Issue No. 19 (September 2021)

Inequality Matters

Quarterly updates on inequality research, LIS micro data releases, and other developments at LIS

Message from New LIS Director

Dear readers,

I am delighted to be writing to you as the new LIS Director. It is enormously exciting to be stepping in the shoes of Daniele Checchi whose five years in charge have left LIS more visible, respected and relevant than ever. I am grateful for the trust and confidence that has been put in me by Daniele, the LIS team, and by Francois Bourguignon and the entire ASBL. I am honored too, to be following in the footsteps of previous directors Janet Gornick, Markus Jantti, and founders Tim Smeeding and Lee Rainwater. I look forward to forging good working relations with the LIS team and hope to take advantage of the reasonably close proximity of my home town, Amsterdam, to Esch-Belval in order to become a regular face also in the corridors of the LIS office in Luxembourg.

The LIS Directorship was perhaps not the most immediately obvious step in my career. After 23 years in the research department of the World Bank, I have been teaching economics at the Vrije Universiteit Amsterdam since early 2015. I am a development economist and have been focused on development issues throughout my career. However, analyzing and measuring economic wellbeing in low income countries has been a central focus of my research throughout the past three decades. I have been heavily exposed to the challenges and opportunities that household survey data embody. I have been involved in the study of issues surrounding data harmonization, and I have actively participated in the exploration of methods to strengthen data comparability. From that perspective, the move to LIS with its mission to disseminate high quality, harmonized, household survey data, makes a good deal of sense. I am thrilled to come on board for this reason.

I hope, moreover, that my experience and background in development may also be helpful given the particular juncture that LIS finds itself at. There is steadily increasing flow of datasets from low and middle-income countries entering into the LIS archives. Whether, and how, to harmonize these with the core LIS data, are important questions. There clearly exists demand for an ability to conduct cross-country comparisons, along a variety of dimensions, involving both developed and developing countries. But as the range of countries in terms of levels of economic development, widens, the underlying data also become increasingly diverse in terms of quality, structure, and composition. Fundamental questions, such as the definition of income, consumption and wealth have to be revisited. New harmonization methods may need to be experimented with. Judgement calls have to be made. We need to reflect on how LIS can best navigate these new opportunities. I hope to contribute to that reflection.

As ideas develop, we will be looking to air them in our LIS newsletter. I hope that you will also convey to us your thoughts and reactions. It promises to be an exciting time!

Enjoy reading! ***Peter Lanjouw***

IN THIS ISSUE

View all the newsletter issues at: www.lisdatacenter.org/newsletter
Subscribe here to our mailing list to receive the newsletter and news from LIS!
Interested in contributing to the *Inequality Matters* policy/research briefs? Please contact us at : neugschwender@lisdatacenter.org

Work-Family Reconciliation Policies: Good or Bad for Gender Employment Inequalities?

Sarah L. Kostecki , (University of Luxembourg)

Introduction

In the work-family policy literature a debate persists about whether generous work-family reconciliation policies (those that help parents, mainly mothers, reconcile tensions between paid and unpaid childrearing, such as leave and early childhood education and care (ECEC)) promotes women's employment, but have adverse consequences for women's attainment. Key to this debate is a growing consensus that these policies may differentially impact women's employment and attainment, by class. Many researchers in this genre utilize policy indicators (hereafter, indicators) in combination with the LIS data to evaluate the links between these policies and women's outcomes (Mandel and Semyonov 2005, 2006; Pettit and Hook 2009; Mandel 2012; Korpi, Ferrarini and Englund 2013; Brady, Blome, and Kmec 2020).

In this article I extend this research. I first address the call for better indicators (Hook and Li 2020; Mandel 2012; Sirén et al. 2020). I present seven new, disaggregated, precise, and multidimensional indicators for leave and ECEC across 24 high-income countries and relationships among the indicators (These indicators are available in a new policy dataset along with detailed documentation here). I then evaluate the relationships between the policy dimensions and gender gaps in employment and annual earnings (hereafter, earnings) for working-aged low-and highly educated men and women using the new indicators and the LIS data for 24 high-income countries around the years 2010 and 2013 (Waves VIII and IX). The goal is to evaluate both parts of the gendered tradeoffs hypothesis (links with employment and attainment) for the two class groups. I do this by using established methods from the literature and the new indicators, to determine relationships among the policy dimensions and outcomes of interest.

"Well-developed" work family reconciliation policies

Leave and ECEC policies are the two most widely studied work-family reconciliation policies in the research on the unintended consequences of family policies (Hook and Li 2020). Though both policies help mothers reconcile work with unpaid childrearing, the policies have different implications for women's attainment – leave promotes mothers' exit from the labor market for one or more periods of time and may negatively affect earnings and status, while ECEC promotes women's time at work and should not be linked with adverse outcomes. Many of the indicators used in past studies are overly simplified, do not include periods of leave for fathers' use, and include only one indicator of leave and ECEC.[1]

I propose seven new indicators, five for leave and two for ECEC, that measure different policy dimensions (see table 1). The seven new indicators show which policies are "well-developed"; those policies that best support women's employment across multiple policy dimensions, based on evidence from earlier scientific studies (Gornick and Meyers 2009; Kostecki 2021).[2] Four of the leave policy indicators measure generosity or "how much" leave is reserved for the mothers' and fathers' use. The index of leave policy universality measures "the breadth of the population covered under leave legislation as well as the accessibility of leave." For ECEC, two indicators measure ECEC availability; young children's enrollment in childcare services (between 0 and up to 2 years) and slightly older children's enrollment in pre-primary education services (between 3 years and up to compulsory schooling), weighted by the dominant mechanism of provision (public or private care).

Relationships among the indicators are shown in Table 2 using Pearson correlation coefficients. The correlation results generally show the policy dimensions are positively and significantly correlated with one another, suggesting leave and ECEC policies that are well-developed along one dimension are well-developed across other dimensions. The exception are the correlations among the unpaid/poorly paid leave indicators and additional indicators. These policy dimensions are generous in terms of length, but not payment, which suggests unpaid/poorly paid leaves may not be "well-developed" in comparison to other leave and ECEC policy dimensions. Long, unpaid/poorly paid leave periods are also negatively and significantly correlated with enrollment of children 0-2 years of age (because mothers stay home to care for children). Overall, the correlations show the importance of measuring leave periods (for both mothers and fathers) at different wage replacement cutoffs.[3] More precise indicators of leave policy especially complicate the issue of how to measure these policies for use in comparative research.

Leave, ECEC and gender gaps: employment and annual earnings

This section assesses the links among the policy dimensions and gender gaps between men and women across measures of employment and annual earnings using correlation analyses and multilevel modelling. The sample is of men and women 25-54 years of age (prime working-aged persons that excludes students and early retirees). The focus is on gender gaps in employment and earnings between low-educated men and women (below a high school education) and highly educated men and women (at least a tertiary education or higher).[4] Education is used as the measure of class. Education is an indicator of skill and harmonized by LIS for comparative research. Earnings is a measure that is a widely used measuring women's attainment.

The methods for the multilevel model analyses were adapted from Mandel (2012) and Brady, Blome, and Kmec (2020). All 24 countries are included in the employment models. The sample is reduced to 18 countries for the earnings models to exclude countries that report net or mixed earnings[5] and where weekly hours worked cannot be used as a control variable.[6] Earnings were first converted to country-year specific percentiles.[7] The same countries are included in the correlation and multilevel analyses.

The questions are whether well-developed leave and ECEC policy dimensions are linked to reduced gender gaps in employment, but unintended, wider earnings gaps between men and women? Do these relationships differ across class groups?

Table 1. Leave and ECEC policy dimension indicators for 24 high-income countries

	Leave Policy[a]					ECEC Policy	
	Generosity[b]				Universality[e]	Generosity[f]	
Country	Unpaid/poorly paid leave (mothers)[c]	Well-paid leave (mothers)[d]	Unpaid/poorly paid leave (fathers)[c]	Well-paid leave (fathers)[d]	Leave universality index	ECEC enrollment (0-2 yrs., weighted)	ECEC enrollment (3 yrs. to comp. school, weighted)
Australia	52	0	0	0	10.5	16.5	51
Austria	82.6	16	4.7	0	53.0	9.5	83
Canada	50	0	0	0	23.2	12	48
Czech Republic	134	28	134	0	45.3	2	81
Denmark	0	50	0	2	67.4	67	94
Estonia	146	20	2	0	53.0	32	88
Finland	0	44	0	3	92.3	27	80
France	146	16	154	2	79.5	48	100
Germany	148	14	8.7	0	77.0	29	95
Greece	17	17	17.3	0.4	45.3	17	95
Hungary	81	24	0	1	60.6	16	95
Iceland	0	26	0	13	79.5	56	96
Ireland	56	0	14	0	46.0	14.5	37
Italy	26	20	17.3	0.2	83.5	23	94
Luxembourg	26	16	26	0	72.4	53	90
Netherlands	26	16	26	0.4	48.6	27	70
Norway	0	35	2	12	73.2	54	98
Poland	156	24	0	2	83.5	4.5	75
Slovak Republic	164	0	0	0	18.7	2.3	73
Slovenia	0	52	10.9	2.1	65.9	45	91
Spain	36	16	50	2.1	73.2	38	98
Switzerland	0	14	0	0	23.5	19	78
United Kingdom	59	6	15	0	62.0	17.5	79
United States	12	0	12	0	14.4	21.6	64

Notes: Countries for which policy data comes from around the year 2009/2010: Australia, Canada, France, Iceland, Ireland, the Slovak Republic. All other countries' policy data is from around the years 2011/2012/2013. For detailed methods of indicator construction, see the documentation for the Leave and ECEC policy dimensions dataset for 31 high- and middle-income countries.

[a]Leave policy includes maternity leave, paternity leave, and parental leave (where applicable).

[b] Indicators measure "how much" leave is available in leave legislation. Reserved (transferrable + nontransferable) + shared leave periods (including any mandatory or optional weeks of pre-birth leave) are included in the two indicators for leave allocated to mothers. Reserved and nontransferable leave periods only are included in the two measures for leave allocated to fathers.

[c]Unpaid/poorly paid leave = leave paid at less than 67 percent of usual earnings or unpaid.

[d]Well-paid leave = leave paid at 67 percent of usual earnings or higher.

[e]The leave universality index was calculated using the multiplicative method and converted to an index with a range between 0-100. The higher the value, the more universal the leave policy legislation in any particular country. The index is constructed using five separate indicators of leave universality: maternity leave coverage, leave financing of paid leave periods, and leave eligibility requirements. For the financing of leave and leave eligibility requirements, both dimensions are measured for leave allocated to mothers and fathers.
ECEC generosity indicators measures the availability of care (childcare services and pre-primary education) for young children in two groups.

[f]Weights for the provision of care as followed: public care = 1 (enrollment rate remains the same), mixed provision = .75, and private care = .50.
Sources: Author's own calculations using various sources; for detailed source information and full bibliographic entries, see the documentation for the Leave and ECEC policy dimensions dataset.

Table 2. Correlations among the seven leave and ECEC policy dimension indicators from Table 1

	1)	2)	3)	4)	5)	6)	7)
1) Unpaid/poorly paid leave (mothers)	1.00						
2) Well-paid leave (mothers)	-0.30	1.00					
3) Unpaid/poorly paid leave (fathers)	0.35*	0.01	1.00				
4) Well-paid leave (fathers)	-0.34	0.41**	-0.11	1.00			
5) Enroll 0-2 yrs. (weighted)	-0.45**	0.54***	0.04	0.56***	1.00		
6) Enroll 3 yrs. to comp. school (weighted)	-0.02	0.59***	0.22	0.37*	0.54***	1.00	
7) Leave universality index	-0.03	0.62***	0.14	0.41**	0.50**	0.65***	1.00

Notes: Pearson correlation coefficients.
Source: Author's own calculations using Stata statistical software and values for the seven policy indicators shown in Table 1.
*P< .10, **P< .05, *** P< .01.

Correlations

Table 3 displays Pearson correlation coefficients among the seven policy dimensions and the two outcomes for the two class groups. A negative correlation implies the policy dimension is correlated with smaller employment or earnings gaps. Positive correlations signal the policy dimension is correlated with wider employment or earnings gaps. Positive correlations therefore imply the unintended consequences of leave or ECEC policy.

The correlation results suggest that different leave and ECEC policy dimensions may have different relationships to women's employment and attainment and that these relationships vary by class. There are no unintended relationships among the policy dimensions and outcomes between highly educated men and women. Any unintended relationships among the policy dimensions and outcomes are for the low-educated group. Unpaid/poorly paid leaves for fathers are moderately and positively correlated with the employment gap between low-educated men and women. Well-paid leaves for mothers and the universality of leave are moderately and positively correlated with earnings gaps between low-educated men and women. The results again point to the importance of measuring leave allocated to both mothers and fathers at different wage cutoffs.

Multilevel models

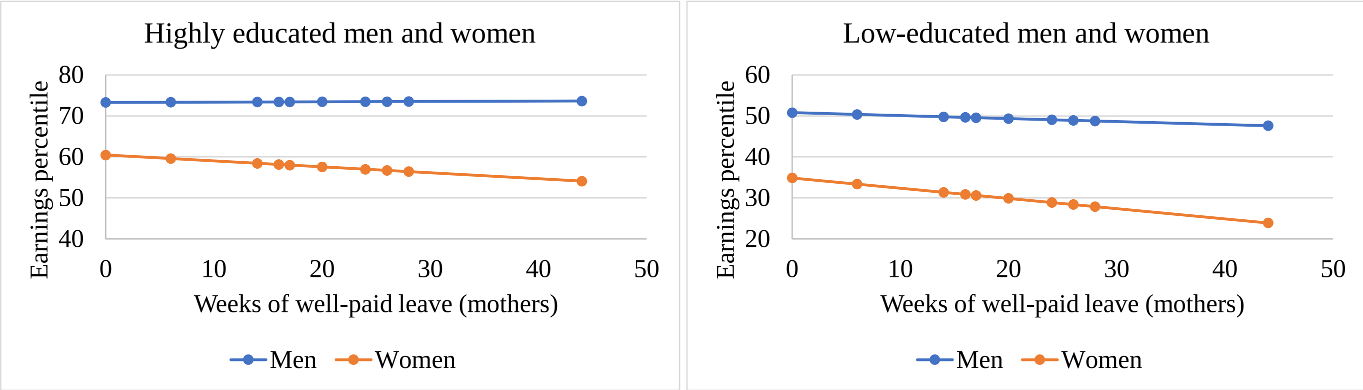
Supporting the correlation results, a one percentage point increase on the universality index improves highly educated women's employment odds by .007, holding the other variables constant at their means (Supplemental Table 1, M1). The models of earnings show the unintended consequences of only well-paid leaves for mothers for both groups (Supplemental Table 2, M3). Increasing well-paid leaves for mothers results in increased gender wage gaps for both low- and highly educated women (y=-.18 percentiles and p<.05, low-educated; y=-.15 percentiles and p<.10, highly educated). The results suggest that only well-paid leave periods may have adverse consequences for the earnings of low- and highly educated women (compared to low- and highly educated men), supporting earlier findings by Mandel (2012).

Table 3. Correlations among the seven leave and ECEC policy dimension indicators and gender gaps, employment and annual earnings, low- and highly educated men and women

	1) Unpaid/poorly paid leave (mothers)	2) Well-paid leave (mothers)	3) Unpaid/poorly paid leave (fathers)	4) Well-paid leave (fathers)	5) Enroll 0-2 yrs. (weighted)	6) Enroll 3 yrs. to comp. school (weighted)	7) Leave universality index
Employment gap (low education)	-0.21	-0.28	0.38*	-0.29	-0.15	-0.04	-0.04
Employment gap (high education)	0.18	-0.37*	-0.01	-0.28	-0.49**	-0.2	-0.36*
Earnings gap (low education)	-0.04	0.45*	-0.05	0.29	0.23	0.18	0.40*
Earnings gap (high education)	0.10	0.24	-0.52**	0.17	-0.12	0.16	-0.07

Notes: Sample includes men and women aged 25-54 years. Pearson correlation coefficients. Countries included in the correlations are the same countries included in the multilevel models. Gaps are unadjusted.
N=24 countries (employment gaps). For the correlations between unpaid/poorly paid leave for fathers and employment gaps, the Czech Republic and France were removed because the leave values had too much an influence on the results (N=22 countries).
N=18 countries (annual earnings). Only those countries with reported gross earnings and hours worked in the LIS data are included in the correlations and multilevel models. In addition, the Czech Republic is removed from the correlations between unpaid/poorly paid leave (fathers) and earnings gaps because the leave value has too much of an influence on the results (N=17 countries).
Source: Author's own calculations using Stata statistical software and values for the seven policy indicators shown in Table 1 and the LIS data Waves VIII and IX.
*P< .10, **P< .05, *** P< .01.

Figure 1. Predicted earnings percentiles by weeks of well-paid leaves (mothers)



Source: Predicted values were derived from Supplemental Table 2, Model 3 (for both low- and high educated men and women). Figure constructed in Microsoft Excel.

Finally, Figure 1 shows the predicted earnings percentiles of both low- and highly educated men and women at the different lengths of well-paid leaves for mothers (ranging from a low of 0 weeks to a high of 44 weeks). The predicted earnings of low- and highly educated men remains largely the same, regardless of the length of well-paid leave for mothers. However, the predicted earnings of low- and highly educated women decline the longer the weeks of well-paid leave for mothers. Therefore, gender earnings gaps between low-educated men and women and highly educated men and women are wider the longer the length of well-paid leave.

Conclusions

First, this short article highlights the importance of constructing and using more precise, disaggregated, and multidimensional measures in work-family policy research. Correlations among the policy dimensions showed the importance of measuring leave policy at different wage cutoffs. Leave policies that are generous in terms of length and payment are not the same policies that are generous in terms of length but not payment. The results suggest singular policy measures used in many past studies may not represent the scope of leave policy more generally. For ECEC, though I constructed two measures of generosity using enrollment rates (like past studies), other dimensions – such as opening hours – continue to be difficult to measure due to data limitations and regulations about how ECEC is set up across high-income countries (Hook and Li 2020; Kostecki 2021; Sirén et al. 2020).

Regarding relationships specifically between leave policies and outcomes, I argue what matters is how different leave policy *dimensions* drive unintended consequences for women in employment and others not at all. Well-paid leaves for mothers may adversely affect the earnings of women, by class, more than unpaid/poorly paid leaves for mothers. More consideration needs to be given to fathers' leave and relationships with women's employment and attainment, by class. Also important is to address the issues of unintended consequences of leave policies for women/mothers of different class groups and to evaluate the link between gender employment inequalities and other policies such as working time regulations.[8]

There is a growing consensus that ECEC is not adversely linked with women's employment or attainment, relative to men (Brady, Blome, and Kmec 2020; Hook and Li 2020; Mandel 2012; Olivetti and Petrongolo 2017). My findings support past research. To study leave and ECEC together, research questions need to be reframed around the specific effects we expect to see of both policies – positive and negative.

No leave policy across the 24 countries provides working mothers and fathers with equal amounts of reserved, nontransferable well-paid leaves (Table 1). Overall, I argue leave policies that treat men and women the same and promote equality in the gendered division of labor can at the very least promote gender equality in employment. However, the understanding that gender employment inequalities may occur because of certain design features of leave policy can help policy makers to continually improve this policy over time to adapt to the needs of working parents.

1 For some exceptions, see Korpi, Ferrarini, and Englund (2013) and Olivetti and Petrongolo (2017).

2 See specifically chapter 3 in my dissertation for past studies that have considered the development and measurement of leave and ECEC policy across different dimensions and relationships to women's employment. I drew on these studies to develop the indicators in my research.

3 Chapter 3 in my dissertation also shows different relationships among the leave and ECEC policy measures used in three studies – Mandel and Semyonov (2005, 2006) and Brady, Blome, and Kmec (2020). In Mandel and Semyonov (2005, 2006), fully paid maternity leave is positively and significantly correlated with enrollment rates of children 0-6 years of age in publicly subsidized childcare (policies used to construct the Welfare State Intervention Index (WSII)). However, in Brady, Blome, and Kmec (2020), the measure of weeks of paid leave is not correlated with the percentage of children 0-6 enrolled in publicly subsidized childcare. Should we expect to see positive relationships among policy indicators?

4 LIS follows the ISCED 2011 standard measure of classification to ensure education is comparable across countries.

5 As gender differences might be less pronounced on net or mixed datasets than in gross datasets, assuming progressive taxation.

6 The LIS database reports three possible current incomes—gross, mixed, and net. For the 24 high-income countries included in this study, the LIS data for France 2010 and Poland 2013 report mixed income—total individual annual earnings do not account for full taxes and contributions. The LIS data for Hungary 2012 and Slovenia 2012 reports net income—total annual earnings does not capture taxes and contributions. These four countries are excluded from the earnings correlations and models. The additional 20 datasets report gross income (taxes and contributions are fully captured, collect, or imputed). In addition, Denmark 2013, France 2010, Norway 2013, Poland 2013, and Slovenia 2012 do not report weekly working hours. Weekly working hours is an important factor to evaluating gender earnings gaps. However, Mandel (2012) utilizes a control of weekly hours worked in her earnings models while Brady, Blome, and Kmec (2020) do not. Summarizing Misra et al. (2011), Brady, Blome, and Kmec (2020) argue "Annual earnings combine pay and quantity of hours, both of which are relevant to evaluating work–family policies." My decision was to include weekly hours worked because in my work with original household income surveys, there is no indication that annual earnings combine information about pay and quantity of hours (though more hours worked generally means higher annual earnings). Hours worked is therefore a necessary control to include in the models. In a future study, hours worked could be excluded as a control (and the countries that do not report weekly hours worked can be re-introduced) to determine how the results change. Countries that report net or mixed income data could also be included to determine how the results change.

7 Before annual earnings were converted to country specific percentiles, negative earnings were bottom coded by converting them to 0. At the top of the distribution, annual earnings were top coded at 10 times above the median. By utilizing country specific percentiles Mandel (2012, 245–246) argues this method is used to "avoid conflating the effect of welfare state policies with the effect of wage-setting institutions. Each respondent's wage is measured by his or her position in their national earnings distribution, irrespective of cross-national differences in the length of the wage ladder."

8 My dissertation research addresses the question of class tradeoffs among women and finds that class employment gaps between low- and highly educated women may be exacerbated by some design features of leave policy.

References

Brady, D.; Blome, A.; Kmec, J. A. (2020). "Work–Family Reconciliation Policies and Women's and Mothers' Labor Market Outcomes in Rich Democracies", *Socio-Economic Review,* 18 (1): 125–61, https://doi.org/10.1093/ser/mwy045.

Hook, J. L. and Li, M. (2020). "Gendered Tradeoffs", in *The Palgrave Handbook of Family Policy*, edited by Nieuwenhuis, R. and Van Lancker, W., Cham, Switzerland: Palgrave Macmillan, 249–66, https://doi.org/10.1007/978-3-030-54618-2_11.

Korpi, W.; Ferrarini, T; Englund, S. (2013). "Women's Opportunities under Different Family Policy Constellations: Gender, Class, and Inequality Tradeoffs in Western Countries Re-Examined", *Social Politics: International Studies in Gender, State & Society,* 20 (1): 1–40, https://doi.org/10.1093/sp/jxs028.

Gornick, Janet C. and Meyers, M. K. (2009). *Gender Equality: Transforming Family Divisions of Labor*. Vol. VI. The Real Utopias Project. New York, NY: Verso.

Kostecki, S. L. (2021). "Work-Family Reconciliation Policies Reexamined: Good or Bad for Gender and Class Inequality in Employment across Twenty-Four High-Income Countries?", CUNY Academic Works, https://academicworks.cuny.edu/gc_etds/4304.

Misra, J., Budig, M. J., and Boeckmann, I. (2011). "Cross-National Patterns in Individual and Household Employment and Work Hours by Gender and Parenthood", Comparing European Workers Part A, Research in the Sociology of Work, 22, 169–207.

Mandel, H. (2012). "Winners and Losers: The Consequences of Welfare State Policies for Gender Wage Inequality", *European Sociological Review,* 28 (2): 241–62, https://doi.org/10.1093/esr/jcq061.

Mandel, H. and Semyonov, M. (2005). "Family Policy, Wage Structures, and Gender Gaps: Sources of Earnings Inequality in 20 Countries", *American Sociological Review,* 70 (December): 949–67.

Mandel, H. and Semyonov, M. (2006). "A Welfare State Paradox: State Interventions and Women's Employment Opportunities in 22 Countries", *American Journal of Sociology,* 111 (6): 1910–49.

Olivetti, C. and Petrongolo, B. (2017). "The Economic Consequences of Family Policies: Lessons from a Century of Legislation in High-Income Countries", IZA Institute of Labor Economics Discussion Paper 10505, January.

Pettit, B. and Hook, J. L. (2009). *Gendered Tradeoffs: Family, Social Policy, and Economic Inequality in Twenty-One Countries*. New York, NY: Russell Sage Foundation.

Sirén, S.; Doctrinal, L.; Van Lancker, W.; Nieuwenhuis, R. (2020). "Childcare Indicators for the Next Generation of Research", in *The Palgrave Handbook of Family Policy*, edited by Nieuwenhuis, R. and Van Lancker, W., Cham, Switzerland: Palgrave Macmillan, 627–55, https://doi.org/10.1007/978-3-030-54618-2_24.

***See full bibliographic information for policy indicators in the Leave and ECEC policy dimensions dataset documentation.

Poverty Monitoring Under Acute Data Constraints: A Role for Imputation Methods? [1]

Peter Lanjouw ,(LIS and Vrije Universiteit Amsterdam)
Nobuo Yoshida , (World Bank)

1 Introduction

Collecting data to estimate monetary poverty in low-income countries and fragile states is a challenging task. Fielding a household survey can easily cost more than a million dollars, and it often takes two or more years between initial preparations and the final calculation of monetary poverty estimates. These constraints often prevent low-income countries and fragile states from collecting data on a regular basis.

Even when survey data are collected their quality is often rather poor. While there are many reasons for the low quality of data, a common factor is that the collection of data is complicated. It requires a well-designed questionnaire, carefully planned logistics, hiring and training of qualified enumerators and supervisors, close supervision of data collection, and accurate processing of data. If all these steps are not properly followed, the quality of the data collected may be so poor that they cannot be used to estimate poverty rates.

This note discusses ways of compiling consumption or income data from the perspective of data quality and sample size. The focus is on low income countries with limited resources and capacity. Given that consumption data are more commonly collected in such countries we couch the discussion largely in terms of consumption data. However the discussion is largely applicable also to income data.

We start with a brief review of the recommendations on sample size provided in the World Bank's Living Standards Measurement Study (LSMS) guidelines (Grosh and Munoz 1996 and World Bank 2000). These LSMS guidelines provide advice on the collection of consumption data for the purpose of poverty measurement and policy analysis. We indicate that survey samples in low income countries have been increasing in size over time – potentially affecting data quality and raising costs. We then describe two approaches that apply survey-to-survey imputation techniques aimed at saving interview time and data collection costs. We limit our focus to these two imputation methodologies, not because others don't exist, but because they provide two examples of alternatives to the conventional approach of direct data collection that have come to receive some kind of official sanction, having been accepted for the World Bank's global poverty monitoring effort.

Our objective in this note is to draw attention to recent exploration of imputation methods as a means to improving the quality and increasing the availability of distributional data at the country level.

2 The Living Standards Measurement Study (LSMS) guidelines

The Living Standards Measurement Study (LSMS) program guidelines recognize the challenge of collecting complex data in developing countries (Grosh and Munoz, 1996 and the World Bank, 2000). They recommend a relatively small sample of around 2000 to 5000 households for a typical LSMS survey.

The guidelines recommend a relatively small sample size in order to limit non-sampling error. They suggest that sample sizes should be large enough to produce reliable statistics at the national level, and possibly at the urban and rural level, but should not aim to produce reliable statistics at the subnational level. Increasing the size of the sample reduces sampling error but makes survey implementation significantly more difficult, resulting in a higher incidence of error in data collection and processing. Such non-sampling error is thought to likely increase in particular if the sample size is expanded and if the National Statistical Office (NSO) in a country has only limited capacity and/or experience in conducting complex surveys.

According to World Bank (2000), in the first 15 years of the LSMS program, many countries followed the general recommendations on sample size. In most countries the sample size was kept below 6000 (see Table 1). Over time, however, the situation evolved – currently, many countries, including many low-income countries and fragile states, collect data from more than 10,000 households. According to the Global Monitoring Database (2021) of the World Bank, the average sample size of all household surveys in sub-Saharan Africa used for the estimation of the international poverty measures is 10,700 households (Table 2). The average sample size in IDA countries is 10,835 households, and that of the fragile and conflict states (FCS) is 8,952 households. Table 2 also shows while sample sizes are large, data collection in most IDA and FCS countries is infrequent. More than half the countries in IDA and FCS groups are six (IDA) and seven (FCS) years old or older. As of March 2021, the most recent survey for the Central African Republic dates back to 2008.

A question is whether sample sizes of more than 10,000 are warranted in low income countries. Large sample sizes allow countries to produce more granular poverty statistics, but plausibly increase non-sampling error and survey implementation costs. Non-sampling error could be substantial, in particular, if local survey capacity is limited, where data collection occurs in settings exposed to conflict and violence, or where physical access and shortages of skilled manpower pose particular constraints. Increasing the sample size of surveys is also associated with higher implementation costs. This latter consideration is particularly important where budgets are sharply binding. Reducing sample sizes could lower survey implementation costs. This could allow surveys to be fielded more frequently, for given a fixed budget.

Table 1. LSMS Surveys in the first 15 years of the LSMS initiative

Country	Year of first survey	Sample size
Albania	1996	1,500
Algeria	1995	5,900
Armenia	1996	4,920
Azerbaijan	1995	2,016
Bolivia	1989	4,330-9,160
Brazil	1996	5,000
Bulgaria	1995	2,000
Cambodia	1997	6,010
China (Hebei and Liaoing only)	1995	800
Côte d'Ivoire	1985	1,600
Ecuador	1994	4,500
Ghana	1987/88	3,200
Guyana	1992/93	1,800
Jamaica	1988	2,000-4,400
Kazakhstan	1996	2,000
Kyrgyz Republic	1994	2,100
Mauritania	1988	1,600
Morocco	1991	3,360-4,800
Nepal	1996	3,373
Nicaragua	1993	4,454
Pakistan	1991	4,800
Panama	1997	4,945
Paraguay	1997/98	5,000
Peru	1985	1,500-3,623
Romania	1994/95	31,200
South Africa	1993	8,850
Tajikistan	1999	2,000
Tanzania-Kagera	1991	800
Tanzania-Human Resource Development Survey	1993	5,200
Tunisia	1995/96	3,800
Turkmenistan	1997	2,350
Vietnam	1992/93	4,800-6,000

Source: World Bank (2000)

Table 2. Summary statistics of the latest household surveys in sub-Saharan Africa

Country Group	sample size (mean)	Survey year Oldest	Median	Newest
All	10,700	2008	2015	2019
Blend	12,079	2011	2015	2019
IBRD	9,125	2014	2016.5	2018
IDA	10,835	2008	2015	2018
FCS	8,952	2008	2014	2019

Source: Global Monitoring Database (2021) World Bank

3 Innovations in the field of official poverty data collection and estimation

Sample size consideration thus have a bearing on both the quality of household survey data and the affordability of regular data collection. It should be recognized, however, that even a modestly sized LSMS survey represents a significant burden for some low-income countries and fragile states. This has prompted exploration of additional options. Two approaches that have recently seen implementation at the World Bank are the Rapid Consumption Survey (RCS) approach developed by Pape and Mistiaen (2018) and the SWIFT 2.0 approach proposed by Yoshida et al (2020). Both aim to further reduce time and cost of data collection. As mentioned above, these two approaches have been formally incorporated into the World Bank's global poverty monitoring efforts.

Both approaches are underpinned by survey-to-survey (S2S) imputation procedures. S2S imputation involves the estimation of an imputation model in a "training" dataset by running regressions of household expenditures or incomes on poverty proxies. Household expenditures and poverty rates are then imputed into an "output" dataset by substituting poverty proxies of the output data into the model.

There are two key assumptions in the standard S2S methodology. First, that the relationship between household income or expenditure and poverty correlates can be expressed in an equation such as (1):

$$lny_{ho} = x_{ho}'\beta_o + u_{ho} \quad (1)$$

where $u_{ho} \sim N(0, \sigma_o)$

$\ln y_{ho}$ refers to a natural logarithm of household income or expenditure of household h in the output data o. x_{ho} is a $(k \times 1)$ vector of poverty correlates of household h in the output data, o. β_o is a $(k \times 1)$ vector of coefficients of poverty correlates (x_{ho}). u_{ho} refers to a residual and is often assumed to follow a normal distribution of $N(0, \sigma_o)$.[2] The output data includes the poverty proxy data $\{x_{ho}\}_{h=1}^{H}$ but do not include household expenditures $\{\, lny_{ho}\}_{h=1}^{H}$, which are to be imputed. For the sake of exposition, the relationship is assumed to be linear, but this can be relaxed.

The second key assumption is that the relationship between household expenditures and poverty proxies follows the equation (1) in the training data *t* as well.

$$lny_{ht} = x_{ht}'\beta_o + u_{ht} \quad (1')$$

where $u_{ht} \sim N(0, \sigma_o)$

The S2S estimates parameters in equation (1') such as $(\hat{\beta}_o, \hat{\sigma}_o)$ with their distributions in the training data, draws them $(\tilde{\beta}^r, \tilde{u}_h^r)$ randomly from their estimated distributions, and substitutes them into equation (1) to impute household expenditures for all households in the output data. The S2S repeats this imputation (say, 100 times), resulting in 100 vectors of household expenditures $(\widetilde{lny}_{ho}^{r})$ in the output data. Poverty and inequality measures are estimated in each of the 100 vectors and the averages are the point estimates of poverty and inequality measures. Also using the 100 estimates, standard errors of all measures can be estimated. [3]

A critical assumption in such S2S techniques is that the models underpinning the imputation from the training data set to the output dataset are stable, in the sense of the parameter estimates being appropriate for both the training and output dataset. If this assumption does not hold, a model estimated in the training data cannot reliably and accurately impute household expenditures and poverty rates into the output data.

A simple solution is to collect the training data and the output data simultaneously. If so, the model stability assumption should hold. It is important to note that the training data needs to include consumption data, which is costly and time-consuming to collect. Therefore, a key question is how to minimize the size of the sample for collecting consumption data. Both RCS and SWIFT 2.0 collect the training and output data simultaneously, but differ in terms of how the consumption data are collected. We describe the SWIFT 2.0 and RCS approaches below.

(i) SWIFT 2.0 and its application in Zimbabwe 2019

SWIFT 2.0 is the second generation version of the original SWIFT approach (Yoshida et al, 2015, 2020). In the original SWIFT approach an income model from the latest available household budget or income survey is developed, and a new data are collected only for those regressors included in that model. Household expenditure or income is then imputed into the newly collected dataset by drawing on parameter estimates from the income model. Although the original SWIFT approach has been frequently found to perform well against official poverty estimates, it depends crucially on the assumption of stability of the underlying income model. In the face of shocks, SWIFT estimates can be unreliable.

SWIFT 2.0 was introduced to overcome the model stability concern. The idea here is to field a typical LSMS survey but collect consumption data only from a small sub-sample of households (see Figure 1). An income model is developed using the small subsample data and household expenditure or income is then imputed into the rest of the sample based on that model. Since the imputation models are created from the subsample collected concurrently, there is no model stability issue.

Figure 1

SWIFT 2.0		
	Subset	**Rest**
modules	Cons	Cons
	Non-cons	Non-cons

The subsample for which full consumption data are collected under SWIFT 2.0 can be very limited in size (typically less than 1000 households). Although collecting consumption from a limited sample may make it possible to ensure higher data quality, sampling error on poverty statistics estimated solely from the subsample will be larger than that obtained with a standard LSMS survey. However, following imputation of household expenditures into the rest of the sample, poverty estimates calculated after combining both actual household expenditures from the subsample with the imputed expenditures from the remaining households, may be quite precise. The ultimate success of the approach will depend on how much new (modelling) error is introduced as a result of the imputation procedure, as well as the extent to which the quality of the consumption data has improved.

A recent application of SWIFT 2.0 occurred in Zimbabwe. Zimbabwe's latest household survey was conducted in 2017, but due to hyperinflation in 2019, there were concerns that poverty incidence might be increasing rapidly. Despite the high demand for poverty data, budget constraints prevented the country from carrying out a traditional household survey. The Zimbabwe National Statistics Agency and the World Bank agreed to apply the SWIFT 2.0 approach in order to update their assessment of poverty and living conditions of the population.

Data collection in Zimbabwe was implemented in May and June 2019. The size of the subsample in which consumption was collected was set at 600, and that of the rest of the sample was set at 3,000. Models were developed for urban and rural areas separately using the subsample data. For the subsample, both models predicted the poverty rates of urban and rural areas well. With actual and imputed data, the final poverty estimates for urban and rural areas were 24.3 and 72.0 percent, respectively. By combining both actual and imputed data, the standard errors declined from 5.1 percent to 4.4 percent for urban areas and from 4.7 percent to 2.5 percent for rural areas. The national poverty estimates for Zimbabwe in 2019, based on SWIFT 2.0, were accepted for the purpose of the World Bank's global poverty monitoring effort.

(ii) Rapid Consumption Survey and applications in Somalia 2017 and South Sudan 2016/17

The Rapid Consumption Survey (RCS) approach introduced by Pape and Mistiaen (2018) and Pape and Wolfgang (2019), attempts to save costs and time in an alternative manner. The idea in RCS is to reduce interview time by collecting data on only a subset of consumption items and imputing the missing consumption components using S2S imputation. Poverty rates for Somalia and South Sudan based on the RCS procedure were accepted as part of the 2021 Spring update of the World Bank's global poverty monitoring effort.

RCS splits the survey sample into three or more subsamples. For the sake of exposition, we focus on the split of 3 here. RCS then separates the consumption module into four categories – a core module and three partitions. Each subsample's consumption module includes the core module and one partition (see Figure 2). Consumption expenditures from the dropped partitions are then imputed by models developed in the other subsamples. In the example in Figure 2, subsample 1 does not include data on partitions P2 and P3. RCS uses subsample 2's data to develop a model by regressing household expenditure in P2 on core-consumption data and selected non-consumption indicators. It then imputes P2 expenditures into subsample 1. The same is done for imputing subsample 1's missing consumption data of P3 by developing a model from subsample 3's data. Similar processes are carried out to fill missing consumption data in subsamples 2 and 3.[4]

Figure 2

RCS			
Modules	Group 1	Group 2	Group 3
Cons	P3	P3	P3
	P2	P2	P2
	P1	P1	P1
	Core	Core	Core
Non-cons	Core	Core	Core

Since RCS does not have to collect two partitions in each subsample, data collection can be completed more quickly than with a traditional household survey. Indeed, Pape and Mistiaen (2018) claim that RCS in Somalia made it possible to collect the necessary consumption data in 60 minutes, rather than the 2 – 3 hours required for a traditional household survey.

RCS collects core consumption data for all households in the survey. Expanding the core module has a clear and positive impact on the accuracy of the final poverty estimates. However, there is a trade-off: expanding the core module increases the data collection time. At the extreme, if the core module is expanded to the full consumption module, RCS does not save any time. Therefore, in RCS, determining how to split the full consumption module into the core module and partitions is important.

(iii) Comparisons between RCS and SWIFT 2.0

While both RCS and SWIFT 2.0 are less exposed to model stability issues than a simple S2S approach by creating models from the same household survey as the imputation data, how they save interviewing time does differ. In RCS, total interview time is reduced because no household is administered the full consumption questionnaire - sampled households are administered only (varying) sections of the consumption module. In SWIFT 2.0, the total interview time is reduced because consumption data are collected only in the subsample. The relative appeal of SWIFT 2.0 over RCS increases as more households are included in the overall survey relative to the consumption subsample, while the relative benefit of SWIFT 2.0 declines as the number of partitions in RCS increases.

(iv) Comparison between traditional data collection and imputation approaches

As noted above, the LSMS guidelines (Grosh and Munoz 1996 and World Bank 2000) recommend that the overall sample size for a household consumption survey should be kept small, in order to balance between sampling and non-sampling errors. An imputation approach such as SWIFT 2.0 would permit a larger overall sample size, as long as the consumption sub-sample is kept modest in size. At first glance, SWIFT 2.0 is more cost-effective than the traditional LSMS because it does not collect consumption data from most of the sample. However, this comparison does not consider the overall statistical accuracy of the final poverty estimates. SWIFT 2.0 does not collect consumption data from a large subsample of the survey but instead imputes consumption data into the non-consumption data. Since the imputations are not perfect, the procedure introduces a certain level of imputation (model) error, to be that needs to be added to the sampling error. While sampling error affects both the traditional and imputation approaches, modelling error enters into only the imputation approaches. To achieve the same level of statistical accuracy as in the traditional approach, an approach such as SWIFT 2.0 would thus require a larger sample than the traditional approach, generating a smaller sampling error that can offset the additional imputation errors. Thus cost-effectiveness of the SWIFT 2.0 approach depends on the savings achieved from collecting non-consumption data only for the S2S projections and how many more observations the SWIFT 2.0 needs to be collected in order to achieve the same level of statistical accuracy as the traditional approach. Fujii and van der Weide (2016) demonstrate that for SWIFT 2.0 to be more cost-effective, the cost of collecting data needed for the S2S poverty projections should be considerably lower than that of collecting the full consumption data.

How realistic it is to assume the cost of collecting the data needed for the S2S projection is only a small fraction of the cost of collecting full consumption data? To impute household expenditures with the S2S projection method, it often suffices to collect only 10 to 15 simple questions, most of which depend only on a yes/no answer. Experience to date indicates that collecting such information may require only three to five minutes. This compares with two hours or more which may be required for collecting full consumption data. However, Fujii and van der Weide (2016) argue that transport costs are also likely to enter into the calculation. Even if the interview time shrinks by adopting the SWIFT 2.0 approach, if transportation costs remain high, the relative advantage of SWIFT 2.0 is attenuated because both the traditional approach and SWIFT 2.0 will incur the same transportation costs per cluster or enumeration area. Moreover, since SWIFT 2.0 needs a bigger sample size to achieve the same level of statistical accuracy, higher total transportation costs are needed, potentially making SWIFT 2.0 less cost-effective than the traditional approach. On the other hand if SWIFT 2.0 is able to collect data via phone interviews or employing local enumerators to collect non-consumption data, then the role of transportation costs might be attenuated.

Moreover, if collecting the non-consumption data for the S2S projections has its own objective, the marginal cost of collecting variables needed for poverty projection declines, making SWIFT 2.0 more attractive. A multi-topic, integrated, household survey like an LSMS has multiple purposes. Estimation of monetary poverty is only one of many objectives. Monitoring education and health outcomes, the coverage of social assistance policies, non-monetary dimensions of deprivation, and employment conditions, are all important facets of the multi-topic household survey. It is often the case that the questionnaire for collecting non-monetary data includes most of variables needed for projecting monetary poverty. If so, the incremental cost of collecting variables for the purpose of poverty projections becomes negligible. However, it is also important to note that while expanding the questionnaire so as to include additional non-monetary indicators significantly reduces the marginal cost of collecting poverty proxies, it does increase total survey costs, and time, considerably.

4 Concluding remarks

Poverty data gaps remain widespread in the developing world. For example, the latest database of the World Bank shows there are 35 countries out of 46 countries in the sub-Saharan Africa region that do not have poverty data in the last 5 years. Since poverty incidence can change quickly, particularly after large shocks such as the COVID-19 pandemic, there is an urgent need to achieve more frequent monitoring of poverty. However, the frequent collection of poverty data is challenging for many developing countries, particularly low income countries and fragile states. This is because the collection of poverty data is costly, time-consuming, and complex. This note revisits recommendations in the traditional approach to collecting Living Standard Measurement Study (LSMS) surveys (Grosh and Munoz, 1996 and the World Bank, 2000) and discusses two methodologies – RCS and SWIFT 2.0. – that have recently seen adoption in the World Bank's official global poverty monitoring effort.

This article notes that the LSMS guidelines recommend relatively small sample sizes in order to minimize the risk of non-sampling error and its implications for data quality and credibility. Recent years have seen a drifting up of sample sizes, with for example the average sample size of the latest household surveys in the sub-Saharan region exceeding 10,000 households. If the LSMS guidelines were returned to, the cost of data collection could be significantly reduced and the risk of non-sampling error significantly curtailed.

This note then discusses two alternatives, both involving survey-to-survey (S2S) imputation. These two methods have recently been implemented to produce official poverty estimates in South Sudan, Somalia, and Zimbabwe. The note assesses under what conditions methodologies such as the SWIFT 2.0 and RCS approaches discussed here, become more cost-effective than traditional data collection. Given that poverty projections based on S2S include imputation errors, a given overall level of precision may entail larger overall sample sizes than the traditional approach. And so cost-reductions are not assured. However, emerging experience suggests that standard survey implementation costs can be substantially reduced if interview time for collecting data for the S2S projections is kept to an absolute minimum and if transportation and lodging costs can be saved via phone interviews or hiring local enumerators. Further research and exploration is warranted.

1 *This note summarizes findings and discussion in Lanjouw and Yoshida (2021). These were also presented by Lanjouw in the LIS Summer Lecture on July 6, 2021. Helpful comments and suggestions have been received from Chris Elbers and Philippe van Kerm, and participants in the Summer Lecture. We are also grateful to participants at the World Bank online presentation on June 16, 2021.*

2 *This normal distribution and linearity can be relaxed. For the sake of exposition, the normal distribution is assumed.*

3 *S2S has been used widely in the field of poverty measurement and monitoring. Deaton and Dreze (2002) and Kijima and Lanjouw (2003) used this approach to estimate poverty rates in India's National Sample Survey Organization survey of 1999-2000. Stifel and Christiaensen (2007) used it to impute poverty into a Demographic Health Survey. Douidich et al. (2013) used it to impute poverty into multiple rounds of Labor Force Surveys in Morocco.*

4 *More details are available in Pape and Mistiaen (2018).*

References

Christiaensen, L., P. Lanjouw, J. Luoto, and D. Stifel. (2012). "Small Area Estimation-Based Prediction Methods to Track Poverty: Validation and Applications." *Journal of Economic Inequality* 10 (2): 267–97.

Deaton, A. and J.P. Dreze. (2002). "Poverty and Inequality in India: A Reexamination" Economic and Political Weekly, September 7, 2002.

Fujii, T. and R. van der Weide. (2020). "Is Predicted Data a Viable Alternative to Real Data?" *The World Bank Economic Review* 34(2): 485–508.

Grosh, M. and J. Munoz. (1996). "A Manual for Planning and Implementing the Living Standards Measurement Study Survey." *Living Standards Measurement Study Working Paper Series* No. 126. World Bank, Washington, DC.

Kijima, Y. and P. Lanjouw. (2003). "Poverty in India during the1990s - a regional perspective," *Policy Research Working Paper Series* No. 3141. World Bank, Washington, DC.

Lanjouw, P. and Yoshida, N. (2021) "Extraordinary Times and Extraordinary Measures: High Frequency Poverty Monitoring in the Face of Data Deprivation", mimeo, Poverty and Equity Global Practice, the World Bank.

Mathiassen, A. (2013). "Testing Prediction Performance of Poverty Models: Empirical Evidence from Uganda," *Review of Income and Wealth, International Association for Research in Income and Wealth*, vol. 59(1), pages 91-112, March.

Pape, U, and J. Mistiaen. (2018). "Household Expenditure and Poverty measures in 60 minutes: A new approach with results from Somalia." *Policy Research Working Paper Series* No. 8430. World Bank, Washington, DC.

Pape, U. and P. Wollburg. (2019). "Estimation of Poverty in Somalia Using Innovative Methodologies." *Policy Research Working Paper Series* No. 8735. World Bank, Washington, DC.

Sohnesen, T. (2015). Tracking Poverty via Consumption Proxies. Mimeo.

Stifel, D., and L. Christiaensen. (2007). "Tracking Poverty over Time in the Absence of Comparable Consumption Data." *World Bank Economic Review* 21 (2): 317–41.

The World Bank. (2000). *Designing Household Survey Questionnaires for Developing Countries – Lessons from 15 yars of the Living Standards Measurement Study*. Edited by M. Grosh and P. Glewwe. World Bank, Washington, DC.

Yoshida, N., R. Munoz, A. Skinner, C. Kyung-eun Lee, M. Brataj, and D. Sharma. (2015). SWIFT Data Collection Guidelines version 2. World Bank, Washington, DC.

Yoshida, N., X. Chen, S. Takamatsu, K. Yoshimura. S. Malgioglio, and S. Shivakumaran. (2020). "The Concept and Empirical Evidence of SWIFT Methodology." Unpublished Manuscript. World Bank, Washington, DC.

Data News / Data Release Schedule

LIS is happy to announce the following data updates:

Georgia – Annualisation of the country series from 2009-2019 for the **LIS** Database (8 new datasets and 3 revised)

Switzerland – CH18 added to the **LIS** Database (1 new dataset)

United Kingdom— UK94 /UK95 (LIS Database**),** minor refinement have been carried out on the household composition variables

Germany- update of previous data points in the **LWS** Database using the latest version of GSOEP data release

Data Releases and Revisions– Luxembourg Income Study (LIS)

Georgia

Eight new dataset from Georgia, **GE09**, **GE11**, **GE12**, **GE14 GE15**, **GE17**, **GE18**, and **GE19**, have been added to the *LIS Database*. The datasets are based on the respective waves of the **Integrated Household Survey (IHS)** carried out by the National Statistics Office of Georgia. As a result, the annual Georgian data now cover the period 2009-2019 in the *LIS Database*. In addition, the annualisation of the series implied a substantial revision to the previously available three data points (**GE10, GE13** and **GE16**), more specifically:

-the construction of the annual sample out of the quarterly subsamples has been refined;

-because of a change in the rotation structure of the survey as of 2017, the construction of the annual income is now based on a reference period of 6 months;

-a few other refinements have been applied (notably the provision of variable *net1* and some minor revisions for the education variables).

Switzerland

One new dataset from Switzerland, **CH18** (Wave XI), has been added to the *LIS Database*. The dataset is based on Income and Living Conditions (SILC) data from the Swiss Federal Statistical Office.

Data Revisions –LIS Database

United Kingdom

Minor refinement have been carried out on the household composition variables; namely *relation, partner, parents, nchildren, ageyoch, hhtype*, and *hpartner.*

Data Revisions –LWS Database

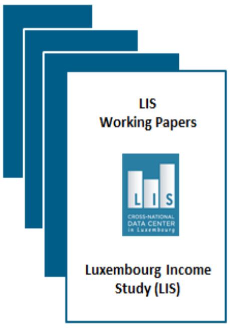
Germany

The LWS datasets (**DE02, DE07, DE12,** and **DE17**) have been harmonised using the latest version of GSOEP data release SOEP-Core v36eu – provided by the German Institute for Economic Research (DIW), and the updated harmonization decisions, notably two major changes: The respondent weight has been implemented; this weight is 0 for non-respondents, but adjustments by DIW have been carried out in such a way that the response sample is representative for the total German society. The education module now provides detailed information on highest education level completed (*educ_c*), separating out general and vocational degrees, as well as various tertiary level degrees.

LIS/LWS Data Release Schedule

	Winter 2021	Spring 2022
	LIS Database	
Australia	AU16/18	
Austria	Annual data AT03-AT18	
Canada		Annual data CA96-CA19
Colombia		Annual data CO07-CO19
Iceland		Annual data IS03-IS17
Japan		JP14/15
Luxembourg		LU15/16/17/18
Paraguay	Annual data PY97-PY20	
Russia	RU19	
Vietnam	VN92/97/01/03/05/07/09	
Uruguay	Annual data UY05-UY19	
	LWS Database	
Chile	CL07/12/14/17	
Japan		JP14/15/16

Working Papers & Publications



Focus on

Pathways toward Inclusive Income Growth: A Comparative Decomposition of National Growth Profiles **LIS WP No.802 by Zachary Parolin (Bocconi University), and Janet Gornick (The Graduate Center, City University of New York),**

Despite rising interest in income inequality, scholars remain divided over the mechanisms underlying inclusive income growth and how these mechanisms vary across countries. This study introduces the concept of national growth profiles, the additive contribution of changes in taxes, transfers, composition, and other factors including market institutions to changes across a country's income distribution. The authors present a decomposition framework to measure national growth profiles for eight high-income countries from the 1980–2010s. The findings adjudicate competing sociological and economic perspectives on rising inequality. First, the authors find that policy-driven changes in taxes and transfers are the dominant drivers of inclusive growth at the tails of the income distributions. Second, rising educational attainment contributes most to income growth across the distribution, but consistently contributes to less-inclusive growth. When changes in education are considered, changes in assortative mating and single parenthood have little consequence for changes in inequality. Third, changes to other factors including market institutions increased inequality in countries such as the U.S., but less so in France and Germany. Had the U.S. matched the changes to Dutch tax policy, Danish transfer policy, or other factors of most other countries, it could have achieved more inclusive income growth than observed.

LIS working papers series

LIS working papers series - No. 813

The Rise of China's Global Middle Class in International Perspective

by Terry Sicular, Xiuna Yang, Bjorn Gustafsson

LIS working papers series - No. 814

Routine-Biased Technological Change Does Not Always Lead to Polarisation: Evidence from 10 OECD Countries, 1995-2013

by Matthias Haslberger

LWS working papers series

LWS working papers series - No. 36

Wealth Accumulation and Retirement Preparedness in Cross-National Perspective: A Gendered Analysis of Outcomes among Single Adults

by Janet Gornick, Eva Sierminska

News, Events and Updates

The closing of the InGRID2 project – a successful collaboration with LIS

After 4 years of successful achievements, the InGRID-2 project runs to its end in October 2021. The project (funded by the European H2020-programme) has integrated research infrastructures to serve the social sciences community to make an evidence-based contribution to a European policy strategy of inclusive growth. Through InGRID-2, LIS has hosted many visiting scholars and provided free of charge virtual data access to LIS non-contributing countries from the European Union and 16 associated countries. In particular, over the last four years, the InGRID-2 project allowed 18 scholars to be hosted in the LIS offices to access directly the LIS and LWS databases, and 82 researchers from 20 non-contributing countries to benefit from the virtual access to the data. Figure 1 demonstrates the number of registrations from countries that were the InGRID-2 eligible but non-contributing to LIS infrastructure. The most significant number of the users came from Spain, Austria, Poland, and Belgium, while researchers from other 16 countries constituted about 35% percent of registrations. Figure 2 displays the number of statistical programs ('jobs') sent to LISSY (LIS remote execution system) to make various statistical and econometric estimations by the same group of researchers. The total number of 'jobs' sent by researchers from these four countries adds up to more than 30,000 from 2017 through 2021. Unfortunately, researchers from InGRID-2 participating countries will lose access to LIS databases at the end of the project; however, LIS hopes that countries from which these researchers come would contribute to the functioning of LIS. Piotr Paradowski presented these numbers together with a description of the benefits of LIS virtual access at the final conference of the InGRID-2 project held online on 9-10 September 2021, the presentation can be found here. The meeting was a very successful event that evaluated the tremendous achievements of the InGRID-2 project in terms of grants, a variety of workshops and summer schools, the development of various datasets, and research projects through the participating research infrastructures. It also stimulated the thoughts on further thoughts enlargement of the research for policy analysis.

Figure 1: Number of registrations by researchers from InGRID-2 countries, 2017-2021

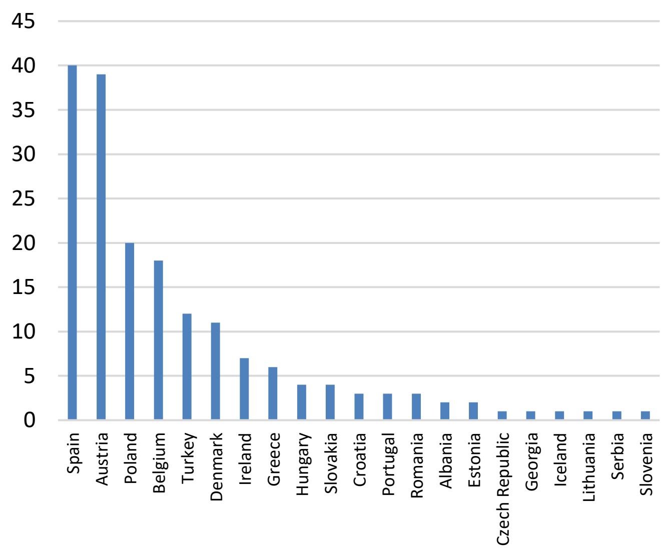
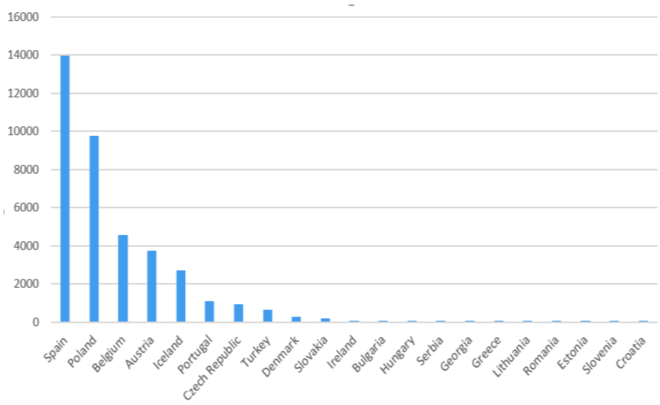


Figure 2: Number of analysis via Virtual Access, 2017-2021



Short Report on 2021 LIS Summer Workshop 2021

For the 29th time in the last 37 years, LIS has run an Introductory Summer Workshop for scholars interested in using the LIS and LWS databases. Like in the past couple of workshops, this year's event was a joint effort with LISER and the University of Luxembourg. Two prominent scholars from these institutions, Professors Louis Chauvel and Philippe Van Kerm taught methods for analyzing inequality with LIS and LWS data. We also hosted Professor Peter Lanjouw from the University of Amsterdam for the 10th LIS Summer Lecture that took place during the workshop. The workshop, which included lectures and practical exercises, took place in the form of online teaching. Furthermore, for the first time, LIS introduced, besides Stata, also R programing language for essential data management and analysis of inequality (readers can access the topics of the 2021 online workshop from the LIS website). We also had an introductory programming session for those who needed a refreshment on programing in Stata. We had 95 registrations coming from 30 countries, but only about 60 researchers actively participated. As to the demographics of participants, the largest participation group was Ph.D. students and post-doctoral fellows. The most significant number of participants came from the field of economics, followed by sociology, public policy, social work, and political science. At the end of the workshop, we also conducted a satisfaction survey. The response rate was about 78%, where 92% of the respondents stated that the LIS workshop met their expectations.

New complementary database: Leave and ECEC Policy Dimensions Dataset

LIS recently added to its complementary databases section a new dataset on Leave and Early Childhood Education and Care (ECEC) policy dimensions for use with the Luxembourg Income Study Database. This dataset, assembled by Sarah L. Kostecki, includes country-level policy indicators on leave and early childhood education and care for 31 countries: 24 high-income countries and seven middle-income countries (Latin America) based on policy and ECEC utilization information compiled from international organizations and country-specific sources. The 21 country-level leave and ECEC policy dimension indicators in this dataset are categorized into four policy dimension groups. The indicators measure three dimensions of leave policy (generosity, gender equality, and universality) and one dimension of ECEC policy (generosity, measured by enrollment rates of children in ECEC). Four country identifier indicators are also included in the

dataset for a total of 25 indicators. The year selections of the country-level policy information used to construct the indicators are from between the years 2009 to 2013.
Users can access the data and its documentation from here.

LIS granted the Aldi Award for 2020 LIS Working Paper

This year's winner of the LIS Aldi Award is Nora Waitkus for the LWS Working Paper No. 33 entitled "The Wealth Inequality of Nations" that has been very recently published in American Sociological Review. Nora is a sociologist currently working at the London School of Economics, International Inequalities Institute; she was twice a visiting scholar at LIS.
Nora co-authored this paper with Fabian T. Pfeffer. The Aldi Award is granted to the writer under age 40, whose LIS or LWS Working Paper from the previous year best demonstrates the qualities of sound scholarship.

Upcoming workshop on "Policies to Fight Inequality: The Case of Work-life Reconciliation and Family Policies", 25-26 November 2021

We are pleased to announce that LIS and LISER convene the second international scientific workshop in the realm of the *(LIS)²ER* initiative. In this workshop we tackle ***"The Case of Work-life Reconciliation and Family Policies"*** which lie at the intersection of the labour market, households and early years of child development, and are crucial in easing the often-competing responsibilities between work and family when young children are present. Acknowledging the diversity of policies and research studying the how, why, and whats of these entitlements this workshop will focus on two interrelated family policies: provision of care for young children and parental leave. It will discuss inequalities as causes and consequences at three levels: inequalities in access due to eligibility rules, inequalities in use due to (un)affordability of the right, and unintended consequences of the given right. Against this background, this workshop aims to offer a space to discuss novel insights on inequalities related to work-life reconciliation policies, to present LIS data as source for comparative research, and provide scholars whose work captures inequalities within the scope of work-life reconciliation policies with an opportunity to unite and exchange ideas.
The workshop will take place from **Thursday November 25th** mid-day through **Friday November 26th** mid-afternoon. It will consist of 8 invited contributions. Thursday evening, a keynote lecture will be delivered by Rense Nieuwenhuis (Stockholm University). A policy roundtable, led by Margaret O'Brian (UCL), will take place Friday early afternoon. It will be possible to follow the event virtually. Registration details and further information will be announced end of October.

Visiting scholars at LIS

This quarter, LIS welcomed four visiting scholars who came to work onsite with the LIS Databases. Two of them came within the framework of the InGRID-2 project, namely Manuel Schlechtl and Rosa Mulé. Two other visitors visited LIS through the collaboration between LIS and LISER, namely Ariane Aumaitre, and Krzysztof Czarnecki.
Rosa is associate Professor of Political Economy in the Department of Political and Social Sciences at Bologna University, where she teaches Political economy of welfare systems. During her stay at LIS, Rosa was using the LIS Database to examine income inequality in different industrialised countries among different groups of women. The aim was to understand the relationship of heterogeneity within women on labour market outcomes in countries with different varieties of capitalism. The results will enable policy recommendations to deal with high gender wage gaps at different levels of the income distribution in different institutional settings.
Manuel Schechtl is a research associate and doctoral student at the Department of Social Sciences at Humboldt University Berlin. During his stay at LIS, Manuel was using the LIS Database to examine the impoverishment of households that occurs due to the tax and transfer systems in rich democracies. The aim was to scrutinise this fraction of people in poverty that is poor because of the welfare system. The results will add a novel perspective to the cross-national literature on redistribution and poverty alleviation.
Ariane Aumaitre – researcher at the European University Institute (Florence) – worked with the LIS data to study the evolution of the living standards of different socio-economic groups during the last decades. More precisely, she was able to analyse in depth gender and generational inequalities across time.
From August 30 to September 3, LIS hosted the first (LIS)²ER visitor, Krzysztof Czarnecki, Assistant Professor at the Institute of Socio-Economics at Poznan University and Associate researcher at the Swedish Institute for Social Research (SOFI). His research interests are higher education, student finance, educational inequalities, the welfare state and public management. With (LIS)²ER research fellow Petra Sauer, he works on higher education stratification across European welfare states.

The Stone Center – Call for Two Postdocs Coming Soon!

The Stone Center will post a call soon for its fourth cohort of postdoctoral scholars. For one position, priority will be given to candidates conducting research on wealth inequality; areas of interest include the determinants and consequences of wealth inequality, and wealth taxation. For the other position, priority will be given to candidates who examine racial/ethnic inequalities in the United States as part of their research agenda, although other topics will also be considered. The two postdocs will be in residence at the CUNY Graduate Center in New York City, from September 2022 through August 2024.
Applications will be due in early November. Detailed information and links to the application will be posted via Twitter (@stone_lis) and on the Stone Center website (https://stonecenter.gc.cuny.edu/). Stay tuned!

Editor: Jörg Neugschwender
Layout and Design: Heba Omar & Jörg Neugschwender
Photo credit: Cover @ ibreakstock/fotolia.com