MEASUREMENT VALIDITY IN COMPARATIVE WELFARE STATE RESEARCH: THE CASE OF MEASURING WELFARE STATE GENEROSITY

Jon Kvist, Simon Grundt Straubinger and Anders Freundt*

Abstract

Examining the generosity of welfare states and individual benefit schemes is a classical task in comparative welfare state studies. Three types of welfare states can be discerned based, in part, on their level of benefit generosity. Although significant advances have been made in the development of measures of welfare state generosity, this progress has not been without its challenges and limitations. In this article, the authors examine two sets of limitations related to measurement validity in comparative welfare state research: securing content validation and ensuring comparability across time and place. Through the use of illustrative examples to compare the situation of the unemployed in five European countries across several income levels and two family types, we demonstrate that, by profiling and stacking public benefits using the OECD Tax-Benefit micro-simulation model, we are able to carry out a more informed analysis of the redistributive strategies of the welfare state.

Keywords: income packaging; micro-simulation; profiling; stacking; unemployment

1. INTRODUCTION

How do we measure the generosity of welfare states? How do we avoid comparing apples and pears? Getting the right answer to these two questions is important in

---

* Jon Kvist, University of Southern Denmark; address: Campusvej 55 5230 Odense M; phone: +45 51718561; e-mail: jon@sam.sdu.dk; Simon Grundt Straubinger, University of Southern Denmark; address: Kragemosen 42 5250 Odense SV Denmark; phone: +45 22564154; e-mail: Sistr08@student.sdu.dk; Anders Freundt, University of Southern Denmark; address: Rådmandsgade 40c, 5, lejl. 133 2200 Copenhagen N Denmark; phone: +45 51683932; e-mail: andersfreundt@hotmail.com.

comparative welfare state studies, not least in studies of their causes and consequences, and for policymakers interested in designing benefit systems that provide adequate benefits without creating unacceptable work and saving disincentives. The purpose of this article is to identify the limitations of conventional measures used in comparative welfare state research on benefit generosity and to propose ways of overcoming these limitations.

The starting point of much comparative welfare state research is the idea of three worlds of welfare capitalism put forward and advanced by Gøsta Esping-Andersen (1990; 1999). The three types of welfare states are characterised in terms of the Liberal, the Conservative and the Social Democratic welfare regimes that are represented by the Anglo-Saxon, the Continental and the Scandinavian countries respectively. When investigating how these types of welfare states came about, or what their consequences are, researchers have often looked at the level of generosity, i.e. how much recipients actually receive in benefits. Access to the benefits and obligations for recipients are other parameters that may be included in the evaluation of social security systems, but that is beyond the scope of this paper (see Kvist 2007: 475 *et seq*.). In short, liberal welfare regimes provide meagre benefits on the basis of need as a last resort when markets and families have failed. Conservative welfare regimes provide generous benefits to labour market insiders and meagre benefits to outsiders. Social Democratic welfare regimes provide benefits on the basis of citizenship. Low-income groups receive generous benefits, while middle-income and high-income groups receive relatively less generous benefits.

In other words, the generosity of the social security system is not equal for every citizen but varies for different socio-economic groups in the different welfare regimes. We can illustrate this heterogeneous impact of the welfare state by referring to four different strategies of equality. The 'Robin Hood strategy' takes from the rich and gives to the poor, while more is given to rich than the poor in the 'St. Matthew strategy'. The '*quid pro quo* strategy' or 'something for something' approach implies that one receives according to what one pays in. Finally, the 'egalitarian strategy' gives equally to everybody, which translates into more generous benefits for the poor than for the rich (Korpi and Palme 2004). Figure 1 shows how these four strategies of equality translate into four distinct profiles, measured in terms of the generosity of benefits compared with previous earnings. Note that the four profiles are ideal types in a Weberian sense; no country is likely to be an exact copy of any one of the ideal types.

Figure 1. Four strategies of equality measured by the generosity of social benefits as a percentage of previous earnings
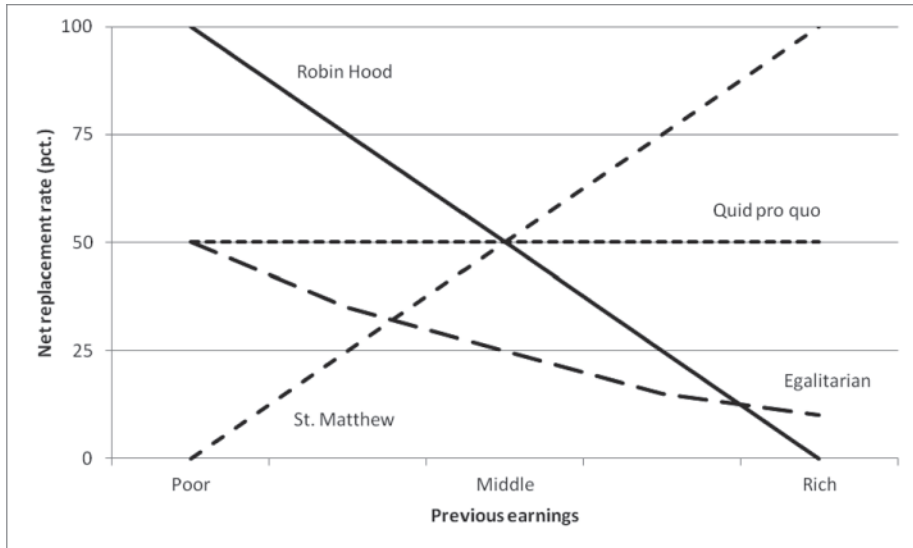
Coupling the redistributive strategies with our welfare regimes we would expect the Liberal welfare model to be associated with an egalitarian strategy, the Conservative model the *quid pro quo* strategy, and the Social Democratic a combination of the Robin Hood strategy for the poor and the *quid pro quo* strategy for middle and high income earners.

We use this theoretical framework as our starting point to examine whether conventional measures on benefit generosity actually reflect these hypothetical strategies. We argue that this is not the case. In the following analysis we find that many of the conventional measures, including some of the measures in the Social Citizenship Indicator Programme (SCIP) and the Comparative Welfare Entitlements Dataset (CWED), are limited with regard to the dual challenge of securing content validation and ensuring comparability. Securing content validation entails understanding how we can adequately capture the meaning of theoretical concepts and analytical constructs. Ensuring comparability entails creating measures that are comparable in different contexts, i.e. in different countries or at different points in time (Adcock and Collier 2001). To overcome such limitations, we use profiling and stacking analysis. This means simultaneously accessing the relevant tax-benefit situations for different relevant groups in the population. With an illustrative application of unemployment insurance in five European countries we suggest how profiling and stacking analysis can address the dual challenge of content validation and establishing comparability over time and space.

The article is set out as follows. In the first section, we set out in more detail the conventional measures used in comparative welfare state research i.e. social expenditure, SCIP and CWED. The second section presents our methodological reflections on profiling, stacking and the OECD tax-benefit micro-simulation model. In the third section, on content validation, we attempt to illustrate the limitations of the single case indicator, the country group indicator and the composite indicator by comparing the tax-benefit position of different socio-economic groups using the example of unemployment insurance. In the fourth section, on comparability, we demonstrate that the conventional measures may not always have the same meaning over time and across countries, and how equivalence may be established by making use of context specific domains of observation and adjusted common indicators. Finally, we offer some concluding remarks on the challenges of measurement validity on welfare benefit generosity in future comparative welfare state research.

2. CONVENTIONAL MEASURES OF BENEFIT GENEROSITY IN COMPARATIVE WELFARE STATE RESEARCH

In broad terms, there are two types of measures on benefit generosity, namely social expenditure data and institutional data. Social expenditure data is based on the monetary input into benefits, whereas institutional data looks at the level of individual benefit schemes. We argue that other types of measures of generosity like the extent of poverty or inequality refer to the consequences or outcomes of benefits rather than express benefit generosity as such.

2.1. SOCIAL EXPENDITURE AS A PROXY FOR BENEFIT GENEROSITY

The most frequently used measure of welfare state generosity is undoubtedly social expenditure as a proportion of GDP. The often implicit assumption is that generous welfare states spend more than non-generous welfare states. Since the 1970s public social expenditure as a proportion of GDP has been by far the most frequently used measure in comparative welfare state research, with empirical data easily available from various sources. Arguably, this measure has been valuable in that it has increased our understanding of differences between countries and among the correlates of variations in social expenditures. There are, however, major shortcomings of public social expenditure data when used in comparative welfare state research as it not only fails to capture all that is relevant to generosity but it also captures elements that are not relevant to generosity and the welfare state.

The first problem is that public social expenditure levels do not take into account the taxation of benefits or fiscal benefits and private social benefits with a social

purpose (see Adema *et al.* 2011). The second problem is that spending as such says little about how, or on whom, the money is spent. In terms of understanding the causes or consequences of welfare state generosity, this distinction can be crucial, both for researchers and for policymakers. Indeed, insofar as the welfare state serves to insure against misfortune *vis-à-vis* the market, actual spending levels are not directly relevant to the protection provided. When interpreted as an indicator of welfare effort, Finland had the biggest welfare state in the mid-1990s, but less so at the start and end of the decade – an observation that squared badly with public perceptions at the time. In the 1990s Finland was hit by an economic crisis in the first half of the decade but recovered in the second half. Public social expenditure varied from 24.1 per cent of GDP in 1990, to 30.7 per cent in 1995 and 24.2 per cent in 2000 (OECD 2013). The measure depends as much on the development of the economy and GDP as on social expenditure. Aggregate social expenditures does not pass the test of content validation.

Do disaggregated social expenditure levels pass the test of content validation? The OECD developed a Social Expenditure database (SOCX) in the 1990s to facilitate social policy analysis that now covers 27 OECD countries for the period 1980 to 2009 but with some missing data for some programmes in some countries in some years. The great advantage compared with the conventional aggregated expenditure data is that the data from SOCX is disaggregated into a number of social policy areas (unemployment, housing, sickness, pensions *etc*.). Thus it is possible to determine which social programmes are the most prominent in the different welfare states (Castles 2002, 2008; Adema *et al.* 2011: 88 *et seq*.). The detailed information on social expenditure items included in SOCX allows for various types of in-depth analysis of the effects of social policy, and several researchers have made extensive use of this opportunity (see Caminada and Goudszwaard 2005; Castles and Obinger 2007). Although SOCX is an improvement, there are still several problems associated with social expenditure data. It is still influenced by economic development and it is still not possible to detect how different socio-economic groups are being affected. In addition, other researchers have pointed to problems regarding missing values and incomparable data between countries and over time (Deken and Kittel 2007). Levels of expenditure are, in other words, a poor measure of welfare state generosity.

2.2.   INSTITUTIONAL MEASURES ON BENEFIT GENEROSITY

Taking into account the measurement problems in comparisons of social expenditure, it was a huge leap forward when a group of researchers started to collect institutional data and established indicators of social citizenship. The Social Citizenship Indicator Programme (SCIP) has been built and used exclusively by researchers at the Swedish Institute for Social Research (SOFI) at Stockholm University. Headed by Walter Korpi, PhD students and other researchers at SOFI have, since the 1980s, collected institutional information for the 'old' 18 OECD countries - most notably on social

insurance schemes for unemployment, sickness, work accidents, and old age (Korpi and Palme 2007). The collected information covers a range of relevant aspects of the above-mentioned insurance schemes: coverage of programmes, conditions for eligibility to benefits, waiting times before benefits are received, duration of benefits, sources of benefit financing, and benefit replacement rates for model households and for different levels of earnings. Net replacement rates are calculated for three model households including a single person, a couple without children, and a four-person household with two minor children and one economically active spouse. Old-age pensions and sickness benefits were covered by first generation SCIPPERS, Joakim Palme (1990) and Olli Kangas (1991), with later PhD students covering broader topics like families with children (Ferrarini 2003; Wennemo 1994), unemployment and work (Carroll 1999; Montanari 2000; Sjöberg 2000), and the poor (Nelson 2003).

Addressing the poor, or low-income groups, Kenneth Nelson has in fact established a whole new dataset, Social Assistance and Minimum Income Protection Interim Dataset (SaMip), which not only complements the social insurance schemes in SCIP but also stands apart in three respects. First, SaMip covers more countries than SCIP (34 countries), including Central and Eastern European countries. Second, SaMip has annual data. Third, SaMip has, from the start, been designed to take into account functional equivalent schemes that provide benefits to low-income persons, which is a *sine qua non* for this target group. These three features allow for analysis of the causes and consequences of contemporary changes in establishing minimum income standards in Europe (e.g. Nelson 2010). SCIP went public in June 2008 and SaMip was publicly accessible from the start.

The Swedes are not alone in producing measures based on institutional data. The Comparative Welfare Entitlements Dataset (CWED) has been built by Lyle Scruggs at the University of Connecticut (Scruggs 2004, 2013). SCIP and CWED contain more or less the same data, but whereas SCIP covers 15 points in time from 1930 to 2005, CWED has annual data for a more recent period (1971–2002). Both databases cover the 'old' 18 OECD countries and focus on unemployment insurance, sickness benefits and old-age pensions, with SCIP also including work accident benefit schemes. Net replacement rates are calculated from previous earnings for the average production worker in two household situations (single person and couple with one earner). Other benefit aspects included in CWED are qualifying conditions (weeks of insurance), benefit duration (weeks), and waiting times. CWED was, in contrast to SCIP, public from the outset.

We are strong supporters of the institutional approach that is embodied in SCIP, SaMip and CWED as a far more appropriate way to evaluate social security systems than social expenditure. We argue, however, that some of these conventional measures are limited with regard to the dual challenge of securing content validation and establishing comparability over time and space (see also Kvist 2011). There has recently been some debate regarding the validity of the data due to substantial disagreement as regards the level and change of benefit generosity. Furthermore,

the divergence between the datasets leads to substantial disagreement about the determinants of welfare state retrenchment, casting doubt on some important results from earlier research (Wenzelburger *et al.* 2013). The creators of SCIP and CWED argue that the two datasets differ in their underlying theoretical approach for policy analysis and therefore capture different aspects of how welfare states secure the livelihood of citizens (Ferrarini *et al.* 2013; Scruggs 2013). In the wake of the debate some recommendations were put forward to improve future data collection, such as looking at several income levels, including other kinds of public benefits (e.g. income dependent benefits), and incorporating other family types (e.g. a single parent or a cohabiting partner) (Danforth and Stephens 2013: 1293 *et seq.*). In this article we have tried to address some of these recommendations.

3.   DATA AND METHODS

The micro-simulation method, proposed in this article as a means of overcoming the limitations of the conventional measures referred to above, has two major advantages: profiling and stacking. *Profiling* refers to examining several income levels at the same time. Instead of national and family averages, an income specific comparison is proposed. Previous studies working with the disaggregation of income have revealed important differences (Gupta *et al.* 2006). Populations do not consist of average citizens' social security and economic incentive structures may be of greater significance for some subpopulations than for others. It is, therefore, necessary to survey the impact of social security at several income levels. In this article we examine the impact on 150 different income levels ranging from half of the average wage (0.5 AW) to twice that of the average worker (2.0 AW). The average annual wage measure, AW, was introduced by the OECD as a replacement of the average production worker measure, APW, used in the databases mentioned above (SCIP and CWED). While the APW is a typical production worker in manufacturing, the AW is close to the average wage in the country (OECD 2012: 11).

Stacking refers to the method by which the aggregate of several taxes/benefits and expenditures is calculated. Jonathan Bradshaw and his colleagues (Bradshaw and Finch 2002; Bradshaw and Mayhew 2006; Bradshaw and Richardson 2009; Bradshaw *et al.* 2007) have paved the way for accounting for multiple benefits at the same time. In this article, this method is taken even further by taking into account benefits, expenditure and taxation (including tax credits) at several income levels at the same time. Previous studies have shown that, whereas one benefit may be of most importance in one country (e.g. unemployment insurance in Denmark), another may be of more significance in another country (e.g. housing benefit in the United Kingdom) (Freundt *et al.* 2013). The benefits and expenditure included here are unemployment insurance, family benefits, advanced child maintenance for the single parent, housing benefits, childcare subsidies, tax credits and allowances, housing

costs, and the expenses relating to childcare. It should be noted that housing benefits and childcare subsidies differ from the other benefit programmes since they are not paid in cash directly to the recipients, but instead reduce certain expenditure, and as such, they cannot easily be used for other types of consumption.

The data used in this article has been retrieved from the OECD Tax-Benefit micro-simulation model, created and updated annually by the OECD, and based on the tax and benefit rules in the OECD countries. The different policies and tax brackets have been validated from independent sources such as national agencies and ministries by the authors of this article. Alterations have been made to correct factual errors and to create a more realistic scenario. For example, the calculation of rent was altered in the model from 20 per cent of an average wage to 20 per cent of gross income. In this way, rent is made variable – the change was made because low-income people usually live in cheaper accommodation and pay less rent than high-income people. This is because housing costs vary positively with income. Housing costs are difficult to compare across countries, and rented accommodation has been chosen to facilitate accurate comparison. Note that it is assumed that the family does not move to cheaper accommodation when one of the adult members becomes unemployed but that it stays in its rented accommodation. However, the calculation of housing benefit follows the situation of unemployment and more generous benefits are provided. It is also assumed that every family uses public childcare, regardless of income and labour market status. These assumptions are made because we want to evaluate how well the social security systems protect the families from major changes in the case of unemployment, e.g. they may be forced to move from accommodation and/or remove their children from public childcare. In some instances, the cost of public childcare exceeds income, which indicates an unsustainable economic situation in which the social security system has not been able to protect the family to a sufficient degree.

In this article two types of family are used: a lone-parent family with one child and a nuclear family with one child. First, we calculated the disposable income for the two family formations in and out of work, and then the net replacement rate (NRR), by dividing the disposable income out of work by the corresponding disposable income in work. Net replacement rate is a commonly used indicator in studying social security generosity and is used by economists, sociologists and political scientists alike, including the creators of SCIP and CWED (Korpi and Palme 2007; Scruggs 2004). Replacement rates can be defined as the benefit share of former wages and net replacement rates are after taxes. The countries considered here are Denmark, Sweden, Finland, the United Kingdom and Germany. In this way, all of the three welfare regimes put forward by Esping-Andersen (1990; 1999) are represented by one or several countries. Including the three Nordic countries allows for the litmus test of categorisation, i.e. that a group of countries belonging to a given type has more in common with each other than with other models and countries.

4. CONTENT VALIDATION OF MEASURES IN COMPARATIVE WELFARE STATE RESEARCH

Conventionally, comparative welfare state researchers use one or more of three types of indicators. The first type of indicator is used in the *situation* where one typical case is compared to the situation of a similar typical case not facing the given social contingency. For example, the generosity aspect of unemployment insurance is measured by the net replacement rate of an average wage, as described in Section 3 above. We call this indicator the *single case indicator*. The second indicator we call the *country group indicator* as it consists of an average score for a number of countries that are grouped by the researcher to represent a given type of country or welfare state. The third indicator we call the *composite indicator*, as it is made up of information on two or more aspects of social security benefits. Often the composite indicator is a simple average of the situation for two or more household types, e.g. the average net replacement rate for a single parent and a nuclear family. But the composite indicator may also be based on an index score of various aspects of the same benefit and typical case, reflecting, for example, accessibility and generosity at the same time. This section addresses the issue of content validation in comparative welfare state studies that are based on single case indicators, country group indicators and composite indicators.

4.1. SINGLE CASE INDICATORS

In the comparative welfare state literature, several studies have relied on single case indicators. The situation of the average production worker (now the average worker) the average is the single case indicator most frequently used in SCIP and CWED, although the two databases also include minimum and maximum benefits (Korpi and Palme 2007; Scruggs 2004). The main problem potentially associated with single case indicators is that policy nuances about social protection and incentive structures for different income groups are lost. This has been acknowledged by the creator of CWED as an important distributive issue in the evaluation of social policy (Scruggs 2006: 363). It is simply not possible to generalise from the single case indicator to other cases, as cross-national differences vary with previous earnings. Figure 2a shows the net replacement rate profiles for an unemployed single parent in Denmark, Sweden, Finland, Germany and the United Kingdom, while Figure 2b shows the net replacement rate profiles for a nuclear family with one dependent child where the spouse becomes unemployed in the five countries.

Figure 2a. The net replacement rates for a single parent with previous earnings between 0.5–2 AW, 2009
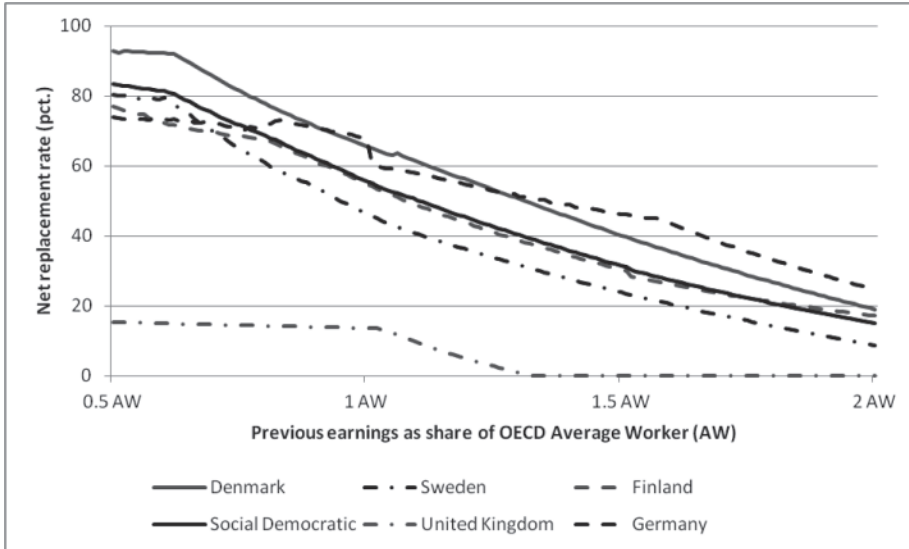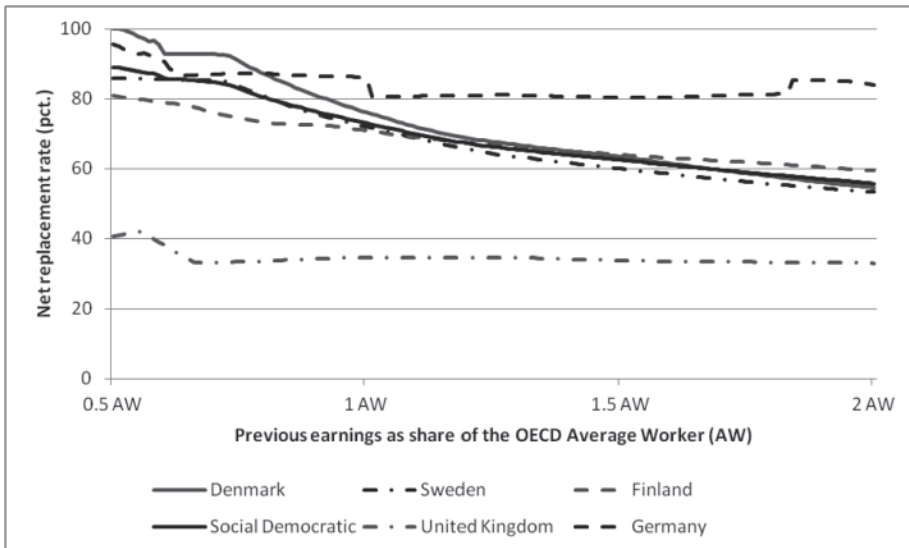

Figure 2b. The net replacement rates for a nuclear family with one dependent child with previous earnings between 0.5–2 AW, 2009

When previous earnings increase we can observe several variations both in, and between, our five countries. In every country, and for both types of family, except for the nuclear family in Germany, the net replacement rate decreases when previous earnings increase.

In Denmark, for example, the single parent with a previous income of the average annual wage will receive a net replacement rate of 66 per cent, whereas the person with a previous income of 50 per cent of an average wage will receive 93 per cent, and the person with a previous income of 200 per cent of an average wage will receive 19 per cent, i.e. 27 percentage points more and 47 percentage points less, respectively. The differences in net replacement rates are less pronounced, but still present, for the nuclear family. The country, with the lowest net replacement rate at 200 per cent of an average wage, is the United Kingdom, the representative of the Liberal welfare state model; whereas Denmark, Sweden, Finland and Germany, which represent the Social Democratic and the Conservative model, have considerable higher net replacement rates. The United Kingdom is the only country with a net replacement rate that reaches zero for the single parent, which indicates an unsustainable economic situation. This is because expenditures on rent and public childcare exceed income for the unemployed single parent, assuming that the unemployed do not move to cheaper accommodation and always make use of public childcare (see Section 3 above). The profiles for the single parent in all five countries resemble the egalitarian strategy for those whose previous earnings are above those of an average worker. For those whose previous incomes are below an average wage, Denmark, Sweden and Finland resemble the egalitarian strategy (maybe even the Robin Hood strategy), while Germany and the United Kingdom come closer to a *quid pro quo* strategy.

The single case indicators paint only a partial picture of cross-national differences and may not provide the most relevant basis for advancing theory or advocating policy. If poverty among the unemployed is the issue, then the situation of persons with low previous earnings may be more relevant than that of those earning the average wage. Are benefits adequate to keep the unemployed out of poverty? If the incentive structures of the unemployment benefit system are the issue, then the situation of persons with low previous earnings may well be the most important to look at. Do the unemployment benefits provide disincentives to work? If sustainability of the unemployment benefit system is the issue, then the situation of persons with high previous earnings may be the most appropriate group to examine. Is the level of public unemployment insurance for high-income people sustainable? To answer each of these questions we have to look beyond the single case indicator of the average wage.

4.2.   COUNTRY GROUP INDICATORS

There have been several attempts in the welfare modelling business to group welfare states according to different indicators or dimensions into a number of distinct welfare typologies with different characteristics (see Bonoli 1997; Castles and Mitchell 1993; Esping-Andersen 1990; Ferrera 1996; Kangas 1994; Kautto 2002; Korpi and Palme 1998; Pitzurello 1999; Siaroff 1994). The main problem with the country group indicator is that calculating group averages masks intra-group differences. To illustrate, Figure 2a and Figure 2b show the profiles for Social Democratic regimes constructed from the average of the three Nordic countries: Denmark, Sweden and Finland.

The graph regarding net replacement rates for the single parent clearly illustrates the problems associated with the country group indicator. First, there are substantial differences between the three Nordic countries in how their profiles develop from low to high previous earnings. The difference in net replacement rate for an average wage earner in Denmark and Sweden, for example, is 19 percentage points. Sweden has the second highest net replacement rate for the lowest incomes, but this position is taken by Finland for higher previous earnings. Second, the Social Democratic country group indicator masks the variations between each of the Nordic countries and the rest of the countries. People with the previous income of an average worker have approximately the same net replacement rate in Denmark and Germany, whereas it is 13 percentage points lower in Finland, and 22 percentage points lower in Sweden. Denmark is in fact closer to Germany than its Nordic peers for every, but the lowest previous income, level. The differences between the three Nordic profiles are less pronounced, but still exist for the nuclear family at lower incomes. In short, countries' tax-benefit systems are hardly ever exact copies of ideal-type welfare state regimes, but usually reflect a mix of several ideal-types, and the grouping of countries into one system involves making too heroic an assumption about its' uniformity and functioning.

4.3.   COMPOSITE INDICATORS

Composite indicators come in two forms. Perhaps the most common composite indicator is the average score across two or more typical cases, e.g. across different types of households. In CWED the average net replacement rate of the single person and the family are given along with the net replacement rate of each of the households. The other type of composite indicator is based on a variety of different types of information, typically put together as an index. Examples relating to welfare state generosity include the de-commodification indicator of Esping-Andersen (1990), the child benefit package indicator of Bradshaw and Finch (2002) and the generosity indicator of Scruggs (2006).

Compared to single case indicators, composite indicators are more robust. Consider the OECD summary measure of entitlement to unemployment benefit that is made up from the average of the gross unemployment benefit replacement rates for two earnings levels, three family situations and three durations of unemployment (OECD 2007: 104 *et seq.*). There is a risk, however, that the composite indicators provide mirror images of a fictional world that is misleading for advancing theory and advocating policy. To illustrate, we compare the net replacement rate profiles from Figure 2a and Figure 2b, i.e. the single parent and the nuclear family with one dependent child in Denmark, Sweden, Finland, Germany and the United Kingdom.

Here, we can observe pronounced differences in generosity between the two family types in each country, as well as in the ways the countries relate to each other depending on the family type under consideration. In all three countries and for every income level the net replacement rate is higher for the nuclear family than for the single parent. This is mainly due to the fact that the spouse in the nuclear family has a lower previous

income than the single parent, which is why the loss in income is more substantial for the single parent than for the nuclear family. This suggests that the net replacement rate in the United Kingdom no longer reaches zero for higher incomes. Besides the level of generosity, the type of family also influences the strategy of equality in some of the countries. The profile of the nuclear family in Denmark, Sweden and Finland is rather like the profile for the single parent, resembling the egalitarian strategy for all income levels. This is different in Germany and the United Kingdom, where the strategy for equality has changed. The profile of the nuclear family now resembles that of the *quid pro quo* strategy for almost every income level, whereas the profile of the single parent resembles that of the egalitarian strategy for previous earnings above the annual average.

Another disadvantage of composite indicators is the loss of information that may be meaningful in theoretical or policy terms. In economic and social policy analysis, the scholar has a substantive interest in the general profile across the income range and the situation of certain groups. Poverty reduction, benefit adequacy and disincentives to work and save are just some of the aspects of major interest for researchers and policymakers alike, where comparisons of profiles are crucial to providing sound answers and policy guidance.

5. COMPARABILITY OF MEASURES OVER TIME AND PLACE

Perhaps the biggest challenge in comparative welfare state research is establishing functional equivalence across different contexts. When does a score of X mean the same in two countries or at two different points in time? This problem is not only prevalent in welfare state research, but in every study that involves a comparison of observations from different contexts. According to Robert Adcock and David Collier (2001), the standard ways of establishing equivalence of meaning in different contexts can be summarised under three headings: context specific domain of observation, adjusted common indicators, and context specific scoring procedures. Under these headings we first describe the conventional approaches taken in comparative welfare state research and then discuss alternatives.

5.1. CONTEXT SPECIFIC DOMAIN OF OBSERVATION

Social protection may not be organised in the same way across welfare states. Indeed, one of the major differences between countries is that social needs are met differently across countries; i.e. their welfare architecture or welfare mix differs. Hence, we want measures that tap into this diversity and allow us to measure the nature and extent of the social security system. Although SCIP and CWED take into account important taxation, they mainly have information on one scheme of social insurance for each contingency. In contrast, SaMip encompasses a wide range of benefits independently

of their labels as social assistance, housing benefit and so on, as long as the benefits aim at providing a minimum income floor for low-income groups. In other words, SaMip includes functional equivalent schemes that may be in different policy domains in different contexts. We argue that often the same approach may strengthen the study of benefits in the case of family, sickness, work incapacity, unemployment and old age.

Coming back to our illustrative case of unemployment, the relevant benefit in SCIP and CWED is unemployment insurance. The sole focus on this particular scheme may only give us a partial picture of the social security system when there are other complementary or supplementary schemes. To compare social insurance between countries like Denmark, where unemployment insurance is indeed the single most important scheme, and the United Kingdom, where housing benefit also plays a very prominent role, would be misleading. The way forward is stacking analysis, where multiple tax-benefits are considered at the same time. Figure 3 shows the net replacement rate profiles for an unemployed single parent with only unemployment insurance and with stacked benefits and expenditures in Denmark, Germany and the United Kingdom.

Figure 3. The net replacement rates for a single parent with only unemployment insurance and with stacked benefits and expenses with previous earnings between 0.5–2 AW, 2009
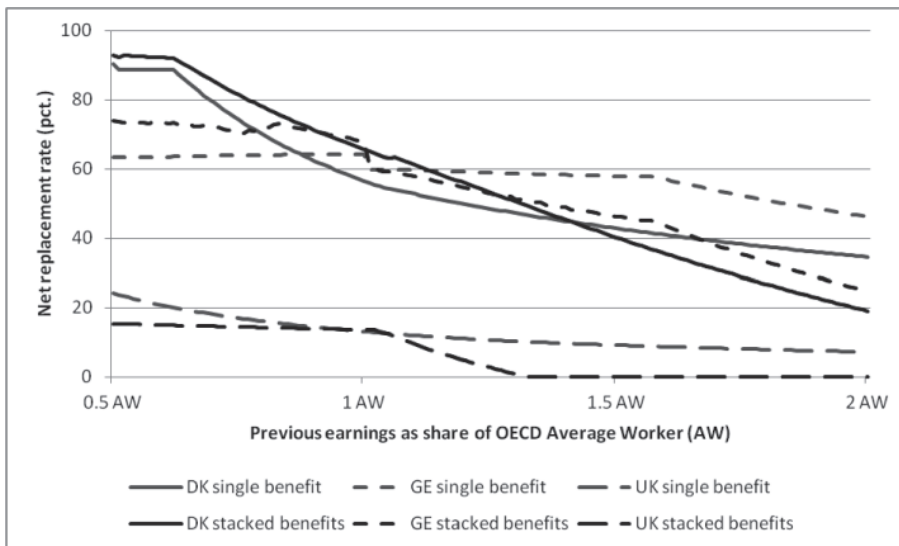

The graph indicates how important it is to use stacking analysis. In Denmark and Germany the strategy of equality is more or less the same, regardless of stacking, with Denmark illustrating an egalitarian strategy, and Germany, a *quid pro quo* strategy for lower incomes and an egalitarian strategy for higher incomes. The change from the *quid pro quo* strategy to the egalitarian strategy in Germany occurs at average wage levels when benefits are stacked, but at a level of 1.5 times the average wage when only unemployment benefit is taken into account, which results in a difference in net

replacement rate between the two scenarios of 21 per cent at the highest income level. In the United Kingdom, the strategy of equality changes drastically from a *quid pro quo* strategy, when only considering unemployment insurance, to an egalitarian strategy from around the level of the average wage earner, when benefits are stacked. This has the implication that the net replacement rate reaches zero because expenditure on rent and childcare exceeds the income of the lone parent, which indicates an unsustainable economic situation. This is because expenditure on rent and childcare exceeds income for the unemployed single parent, assuming that the unemployed do not move to cheaper accommodation and always make use of public childcare (see Section 3 above).

5.2. ADJUSTED COMMON INDICATORS

Adjusting common indicators by converting measures in national currencies into a common unit is standard practice in most comparative studies. When expressing the generosity of benefits, both the OECD and many individual researchers use net replacement rates that express the share of previous earnings instead of giving the net amount in the national currency. This is a relative measure. Another strategy is to convert the amounts into a common currency, for example, US dollars, to show how much benefit claimants receive. However, consumption options are not only a function of income levels, but also of expenditure levels. Fortunately, the OECD publishes purchasing power parities whereby the national currency amount can be translated into a common currency, e.g. dollars, that takes into account how much can be bought. Figure 4 shows the disposable income of a single parent in Denmark, Sweden, Finland, Germany and the United Kingdom.

Figure 4. The net income for a single parent in USD PPP with previous earnings between 0.5–2 AW, 2009
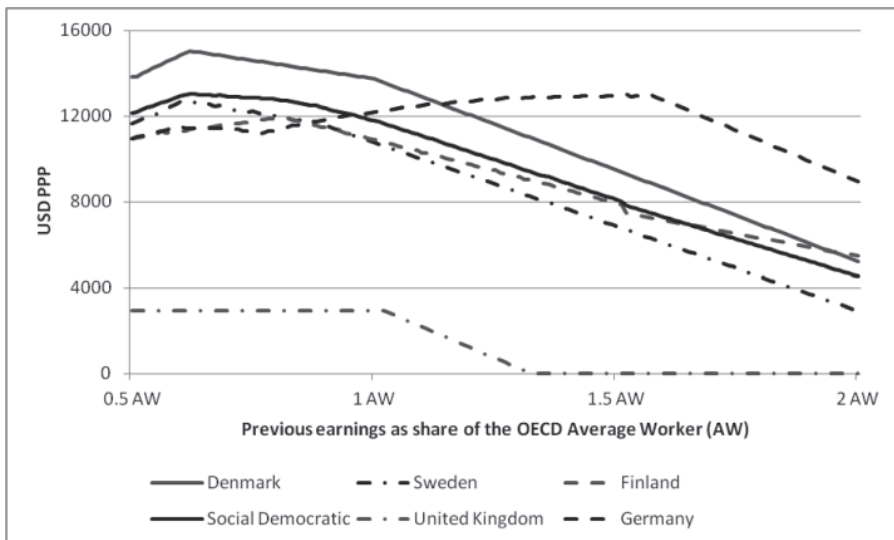
The graph shows that German unemployment insurance comes close to adopting a *quid pro quo* strategy, with a touch of the Matthew principle, from half an average wage to 1.5 times the average wage. In fact, the unemployed who were middle-income to high-income earners are better off in Germany with regard to benefit generosity than in the other countries. The unemployed of low income earners are best off in Denmark. The profiles of the Nordic countries follow each other within a band. Finally, Figure 4 confirms that not only does the United Kingdom pay the smallest benefits, but their purchasing power is the smallest of the countries included. With the growth of indicators on well-being and inequality, we are likely to see purchasing power parities used more frequently in comparative studies.

Another type of adjustment of common indicators that we can imagine in the longer term is a move away from the national measure of the average wage. With globalisation of labour markets, we may see a move towards using stylised cases with the same income gross or net of taxes, thereby indicating that workers or, more appropriately, earners, with the same human capital receive the same wage, independent of whether they are living in Berlin, London or Stockholm. However, this type of adjustment is probably a few steps down the road.

6. CONCLUDING REMARKS

In this article we have highlighted two sets of issues relating to measurement validity that researchers must pay attention to when undertaking comparative studies of welfare state generosity, whether they are economists looking at disincentives to work or political scientists looking at the strength of social rights. The first set of issues relates to content validation of conventional indicators like single case indicators, country group indicators and composite indicators. Using single case indicators to study, say, the situation of workers based on an average wage does not capture the situation of low-income and middle-income workers. Country group indicators are typically used when researchers compare a large number of countries or when they focus on welfare state regimes. However, as we have shown by comparing the Nordic countries representing the Social Democratic welfare state regime with countries belonging to other welfare state regimes, country group indicators may mask intra-group variation between countries that is larger than inter-group variation between different welfare state regimes. Composite indicators may help to increase robustness of indicators, but again this comes at a price. Comparing the benefit of a person who becomes unemployed in two different family contexts showed that different socio-economic groups may be treated differently. We recommend the profiling of redistributive strategies for different socio-economic groups, at least in the initial stages of the research, since this enables the researcher to become familiar with the adequacy of issues benefits and the structure of incentives for different groups.

The second set of measurement validity issues to which researchers should pay attention concerns the comparability of measures over time and place. Here, we recommend searching for functional equivalent schemes and undertaking stacking analysis, i.e. including all the relevant benefits and expenditures a person or family are likely to be subject to in a given situation. In an example of a family receiving only one benefit, we showed how some countries actually appear to change their redistributive strategy as shown by the different profiles in the two situations.

We gather that more work will be undertaken on how to adjust common indicators to get the most meaningful comparison for a given question. Although the move from an average production worker to an average wage earner has led to a better capturing of the situation in many countries, we argue that there may also be a good case for taking into account consumption possibilities, as a dollar in two countries rarely buys the same amount of goods and services. Purchasing power parities are now published on an annual basis by the OECD, and may prove valuable in many studies.

In any case, the availability of the OECD Tax-Benefit model in still more areas of social protection opens up the possibility of more informed analysis, just as the public availability of CWED, SCIP and SaMip ought to spur on researchers to undertake more comparative studies. For historical analysis CWED, SCIP and SaMip may be the preferred starting point. For contemporary studies, the OECD Tax-Benefit model has greater potential for stacking benefit analysis combined with profiling across socio-economic groups that can take into account the more complex nature of benefit packages and the diversity of benefit populations. In both historical and contemporary studies we hope that researchers will not underestimate the problems of measurement validity, as these have indeed become more complex than in the days when social expenditure was the only, and thereby a legitimate, second-best solution.

REFERENCES

Adcock, R. and Collier, D. (2001) 'Measurement Validity: A Shared Standard for Qualitative and Quantitative Research', *American Political Science Review*, 95(3), 529–546.

Adema, W., Fron, P. and Ladaique, M. (2011) 'Is the European Welfare State Really More Expensive? Indicators on Social Spending, 1980–2012; and a Manual to the OECD Social Expenditure Database (SOCX)', *OECD Social, Employment and Migration Working Papers*, No. 124, Paris: OECD Publishing.

Bonoli, G. (1997) 'Classifying Welfare States: A Two-dimension Approach', *Journal of Social Policy*, 26(3), 351–72.

Bradshaw, J. and Finch, N. (2002) 'A Comparison of Child Benefit Packages in 22 Countries' *Department for Work and Pensions Research Report*, No 174, Leeds: Corporate Document Services.

Bradshaw, J. and Mayhew, E. (2006) 'Family Benefit Packages' in Bradshaw, J. and Hatland, A. (eds.) *Social Policy, Employment and Family Change in Comparative Perspective*, Cheltenham: Edward Elgar Publishing.

Bradshaw, J. and Richardson, D. (2009) 'An Index of Child Well-Being in Europe', *Child Indicators Research*, 2(3), 319–351.

Bradshaw, J. Richardson, D. and Ritakallio, V.M. (2007) 'Child Poverty and Child Well-Being in Europe', *Journal of Children's Services*, 2(1), 18–36.

Caminada, C.L.J. and Goudswaard, K.P. (2005) 'Are Public and Private Social Expenditures Complementary?', *International Advances in Economic Research*, 11(2), 175–189.

Carroll, E. (1999) *Emergence and Structuring of Social Insurance Institutions: Comparative Studies on Social Policy and Unemployment Insurance*, Stockholm: Swedish Institute for Social Research.

Castles, F.G. (2002) 'Developing New Measures of Welfare State Change and Reform', *European Journal of Political Research*, 41(5), 613–641.

Castles, F.G. (2008) 'What Welfare States Do: A Disaggregated Expenditure Approach', *Journal of Social Policy*, 38(1), 45–62.

Castles, F.G. and Mitchell, D. (1993) 'Worlds of Welfare and Families of Nations', in Castles, F.G. (ed.) *Families of Nations: Patterns of Public Policy in Western Democracies*, Aldershot: Dartmouth Publishing Company.

Castles, F.G. and Obinger, H. (2007) 'Social Expenditure and the Politics of Redistribution', *Journal of European Social Policy*, 17(3), 206–222.

Danforth, B. and Stephens, J.D. (2013) 'Measuring Social Citizenship: Achievements and Future Challenges', *Journal of European Public Policy*, 20(9), 1285–1298.

Deken, J.D. and Kittel, B. (2007) 'Social Expenditure Under Scrutiny: the Problems of Using Aggregate Spending Data for Assessing Welfare State Dynamics', in Clasen, J. and Siegel, N.A. (eds.) *Investigating Welfare State Change: The 'Dependent Variable Problem' in Comparative Analysis*, Cheltenham: Edward Elgar.

Esping-Andersen, G. (1990) *Three Worlds of Welfare Capitalism*, Princeton: Princeton University Press.

Esping-Andersen, G. (1999) *Social Foundations of Post-Industrial Economies*, Oxford: Oxford University Press.

Ferrarini, T. (2003) *Parental Leave Institutions in Eighteen Post-war Welfare States*, Stockholm: Swedish Institute for Social Research.

Ferrarini, T., Nelson, K., Korpi, W. and Palme, J. (2013) 'Social Citizenship Rights and Social Insurance Replacement Rate Validity: Pitfalls and Possibilities', *Journal of European Public Policy*, 20(9), 1251–1266.

Ferrera, M. (1996) 'The "Southern" Model of Welfare in Social Europe', *Journal of European Social Policy*, 6(1), 17–37.

Freundt, A., Straubinger, S.G. and Kvist, J. (2013) 'Modernising Social Security for Lone Parents: Avoiding Fertility and Unemployment Traps when Reforming

Social Policy in Europe', in Ramia, G., Farnsworth, K. and Irving, Z. (eds.) *Social Policy Review 25: Analysis and Debate in Social Policy*, Bristol: The Policy Press.

Gupta, N.D., Oaxaca, R.L. and Smith, N. (2006) 'Swimming Upstream, Floating Downstream: Comparing Women's Relative Wage Progress in the United States and Denmark', *Industrial and Labour Relations Review*, 59(2), 243–266.

Kangas, O. (1991) *The Politics of Social Rights: Studies on the dimensions of sickness insurance in OECD countries*, Stockholm: Swedish Institute for Social Research.

Kangas, O. (1994) 'The Politics of Social Security: on Regressions, Equality Comparisons and Cluster Analysis', in Janoski, T. and Hicks, A. (eds.) *The Comparative Political Economic of the Welfare State*, Cambridge: Cambridge University Press.

Kautto, M. (2002) 'Investing in Services in West European Welfare States', *Journal of European Social Policy*, 12(1), 53–65.

Korpi, W. and Palme, J. (1998) 'The Paradox of Redistribution and Strategies of Equality: Welfare State Institutions, Inequality and Poverty in the Western Countries', *American Sociological Review*, 63(5), 661–87.

Korpi, W. and Palme, J. (2004) 'Robin Hood, St Matthew, or simple egalitarianism? Strategies of equality in welfare states', in Kennett, P. (ed.) *A Handbook of Comparative Social Policy*, Cheltenham: Edward Elgar Publishing.

Korpi, W. and Palme, J. (2007) *The Social Citizenship Dataset (SCIP)*, Swedish Institute for Social Research, Stockholm University.

Kvist, J. (2007) 'Fuzzy Set Ideal Type Analysis', *Journal of Business Research*, 60(5), 474–481.

Kvist, J. (2011) 'One-Size-Fits-All? Measurement issues in Medium-N comparative welfare state analysis', *Qualitative and Multi-Method Research*, Fall, 20–25.

Montanari, I. (2000) *Social Citizenship and Work in Welfare States: Comparative Studies on Convergence and on Gender*, Stockholm: Swedish Institute for Social Research.

Nelson, K. (2003) *Fighting Poverty: Comparative studies on social insurance, means-tested benefits and income redistribution*, Stockholm: Swedish Institute for Social Research.

Nelson, K. (2010) 'Social Assistance and Minimum Income Benefits in Old and New EU Democracies', *International Journal of Social Welfare*, 19(4), 367–378.

OECD (2007) *Benefits and Wages*, Paris: OECD Publishing.

OECD (2012) *Taxing Wages 2011*, Paris: OECD Publishing.

OECD (2013) *Social Expenditure Database (SOCX)*. Retrieved July 2013 from www.oecd.org/els/social/expenditure.

Palme, J. (1990) *Pension Rights in Welfare Capitalism. The development of old-age pensions in 18 OECD countries*, Stockholm: Swedish Institute for Social Research.

Pitzurello, S. (1999) *De-commodification and the Worlds of Welfare Capitalism: a cluster analysis*, Florence: European University Institute.

SCRUGGS, L. (2004) *Welfare State Entitlements Dataset: A Comparative Institutional Analysis of Eighteen Welfare States*, Version 1.1.

SCRUGGS, L. (2006) 'The Generosity of Social Insurance, 1971–2002', *Oxford Review of Economic Policy*, 22(3), 349–364.

SCRUGGS, L. (2013) 'Measuring and Validating Social Program Replacement Rates', *Journal of European Public Policy*, 20(9), 1267–1284.

SIAROFF, A. (1994) 'Work, Welfare and Gender Equality: a New Typology', in SAINSBURY, D. (ed.) *Gendering Welfare States,* London: Sage.

SJÖBERG, O. (2000) *Duties in the Welfare State. Working and Paying for Social Rights,* Stockholm: Swedish Institute for Social Research.

WENNEMO, I. (1994) *Sharing the Costs of Children. Studies on the Development of Family Support in the OECD Countries*, Stockholm: Swedish Institute for Social Research.

WENZELBURGER, G., ZOHLNHÖFER, R. and WOLD, F. (2013) 'Implications of Dataset Choice in Comparative Welfare State Research', *Journal of European Public Policy*, 20(9), 1229–1250.