Gayle, Vernon, and Paul Lambert. "Temporal analysis with cross-sectional data." *What is Quantitative Longitudinal Data Analysis?*. London: Bloomsbury Academic, 2018. 31–38. *Bloomsbury Collections*. Web. 25 Apr. 2023. <<http://dx.doi.org/10.5040/9781472545268.0010>>.

Downloaded from Bloomsbury Collections, www.bloomsburycollections.com, 25 April 2023, 23:43 UTC.

Copyright © Vernon Gayle and Paul Lambert 2018. You may share this work for non-commercial purposes only, provided you give attribution to the copyright holder and the publisher, and provide a link to the Creative Commons licence.

3 Temporal analysis with cross-sectional data

Introduction

At the beginning of Chapter 1 we stated that temporal data can be collected in cross-sectional surveys and therefore it is mildly misleading to assume that longitudinal analyses can never be undertaken with cross-sectional survey data. We begin this chapter with a simple example. There is no doubt that Welsh is one of Europe's most robust minority languages and its cultural influence and traditions remain relevant today, and they are embraced by new generations learning and using the language (Welsh Language Unit, 2012). A comparison of the 1891 UK Census and the 1981 UK Census shows that the percentage of Welsh speakers in Wales had fallen from 54 per cent to 19 per cent. In this example, data collected at two distinctive time points allow us to understand change over time. Comparing the 1981 UK Census with the 1991 UK Census we find that the percentage of Welsh speakers remained at 19 per cent and then increased slightly to 21 per cent in the 2001 UK Census. This simple example should illustrate that using multiple sources of cross-sectional data we have been able to investigate social change. Examining patterns using data collected at different time points should not be overlooked because results can be useful in determining empirical regularities.

In applied research, relatively straightforward analytical techniques can even be used to explore temporal patterns within one-off cross-sectional social surveys. One notable example that has supported a large amount of analysis is the famous 1972 Oxford Mobility Study (OMS) (Goldthorpe and Mills, 2008). The original OMS was a cross-sectional data collection exercise that contributed to an influential analysis of social change in two ways. First, it summarized data on social mobility based upon comparing data about its respondents' current jobs with retrospective information about jobs from earlier time periods (e.g. the respondents' fathers' jobs or the respondents' first jobs). Second, it uncovered trends in social mobility

rates, by comparing the social mobility experiences of respondents from different birth cohorts contained within the study.

More generally, using multiple datasets from repeated cross-sectional surveys is an effective way to study trends (or stability) over time. This is the main focus of this chapter.

When analysing repeated cross-sectional data, social scientists sometimes concentrate upon summarizing trends at the macro-level. This might involve deriving summary statistics for each different repeated cross-sectional dataset, and then presenting a summary of those statistics in relation to the time point of the survey. An important piece of advice is that when this approach is used a research study will always benefit from basing its findings on as many data points as possible. In the area of social mobility research, where analysts summarize trends in social reproduction based on summary statistics from different datasets, researchers have stressed that increasing the number of data points substantially reduces the risk of erroneous conclusions that might arise from measurement or sampling errors associated with any one specific dataset (see Ganzeboom, Luijkx and Treiman, 1989; Breen, 2004; Lambert, Prandy and Bottero, 2007; Bernardi and Ballarino, 2016).

Repeated cross-sectional data are also often analysed at the micro-level, using pooled datasets that combine records from different years. In most situations, methods that are common in the analysis of cross-sectional data are used. The most notable difference is that the time period associated with the data becomes a particularly important explanatory variable within the analysis (an example is provided below). Firebaugh (1997) provides a useful reference text on analysing repeated cross-sectional data in this context.

Comparing survey measures

The main challenge involved in making meaningful comparative statements with survey measures that are collected at different time periods, or in different countries, centres on the 'equivalence' of measures. Does completing a degree in engineering have the same social significance for a woman in 2016 as it would have had in 1967? Or can the ethnic group category in the UK 2001 Census 'Black Caribbean' reasonably be compared with the category 'Coloured' in the 1972 General Household Survey?

Johnson (1998) provides a comprehensive review of alternative concepts of equivalence. The term 'measurement equivalence' is used to describe

data that have been purposively collected and organized to easily facilitate direct comparisons (Ehling, 2003). The term ‘meaning equivalence’ invokes the comparison of locations according to their relative, contextual meaning (see Van Deth, 2003). As an illustration, the relative social advantage associated with having an MBA may be substantially less in the UK than in Burkina Faso. Therefore, an analysis prioritizing ‘meaning equivalence’ would regard MBAs held in the two countries as marking different relative levels of educational attainment. It is important to understand that the source of equivalence problems for longitudinal data often lies not with a change in how something has been recorded over time, but with a change in the underlying distribution of a measure in the population. This means that the appearance of ‘measurement equivalence’ may not be enough to ensure ‘meaning equivalence’.

Researchers using measures from repeated cross-sectional surveys should place effort into understanding the equivalence of their data. This will inevitably draw upon both their own and others’ expert knowledge within the substantive field, and it is therefore impossible to provide detailed advice on the equivalence of measures in specific surveys. The terms ‘standardizing’ and ‘harmonizing’ are used to describe the activities associated with making measures comparable (see Ehling, 2003). Hoffmeyer-Zlotnik and Warner (2014) provide an extended account of harmonizing variables, and Connelly, Gayle and Lambert (2016a) provide practicable advice on how to organize and analyse key variables in social survey research.

An example of analysing pooled cross-sectional surveys

In this section we provide a simple example of using pooled cross-sectional surveys. The data are from the YCS, which we introduced in Chapter 2. In this example we pool data from the first sweeps of the 1990, 1993, 1995, 1997 and 1999 cohorts (using data from UK Data Archive Study Number SN 5765). The YCS is actually a repeated contacts cohort study, but in this example we are using only the first record from its respondents in each cohort, and we are analysing these data as repeated cross-sections. While this might seem to complicate the definition of datasets, it is in fact quite common for influential social research to be undertaken through methods of repeated cross-sectional analysis while using data that was itself drawn from a wider, repeated contacts longitudinal design (Breen and Goldthorpe, 2001).

Historically, school pupils in England and Wales have undertaken a diet of examinations called the General Certificate of Secondary Education (GCSE) at the end of compulsory education (see Playford and Gayle, 2016). In this analysis the outcome variable of interest is the pupils' GCSE points score in Year 11 (which was the end of compulsory schooling). We also have information on the pupils' gender and their parents' social class, which is measured by a three-category version of the official UK socio-economic classification NS-SEC (see Rose, Pevalin and O'Reilly, 2005).

An attractive feature of undertaking analyses on pooled cross-sectional datasets is that standard techniques (e.g. linear regression models and logistic regression models) can be used. It is common to include variables that identify the source of the data (i.e. the cross-sectional survey) or the time point within the analysis. An obvious extension to simpler approaches is to estimate statistical models that include interaction effects to assess changes in the impact of explanatory variables at different time points (e.g. in different surveys).

A linear regression model estimated on the pooled data from the five YCS cohorts is reported in Figure 1. The results in Figure 1 show that across the time period the overall profile of school GCSE attainment has increased considerably. This is evident from the positive main effect for the school cohorts. Pupils that were surveyed in 1999 on average achieved a GCSE points score that was 13.6 points higher than that of their counterparts in

Source	SS	df	MS	Number of obs	=	62,910
Model	1359059.45	5	271811.89	F(5, 62904)	=	948.14
Residual	18033188.1	62,904	286.677924	Prob > F	=	0.0000
				R-squared	=	0.0701
				Adj R-squared	=	0.0700
Total	19392247.6	62,909	308.258716	Root MSE	=	16.932

t0score	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
cohort_1993	5.143661	.2080184	24.73	0.000	4.735944	5.551377
cohort_1995	9.365432	.2157778	43.40	0.000	8.942508	9.788357
cohort_1997	8.334607	.2166697	38.47	0.000	7.909934	8.75928
cohort_1999	13.63625	.225169	60.56	0.000	13.19492	14.07758
male	-3.119253	.1354489	-23.03	0.000	-3.384733	-2.853773
_cons	31.93301	.1675295	190.61	0.000	31.60465	32.26136

Figure 1 Stata output: Regression model (OLS) General Certificate of Secondary Education (GCSE) points score School Year 11 Youth Cohort Study of England and Wales

Source	SS	df	MS	Number of obs	=	62,910
Model	1362275.97	9	151363.996	F(9, 62900)	=	528.05
Residual	18029971.6	62,900	286.645018	Prob > F	=	0.0000
				R-squared	=	0.0702
				Adj R-squared	=	0.0701
Total	19392247.6	62,909	308.258716	Root MSE	=	16.931

t0score	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
cohort_1993	5.511817	.2834742	19.44	0.000	4.956207	6.067427
cohort_1995	9.901535	.2939396	33.69	0.000	9.325413	10.47766
cohort_1997	8.92802	.2970979	30.05	0.000	8.345708	9.510333
cohort_1999	14.16292	.3062392	46.25	0.000	13.56269	14.76315
male	-2.249934	.3110738	-7.23	0.000	-2.859639	-1.640229
male*cohort_1993	-.7883375	.4172608	-1.89	0.059	-1.606169	.0294943
male*cohort_1995	-1.156655	.4328531	-2.67	0.008	-2.005048	-.3082621
male*cohort_1997	-1.269568	.4341967	-2.92	0.003	-2.120594	-.4185414
male*cohort_1999	-1.137471	.4518282	-2.52	0.012	-2.023055	-.2518865
_cons	31.52792	.212347	148.47	0.000	31.11172	31.94412

Figure 2 Stata output: Regression model (OLS) General Certificate of Secondary Education (GCSE) points score School Year 11 Youth Cohort Study of England and Wales (cohort and gender interactions)

the 1990 survey (the reference category). On average male pupils scored 3.1 points lower than female pupils. The model reported in Figure 2 includes an interaction effect to assess changes in the impact of gender at each of the five time points (i.e. the five cohorts of pupils surveyed in the YCS).

Cross-national comparative research using repeated cross-sectional surveys

Many social surveys are deliberately coordinated internationally to aid comparative research, other existing survey data can be brought together at a later stage for the purpose of making cross-national comparisons. Hoffmeyer-Zlotnik and Warner (2014 Chapter 4) provide a review of a number of important examples. Many of these surveys can also make a contribution to longitudinal research. Figure 3 lists a number of examples of large-scale survey projects that provide opportunities for longitudinal analysis through repeated cross-sectional surveys. These examples encompass considerable volumes of micro-social data. For example, IPUMS-I boasts on its website that it provides access to data from 1960 covering 614 million people in 277 censuses from 82 countries! The examples listed in Figure 3 are selective, and many other relevant studies exist. While some of the resources that are listed are intended to be general

<i>(1) Studies with an orientation to socio-economic and socio-demographic data</i>	
IPUMS-I	https://international.ipums.org/international/
NAPP	https://www.nappdata.org/napp/
Luxembourg Income Study	http://www.lisdatacenter.org/
ISMF	http://www.harryganzeboom.nl/ismf/index.htm
CASMIN	http://www.nuff.ox.ac.uk/Users/yaish/npsm/
<i>(2) Studies with an orientation to social and political attitudes</i>	
European Social Survey	http://www.europeansocialsurvey.org/
International Social Survey Programme	http://www.issp.org/
World Values Survey	http://www.worldvaluessurvey.org/
Eurobarometer	http://www.gesis.org/eurobarometer-data-service/home/
<i>(3) Studies in educational research</i>	
TIMSS / PIRLS	http://timss.bc.edu/
PISA	http://www.pisa.oecd.org/

Figure 3 Selected examples of large-scale cross-nationally harmonized repeated cross-sectional survey projects

purpose or ‘omnibus’ resources suitable for a variety of uses, some of the examples are illustrative of data that emerge from more specialist studies. Such studies are typically academic projects that construct comparable data with a specific research aim.

The studies listed in Figure 3¹ are examples that have already been subject to some degree of harmonization. An important example is the annual Labour Force Surveys, which have been carried out in standardized ways across countries since the 1980s. They have coverage in around 200 countries and territories across the world and are coordinated by the International Labour Office.² The Luxembourg Income Study offers coordinated access to micro-data for the analysis of selected Labour Force Surveys from different countries after harmonization work has already

1 Websites (accessed 1 June 2016).

2 See <http://www.ilo.org/dyn/lfsurvey/lfsurvey.home> (accessed 1 June 2016).

been undertaken (see Smeeding, Jesuit and Alkemade, 2002). A standard approach is to derive aggregate summary statistics from different Labour Force Surveys and explore trends in those aggregate records.

There are some important attractions to using cross-national surveys for longitudinal research. There are many surveys available for studying trends over time. There are well-known challenges of comparability and standardization of measures when working with cross-national repeated cross-sectional survey datasets (see Johnson, 1998; Burkhauser and Lillard, 2005; Hoffmeyer-Zlotnik and Warner, 2014; Connelly, Gayle and Lambert, 2016b). Despite these challenges many projects have successfully analysed trends over time (see Mayer, 2005). It is reasonably easy for academic researchers to gain access to international survey data from national or international data archives³ or from the online resources that projects often provide (see Hoffmeyer-Zlotnik and Warner, 2014). The majority of studies listed in Figure 3 offer access to their micro-data for secondary research free, or at nominal cost, for non-commercial and not-for-profit uses.

Conclusions

Considerable progress can be made by investigating longer term social trends using repeated (or pooled) cross-sectional surveys. A clear benefit of analysing pooled or repeated cross-sectional surveys is that well-understood standard techniques such as linear regression and logistic regression models can be used. The most serious challenge when dealing with repeated cross-sectional surveys is ensuring that measures are sufficiently equivalent over time in order to allow realistic comparisons to be made. Researchers should also consider how best to represent temporal dimensions in their analyses.

³ See UK Data Service <https://www.ukdataservice.ac.uk> (accessed 1 June 2016).