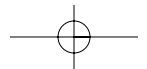
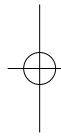
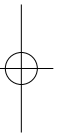


PART III

Beyond spending: welfare state generosity, social rights and obligations



7. Welfare state generosity across space and time

Lyle Scruggs

INTRODUCTION

Comparative analyses of welfare state reforms have relied overwhelmingly on public spending data as the indicator of welfare state commitment and change. However, scholars have long emphasized the problems with spending outputs and also stressed the importance of programmatic elements of welfare state policies. One particular focus has been on national commitment to social citizenship rights. This general line of research has offered few alternative measures that are compatible with comparative analysis across many countries and long periods of time. Those that do exist, such as the Social Citizenship Indicators project at the Swedish Institute of Social Research, are not generally available to scholars (see Kangas and Palme, Chapter 6, this volume).

This has led to two reinforcing cleavages in welfare state research: large-n comparisons of many countries testing general theories but relying on spending data as proxies of welfare state generosity (or effort), and small-n comparisons with *sui generis* data sets that propose and demonstrate, but seldom really test, hypotheses. Both sides of the cleavage are subject to specific scientific limitations and advantages. For its part, the 'large-n camp' focuses on honing a well developed (if not uniform) set of empirical data to test analytical models and theories, but may be consistently misspecifying what the welfare state is. The more qualitative camp, by contrast, can often claim a more nuanced set of concepts, but lacks quantifiable leverage for scientific verification, or, in the case of single country studies, any direct comparisons. Qualitative researchers are probably right about aspects of conceptual validity, but the quantitative approach is probably right about the requirements for theory validation.

By both being half right, both approaches cannot help but be half wrong. Moreover, neglected in both approaches is the construction of valid empirical measures of concepts that are to be tested. Qualitative research generally fails to get beyond defining the concept and applying it to a case

or two, thus failing to direct research to comparable, systematic data collection. Quantitative research often fails to get beyond the technical aspects of the empirical analysis, thus failing to ask whether the most used, or available, data sets are worth analysing in the first place (see also Chapter 2 by Green-Pedersen in this volume).

Collecting comparable data on non-spending aspects on welfare state programmes is an important step in bridging this gap. Such data can add fundamentally to the evaluation of changes in welfare state policy. One of the main goals of this chapter is to introduce and discuss one data set, the Comparative Welfare Entitlements Dataset (CWED), which details changes in the major components of key social insurance programmes in 18 countries from the 1970s to 2002. The other major goal of the chapter is to suggest ways in which data collection efforts on 'qualitative' dimensions of the welfare state should be augmented in order to promote the development of more general theories about welfare state development and change.

The rest of the chapter is divided into five sections. The next section addresses why cross-national and historical data sets of social entitlements are necessary in order to think theoretically about welfare state change. This section also critiques the widespread use of comparative spending data. The third section presents the CWED data on trends in social insurance programmes. The fourth section develops a benefit generosity index, based on a refinement of Esping-Andersen's popular declassification index, and using the data presented in the third section. The evidence in the third and fourth suggests that welfare states became more generous during the 1970s and 1980s, but experienced some considerable retrenchment in the 1980s and beyond. That retrenchment has been greatest in the more generous countries. The fifth section acknowledges that the CWED data set is incomplete, discusses possibilities for expanding it and, reinforces the central theme of this volume: the importance of reliable and systematic data in developing and evaluating explanations of welfare state development and change.

CROSS-SECTIONAL, TIME SERIES QUANTITATIVE MEASURES OF BENEFIT GENEROSITY AND WELFARE REFORM

Time and Space in Theories of Reform

Any comparative empirical analysis of reform requires detailed data in both *space* and *time*.¹ The importance of a spatial dimension in comparative

public policy research almost goes without saying: comparative analysis requires multiple objects of comparison. Nevertheless, many 'comparative' studies focus one country, or even one programme in one country. The value of this latter type of approach, *when nested among other case studies that isolate similar programmes in other countries, and investigate common causal processes*, is beyond dispute. Most of the time, however, case studies (or other small-n comparisons) are undertaken with only loose attention to comparability with other studies. So case studies often just accumulate.

Attempts to compare findings from such studies exacerbate a classical data problem in social science. Tests of social theories are usually based on data collected for purposes other than testing social theories; theory evaluation with 'data' comprised of a hodgepodge of accumulated case studies further compounds the problem.²

More qualitatively oriented researchers *do* have a point when suggesting that their work *can* avoid a fundamental social science data problem: it can derive and collect the appropriate data to test a theory. Because analytical comparisons (particularly quantitative comparisons) must assume that their data constitute the correct measurements of the underlying phenomena they are evaluating, 'qualitative research' is essential to the comparative evaluation of theories. Anyone with doubts about how critical correct measurement of concepts is for statistical inference should review contemporary texts in econometrics (e.g. Gujarati, 2003).

So there are really two problems in comparison. One is using invalid conceptual measures, the second is ensuring that you have consistent measurement across units. Tackling these two problems is important in developing a 'large-n' data set to study any question. In many comparisons, this probably does require a set of heavily structured 'case studies', demanding detailed knowledge about specific features of individual units. But this has to be undertaken systematically in order to maximize comparability across the cases. Such work is surprisingly rare in social policy research. Even for topics like comparing national incomes or price levels – big topics with big budgets – there are surprisingly few. In social policy (and many other questions in comparative politics) there are precious few data sets that really attempt to do this in order to answer research questions.

The importance of *time* when comparing welfare state reforms is also obvious, yet absent from much social policy work. Identifying a reform requires at least two time points. Korpi's (1989) discussion of social rights and welfare development, for example, is quite explicit in its call for considering welfare state developments in time. And, as historically oriented researchers often point out, theories of macrosocial phenomena imply a lot of historical inertia with change in the presence of much noise. But precisely because there is so much noise in historical processes, social scientists

must look at many similar histories in order to be able to verify empirical regularities. The data demanded to evaluate historical path dependence expand exponentially if we want to have much confidence in any claim (de Marchi, 2006: 49–53). History makes information across both time *and* space (lots of it) vital.

Comparative Social Spending

The preceding discussion is perhaps too abstract. Knowing that information across time and space is essential to understanding reform tells us nothing about the content. There are, after all, a few widely used cross-sectional time series indicators of welfare states available, and they are widely used in the literature. Why are they inadequate? Why is a measure of welfare programme generosity or social rights necessary? The most widely used series of data on welfare reform are probably national time series of social spending. Probably the most popular measures are general government spending, transfer spending and consumption spending (Garrett and Mitchell, 2001). Such measures of spending enjoy a long history in the comparative analysis of social policy and political economy (Wilensky and Lebeaux, 1958; Wilensky, 1975; Cameron, 1978; Korpi, 1983; Garrett, 1998; Swank, 2002b). More recently, the OECD's social expenditure database represents a comprehensive effort to make comparable specific categories of social spending (Castles, 2002; Lindert, 2004; cf. De Deken and Kittel, Chapter 5, this volume).³ While it is true that reliance on this spending data is prominent in statistical analyses, many qualitative accounts also utilize a great deal of spending based quantitative information. Undoubtedly, researchers rely on spending measures because they are available, (relatively) comparable, and vary across countries and time.⁴ Yet researchers widely acknowledge that spending data have significant drawbacks, in particular in evaluating welfare retrenchment (Gilbert and Moon, 1988; Esping-Andersen, 1990; Castles and Mitchell, 1993; Clayton and Pontusson, 1998; Goodin et al., 1999). As a guide to understanding the impact of the welfare state on individual life-chances, for example, total spending reveals little about who benefits. As Esping-Andersen (1990: 21) famously remarked, 'it is difficult to imagine that anyone struggled for spending per se.' Several other shortcomings of spending data are also worth considering.

Welfare Dependency Structure

Spending measures do not typically account for the size of the dependent population (but see the contribution of Kangas and Palme, Chapter 6 of

this volume). This is a problem in cross-national comparisons as well as in historical accounts of welfare state commitments. To take the example cited by Esping-Andersen (1990: 20), expenditures on unemployment benefits in the United Kingdom grew sharply in the 1980s under Thatcher, while replacement rates were sharply cut. Rising expenditure was a product of unemployment rising faster than spending cuts. The old-age ratio is perhaps even more relevant in terms of overall social spending. Every OECD country, except Ireland, experienced considerable growth in the over 65 to working-age ratio in recent years. Faster growth in the age ratio implies higher aggregate spending growth *and* lower aggregate GDP growth, *ceteris paribus*. Indeed, those studies that have adjusted spending for some indicator of the dependent population are less sanguine about the future generosity of the welfare state; though they often remain sanguine about its overall size (Clayton and Pontusson, 1998, Siegel, 2002; Castles, 2004; Lindert, 2004).

Different Rates and Sources of Economic Growth and Tax Structures

Another problem with most spending ratio data is that differences in economic growth rates lead spending ratios (spending/GDP) to diverge from real spending itself. Public spending ratios thus underestimate or overestimate inflation adjusted welfare expansion. This effect may be biased towards lowering spending ratios in less developed countries (Abramowitz, 1984).

Differences in the tax treatment of transfers (either due to special credits and exemptions, or simply to different tax rates) can distort the degree to which spending ratios translate into different real levels of state commitment to welfare. The role of taxation as an avenue for social transfers has been given more attention in recent years, but while the tax system is increasingly being used as a transfer mechanism—the United States' Earned Income Tax Credit and the United Kingdom's Working Families Tax Credit being notable examples—it can also be used to claw back social spending (Howard, 1997). Adema and Ladaique (2005) discuss extensively how gross expenditure can be offset considerably by the tax treatment of transfers (e.g. making benefits taxable), or by increasing consumption taxes.

Ireland and Sweden Compared

Based on various spending ratios, Ireland has experienced *much greater* welfare retrenchment after 1980 than has Sweden. The total spending/social expenditure ratio between 1980 and 2002 in Ireland fell from 47/17

per cent of GDP to 29/14 per cent. In Sweden the total spending ratio fell from 58 to 53 per cent of GDP, and the figure for social expenditure was stable at around 29 per cent of GDP. Important changes in demographics, economic performance, and tax structure suggest just how distorted spending data may be as an indicator of generosity.

First, in Ireland the dependency ratio dropped dramatically, while it increased dramatically in Sweden. The share of the Irish population over 65 was unchanged over the period 1980–2002 (at 11 per cent), while it grew slightly in Sweden (from 16 to 17 per cent). The share of under 15s in the population, moreover, declined from around 30 per cent to around 20 per cent in Ireland, and was largely unchanged in Sweden (around 19 per cent). Meanwhile, the employment rates of the two countries converged dramatically. In Ireland, *non-employment* as a share of the 15–64 population decreased from around 43 per cent to 34 per cent between the early 1980s and 2002, while in Sweden it increased from 21 per cent to 27 per cent. Thus, just adding these percentages together (admittedly a very rough gauge), Ireland went from experiencing an 84 per cent (!) non-employment rate in 1980 to 65 per cent non-employed in 2002, while in Sweden this proportion rose from 56 per cent to 63 per cent. So, for a constant share of GDP per non-employed person (again a crude indicator), we would expect the Irish spending ratio to fall by about one-third, while we would expect the Swedish ratio to rise by around 12 per cent. What we observe is a 38 per cent drop in total spending in Ireland (18 per cent social) and an 8 per cent drop in total (nil social) spending in Sweden.

Second, the difference in economic performance between the two countries during this period is also very large. The Irish economy grew at a *real* rate of 5 per cent per year in this period, while the Swedish economy grew at around 2 per cent per year. This partly follows from the demographic and employment trends cited above – a rising share of engaged labour should increase the economic growth rate – but growth has its own implications for what one might call fiscal capabilities. Compared to 1980, real public spending in Ireland was about 180 per cent higher in 2002, while it was just 140 per cent higher over the same period in Sweden. Hence, real spending actually grew twice as fast in Ireland, but since the economy grew about 2.5 times faster, the spending ratio declined.

Finally, the evolution of the tax burden on welfare benefits differs somewhat across the two regimes. Effective income tax rates and VAT rates are lower in Ireland, and those rates have declined since 1980 on lower earners (which is assumed to include benefit recipients). In Sweden, income tax rates on earners have also declined, though the overall system is somewhat less progressive than in the early 1980s. While VAT was reduced from 25 per cent to 21 per cent in Ireland, it has increased slightly (from 23.5 per cent

to 25 per cent) in Sweden. Reduced or zero VAT rates apply to many more necessities (e.g. food, social housing, utilities, public transit) in Ireland than Sweden (European Commission, 2005b). Adema and Ladaique's (2005) analysis seems to suggest that net spending in Sweden is considerably lower than gross spending due to tax clawbacks, while there are fewer clawbacks in Ireland.

The point of this example is not to suggest that Ireland has or will become a more generous welfare state than Sweden. It is worth noting that the gap between gross and net public transfer spending in the two countries has closed, particularly in Ireland, which may suggest some reduced generosity. These facts do suggest, however, that very large biases may be hidden in an indicator that is so often implicitly accepted as synonymous with the generosity of the welfare state.

From Welfare Spending to Welfare Rights

An alternative way to evaluate welfare commitments is an entitlement, or social rights, approach. This is essentially the approach advocated by Korpi (1989), promoted in Esping-Andersen's *Three Worlds of Welfare Capitalism*, and which has been widely embraced in comparative social policy (Hicks, 1999; Kitschelt, 2001; Castles, 2002; Green Pedersen and Haverland, 2002; Armingeon and Griger, 2006; Kenworthy, 2007). The approach is critical of spending as too unidimensional. However, it is still compatible with systematic data collection on more qualitative dimensions of welfare policies.

The essence of the rights or entitlements approach is its consideration of the generosity and universalism of social insurance. One can think of social insurance as a political commitment which alters the balance of power between segments of the population or economic classes. Such commitments have behavioural effects that extend beyond government outlays at any point in time. This is an important reason for presuming that spending captures only a small part of how welfare policy structures political outcomes. Welfare state generosity could thus be considered more important than spending when trying to understand the political economy of the welfare state. Viewing social insurance as a commitment is also useful for understanding more traditional questions in the sociology and economics of labour markets. Generosity of welfare commitments may affect things like quit rates, unemployment duration, labour market matching, and other micro-level phenomena that the level of spending will not.

There are various sources of data which systematically compare welfare state commitments at some level. However, all contain some important shortcomings compared with the approach in CWED. First, the OECD has

collected data on unemployment benefit programmes, starting with its *Jobs Study* in the mid 1990s (OECD, 1994). One of its data sets contains estimates of unemployment insurance replacement rates for a typical worker under a variety of scenarios. These data have been used in a number of large-n empirical studies of welfare retrenchment (e.g. Huber and Stephens, 2001). A major shortcoming of this data set is that the replacement rates ignore income and social security taxes, which considerably distorts comparative generosity. This data set also does not take into account important features of benefits, like population coverage or the qualifying conditions. More recently, the OECD has compiled net-of-tax replacement rate data, but these are available only for the mid to late 1990s onwards (OECD, 2005b). These data are also limited to unemployment insurance only, and thus provide a limited, albeit economically important, perspective on welfare insurance programmes.⁵

Other sources of replacement rate data are scattered in terms of countries and coverage. Hansen's *Elements of Social Security* (2002) includes other programmes (e.g. work injury) and income profiles, but only covers a few countries and only since the early to mid 1990s. The Nordic Social Statistical Committee publishes *Social Protection in the Nordic Countries* but this only covers relatively recent years, and only for the Nordic countries. A comparative study of old-age pension replacement rates was provided for the late 1980s by the European Commission (1993). Bamba (2005) also produces an index of 'decomposition', circa 1998, that incorporates recent data on replacement rates and qualifying conditions plus health data. Finally, SOFI's *Social Citizenship Indicators Project* covers the same programmes as CWED, plus work injury programmes. It is more historical, going back to the 1930s, but contains data only for every fifth year. Most importantly, however the data are not available to the scholarly community, despite the fact that they have been in existence for more than two decades.

The Comparative Welfare Entitlements Dataset, in contrast to these, contains annual information on replacement rates for unemployment, sickness and public pensions, as well as other eligibility criteria and the size of the insured population. (Plans for the future are to extend the data to cover child care and maternity benefits.) It currently covers 18 countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Ireland, Italy, Japan, the Netherlands, New Zealand, Norway, Sweden, Switzerland, the United Kingdom, and the United States, from around 1970 to 2002. Additional information for most countries is available for the 1960s, except for net replacement rates. The data are collected annually, although some extrapolations from bi-annual data have been made where annual data are not present.⁶ Table 7.1 provides a more specific breakdown of the particular characteristics in the data set that are used in the rest of the chapter.

Table 7.1 *Dimensions of the decommodification/generosity index*

Core programme	Programme characteristic	Definition
Unemployment	Single replacement rate	After tax benefit for single, fully insured 40-year old earning average production worker (APW) wage divided by after tax wage of employed APW
	Family replacement rate	After-tax benefit for a family of four (one APW earner, non-working spouse and two children) divided by after tax wage of employed APW
	Qualifying period	Weeks of insurance/employment required to qualify for benefit
	Waiting days	Number of days before benefits start
	Duration of benefit	Weeks of benefits payable for fully insured (single) 40 year old
Sickness benefit	Coverage ratio	Percentage of the labor force covered by unemployment insurance
	Single replacement rate	See definitions under unemployment insurance
	Family replacement rate	
	Qualifying period	
	Waiting days	
Retirement pension	Minimum replacement rate (single)	After tax replacement rate at retirement for single with no work history (or income)
	Minimum replacement rate (couple)	After tax replacement rate at retirement for couple with no work history (or income)
	Standard replacement rate (single)	After tax replacement rate for a single with a full work history (max 45 years) at APW wage

Table 7.1 (continued)

Core programme	Programme characteristic	Definition
	Standard replacement rate (couple)	After tax replacement rate for a couple with one full work history earner and spouse without work history
	Qualifying period	Years of insurance needed to qualify for single standard pension (defined above)
	Contribution ratio	Employee/Employer + Employee ratio of payroll taxes (at time pension is claimed)
	Take-up rate	Portion of population above retirement age receiving pension

The indicators included in the data set are combined to form indices following an aggregation procedure discussed later in the chapter. Conceptually, the measure is very similar to Esping-Andersen's decommodification index, but also contains some important modifications to the scoring which are detailed below.⁷

MEASURING WELFARE STATE GENEROSITY

This section outlines the basic approach to measuring generosity in CWED by discussing specific elements outlined in Table 7.1.

Programme replacement rates One of the most popular institutional indicators of the generosity of programme benefits is the income replacement rate, that is, the portion of income replaced by a social welfare programme (Esping-Andersen, 1990; cf. Whiteford, 1995; Korpi and Palme, 2003; Allan and Scruggs, 2004; OECD, 2004). Replacement rates provide some measure of the level of well being that is compensated by income transfer programs. CWED defines replacement rates as

$$(\text{Cash Benefits} - \text{Income Taxes})_{\text{out of work}} / (\text{Wages} - \text{Income Taxes})_{\text{in work}}$$

where income taxes include net social charges (compulsory contributions to social insurance programmes less cash transfers), and APW refers to the 'average production worker'.⁸ Since most working people in OECD countries live in families, CWED provides benefit replacement rate for a single worker and for a family, the latter defined as a household with a dependent spouse, two children and a head of household drawing the specified benefit. Benefits for families include child benefits, including means tested benefits.

Unemployment replacement rates Table 7.2 shows the development of replacement rates for unemployment benefits. The family types are (a) a fully insured single APW wage and (b) a fully insured head of household as described above. The worker receiving the benefit is assumed to be 40 years old, with 20 years of social insurance contributions. The table organizes countries based on (more or less) their traditional classification into Esping-Andersen's Three Worlds typology (cf. Scruggs and Allan, 2006a). In the 1970s and into the 1980s, there is a great deal of diversity across countries, not only for the level of the replacement rates, but also with respect to long-term trends. After the 1980s, replacement rates tend to be stable or decline as a share of APW wages. It is notable that the traditional *Three Worlds* perspective is not a great predictor of replacement rates (see Hu et al., 2006). There is considerable variation within regimes types (calling into question the idea of regimes).

Table 7.2 Unemployment benefit replacement rates

	Single replacement rates				Family replacement rates			
	1971	1980	1990	2002	1971	1980	1990	2002
<i>Liberal</i>								
Australia	21	26	30	26	40	52	66	65
Canada	32	60	66	60	44	64	70	72
Ireland	24	60	35	29	47	85	64	58
New Zealand	26	31	34	26	53	62	72	57
United Kingdom	55	46	20	18	72	63	36	55
United States	66	69	58	58	59	61	60	55
Japan	69	69	57	64	68	65	53	61
Switzerland	25	69	72	72	38	83	82	82
Mean	40	54	47	44	53	67	63	63
<i>Conservative</i>								
Austria	52	58	58	55	61	74	72	67
Belgium	53	67	64	66	63	67	60	61
France	43	68	70	70	35	61	63	73
Germany	63	68	63	60	78	70	70	72
Italy	10	4	20	45	20	13	29	61
Mean	44	53	55	59	51	57	59	67
<i>Social democratic</i>								
Denmark	87	78	68	59	90	81	73	64
Finland	40	34	63	57	48	48	73	67
Netherlands	87	86	74	78	94	89	78	77
Norway	52	70	68	65	68	75	73	72
Sweden	75	82	85	75	82	85	81	78
Mean	68	70	72	67	77	75	75	72

For the single-worker replacement rate, trends for the conservative and social-democratic welfare states converge, while both diverge from the liberal regime average.⁹ For family replacement rates, there is convergence among all three regimes, with considerable declines in the social-democratic countries in recent years, small declines in the liberal countries, and an increase in income replacement in the conservative countries (mostly due to Italy).

Sick pay replacement rates Table 7.3 shows trends in sick pay replacement rates for singles and families which correspond to those provided in the previous paragraph.¹⁰ A number of countries, especially among the social-democratic and liberal welfare regimes, have sickness replacement rates similar to those for the unemployed. However, it is generally true that the liberal welfare states tend to have lower sickness replacement rates than the

Table 7.3 *Sickness benefit replacement rates*

	Single replacement rates				Family replacement rates			
	1971	1980	1990	2002	1971	1980	1990	2002
<i>Liberal</i>								
Australia	21	28	30	26	40	56	66	65
Canada	0	60	66	60	0	64	70	72
Ireland	24	60	35	29	47	85	64	58
New Zealand	26	36	39	26	53	62	72	57
United Kingdom	55	46	28	22	72	63	32	26
United States	0	0	0	0	0	0	0	0
Japan	50	52	57	63	47	48	53	60
Switzerland	80	83	81	79	82	83	81	79
Mean	32	46	42	38	42	58	55	52
<i>Conservative</i>								
Austria	71	75	78	79	90	88	89	86
Belgium	66	88	89	85	66	84	90	88
France	58	59	63	62	61	61	66	63
Germany	100	100	100	92	100	100	100	93
Italy	74	68	74	77	80	72	80	87
Mean	74	78	81	79	79	81	85	84
<i>Social democratic</i>								
Denmark	65	78	68	59	70	81	73	64
Finland	68	41	88	72	83	54	88	75
Netherlands	87	86	74	78	94	89	78	77
Norway	52	100	100	100	68	100	100	100
Sweden	84	97	84	82	84	97	86	84
Mean	71	80	83	78	80	84	85	80

other countries, particularly if we count Japan, as Scruggs and Allan (2006a) and Bamba (2005) suggest, as a liberal country. While the majority of other countries have higher replacement rates for sick pay (e.g. Austria, Belgium, Finland, Germany, Italy, Norway, Sweden and Switzerland), a few have higher unemployment replacement rates.

Over time, trends are also similar to trends for unemployment replacement rates. There are a few cases where benefits increased in the 1970s (Canada, Norway, Ireland, Finland), but since the mid to late 1980s, declining replacement rates dominate. Not surprisingly, patterns for family replacement rates also resemble what we saw for unemployment. Preferential benefits for male breadwinner households are no more predominant in conservative welfare states than they are in the countries in the other two regime types.

Table 7.4 Evolution of social pension replacement rates

	Single replacement rates				Couple replacement rates			
	1971	1980	1990	2002	1971	1980	1990	2002
<i>Liberal</i>								
Australia	24	28	32	30	39	42	49	45
Canada	28	30	42	40	49	49	59	57
Ireland	23	31	36	35	20	36	43	45
New Zealand	6	19	17	23	11	36	32	45
United Kingdom	26	40	41	39	43	57	65	59
United States	37	39	40	38	43	44	45	43
Japan	31	37	31	37	35	43	36	42
Switzerland	0	40	42	39	0	47	49	45
Mean	22	33	35	35	30	44	47	47
<i>Conservative</i>								
Austria	47	42	41	50	51	49	47	56
Belgium	15	27	37	39	19	30	36	39
France	23	35	45	42	37	60	68	65
Germany	18	17	18	18	29	27	26	23
Italy	12	19	27	30	24	34	51	53
Mean	23	28	33	36	32	40	46	47
<i>Social democratic</i>								
Denmark	47	42	51	44	53	57	61	57
Finland	33	42	38	31	48	61	53	48
Netherlands	46	55	48	51	55	67	58	60
Norway	39	42	43	44	52	56	56	66
Sweden	35	43	43	35	45	53	53	43
Mean	40	45	45	41	51	59	56	55

Social pension replacement rates Table 7.4 shows trends in the replacement rates for social pensions. Social pensions here are defined as the pension benefits for someone without a work history. Where there is not an explicit minimum-age pension, we substitute social assistance. We assume that a 'family' consists of two people of retirement age with no children at home. The results indicate that social-democratic regimes have generally provided the most generous social pensions. However, the generosity of these benefits has fallen considerably since the 1980s in Finland and Sweden, and, to a lesser extent, in Denmark and the Netherlands. (Norway is the exception: benefits went up in the late 1990s.) In the early 2000s, Canadian and New Zealand's social pensions replace as large a percentage of work income as do the social pensions in four of five social-democratic countries.

Among the conservative regimes, replacement rates do not uniformly support the idea of strictly income linked benefits. While social pensions are low in some countries (Germany, Italy and Belgium), they are among the highest in others (Austria and France, particularly in recent years). Except for Germany, replacement rates have tended to either remain stable or increase. Replacement rates in countries which belong to the liberal regime are also not uniformly low. As mentioned, Canada and New Zealand have comparatively high replacement rates for those relying solely on public pensions. Differences between single's and couple's rates are large in most social democratic countries, as well as in Canada, France and the Antipodes.

Across regime types, there is *convergence* in replacement rates. Benefits in conservative and liberal countries have tended to rise toward the social-democratic countries. Replacement rates in social-democratic countries have declined since the 1980s.

Standard pension replacement rates The last set of replacement rates presented is for standard public pensions (Table 7.5). These benefits are typically based on a long history of earnings, and are computed using rules governing the treatment of past wages, accrual rates, and calculation formulas. We make a common set of assumptions about career earnings. The basic assumptions are that the notional worker retired on 1 April of the year in question, worked a full career (or from age 21 to retirement age), and earned the APW in each year.¹¹ Family benefits refer to a retired couple with one lifetime APW wage beneficiary and a spouse with no work history. Replacement rates and conditions refer only to public pensions, which excludes some pension benefits that are based on sectoral or industrial relations agreements. (For example, supplementary pension regimes in France are excluded.)

With a few exceptions, such as the United States, Italy, and Norway, public pension systems fully matured in the 1980s. Since then, replacement rates have either declined or levelled off. Conservative countries are typically the most generous. Liberal countries have converged only slightly up toward social-democratic countries. This is due to Japan and Switzerland, which both start the period with very small, public pensions. In the social-democratic regimes, there was general replacement rate growth in Norway and Finland, and some expansion of couple's pensions, specifically in the Netherlands. In these countries we can observe that first retirees who have paid the expected number of years into second pension pillars have started to be rewarded by higher replacement ratios in the more recent past.

Across the three regimes, standard pension replacement rates for singles have not converged much over time. The picture with respect to couples,

Table 7.5 Evolution of standard pension replacement rates

	Single replacement rates				Couple replacement rates			
	1971	1980	1990	2002	1971	1980	1990	2002
<i>Liberal</i>								
Australia	24	28	32	30	39	42	49	45
Canada	33	40	54	52	54	58	69	67
Ireland	27	36	42	38	40	49	58	53
New Zealand	16	51	51	50	21	72	73	77
United Kingdom	26	40	41	39	43	57	65	59
United States	22	41	39	42	33	55	51	55
Japan	27	37	47	55	35	48	56	57
Switzerland	44	56	57	60	59	74	78	77
Mean	27	41	45	46	40	57	62	61
<i>Conservative</i>								
Austria	85	80	82	85	75	68	70	67
Belgium	76	81	80	74	79	82	77	72
France	46	59	60	52	37	68	69	58
Germany	75	77	75	72	68	65	62	52
Italy	69	61	72	93	64	55	69	81
Mean	70	72	74	75	65	68	69	66
<i>Social democratic</i>								
Denmark	51	46	54	52	58	59	64	63
Finland	40	57	67	63	42	57	64	64
Netherlands	42	55	48	51	51	67	58	60
Norway	42	55	63	61	51	56	61	65
Sweden	50	64	65	59	61	68	68	61
Mean	45	55	59	57	53	61	63	62

however, is different. There has been considerable convergence. This can be explained by the absence of dependant's supplements in conservative countries, and an increase in such supplements in liberal and social-democratic countries.

Despite its liberal welfare state label, the US Social Security programme is a curious blend of non-liberal attributes. APW replacement rates compare favourably with other countries, the benefit formula is very progressive (those with lower lifetime earnings receive larger replacement rates), and the supplement for a non-insured spouse is 50 per cent of the breadwinner's pension. (This supplement is *much* larger than the supplement in any other country.) The main shortcoming of Social Security is that pensions in force have always been indexed to prices rather than wages, something that many pension reforms are moving increasingly towards. The progressive

replacement rate formula and earnings ceiling combine to provide a relatively low maximum (compared with many continental systems), which is consistent with company schemes for middle and upper-income households.

Across these four programme and two household types, there are some obvious patterns. First, there is considerable convergence in replacement rates for single-earner families. In general, benefits have converged upwards over the entire period since the 1970s, but with clear evidence of cuts after the mid to late 1980s, particularly in previously high-benefit systems. Whether programmatic cuts constitute true retrenchment is, of course, partly dependent on the definition of retrenchment and the historical time scale used. For example, virtually all programme replacement rates were more generous in 2002 than they were in the early 1970s in terms of expected worker pay. In terms of real purchasing power (adjusted for inflation), they are even more generous. Finally, the results consistently suggest little evidence for the conventional wisdom that social insurance is much more 'family friendly' in conservative, continental European countries.

Social Insurance Coverage

Coverage is considered an important indication of the universalization of benefit eligibility, something that Korpi (1989) argues characterizes mature welfare states. In their seminal article on the development of welfare states, Flora and Alber (1982) included the portion of workers covered by social insurance (see also Flora, 1987b). This subsection discusses coverage trends for unemployment and sickness schemes, and the portion of the population above retirement age in receipt of a public pension. The available data for many countries go back to 1960, so we can assess longer trends than we could when discussing replacement rates. Generally speaking, the results suggest that coverage increased substantially in the 1960s and 1970s, and has not declined since then (with the exception of two aberrant cases).

Unemployment and sickness insurance coverage Table 7.6 provides trends in the coverage of unemployment and sickness insurance as a portion of the labour force in 1960 and 2000, as well as changes by decade. A few important notes about these data are worth mentioning. First, coverage for the Antipodean countries is listed as 100 per cent here even though these benefits are mean tested and subject to a residency requirement.¹² Second, coverage in some countries with Ghent unemployment insurance systems (Denmark, Finland, and Sweden) is based on enrolment in union unemployment funds, even though these are not technically

Table 7.6 Evolution of unemployment and sickness insurance coverage

	1960		2002		Unemployment Change:				Sickness Change:			
					1960-1970	1970-1980	1980-1990	1990-2002	1960-1970	1970-1980	1980-1990	1990-2002
	1960	2002	1960	2002	1960-1970	1970-1980	1980-1990	1990-2002	1960-1970	1970-1980	1980-1990	1990-2002
Australia	100	100	0	0	0	0	0	0	0	0	0	0
Canada	63	79	15	1	16	-3	2	2	0	80	-3	2
Ireland	58	100	42	9	2	4	29	4	58	100	42	27
Japan	31	50	19	11	3	5	0	41	58	17	13	-1
New Zealand	100	100	0	0	0	0	0	100	100	0	0	0
Switzerland	23	84	61	-6	75	-4	-4	93	18	-75	16	-76
United Kingdom	81	86	5	-5	15	-10	5	81	86	5	-5	6
United States	61	88	27	4	17	3	5	16	16	1	0	-1
Mean	65	86						66	56			
sd	29	17						36	38			
COV	0.44	0.20						0.55	0.69			
Austria	55	67	12	7.0	7	-1	-1.6	75	84	9	4	-1
Belgium	75	84	10	7	3	-1	0	85	83	-2	5	-23
France	40	58	18	17	0	-1	2	95	95	0	0	0
Germany	59	68	9	12	4	-2	-2	82	88	6	3	1
Italy	37	59	22	8	1	6	6	56	66	10	8	2
Mean	53	67						78	83			
sd	15	11						15	11			
COV	0.29	0.16						0.19	0.13			
Denmark	35	83	49	0	27	10	12	77	99	22	8	12
Finland	13	74	61	23	26	5	7	0	100	100	100	0
Netherlands	82	89	7	5	1	1	-1	82	89	7	5	1

Norway	59	93	34	3	24	3	4	74	100	26	6	20	0	0
Sweden	36	85	49	19	17	6	8	94	93	-1	8	-1	2	-10
Mean	45	85						66	96					
sd	26	7						37	5					
COV	0.59	0.09						0.57	0.05					
All countries														
Mean	49	68						59	68					
sd	28	31						36	35					
COV	0.58	0.46						0.61	0.51					

compulsory insurance systems. Third, the measure of 'coverage' in the population used here is often higher than the portion of the unemployed receiving benefits. While it does reflect the degree of universalism, it can be criticized for being insensitive to the fact that being in unemployment is not a random draw from the covered population, and those in unemployment may have exhausted insurance benefits or failed to qualify due to incomplete contribution records.

Unemployment insurance coverage increased considerably during the 1960s and early 1970s in most countries, and has gradually converged towards full coverage over time. This can be seen in the rising average coverage rates and falling coefficients of variation (standard deviation/mean) in the table. The extension of coverage was typically the result of major legislative changes. Coverage in a number of conservative regimes tended to stagnate after the 1970s.

Coverage for sickness cash benefits resembles unemployment insurance coverage in the liberal countries (recall that the two systems have the same or similarly structured replacement rates), except for the United States.¹³ In the non-liberal countries, coverage is generally higher, sometimes much higher, than unemployment insurance coverage. Finland instituted sickness insurance only in the 1960s, and Canada did so only in 1971. Switzerland and Belgium both saw considerable declines in coverage in the 1990s. The reason for the decline in coverage in Belgium is not clear. In Switzerland, excluding sickness coverage in health insurance policies became easier, and coverage rates (official ones at least) declined considerably.

Pension take-up rates Unlike coverage of sickness or unemployment insurance, the indicator of the universalism of the pension system is the portion of people above retirement age drawing a public pension.¹⁴ Surprisingly, estimating the portion of those over retirement age who are in receipt of a pension presents considerable challenges in many countries. Most countries report the number of pensions paid in their statistical year-books. However, many countries have more than one large public pension scheme, and statistics are often compiled for the number of pensions being paid, not the number of people getting pensions. This problem is considerable in conservative regimes such as Italy and France (and even Germany) which have multiple sectoral pension funds.

A second complication in determining pension take-up rates is that some countries have administratively separate pension systems for civil servants. The number of civil servant pensions is published for most of these countries, but civil servants are often allowed to retire early and may be entitled to other public pensions (through different jobs), making it difficult to consistently avoid under- or over-counts. For example, in the United States, federal employees hired since 1984 have been insured in

Table 7.7 Pension take-up rates

1960	2000	Change 1960–2000	Change 1960–1970	Change 1970–1980	Change 1980–1990	Change 1990–2000
52	68	16	11	21	–15	–1
59	87	28	19	8	0	1
	101	101	75	15	–1	12
99	97	–2	–7	3	0	1
82	101	19	2	14	2	1
97	100	3	0	4	0	–1
	0				–2	–96
	100				5	8
87	96	9	1	–2	8	2
	99					0
32	127	95	23	39	–3	35
80	107	27	0	2	24	1
88	102	14	15	–4	1	2
71	95	24	2	22	2	–3
100	100	5	3	–2	1	3
100	100	0	0	0	0	0
79	104	25	10	5	6	5
72	93	21	18	4	0	–1

the Social Security system, and many (but still not all) state and local government pension systems have also been integrated into the national Social Security system.

The final complication involves the treatment of spouses. Where second earners have had sufficient labour force participation, they will receive a public pension like any other worker. Where spouses have not worked, they would presumably be entitled to a social pension save for the fact that their household pension income (i.e. from their spouse) is higher than the income test. Alternatively, they may be (indirect) beneficiaries of a pension supplement in the name of their spouse.

These complications should generate some caution at the margins when interpreting the results in Table 7.7. We attempted to compile total pensioner information that includes as much information as possible and avoids double counting. In some cases, we rely on information from public surveys that ask about sources of income. Our results suggest that most of the countries considered here attained something approaching universal take-up, once social pensions, public pensions and civil service pensions are all considered. With a couple of exceptions – e.g. Australia moved towards and then away from making its state pension universal – pensions

are more universal today than they were 30 or 40 years ago. Most of the coverage expansion occurred prior to 1980.

Other Eligibility Conditions

Here, briefly, I present and introduce the definitions of the other main conditions of eligibility for social insurance included in CWED.

Duration of benefits for sickness and unemployment refers to the number of weeks for which a qualified worker is allowed to collect regular benefits. In some countries, the duration of benefit is a function of the length of insurance coverage and age. The period given in the data refers to benefits available to the 'notional worker': age 40, and 20 years of insurance. There are a few cases where the unemployed are entitled to smaller benefits after the full benefit period (e.g. France). Coding refers to the full duration, including the 'reduced benefit' period.

Qualifying conditions for unemployment and sickness benefits refer to the number of weeks one must be insured to qualify for cash benefits of the specified duration. In some cases, this qualifying period is longer than the minimum insurance period required to receive *some* benefit.¹⁵

Waiting days refers to the number of weekdays after becoming unemployed/sick that one must wait for specified benefits. Thus, for example, it counts the waiting period for earnings related benefits in the UK in the 1970s, even though the (much smaller) flat rate portion of the benefit was payable earlier.

The *qualifying conditions* for pension benefits are based on a slightly different set of criteria than those for unemployment and sickness. First, the *qualifying period* refers to the number of years of insurance necessary to qualify for the standard pension. This is coded to be consistent with the maximum number of years of actual contributions (assuming a full working life and the age of the public pension scheme), since pension *replacement rates* are based on what one actually receives based on the programme rules.

An example may be helpful. A pension system pays 2 per cent of final salary for each year of coverage up to 40 years of coverage. The pension law is passed in 1960, going into effect in 1961. Someone retiring in 1970 would retire with 18 per cent of final salary (nine years at 2 per cent per year), and has a qualifying period of nine years. Someone retiring in 1971 has a 20 per cent replacement rate and a qualifying period of ten years. Someone retiring in 2002 would have an 80 per cent replacement rate and a maximum (40-year) qualifying period. If the pension system grandfathered in workers (a common occurrence), it might, for example, guarantee 20 per cent of final salary for anyone paying in all years that the programme

existed. Applying that rule to the above example, someone retiring in 1970 would have a 20 per cent replacement rate, and a nine-year qualifying period, while someone retiring in 1971 would have 20 per cent replacement rate and a ten-year qualifying period.

The final eligibility condition is the *funding ratio*: employee social contribution to the combined employee/employer contribution. If the contribution is borne solely by employees, the funding ratio is 1, if it is borne equally by employees and employers it is 0.5.¹⁶

COMPARATIVE INDICES OF WELFARE GENEROSITY

A widely used empirical indicator of the generosity of welfare benefits is Esping-Andersen's (1990) decommodification index. It is based on a several important characteristics of unemployment, sickness, and pension programmes, including the replacement rate for singles and coverage/take-up ratios discussed in the last two sections of this chapter. The decommodification index has been used as an indicator of benefit generosity in numerous empirical studies (Huber et al., 1993; Pampel and Gartner, 1995; Rosenfeld and Birkelund, 1995; Mitchell, 1996; Western, 1996; Messner and Rosenfeld, 1997; Crepaz, 1998; Iversen, 1998; Hicks, 1999; Kenworthy, 1999; Huber and Stephens, 2001; Pampel and Williamson, 2001; Radcliff, 2001; Swank, 2002b).

The original index was created by assigning each country a score for each characteristic in Table 7.1 based on the cross-country mean and standard deviation for that characteristic, except for coverage and take-up: '1' if less than a standard deviation from the mean, '2' if within one standard deviation, and '3' if greater than one standard deviation. (The points are reversed for characteristics where a higher score implies a more stringent condition.) For each of the three main social insurance programmes, replacement rate scores are doubled, and the other characteristic scores summed. That total is multiplied by the coverage (take-up) rate.

CWED data allow us to compute a decommodification score for each year: 1971–2002. For each characteristic, the 18-country mean and standard deviation in 1980 is used as the benchmark against which values for all other country-years are scored. The choice of 1980 as a benchmark is ultimately arbitrary, but since this is the year used in the original index, it is convenient for comparative purposes. The overall index, and the index for each of the three component programmes, are presented in Figure 7.1 (the overall score uses the right-side scale). The figure suggests remarkable stability over the period with little evidence of retrenchment anywhere except in Switzerland,

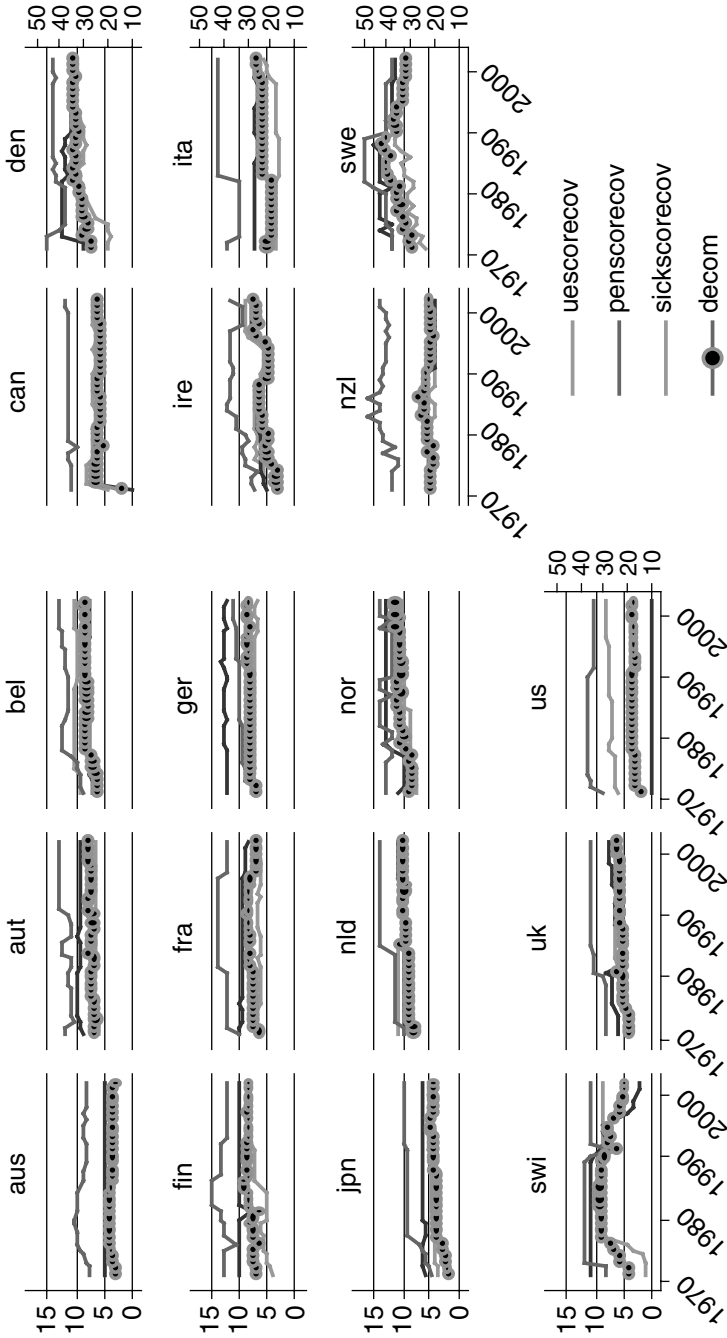


Figure 7.1 Comparative welfare state entitlements decommodification indices (Esping-Andersen scoring method)

Sweden and New Zealand. This stability would seem to back up claims that welfare states have been very resilient to pressures for retrenchment.

How do we explain this apparent stability? One explanation is simply that welfare state programmes have, in fact, not changed much. Another is that the system of scoring welfare state characteristics is flawed. The scoring is, in fact, insensitive even to very large changes in the values for underlying characteristics. For example, Belgium's unemployment replacement rate moved up from 55 per cent to 70 per cent (0.8 standard deviations (sd) spanning the mean) between the early 1970s and mid 1980s. It then falls to around 62 per cent by 2001 (a decline of about 0.4 sd). This would seem like a clear case of expansion and retrenchment over time. However, because the replacement rate never exceeds one standard deviation above or below the cross-country mean in 1980, the scoring procedure from *Three Worlds* scores it a '2' for all years.

A related problem with the scoring procedure is that, in a given year, a country with consistent, but moderately generous, programme values is scored the same as a country with consistently low values. For example, assuming full insurance coverage for the working population, a country that is -0.9 sd for all of four characteristics of sickness insurance would receive a sickness declassification score of 10; this is the same score that a country scoring $+0.9$ sd from the mean on all characteristics would receive. Meanwhile, two countries scoring only marginally higher/lower – e.g. 1.05 sd from the mean – for all characteristics would receive the maximum/minimum score (15 and 5 respectively).

A more appropriate, and perhaps more intuitive, scoring method is to use the standardized values for each characteristic (except for coverage). Comparing countries on a continuous scale eliminates most of the problems referred to above. First, it avoids arbitrary assignment of cut-off points. While one might argue that this *artificially* creates a continuum, this is superior to arbitrary assignment of cut-off points at $+/-$ one standard deviation, and prevents any accidental bias in the creation of cutpoints (see Scruggs and Allan, 2006a, b). The revision scores a country that is 'moderately high' on many characteristics higher than one that is moderately low on all, while countries that are 'a bit high' on some and 'a bit low' on others score like those with a programme that is in the middle on all attributes. The upper and lower bounds of the scores are also truncated to be $+/- 2$ in order to prevent an outlying score from distorting the overall distribution of results. Furthermore, in order to make all scores positive, we add 2 to each characteristic score so it varies between 0 and 4 rather than -2 and 2.

The final major change to the declassification scoring is the use of replacement rates for the family household type. This is justified largely on the basis that the vast majority of workers are members of families, not

Table 7.8 *Decommodification and generosity scores, 1980*

Decommodification score from <i>Three Worlds</i> *		Benefit generosity & difference in ranks	
Sweden	39.1	Sweden	42.3 (0)
Norway	38.3	Norway	38.4 (0)
Denmark	38.1	Denmark	37.2 (0)
Belgium	32.4	Netherlands	35.9 (1)
Netherlands	32.4	Belgium	31.3 (-1)
Austria	31.1	Switzerland	31.2 (1)
Switzerland	29.8	France	30.3 (3)
Finland	29.2	Germany	29.1 (1)
Germany	27.7	Austria	27.8 (-3)
France	27.5	Finland	27.4 (-2)
Japan	27.3	New Zealand	26.2 (5)
Italy	24.1	Canada	21.2 (3)
United Kingdom	23.4	Ireland	21.2 (1)
Ireland	23.3	Australia	19.3 (4)
Canada	22.0	United States	19.3 (2)
New Zealand	17.1	United Kingdom	18.7 (-3)
United States	13.8	Italy	17.8 (-5)
Australia	13.0	Japan	17.4 (-7)

* Scores in our replication of the decommodification index differ from these.

single workers. To make the relative weights of replacement rate scores similar to those in the decommodification index, the replacement rate scores for singles and couples are no longer weighted twice.

It is not clear whether *Three Worlds* relied only on single-worker replacement rates because the family rates were not available in the SCIP data at the time, or due to a fear that generous spousal supplements in 'conservative' welfare systems in continental Europe would conflate state support for traditional family structures with decommodification. In the latter case, there are two problems. First, it is not clear that support for traditional families does not in fact constitute decommodification, albeit not the post-feminist social-democratic one. Second, based on the replacement rate data in the previous Tables 7.2–7.5, there is really no evidence that conservative regimes have higher family replacement rates due to family supplements.

The revised decommodification scores, which we label the benefit 'generosity index', are plotted with scores computed using the original methodology (right-axis scale), and as presented in Figure 7.2. Table 7.8 shows the generosity index scores in 1980 alongside Esping-Andersen's decommodification data in order to provide a direct comparison between the CWED

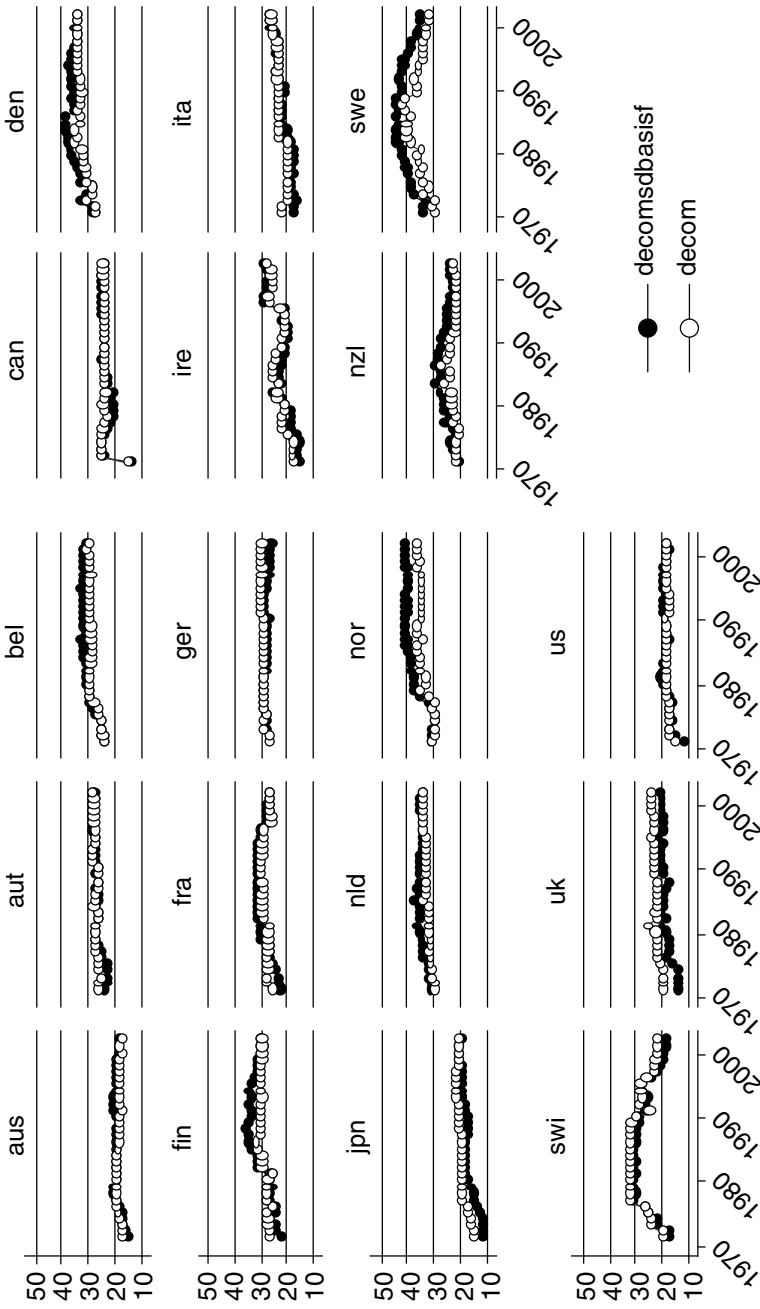


Figure 7.2 Revised (and original) decommodification scores for 18 OECD countries

results and the original decommodification index from the *Three Worlds of Welfare Capitalism*.

Table 7.8 suggests, unsurprisingly, some basic correspondence with the original decommodification index. First, three paradigmatic cases – Germany, Sweden, and the United States – all fit squarely with the *Three Worlds* ranking. A number of other cases also fit: Australia, Austria, France, and the other Scandinavian countries. Several anomalies in the original text – Switzerland and Finland – also appear anomalous in our ranking. On the other hand, Japan and Italy emerge in our rankings as extremely liberal countries, with low decommodification scores, while New Zealand appears much more decommodifying than suggested by the decommodification index. The Netherlands' score is also much closer to those in the Scandinavian countries in the generosity index than it is to scores among the conservative regimes like those of France and Germany. Finally, the overall variation among countries, particularly the distance between the best and worst, is much less in the generosity index than the decommodification index. Such an exaggerated variance between liberal and social-democratic countries suggests that existing empirical relationships employing decommodification as an explanatory variable, or as a variable thought to determine decommodification, are overstated.

Are these differences very significant substantively? Consider that the original decommodification index suggested anomalous results for two of the 18 cases – Switzerland and Finland. Our results suggest anomalous results for at least four additional countries – Japan, Italy, the Netherlands and New Zealand. Since our approach replicates the same underlying data that were used in *Three Worlds*, the results are not a case of employing different indicators. Presumably, our results should be almost identical. This implies a much more tenuous fit between theory and reality than comparative social policy scholars have been led to believe.

Compared with the original scoring method, our revised measures show a much greater degree of variation *within many countries* over time. Figure 7.2 suggests that the 1970s were an important period of increased benefit generosity in many countries, that the 1980s saw somewhat more expansion in generosity, and that the 1990s were a period of reduced generosity. All of the social-democratic regimes, except for oil-fed Norway, were considerably less generous in 2002 than they were in the mid 1980s. Germany, France, and, to a lesser extent, Belgium, also appeared to be substantially less generous in 2002 than they were in the mid 1980s. This contrasts with spending figures; which were higher in 2002 in all three countries. The trends also indicate that a number of countries were less generous in 2002 than at any other time in the last generation. Indeed, Germany, despite its reputation as a welfare reform laggard, is probably the only country that has been consistently less generous

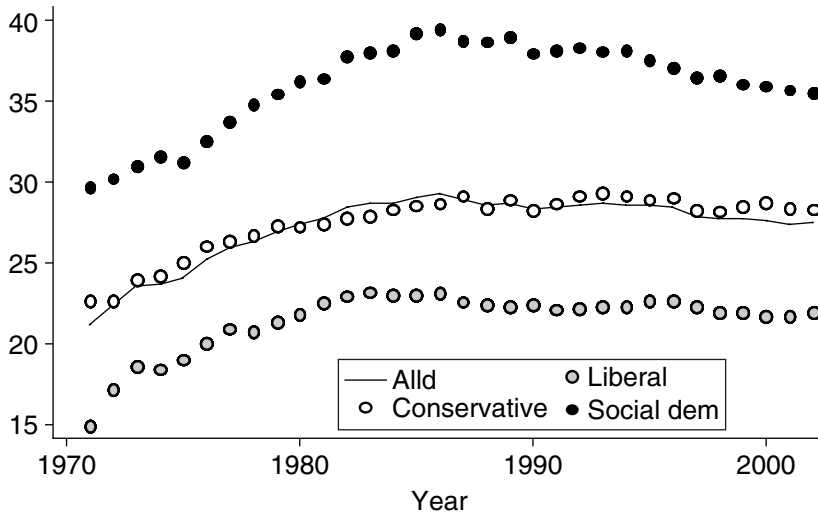


Figure 7.3 Revised decommodification scores: mean values, 1970–2001

over the period covered (see Siegel, 2004 for corroborating evidence). Explanations of this greater within-country variation should be of intrinsic interest to researchers. One immediate implication for the comparative social policy field is that one should be careful in assuming that cross-sectional variations at a single point in time are valid much earlier or later in time. Unfortunately, a great deal of work since *Three Worlds* has tended to do that.

Figure 7.3 provides the average annual scores for all countries and annual averages in each welfare regime (which we have defined here according to the conventional classification, not the one implied by our results in Table 7.8). It mirrors patterns described in earlier sections of the chapter. By the early 21st century, generosity is higher than it was in the 1970s. Programme generosity tends to increase until the mid to late 1980s, after which there are some clear signs of retrenchment. This is especially true in the more generous social democracies. Finally, after diverging throughout the 1980s, scores seem to converge through 2002.

CONCLUSION: WHY AND HOW TO IMPROVE THE CONCEPTUALIZATION AND MEASUREMENT OF WELFARE STATE PERFORMANCE

The CWED project data set is far from a complete or ideal set of data on welfare programme commitments. What the project has attempted, and

what should be repeated in all future empirical research on comparative welfare states, is to expand the scope of comparison by trying to measure things in a systematic way across as many countries and years as possible. Here we make several specific suggestions that can further improve the depth, quality, and comparability of social rights and welfare state institutions across countries and time.

First, social entitlements need to consider two-earner households and single-parent households. Two-earner households have become the norm in most countries, and single-parent households are a particularly vulnerable group in all countries. We have a very patchy view of how the tax and benefit systems have interacted to affect such groups over time. In addressing this issue, there are several complications that have to be confronted. The experience with the CWED suggests that computing replacement rates and benefit conditions for these alternative households will be a considerable task, but that it can be accomplished more easily if it builds on the current project.

Second, the distributional effects of welfare programmes across income and social strata are not well examined. The generosity measure is computed for a single point in the wage distribution (the average production worker). However, two policies affecting a typical worker in the same way may affect lower (or higher) earners differently.¹⁷ Are cuts affecting some income groups and not others? Are those affected actually groups that should be targeted for cuts?

Third, there is still a paucity of systematic comparative data on alternative types of social rights. Indeed, virtually none of the literature considers the adequacy of social assistance, arguably the ultimate social floor of the welfare state. Other substantive improvement in data might stress the role of non-cash benefits and services. Still other improvements would take up programmes that provide a more balanced gender orientation to welfare state programmes, such as assistance programmes targeted at single parents.

While most of the above criticisms are not new, work on the next step – specifying new conceptions and associated observables – tends to lag. There is, for example, little work comparing social assistance programmes in the context of social rights and income replacement (Gough et al., 1997 is an exception, but even it primarily looks at spending levels). ‘Universal services’ are often casually asserted as affecting outcomes, but a systematic comparative evaluation of the distributive implications of such services is needed to include them in the analysis alongside cash benefits.

Finally, and apropos of the three previous suggestions, we simply need more data across space and time using the main concepts that we already focus on in the literature. Too often, interesting contemporaneous indicators

are developed to investigate (or define) some 'new or 'nationally unique' social problems. Appellations such as 'new' and 'unique' imply comparison to other things, yet those comparisons tend to stop at the preliminary stages, just when the necessary work to be done is identified.

The importance of establishing appropriate measures of generosity is vital. Social scientists tend to focus overwhelmingly on creating 'new theory' (or modifying old theories). Developing tests of theories is typically an afterthought, or worse, 'test development' becomes an exercise in a search for measures to show that this is consistent with a theory. Why it is so would fill its own volume (or, on second thoughts, may be all too obvious) but it is scientifically pathological. *If concepts in a theoretical model are poorly operationalized and measured, then we are not testing the theoretical model.* One might argue that social science contains very abstract, hard to operationalize concepts and theories, so asking for precise correspondence (compared with, say, physics) is unreasonable. This has it completely backwards. Errors probably proliferate exponentially: sloppily operationalizing sloppily theorized relationships does not make a 'good test'. The time spent conceptualizing and operationalizing should vary *inversely* with theoretical specificity (allowing of course for some minimal standard of both). Vague theories demand more attention to concept validity and better measurement, they do not justify worse concept validity and measurement.

Mis-measurement produces well known problems in statistical analysis, including bias and inconsistency of parameter estimates (Gujarati, 2003: 526–7). Since many claims about causal relationships are based on inferences drawn from statistical analysis, measurement problems may fundamentally undermine existing 'facts'. For example, a great deal of recent empirical research relies on spending data to infer that the relationship between political partisanship and welfare state generosity has largely disappeared in the 'era of permanent austerity'. If, however, as many (perhaps most) social policy scholars suggest, spending data have always been poor indicators of welfare generosity – or if previous non-spending indicators of generosity (e.g. the decommodification index) are mis-measured – then empirical results are 'biased' against finding any effects. Indeed, results with non-spending programme data as discussed by Allan and Scruggs (2004) or Korpi and Palme (2003) suggest that partisanship continues to play a big role in explaining variations in welfare state generosity in the era of austerity.¹⁸ The emperor has no new clothes.

Mis-measurement presents even more serious problems when it occurs on the right-hand side of a regression equation. There it produces not just inefficient but also biased estimates of the effects of generosity as a causal variable. This suggests that the estimated effects of social spending in numerous statistical models throughout comparative political economy

more broadly may provide biased and inconsistent impressions of the actual effects of welfare generosity on a wide variety of social processes. The size and direction of these biases is unknown until those results are corrected with more a more accurate operationalization of generosity.

NOTES

1. The most obvious shortcoming of current data sets and theory evaluation (including the current one) is that the 'sample' remains so selective: less than 20 of the same Western, industrial countries. This tends to make the idea that any new explanation is really 'tested' (rather than simply derived from them given how much we already know) quite problematic. There is, furthermore, almost no effort to use results to make any predictions which are evaluated against evidence. If the explanations were very good, for example, then they should help us (a) predict results in other countries, (b) predict future trends in the current countries under study or even (c) 'postdict' the past.
2. The problem is more than simply that the social science data are observational; the available data are generally collected for some alternative purpose than testing the theory that the social researcher wants to test.
3. Pierson (1994), to cite one of the most influential examples, exemplifies the practices alluded to here. First, his argument often relies considerably on spending trends to bolster his central argument that there has not really been welfare state retrenchment. Second, the book provides 'new' political explanations with a selective consideration of 'old' explanations which might predict results with the same success (Scarborough, 2000). Third, and perhaps most problematic, this line of explanation bears no resemblance to a progressive research programme. The 'new politics' explanation makes no real effort to explain both past and present under the same model, it asserts that the present requires a different one.
4. Kittel and Winner (2005) suggest that this variation is limited, making pooled statistical analysis of these data potentially problematic. The need for cross-sectional, historical data sets exists independently of questions about how to analyse them.
5. One advantage of the OECD indices is that, in recent years, they incorporate housing benefits and social welfare (the latter available after insurance benefits are exhausted) replacement rates.
6. Detailed sources and procedures are available in the data set codebook. Some uses of the SCIP data (e.g. Sjöberg, 2000b) extrapolate between the five-year interval data points to get 'annual' data, or else (e.g. Korpi and Palme, 2003) rely on some other extrapolation procedure to get annualized data.
7. A comparison of our and Esping-Andersen's results is provided in Scruggs and Allan (2006a). Analyses using the index are Armingeon and Griger (2006), Pontusson (2006), Hu et al. (2006), Scruggs and Allan (2006b), Scruggs (2005).
8. APW wages and taxes used here are provided, with modifications, by OECD publications like *Taxing Wages*. We use net benefits payable in the first six months of receipt. Further details of the calculation of replacement rates computed in each country and year are provided online in the data set codebook and replacement rate spreadsheets at the CWED website.
9. The current placement of some countries – especially Ireland, Japan, the Netherlands and Japan – is very ambiguous (compare, for example, Esping-Andersen, 1990; Castles and Mitchell, 1993; Esping-Andersen, 1999; Scruggs and Allan, 2006a; Shalev, forthcoming) but not substantively important here.
10. It is important to note that we include periods in which employers are legally obligated to pay wages to absent employees as insurance payments in computing replacement rates. We do so based on the assumption that such payments constitute legal rights.

11. Using the wage in the year of retirement as the 'historic insured wage' leads to a considerable overstatement of actual benefits. Many pension systems were designed to be based on a long earnings history, a fact obscured by comparatively long transitional periods which were especially favorable to transitional workers. This describes *inter alia* the (pre-reform) Swedish and American systems. Further details of specific cases are available in the CWED data (also see Scruggs and Allan, 2006a).
12. When we discuss the generosity index, we will assume coverage is 50 per cent to account for the fact that the benefits are purely means tested.
13. The non-zero values for the United States in the table reflect the fact that five states, including California and New York, have sick pay insurance modelled on their unemployment insurance systems. Another important factor to consider in interpreting all of the sickness benefit data is that most (though certainly not all) Americans receive some paid sick leave each year from their employer.
14. Coverage rates for public pensions (the proportion of people in the work force who are insured) are hard to evaluate, because many people not in the labor market at a given time can be entitled to a pension based on their previous work history. Thus, the number of people paying social contributions for pensions in a given year (or counted as entitled to a pension) may be a considerable undercount.
15. Empirically, our coding choice tends to give 'length of service' rules comparatively long contribution periods and benefit durations. This seems to be consistent with how such systems are comparatively characterized in the literature. However, this obviously fails to fully capture some differences in the way the 'length of service' features of this insurance model function. Choosing differently, e.g. taking the minimum qualifying period and duration would also be imperfect and correspond less closely with distinctive features according to social policy scholars.
16. This procedure was adopted to be consistent with Esping-Andersen's decommodification index. The measure has two shortcomings. First, it ignores the general tax system's role in funding pensions. Second, it assumes a certain 'paycheque illusion'. Whether a social contribution is paid out of each worker's salary or is based on a percentage of the employer's total wage bill makes no real difference to take-home pay.
17. Some of this type of work has been done (e.g. Hansen, 2002). However, the countries and years of coverage are limited, making it difficult to discuss very broad trends.
18. Because generosity measures, like spending, trend upward during the 1980s they are theoretically and empirically compatible with a role for partisanship in explaining welfare expansion *and* retrenchment. Unfortunately, the 'new politics' literature fails to explain the rise of the welfare state via the same causal processes invoked to explain retrenchment. If new theories only need to explain new events – and not also 'cover' new and old events – there is not much recourse to external verification or theory building: one can safely have a different theory of each event. Not an impossible state of affairs, but one which seems to command that we give up the candle of social science.