

Luxembourg Income Study Working Paper Series

Working Paper No. 494

Subnational Analyses Using the Luxembourg Income Study (LIS) Data Archive

David K. Jesuit

September 2008



Luxembourg Income Study (LIS), asbl

Subnational Analyses Using the Luxembourg Income Study (LIS) Data Archive

David K. Jesuit, Central Michigan University (david.jesuit@cmich.edu)

This brief chapter introduces researchers to the possibilities for subnational research using the harmonized data sets made available via the Luxembourg Income Study (LIS) (<http://www.lisproject.org>). We first offer a brief overview of the LIS and discuss specific challenges for subnational research using LIS microdata. Next, we review some of the contributions LIS data and researchers have made in the comparative study of subnational poverty and income inequality. Finally, we conclude with a brief presentation of the most recent poverty rates for German Länder and Italian Regions derived from the LIS. It is hoped that this introduction will provide scholars with a resource that will foster subnational examinations of economic well-being.

A Brief Overview of the Luxembourg Income Study

The LIS is an independent not-for-profit research institution and data archive located in the Grand Duchy of Luxembourg.¹ As a data-provider, the LIS project staff does not collect primary data. Rather, LIS microdata experts gather national-level household income, budget or labor-force surveys, which include income, labor market, and demographic variables at the household- and person-level, and harmonize these data so that comparative studies are facilitated. Since many of the countries participating in the project restrict direct access to such data for confidentiality purposes, the LIS requires researchers to submit an application before being granted access to the data and employs an automated system based upon electronic mail to make the harmonized data available to researchers around the globe.² At present, there are more than

¹ LIS was founded in 1983 under the joint sponsorship of the government of the Grand Duchy of Luxembourg and the Centre for Population, Poverty and Policy Studies (CEPS). The LIS project's research center and data archive are cooperatively funded by the national science and social science research foundations of approximately 30 member and associated countries from four continents.

² The application form may be downloaded from the LIS Web Page at the following link: <http://www.lisproject.org/introduction/userform.htm>.

100 data sets contained in the LIS data archive, including 30 countries and covering seven points in time spanning almost 40 years.³

In the main, LIS data have been used to examine economic and social policies and their effects on outcomes including poverty, income inequality and employment status. Others have employed LIS data to analyze immigration, family formation, wage patterns, health status, political attitudes and voting behavior. For our purposes, what is most relevant is that LIS data also often include information about a respondent's region of residence, allowing researchers to aggregate observations at the subnational (regional) level rather than at the national level.

Subnational Analyses Using LIS Data

The majority of the national-level surveys included in the LIS report the respondent's state/province/municipality of residence.⁴ For the most part, these units are well defined politically, territorially, and culturally. Moreover, the geographic classifications used within the European surveys conform to the Nomenclature of Territorial Units for Statistics (NUTS) adopted by the European Statistical Agency, EUROSTAT. For example, table 1 provides details for the 18 European countries where subnational analyses may be accomplished.⁵ By examining the third column of this table, it is evident that the level of geographic detail included within each survey ranges from districts and municipalities, which are sometimes referred to as NUTS Level 4,⁶ in the Czech Republic, Denmark and Estonia, to NUTS Level 1 in Austria, Belgium, Germany, Greece, Spain and the UK. Accordingly, the number of subnational units ranges from the 86 *okres* of the Czech Republic to the three Austrian and Belgian States. In between these levels one

³ Two "historical" waves covering the late-1960s and mid-1970s are available for a handful of countries. The five more recent waves include data from 1979 until 2002. See <http://www.lisproject.org/techdoc/datasets.htm> for more details.

⁴ With a few exceptions, this information is reported in the LIS variable "D7." The other LIS variable related to geography is "D20," which most often contains information about the population density of a respondent's residence but may also report region of residence.

⁵ The following EU-25 countries are currently not participating in the LIS project: Cyprus, Latvia, Lithuania, Malta and Portugal. Furthermore, an additional four European countries included in the LIS do not allow for regional aggregations of households: Luxembourg, the Netherlands, Slovenia and Norway.

⁶ These territories are classified as "Local Administrative Units" (LAUs). They are not regulated by EUROSTAT and thus may vary by member state. See European Communities Commission (2006).

also has the opportunity to identify households at other meaningful geographic levels, such as Italian Regions (NUTS Level 2). However, there are at least two reasons why many researchers may wish to aggregate households together in geographic territories that are larger than the lowest level reported in a particular survey.

TABLE 1 ABOUT HERE

First, most LIS surveys do not include a sufficient number of observations within NUTS levels 3 or 4 to permit statistically meaningful aggregate measures to be calculated. More specifically, confidence intervals for estimates based on too few observations will be too large to permit us to make any substantive conclusions (see e.g., Osberg and Xu, 1999; Stewart, 2002; Moran, 2006). For example, although the Czech, Danish and Estonian surveys include a substantial number of observations such that even the most detailed geographic information may be used with some statistical confidence, with an average of 100 observations included in the survey at the NUTS level 3 classification, Hungarian households should be aggregated at NUTS level 2 or higher so that statistically meaningful comparisons may be made between Hungarian regions.

Second, researchers may also wish to aggregate household geographic locations at a higher NUTS level in order to work with territorial definitions that are more meaningful historically, politically, culturally and economically. For example, aggregating households at the level of Italian Regions, which is NUTS Level 2, rather than in groups of regions such as “Islands” or “Center,” which is appropriate to NUTS Level 1, allows researchers to measure poverty using a more relevant social reference category: the Region. In short, Italian Regions have significance while “groups of Regions” do not. Thus, many researchers might prefer the trade-off between numbers of observations gained by aggregating several regions together over the theoretical justification for working with territorial units that have some historic or political significance.

Research Adopting a Subnational Approach

In an investigation of subnational poverty using the LIS data, Jesuit et al. (2003) examine disposable income poverty rates for LIS waves 3 (around 1990) and 4 (mid-1990s) for eight countries: Australia, Canada, Finland, France, Germany, Italy, the United Kingdom, and the United States. Adopting both national and local poverty lines, one of their main conclusions is that “...the regional dimension is vitally important in measuring poverty. Studies at the national level of analysis mask intracountry variance in the rate of poverty and do not allow us to identify geographic concentrations of individuals living in dire economic straits” (2003: 365. See also Coder et al., 2001; Stewart, 2002; Berthoud, 2004; Ritakallio and Kangas, 2004).

More recently, Förster et al. (2005) examine subnational poverty and income inequality in Central and Eastern European countries. With respect to poverty, they find that “...capital cities and major urban areas are mainly winners, while regions which are longer distances from central cities and which are further from their richer western neighbors characterize losers” (336). In addition, this study seeks to determine the relative contributions of intra- and inter-regional inequality to overall inequality within nations. The authors conclude that “...the contribution of intra-regional inequalities to overall inequality largely outweighs the inter-regional contribution and, contrary to conventional wisdom, the latter is less important in CEE countries than in some of the Western EU countries” (Ibid).

Besides being able to measure poverty and income inequality for subnational units, as one would for a nation-state, researchers have also used the LIS data to aggregate households from regions in one country together with households in a neighboring country's (or countries') regions. For example, Allegrezza et al. (2004) combine household data from Walloon (Belgium); Lorraine (France); Rhineland-Palatinate and Saarland (Germany); and Luxembourg in order to perform intra- and inter-regional poverty and income inequality analyses for the so-called *Grande Région*. Using Purchasing Power Parities (PPPs), one of the authors' main conclusions is that “...both in relative and absolute terms, the income gaps between Luxembourg and the regions

comprising the *Grande Région* are even wider than the gap between northern and southern Italy” (2004: 280). In sum, such subnational examinations of poverty and income inequality have contributed much to our understanding of the spatial distribution of economic well-being in Europe and to our knowledge about the extent to which regional disparities in income contribute to overall inequality.

A Brief Application: Regional Poverty in Italy and Germany in 2000

The final contribution this chapter hopes to make is to offer subnational poverty rates for the most recent LIS surveys made available for Germany and Italy. These countries were selected because they illustrate many of the most important issues concerning regional aggregations of LIS surveys previously discussed. For example, as shown in the final column of table 2 the number of households included within Regions and Länder in these surveys presents a dilemma to researchers. On the one hand, fewer observations tend to yield larger standard errors of the estimates, making it more difficult to conclude that observed differences in poverty rates are genuine. On the other hand, if we were to aggregate these territories at NUTS Level 1 we would not be able to compare territories that have any historic or political significance. As such, we offer a compromise by combining neighboring regions, such as Basilicata and Calabria in Italy and Bremen and Lower Saxony in Germany, where a territory had fewer than 200 observations. In order to assess whether the reported differences in poverty rates are statistically significant, we compute confidence intervals using the bootstrap method.

TABLE 2 ABOUT HERE

As shown in this table, despite limitations in the number of observations in many regions, there was substantial variation in the rates of poverty within Germany and Italy in 2000 and one is able to make several statistically meaningful conclusions based upon these estimates and their confidence intervals. As much previous research has indicated, subnational disparities in Italy, known for being among the largest in the developed world, were considerably greater than those in Germany. For example, while only 2.0% of the population fell below the national poverty line

in Emilia-Romagna, just over 40% of Sicilians were below the poverty threshold in 2000.⁷ Furthermore, about one-quarter of Italians living in the southern regions of Basilicata/Calabria and Campania fell below the national poverty line while fewer than 5% of Italians living in the central and northern regions of Tuscany, Umbria, Lombardy and Veneto did so in 2000. In sum, as other research has indicated, regional disparities in Italy are exceptionally wide and statistically significant.

Subnational disparities in economic well-being were much lower in Germany, however, as the difference between the Länder reporting the highest and lowest poverty rates, Berlin and Baden-Württemberg, equaling 13.6% and 6.7%, respectively, was about 7 percentage points. Although we cannot conclude with 95% statistical certainty that this reported gap is genuine, the upper and lower bounds of the confidence intervals overlap considerably suggesting that the reported difference is very likely actual. Regardless, overall the results for Germany indicate that the economic gap between eastern and western Länder was relatively narrow by 2000, suggesting that the decade after reunification has been largely successful in reducing the wide economic disparities found upon reunification (see also Sinn and Westerman, 2000). Finally, it is interesting to note that despite the fact that the national rate of poverty in Germany is about 4% lower than the Italian national rate (8.3% and 12.7%, respectively), the regions having the lowest poverty rates are found in central and northern Italy rather than in Germany.

Conclusion

This brief chapter introduced readers to the possibilities for subnational research using the LIS data archive and reviewed a selection of recent research adopting this approach. A brief examination of regional poverty rates in Germany and Italy in 2000, the most recent LIS data available for these countries, demonstrated how a subnational approach to measuring poverty enables researchers to identify pockets of poverty that would otherwise be overlooked in national studies.

⁷ See Jesuit et al. (2003) for a discussion of the merits of national versus local poverty lines.

References

- Allegrezza, Serge, Georges Heinrich and David Jesuit. 2004. "Poverty and Income Inequality in Luxembourg and the *Grande Région* in Comparative Perspective." *Socio-Economic Review* 2: 271-91.
- Beblo, Miriam and Thomas Knaus. 2001. "Measuring Income Inequality in Euroland." *Review of Income and Wealth* 47: 301-320.
- Berthoud, Richard. 2004. *Patterns of Poverty Across Europe*. Bristol, UK: The Policy Press.
- European Communities Commission. 2006. "Basic principles of the NUTS." Luxembourg: Author. Retrieved October 1, 2006 from:
http://ec.europa.eu/comm/eurostat/ramon/nuts/basicnuts_regions_en.html
- Förster, Michael, David Jesuit and Timothy Smeeding. 2005. "Regional Poverty and Income Inequality in Central and Eastern Europe: Evidence from the Luxembourg Income Study." In *Spatial Inequality and Development*, Ravi Kanbur and Anthony J. Venables (eds.), Oxford, England: Oxford University Press, 311-347.
- Jesuit, David, Lee Rainwater and Timothy Smeeding. 2003. "Regional Poverty within Rich Countries." In *Inequality, Welfare and Poverty: Theory and Measurement*, John A. Bishop and Yoram Amiel (eds.), New York: Elsevier Science, 345-377.
- Moran, Timothy Patrick. 2006. "Statistical Inference for Measures of Inequality With a Cross-National Bootstrap Application." *Sociological Methods and Research*. 34: 296-333.

Osberg, L. and K. Xu. 1999. "Poverty Intensity: How Well do Canadian Provinces Compare?"

LIS Working Paper No. 203. Center for Policy Research, Syracuse University.

Rainwater, Lee, Timothy M. Smeeding and John Coder. 2001. "Child Poverty Across States, Nations and Continents." In K. Vleminckx and T.M. Smeeding (eds.), *Child Well-Being, Child Poverty and Child Policy in Modern Nations: What Do We Know?* UK: The Policy Press.

Ritakallio, Veli Matti and Olli Kangas. 2004. "Relative to What? Cross-national Picture of European Poverty Measured by Regional, National and European Standards." Luxembourg Income Study Working Paper No. 384. Center for Policy Research, Syracuse University.

Sinn, Hans W. and Frank Westermann. 2000. "Two Mezzogiornos." CESifo Working Paper #378. Munich: CESifo.

Stewart, Kitty. 2002. "Measuring Well-Being and Exclusion in Europe's Regions." Luxembourg Income Study Working Paper No. 303. Center for Policy Research, Syracuse University.

Table 1. European LIS surveys and Subnational Detail

Country	Year	NUTS Level	Number of subnational units	Mean observations
Austria	2000	1	3	833
Belgium	2000	1	3	908
Czech Rep.	1996	4	86	327
Denmark*	1992	4	32	379
Estonia	2000	4	16	379
Finland	2000	3	20	521
France	1994	2	22	513
Germany ^a	2000	1	16	687
Greece	2000	1	4	957
Hungary	1999	3	20	101
Ireland	2000	3	8	319
Italy	2000	2	20	400
Poland	1999	2	16	1964
Slovakia	1996	3	8	2042
Spain	2000	1	7	688
Sweden	2000	3	21	690
UK ^b	1999	1	10	2272
Switzerland	2002	2	7	532

^a Rhineland-Palatinate and Saarland are combined into one region in the LIS survey. In addition, residence in east and west Berlin is reported (these may be combined by a researcher, however).

^b Northern Ireland excluded.

*Regional information located in LIS variable D20.

Table 2. Regional Poverty Rates, 95% Confidence Intervals and Number of Observations, 2000

Country	Territorial Unit	95% c.i.			N
		Estimate	Lower	Upper	
Germany	Nation	8.3	7.6	9.1	10982
	Baden-Württemberg	6.7	4.9	8.5	1260
	Bavaria	7.6	5.8	9.3	1515
	Berlin	13.6	8.0	19.2	443
	Brandenburg	8.5	5.0	12.1	469
	Hessen	7.7	5.1	10.2	742
	Lower Saxony + Bremen	7.5	4.7	10.3	515
	Mecklenburg-Western Pomerania	9.5	4.0	13.9	281
	North Rhine Westphalia	8.8	7.3	10.3	2332
	Rhineland-Palatinate + Saarland	9.2	5.4	13.0	663
	Saxony	7.6	5.3	9.9	825
	Saxony-Anhalt	11.1	7.2	15.1	475
	Schleswig-Holstein + Hamburg	7.1	5.3	9.0	985
	Thuringia	10.2	5.9	14.6	477
Italy	Nation	12.7	11.5	13.8	7925
	Abruzzi + Molise	13.2	7.9	18.4	310
	Basilicata + Calabria	26.9	20.3	33.5	302
	Campania	26.0	22.2	29.8	780
	Emilia-Romagna	2.0	1.0	2.9	750
	Friuli-Venezia Giulia	5.4	1.0	9.9	255
	Lazio	7.4	3.6	11.2	423
	Liguria	6.2	3.1	9.2	314
	Lombardia	4.0	1.6	6.4	858
	Marche	8.5	3.4	13.5	328
	Piemonte + Val D'Aosta	5.4	3.3	7.6	756
	Puglia	19.3	14.8	23.7	460
	Sardegna	17.5	11.5	23.5	308
	Sicilia	40.4	33.2	47.6	614
	Toscana	4.8	2.7	6.9	596
	Umbria	4.2	1.7	6.6	271
	Veneto + Trentino Alto Adige	4.3	6.2	2.3	600

Source: Author's calculations using LIS data.

Notes: Poverty line is defined as 50% of median national equivalent income.
Confidence intervals computed using 300 iterations of the bootstrap method.