explaining the cross-national variation in democraticness as well as possibly accounting for certain societal performances. We shall employ the 'variable oriented' approach for a global universe of discourse because this type of analysis with a high number of cases and few variables is crucial for the development of a 'middle-range' theory regarding the democraticness of political systems (see Lane and Ersson, 1999).

Surveying the literature on explaining democracy as a system and its development (i.e., the process) the following answers have been offered.

- Economic development and socio-economic circumstances influence both its development and working (e.g., Berg-Schlosser and de Meur, 1996).
- Modernization of society and the extension of public welfare are conducive to (further) democratization of the national state (e.g., Dahl, 1998).
- Institutionalization of democracy as a regime in relation to its viability which over time enhances the level of democraticness (e.g., Diamond and Plattner, 1994).
- Organized political action in terms of participation and opposition, which 'makes democracy work' (in whatever fashion or way) is an important and often neglected facet of democratic politics (e.g., Norris, 1999).

To what extent do these factors account for the cross-national variation regarding the extent of pluralism and polyarchy? Table 2.1 reports four regression models incorporating explanatory variables for occurrence and viability of democracy. The four models are all, but for two factors, statistically significant (the rate of urbanization and the size of the public sector appear irrelevant in this context) and thus all lend support to the answer as to why democracies are dependent on certain factors to develop and remain viable as democracies. Most of the results are unsurprising and underwrite extant knowledge (Landman, 2003). Yet, it is also clear

Table 2.1   Regression analysis of factors explaining democracy

Independent variables		Dependent variables			
		Pluralism		Polyarchy	
Economics	α	−14.1		−25.8	
GNP per Capita	β	0.45	(3.95)	0.42	(3.91)
Government expenditures per Capita	β	0.12*	(1.08)	0.25	(2.33)
	R^2	25.5%		33.1%	
Society	α	−17.7		−18.6	
Urbanization	β	−0.12*	(−1.13*)	0.07*	(0.70)
Human Development Index	β	0.66	(6.49)	0.56	(5.64)
	R^2	32.4%		36.9%	
Institutions	α	−7.3		−7.2	
Presidentialism	β	0.34	(4.48)	0.34	(4.51)
Parliamentarism	β	0.74	(10.49)	0.73	(9.76)
	R^2	40.3%		39.3%	
Politics	α	−16.5		−19.44	
Electoral turnout	β	0.35	(3.37)	0.38	(3.84)
Central government expenditures	β	0.19	(1.81)	0.24	(2.39)
	R^2	16.8%		22.9%	

Note: OLS procedure has been employed; number of cases is 82 and 110; *t*-values are in parentheses; insignifcant results are flagged:*. Source: Keman, 2002 (adapted).

that none of the models is superior to the others: neither in terms of explained variance (adjusted R^2), nor in the magnitude of influence.

The first model, depicting the working of the market as well as the state, demonstrates that the 'wealth of a nation' is certainly an incentive for democratization. However, this is not the case for the size of the public sector. Yet, at the same time it is also clear that this is an insufficient condition per se. There are many outliers that prove the contrary. For example, many non-democratic nations have also considerable levels of public expenditure. Likewise a number of states with aggregated economic riches spring to mind that are close to dictatorship or autocracy (e.g., some of the Arabian countries). In short, we hold the view that

economic wealth certainly can help to foster democracy and is more often than not associated with higher level of democraticness, but is not the driving force as many political scientists and economists in the period directly after the World War II claimed (Castles, 1998).

The same can be said of the societal forces (the second model). Although much of the literature claims that the composition of society and its consequences for inter-class rivalry are important for understanding the process of democratization as well as the stability of a democratic regime, this hypothesis is not supported by our analysis. From our analysis it appears that urbanization – used as a proxy for modernization – is unrelated to the indicators for democracy. Hence, it is either an invalid proxy indicator

or the modernization thesis is not valid. We think both explanations are plausible (and this is supported in much of the literature; see: Landman, 2003; Rueschemeyer et al., 1992).

Conversely, the quality of life as expressed by the Human Development Index (http://hdr.undp.org/) is an important asset for developing and sustaining democracy. Yet, again as with economic factors, we can only go along with this claim as far as it implies a necessary condition; but – judging by an explained variance of approximately 36.8% – it is an insufficient condition for improving the level of democraticness of a nation. In addition, it should be noted that both explanations – the economy and society – tend to become functional ones. If so, and we think this is correct, the causality of the argument is weak if not absent. Rather we would go along with those who advocate a more 'case oriented' approach that enables researchers to disentangle the subtle variations within a society and to develop 'path dependent' explanations (e.g., Putnam, 1993).

The third model concerns the impact on the level of democraticness of the organization of the democratic polity. Too often the institutional fabric of democracy has been considered as the end-result of democratization. We think this view is biased if not wrong because institutions are not static, but are continuously modified by actors. The coming about of a democracy, whether it is 'old' (and now established, as in the OECD-world) or 'new' (hence recently established, as in Central and Eastern Europe), the struggle for more democracy is mainly fought out over institutions which explains why they are not constants (in the long run).

The last model reported in Table 2.1 concerns the active use of designated powers by the people and by the state. On the one hand, we examined the use of the ballot box, and on the other hand, we scrutinized the idea that central government is strongly associated with democraticness: a democratic state will be conducive to greater state intervention (by popular demand). Both contentions are only weakly supported, and – as was the case with economics and society – we can only repeat our observation that, although there is a relationship, it is not convincing and cannot be considered as a major factor for democratization and democraticness as such (Keman, 2002).

In summary: the cross-national analysis of factors promoting pluralism and polyarchy demonstrates (*ceteris paribus*) that favourable economic conditions and high(er) levels of human development are incentives for achieving higher levels of democraticness. However, like political factors, they are not crucial per se, nor functional under all circumstances. It appears rather that the interplay of these factors benefits further democratization and may well enhance the level of democraticness of a nation. Hence there is not a definitive set of factors, conditions or prerequisites (although their absence may certainly harm the level of democraticness attained!) that allows for a successful development and extension of democracy.

TIME SERIES ANALYSIS

Until now we have discussed synchronic research designs. By introducing the element of time one can analyze more cases and study developments over time. In doing so, the problem of 'too many variables and too few cases' becomes less, but it also introduces a new problem, namely that the cases are not independent. This problem may invoke a large number of statistical complications of which the most important ones are discussed below. The discussion and examples are derived from Pennings et al. (2006).

Time series analysis is discussed here only in the sense of ordinary regression analysis with points or periods in time as the units of analysis. The dependent variable y_t is measured at point t. Since it takes some time before effects come into place, the independent variables in time series analysis are often

measured in earlier points in time than the dependent variable. Time series regression analysis is a powerful tool for causal analysis, since the timely order of a cause and its consequence can be expressed with a time lag between the independent variables and the dependent variable. A time lag is the difference k between a time t in a series and a time $t-k$ in the same or another series. Often this means that the current Y-values are due to changes in X in the previous period. The introduction of this time element is an important added value compared to the cross-sectional analysis that we discussed before.

The availability of time series data allows one to construct an autoregressive model. In an autoregressive process the current value depends on one or more of the (usually immediately) preceding values. The basic idea of an autoregressive model is that the current state of affairs y_t is dependent primarily on the state of affairs in the immediate past (y_{t-1}), although external influences (effects of x_t and z_t) and random shocks (ε_t) together with an autonomous trend (b_0) may sum up to a change. The resulting R^2 from an autoregressive model is not to be compared with the R^2 in an 'ordinary' model. Especially when almost nothing changes as compared to the previous point in time the R^2 of an autoregressive model will be high, since a lack of changes (due to slowness of social changes and rigidities in social structures) will result by definition in a close correspondence between y_t and y_{t-1}. This contradicts the intuitive meaning of 'explained variance' of many social scientists.

Autocorrelation is defined as serial correlation between residuals. It occurs when the residuals in a given time period carry over into a later time period. *First order* serial correlation is correlation between immediately successive points in time (between observations at time points t and $t-1$), for example, when an overestimate in one year is likely to lead to an overestimate in the next year. False predictions for one point in time will result in false predictions for the next points in time. If autocorrelation is present, then it is

misleading to think of the consecutive time points as *independent* observations. Autocorrelation implies that the number of independent observations is smaller than the number of time points. Whereas the computation of standard errors of regression estimates in ordinary least squares (OLS) is based on the available number of time points, this computation should be based – less optimistically – on the (unknown) number of independent observations. In the presence of autocorrelation OLS estimates of regression coefficients in non-autoregressive models are *inefficient*, although still *unbiased*. Autocorrelation in autoregressive models makes things even worse. Estimates will not only be inefficient but also biased.

A straightforward diagnostic of first order serial correlation would be the correlation coefficient $r_{t,\,t-1}$ between residuals in successive points in time. The *Durbin-Watson statistic DW* is based on this serial correlation coefficient between residuals. It is roughly equal to $2 - 2r_{t,t-1}$; it thus takes values between 0 and 4 rather than between -1 and $+1$. Since DW is roughly equal to $2 - 2\,r_{t,\,t-1}$ its value range is 0.4 instead of $-1 + 1$. $DW = 2$ corresponds with $r = 0$, $DW = 0$ with $r = +1$, and $DW = 4$ with $r = -1$. DW-values in the neighbourhood of 2 indicate the absence of autocorrelation. Values near 0 indicate the presence of autocorrelation: it is likely that a deviation from the regression line at time t will be followed at time $t + 1$ by a deviation in the same direction. If errors are positively correlated, as they usually are, standard errors are underestimated and the R^2 and t-values have an upward bias so that they present an overly optimistic view about the accuracy of the coefficients. This bias makes us reject the null hypothesis of no relationship far more often than we should. Values of DW between 2 and 4 indicate an oscillating pattern: if the actual value at time t is higher than one would expect on the basis of the regression equation, then it is likely that the actual value at time $t + 1$ is lower than one would expect on the basis of the regression equation. Negatively correlated errors

($DW = 4$) are associated with overestimated standard errors.

Durbin-Watson-values are computed by most statistical packages, but usually it is still necessary to consult a table with DW-values to find out whether a specific DW-value indicates autocorrelation, no autocorrelation, or doubt given a specific number of time points as units of analysis and a given number of independent variables. Regrettably the DW-tables have a region of doubt, depending on the number of independent variables and the number of observations, in which it is undecided from a statistical point of view whether autocorrelation is present or not.

The Durbin-Watson DW-test applies to non-autoregressive time series regression models, but should not be applied to autoregressive models. To indicate whether autocorrelation in the residuals from an autoregressive equation is absent, one should not use the ordinary Durbin-Watson-test, but for example the Durbin's h test. If the usual 5% criterion is used, the assumption that serial autocorrelation is absent is tenable when $h < 1.645$. If Durbin's h-test indicates that autocorrelation is present in an autoregressive regression equation estimated with OLS, then the OLS conclusion should be that should not have been used. One must resort to generalized least squares estimation procedures which are implemented in most statistical packages. OLS estimates of regression coefficients can be used in autoregressive models, however, when Durbin's h-test indicates the absence of autocorrelation.

An often used, rather intuitive solution to obtain independent observations would be to diminish the number of time points in the regression analysis, for example by aggregating quarterly data to yearly data, or by aggregating yearly data to five years data, or by aggregating all time points before and after important historical events (e.g., World War II, the 1973 oil crisis, the 1989 velvet revolution). Two procedures may be used: simply pick out one time point per period or smooth the data within each time period (for example by computing average values for each time period). This intuitive solution is flawed, however. Meaningful variation within the aggregated time spans is easily ignored. Moreover the periodization is often arbitrary, because each variable tends to have its own periodicity, its own rhythm of change. Here we will stick to solutions which retain all data points in the regression equation.

Let us first consider a non-autoregressive model which exhibits autocorrelation according the DW test (DW far lower than 2). This indicates that the process being studied remains by and large in the same state as in the previous point in time. It may still be possible to explain changes, however. To explain changes relative to the status quo either a simple first-order-difference regression model or a more advanced autoregressive model should be used.

In the first-order-difference model the dependent variable is the change $Dy_t = y_t - y_{t-1}$ in Y (the zero order dependent variable) as compared to the previous point in time. Regardless of the previous level y_{t-1} of the dependent variable Dy_t will become zero whenever $y_t = y_{t-1}$. The difference model $Dy_t = b_0 + b_2 x_{t-1} + b_3 z_{t-1} + \varepsilon_t$ is equivalent to a model $y_t = b_0 + b_1 y_{t-1} + b_2 x_{t-1} + b_3 z_{t-1} + \varepsilon_t$ with y_t as the dependent and the lagged dependent variable y_{t-1} as an independent variable with b_1 constrained to 1. In a first-order-difference model the *motion* of an object is the dependent variable, whereas in a zero order model the *position* of an object is the dependent variable.

In an autoregressive model $y_t = b_0 + b_1 y_{t-1} + b_2 x_t + b_3 z_t + \varepsilon_t$ the regression coefficient for the lagged dependent variable y_{t-1} is not constrained to 1, but empirically estimated. The autoregression coefficient b_1 gives information about what exactly is being influenced by the remaining independent variables. An estimate of $b_1 = 0$ is equivalent to an ordinary regression model with y_t as the dependent variable. An estimate of $b_1 = 1$ is equivalent to the first-order-difference model. Empirical estimates of b_1 will often result in between 0 and 1. An estimate of $b_1 = 1/2$ would indicate that the remaining independent variables in the model have an influence on $y_t - 1/2 y_{t-1}$.

To compare a non-autoregressive model ($b_1 = 0$), a first difference model ($b_1 = 1$) and an autoregressive model (say, with $b_1 = 1/2$) it is helpful to think of the 'shocks' required from the remaining independent variables to keep y at an extreme high (or low) level. In a first-order-difference model a continuation of the shocks which brought about today's level of y_t is superfluous to preserve the status quo. For this reason a first-order-difference model is also known as the *random walk* (RW) which is the most well-known non-stationary process. This process resembles a walker who time and again takes a step so as to keep a tail wind from the independent variables, regardless of where he came from or where he wants to go. He will stay where he is when it is dead calm. In a non-autoregressive model our walker will return home immediately once there is not a breath of wind. This property of non-autoregressive model is known as regression towards the mean, which means that without continued external shocks the mean will be restored. An autoregressive model with an autoregressive parameter of 0.5 resembles a walker who returns half way home when the wind drops.

The solution for autocorrelation in an autoregressive regression equation (as indicated by Durbin's h) or in a first-order-difference model (as indicated by the ordinary *DW*-test) is subject to debate, both from a theoretical as from a statistical point of view. One solution would be to develop a second-order-difference model, which has as the dependent variable the rate of the change of the change of the original dependent variable. A second-order model from physics would be a model with the acceleration of an object – rather than its position (zero order) or its motion (first order) – as the dependent variable.

Time series data are a perfect means to assess causality because of their timely order and therefore might be superior to cross-sectional models. However, this is only the case when they are handled properly given a large number of statistical complications that are likely to impact on the results.

The regression models based on the original variables typically suffer from the autocorrelation defect. Difference models and/or autoregressive models will usually cure the autocorrelation disease, but difference models and autoregressive are usually not robust. At least three origins of this lack of robustness can be mentioned.

Autoregressive models will usually leave only a small portion of the variance in the dependent variable unexplained. Exogenous influences are hard to verify when the remaining unexplained variance is small, especially when measurement errors are present. A second reason why autoregressive models and difference models often fail to retrieve the obvious is their fixation on short term changes. Long term shocks in exogenous variables which have already influenced the lagged dependent variable will not be attributed to exogenous variables but to the endogenous lagged dependent variable. In the last decades error correction models or co-integration models have been developed to account for long-term effects of exogenous variables, without introducing autocorrelation once more. These models will be left aside here.

The third, and most important reason, is simply the limited number of time points. Data on 25 consecutive years is almost nothing, especially when autocorrelation is present. Twenty-five years may shrink to five 'independent' years when most years are almost perfect copies of their predecessors. Data on short time series cursed with autocorrelation are compatible with many simplistic rivalling theories, but they are simply insufficient to estimate the parameters of any complex theory.

Pooled time series analysis

One way out of this difficulty in time series analysis is to test elaborated theories for many time series simultaneously, which brings us to pooled time series analysis. The advantage of time series analysis is its ability to assess the

time dependency of causal relationships. Often the data available do mount up to short time series only (e.g., 40 points in time or even less). More often than not various plausible models will account for the data on such a short time series. One way out is to increase the quantity of the data used for testing.

Pooled time series analysis (or panel analysis) combines time series for several cross-sections. The data are stacked by cross-section and time points. A classical example is a pooled time series database of 828 units stacked by 18 OECD countries over 46 years (1960–2006). Instead of studying the effects of various variables on public expenditures in each country through time, these effects may be studied for a number of countries simultaneously. Instead of testing a time series model for one country using time series data, or testing a cross-sectional model for all countries at one point in time, a pooled time series model is tested for all countries through time. Much more refined tests of theories will become possible, since the available units of analysis increase from T (number of time points) to NT (number of cross-sections times number of time points). Pooled time series analysis captures not only variation that emerges through time, but variation across different cross-sections as well. Note that not all global methods are necessarily highly complex. Most available studies can be situated between the advanced statistical analysis of Przeworksi et al. (2000) on the relationship between democracy and development (1950–1990) and the more descriptive approach of Lane and Ersson (2002) who study the size of government in all countries on the basis of aggregated data.

Regrettably pooled time series analysis also has a serious drawback. Since pooled time series analysis is still time series analysis, the problem of autocorrelation must still be dealt with. But in addition to autocorrelation per cross-section heteroscedasticity between cross-sections comes in. Heteroscedasticity is the unequal distribution or variance of the error term which invalidates significance tests. It is a common problem in cross-sectional analyses, especially in aggregate data. Heteroscedasticity will usually arise because the appropriate models for the various cross-sections will not be precisely identical. Therefore a model to explain all cross-sections will usually do better for some than for others, which amounts to unequal variances of the residuals for the cross-sections. In our example on expenditures heteroscedasticity means the following. The tendencies which led to higher public expenditures in the seventies manifested themselves in all capitalist countries. Nevertheless the precise effect of an increasing percentage of elderly on public expenditures may depend on polity variables such as the electoral system, and on policy and legislation with respect to health care technology, health care insurances and pensions for the elderly. If one model is tested for all cross-sections at all time points, then heteroscedasticity comes in since the residuals for 'extreme' countries will be large as compared to the residuals for mainstream countries.

The combination of autocorrelation and heteroscedasticity in sample data may result in extremely inefficient, although unbiased, estimates of the true population parameters. The diagnosis of autocorrelation and heteroscedasticity in pooled time series analysis is fairly straightforward, although statistical software packages are usually not ideally suited for its implementation. The degree of heteroscedasticity due to pooling is to be obtained by examining the residual variances of the pooled model per cross-section. A sequence plot of the residuals for the various cross-sections will give a first visual impression. Ideally the average of the residuals within each cross-section should be equal to zero. If an inspection of the sequence plot suggests that the mean residual varies from cross-section to cross-section then the conclusion should be that crucial variables that explain the differences between cross-sections (regardless of the precise time point being looked at) are still lacking.

A simple diagnostic test on the robustness of the pooled model is to run the same model

on its residuals for each cross-section through time, and on its residuals for each time unit over cross-sections. If the same model holds for all cross-sections and all time points, then the pooled model will not be able to explain its own residuals split up by cross-section and time unit. Thus, for a regression model tested on 80 units stacked by 8 cross-sections over 10 years, $8 + 10 = 18$ regressions should be performed on the residuals from the pooled model. The model should not be able to explain significant proportions of the variance within its own residuals in more than 5% of the cases. Thus, the pooled model from our example should not be able to produce significant regression estimates within its own residuals in more than four time units or cross-sections. If the model is able to explain additional variance in its own residuals for a large number of time units or cross-sections (more than 4 in our example) then the suspicion should be that the original model does not hold for all cross-sections and time units equally well.

A proper diagnosis of autocorrelation in pooled time series analysis is cumbersome, because of its statistical relatedness with cross-sectional heteroscedasticity. If there is cross-sectional heteroscedasticity there will be autocorrelation almost by definition: if the predictions for the complete cross-section are wrong, then the mispredictions for each of its successive time points will be serially correlated. Model improvements to reduce the cross-sectional heteroscedasticity will therefore usually also diminish autocorrelation. The formulas of the Durbin-Watson statistic and Durbin's h allow for a computation over time series for several cross-sections. One technical warning is probably not superfluous: the lag of the first time point for a specific cross-section is missing (and not equal to the last time point for the preceding cross-section in the data file). It is a pitfall to rely on autocorrelation diagnostics per time series. Precisely because the separate time series in pooled time series analysis are too short, Durbin-Watson-tests per cross-section produce chaotic results.

The solutions to the problems raised by pooled time series analysis might be divided in two groups. The first group of solutions is directed at the improvement of the models to fit pooled time series data. The second group of solutions is directed at the development of statistical estimation procedures to improve on OLS deficiencies when a combination of autocorrelation and heteroscedasticity is present.

Let us start with model improvements to get rid of heteroscedasticity between cross-sections. When the mean of the residuals for one or more specific cross-sections is unequal to zero, then one should add variables to the model so as to explain these cross-sectional differences better. A non-theoretical model to get rid of heteroscedasticity between cross-sections completely would be to add one dummy variable to the model for each cross-section, except one. This model is called the least squares dummy variable (LSDV) model in the jargon of pooled time series analysis. The LSDV-model accounts for different Y levels by estimating different intercepts for each cross-section. A more advanced variant would be to assume that each cross-section has a randomly distributed intercept associated with it (the random coefficients model). We would advise against these non-theoretical solutions, since a-theoretical dummies and random intercepts that are added to a regression model will usually be collinear with some variables of theoretical interest. The explanatory power of the variables of theoretical interest will easily get obscured. It is far better to include a few variables which account for the major differences between the cross-sections, than to include every separate cross-section (except one) as a dummy-variable. The LSDV model and the random coefficients model should only be used when the available theory gives no cues at all with respect to differences between processes in the cross-sections being studied.

To get rid of serial autocorrelation the same model ramifications (first-order-difference model, autoregressive model) should be

considered as in ordinary time series analysis. A rather different question is which estimation technique should be used when autocorrelation and heteroscedasticity have not been banned completely. How should we deal with the fact that OLS estimates will be inefficient and therefore usually underestimate the standard errors of the regression estimates? Econometricians have proposed various estimation techniques for this purpose. The most widely applied is the Parks-Kmenta method, a specimen of the generalized least squares (GLS) family of estimation techniques (White, 1994: 245–254). These estimation techniques guarantee that the estimates asymptotically hit the mark. They are unbiased when sample sizes draw near infinity. Recently Beck and Katz (1995) have shown that the Parks-Kmenta estimation technique produces quite chaotic results when time series are as short as in comparative political science (usually less then 50 years per cross-section). Katz and Beck showed also that OLS estimates of regression coefficients are more robust than Parks-Kmenta estimates when sample sizes are small. Katz and Beck have developed a formula to compute panel corrected standard errors (PCSEs) which encompass autocorrelation and heteroscedasticity in the computation of the standard errors of the OLS-regression estimates.

The use of panel data has become quite common in quantitative comparative research. Unfortunately this is accompanied by a number of problems and pitfalls. These are discussed by Kittel and Winner (2005) and Plümper et al. (2005) in their critique on the study of Garrett and Mitchell (2001) on the relationship between total government expenditure and the partisan composition of government as well as economic internationalization. In their discussion on PCSE they argue that autoregressive models with panel corrected standard errors should not be used as a universal remedy for problems in panel data analysis. If the assumption on the error terms are not tested before PCSEs are calculated and/or problems with non-stationary data are not recognized, the conclusions will always be highly problematic (Plümper and

Troeger, 2007). In addition, both the size and the sign of the estimates may strongly depend on the exclusion of particular countries.

From this overview of the problems and pitfalls of pooled analysis follows that purely cross-sectional analyses are still necessary and useful since they are not disturbed by the problems inherent in time series. They can be used to validate the results of pooled time series analysis. If an analysis includes institutional and political variables that hardly vary over time, there is not much use for pooling repeated observations over time, unless efficient estimation techniques can be utilised (Plümper and Troeger, 2007). Pooling data is especially useful if an effect is assumed to be equal across space and time or when the research focuses on short term effects. When these conditions are not fulfilled, statistical problems are likely to make the results meaningless.

We end our discussion with the same example as in a previous section on synchronic analysis, but we now introduce the element of time. Burkhart and Lewis-Beck (1994) have analysed the economic factors that may boost democratization. Their data set is an adapted and extended version of the Freedom House democracy indicators. Burkhart and Lewis-Beck added to this data dummies for the position of countries (c = core, m = semiperiphery, p = periphery). They also employ the energy consumption per capita (logged) as an economic development measure (that correlates 0.9 with gross national product per capita). Burkhart and Lewis-Beck test the 'economic development thesis' with the following model:

$$D_t = a + bD_{t-1} + cE_t + d(M \times E_t) + e(P \times E_t) + u$$

where

D_t is the democracy index at time t;
D_{t-1} is the democracy index from the year before;
E_t is energy consumption per capita (logged to the base 10) at time t;
$(M \times E_t)$ is the dummy variable for semiperiphery status multiplied by E_t;

The Burkhart and Lewis-Beck's model is an autoregressive model having the lagged dependent variable at the right hand side of the equation (like in $Y_t = Y_{t-1} + X_t$). This type of modelling is not without complications as it may well boost the R^2 and Beta-weight. D_{t-1} acts to control for omitted independent variables: as the other forces acting on democracy are uncertain, they will be essentially summarized in the democratic performance of the nation during its previous year. Their estimation procedure is GLS-ARMA which avoids first order autocorrelation and cross-sectional heteroscedasticity. Their model throws up a pseudo-R^2 of 0.71 and the b-scores are 2.49 (for E_t), −1.33 (for $M \times E_t$) and −1.54 (for $P \times E_t$). Their conclusion is that economic development matters most for nations in the core, it still matters, but about half as much, in the semi-periphery. For nations in the periphery, the economic effect is just a bit less. Taken together, economic factors, both international and domestic, appear decisive in shaping a nation's democratic future.

In order to show the complications of this type of diachronic analysis, we will replicate the analysis synchronically, using OLS regression on a 1988 cross-section. The results of our analysis match with that of Burkhart and Lewis-Beck, be it that our estimates indicate moderate effects. This outcome confirms our suspicion that the autoregressive model might not throw up a reliable R^2, namely an adjusted R^2 of 0.36 (compared to 0.71 in the original analysis). A theoretical, instead of statistical, explanation of the moderate performance of the Burkhart and Lewis-Beck model is provided by Vanhanen (1990). He proposed an alternative for the socio-economic hypothesis of democratization, by hypothesizing that democratization takes place under conditions in which power resources have become so widely distributed that no group is any longer able to suppress its competitors or to maintain its hegemony (Vanhanen, 1990: 66). The main difference with Burkhart and Lewis-Beck is that Vanhanen not only looks at the level of welfare but also, and more importantly, at the distribution of a wider range of power resources. Vanhanen's conceptualization and operationalization of the index of power resources indeed results in a much higher explained variance of 0.71.

This example shows us that a high explained variance is only to be trusted when both the theoretical and statistical specifications of the model are correct. The diachronic Burkhart and Lewis-Beck model is far more complicated than our synchronic replication. But by reducing its complexity and by comparing its results with other research outcomes, it becomes clear what the weaknesses of this model are. In that sense we can conclude that, although diachronic methods are more advanced, they cannot replace synchronic methods.

CONCLUSIONS

Global comparative methods are potentially capable to incorporate many countries and extensive time series in the analysis. The strengths and weaknesses are closely related to those of quantitative methods in general. Their main strength is that the scope of comparison is widened across time and space. This opens up new possibilities for strong inferences and theory-building and the identification of deviant cases. Their main weakness is that they may easily lead up to misleading results due to their complexity. In addition, global methods are often applied in a case-blind manner by focusing on the interrelationships between the variables which are included in the statistical models.

These pitfalls or weaknesses do not make 'global methods' worse or better equipped for comparative studies than other approaches in social science since their usefulness for comparative research depends on how they are applied. Anyone applying 'global methods' should be aware of the methodological trade-offs which are involved in doing this type of research. In particular there is a

trade-off between reliability (which improves with the increase of cases) and validity (which is hampered by a large number of cases).

During the last 20 years several new techniques have been introduced which enable the statistical analysis on data relating to many countries and time points that are integrated into a single pooled data set (also referred to as panel data). The main problem with panel analysis is the lack of robustness, since the estimates are highly dependent on the model specification. For this reason it is often necessary to compare the results of panel analysis with those of cross-sectional analysis in order to determine whether they point into the same direction. This brings us to the conclusion that, although pooled time series analysis is often seen as a methodological advancement compared to cross-sectional regression analysis, it does not make the latter useless. In addition, cross-sectional analysis is still to be preferred to panel analysis if the variables vary little over time, as is often the case with institutional variables.

REFERENCES

Alvarez, M., Cheibub, J. A., Limongi, F. and Przeworski, A. (1996) *Classifying political regimes*. Unpublished manuscript. University of Chicago.

Beck, N. and Katz, J. (1995) 'What to do – and not to do – with time-series cross-section data in comparative politics', *American Political Science Review*, 89 (3): 634–47.

Berg-Schlosser, D. and de Meur, G. (1996) 'Conditions of authoritarianism, fascism, and democracy in interwar Europe: Systematic matching and contrasting of cases for "small N" analysis', *Comparative Political Studies*, 29 (4): 423–68.

Bollen, K.A. (1993) 'Liberal democracy: validity and method factors in cross-national measures', *American Journal of Political Science*, 37 (4): 1207–30.

Bollen, K.A. and Paxton, P. (2000) 'Subjective measures of liberal democracy', *Comparative Political Studies*, 33 (1): 58–86.

Burkhart, R.E. and Lewis-Beck, M.S. (1994) 'Comparative democracy: The economic development thesis', *American Political Science Review*, 88 (4): 903–10.

Castles, F.G. (1998) *Comparative public policy. Patterns of post-war transformation*. Cheltenham: Edward Elgar.

Coppedge, M. and Reinicke, W. H. (1990) 'Measuring polyarchy', *Studies in Comparative International Development*, 25 (1): 51–72.

Dahl, R.A. (1984) 'Polyarchy, pluralism, and scale', *Scandinavian Political Studies*, 7 (4): 225–40.

Dahl, R.A. (1998) *On democracy*. New Haven, CT: Yale University.

Diamond, L. and Plattner, M.F. (eds) (1994) *Nationalism, ethnic conflict, and democracy*. Baltimore, MD: Johns Hopkins University Press.

Freedom House (2007) *Freedom in the World 2006. The Annual Survey of Political Rights and Civic Liberties*. Retrieved June 6, 2007, from www.freedomhouse.org.

Garrett, G. and Mitchell, D. (2001) 'Globalization, government spending and taxation in the OECD', *European Journal of Political Research*, 39 (2): 145–177.

Jaggers, K. and Gurr, T.R. (1995) 'Tracking democracy's third wave with the Polity III data', *Journal of Peace Research*, 32 (4): 469–82.

Keman, H. (ed.) (2002) *Comparative democratic politics. A guide to contemporary theory and research*. London: Sage.

Kittel, B. and Winner, H. (2005) 'How reliable is pooled analysis in political economy? The globalization-welfare state nexus revisited', *European Journal of Political Research*, 44 (2): 269–93.

Landman, T. (2003) *Issues and methods in comparative politics. An introduction* (2nd edition) Oxford: Oxford University Press.

Lane, J.E. and Ersson, S. (1999) *Politics and society in Western Europe*. London: Sage.

Lane, J.E. and Ersson, S. (2002) *Government and the economy. A global perspective*. London: Continuum.

Norris, P. (ed.) (1999) *Critical citizens: Global support for democratic government*. New York: Oxford University Press.

Pennings, P., Keman, H. and Kleinnijenhuis, J. (2006) *Doing research in political science. An introduction to comparative methods and statistics*. London: Sage.

Plümper, T. and Troeger, V. (2007) 'Efficient estimation of time-invariant and rarely changing variables in finite sample panel analyses with unit fixed effects', *Political Analysis*, 15 (2): 124–139.

Plümper, T., Troeger, V. E. and Manow, P. (2005) 'Panel data analysis in comparative politics: Linking method to theory', *European Journal of Political Research*, 44 (2): 327–54.

Przeworski, A., Alvarez, M. E., Cheibub, J. A. and Limongi, F. (2000) *Democracy and development. Political institutions and well-being in the world, 1950–1990*. Cambridge: Cambridge University Press.

Putnam, R.D. (1993) *Making democracy work: Civic traditions in modern Italy*. Princeton, NJ: Princeton University Press.

Riggs, F.W. (1998) 'Presidentialism vs. parliamentarism: Implications for representativeness and legitimacy', *International Political Science Review*, 18 (3): 253–278.

Rueschemeyer, D., Huber, E. and Stephens, J.D. (1992) *Capitalist development and democracy*. Cambridge: Polity Press.

Schmidt, M.G. (2000) *Demokratietheorien* (3rd edition) Opladen: Leske and Budrich.

Vanhanen, T. (1990) *The process of democratization. A comparative study of 147 states, 1980–1988*. New York: Crane Russak.

Vanhanen, T. (1997) *Prospects of democracy: A study of 172 countries*. London: Routledge.

Vanhanen, T. (2003) *Democratization. A comparative analysis of 170 countries*. London: Routledge.

White, L. G. (1994) *Political analysis: Technique and practice*. Belmont, CA: Wadsworth.