InGRID

Integrating expertise in inclusive growth

www.inclusivegrowth.be

Deliverable 2.2

STATISTICAL RESEARCH AGENDA FOR AN EU INCLUSIVE GROWTH RESEARCH INFRASTRUCTURE: SOME KEY ISSUES

FUTURING BRIEFING NOTE 2

Simon Lenau, Philippe Liégeois,
Eric Marlier, Ralf Münnich and Guy Van Gyes

September 2016

This project has received funding from the European Union's Seventh Programme for Research, Technological Development and Demonstration under Grant Agreement No 312691

Abstract

The aim of this *Futuring briefing note* is to highlight and discuss the most relevant information from the InGRID-Delphi survey (cf. SZEKÉR and VAN GYES, 2015) with regards to current developments in official statistics.

This note focuses in particular on the issues of new data and non-probability sampling; micro-simulation and cross-border statistics as future arenas of innovation.

This constitutes Deliverable 2.2 'Futuring briefing note 2: Statistical research agenda for an EU inclusive growth research infrastructure: Some key issues', for Work Package 2 'Research Infrastructure strategic forum' of the InGRID project.
August 2016

General contact: inclusive.growth@kuleuven.be
p.a. InGRID
HIVA - Research Institute for Work and Society
Parkstraat 47 box 5300, 3000 LEUVEN, Belgium

For more information muennich@uni-trier.de

Please refer to this publication as follows:
Lenau, S., Liégeois, P., Marlier, E., Münnich, R. and Van Gyes, G. (2016): Statistical research agenda for an EU inclusive growth research infrastructure: Some key issues. Futuring briefing note 2 (InGRID deliverable 2.2). Trier: Trier University.

This publication is part of the InGRID project, this project has received funding from the European Union's Seventh Programme for Research, Technical Development and Demonstration under Grant Agreement No 312691.

Contents

1 The InGRID "Futuring survey" and the EU "IESS" Initiative as Starting Point

The aim of this Futuring briefing note is to highlight the most relevant information from the InGRID-"Futuring survey" with regards to current developments in official statistics (cf. Szekér and Van Gyes, 2015). The central question of this futuring survey was: "What are the needs and challenges of the European research infrastructure InGRID which will foster evidence-based policy making on the European inclusive growth strategy?" In plain words: what is necessary for researchers to develop more and better comparative evidence which can be used by policy makers in their decision-making process? What goes wrong today? What should be improved? Can we find some common ground on these challenges across Europe? Aiming to get a broad view on all existing perspectives of relevant stakeholders to this issue, a sample of European experts within different fields related to issues of inclusive growth was surveyed. In doing so, the diversity of the sample was the main goal, while representativeness was no requirement. The mainly qualitative analysis of the survey responses led to the identification of four main challenges:

1. More and better data should be the main priority: more and better data are necessary for high-quality comparative research in Europe. Regional and cross-border data represent a particularly urgent challenge.

2. The improvement of methods and researchers' (methodological and analytical) skills must go hand in hand with more and better data. Micro-simulation models have a key role to play in this context.

3. Bridging the gap between policy and research is a challenge to be tackled both by policy makers and researchers.

4. A research context that encourages cooperation and innovation and provides the necessary resources should be more strongly developed.

In this second Futuring briefing note on the InGRID research infrastructure, a particular part of the survey results is discussed. A more in-depth reflection is made on the raised methodological challenges and needs by focusing on "official statistics", i.e. statistics produced by National Statistical Institutes and other (sub-)national statistical bodies. In the research community that the InGRID tries to serve and facilitate, the use of social statistics and data collected by governments is very important. With a view to anchoring it in the most recently planned EU developments in social statistics, the note also takes account of the major EU "Integrated European Social Statistics" (IESS) initiative.

2 A Few Words about the Most Recent Planned Developments in the EU Agenda for the Modernisation of Social Statistics

Social Statistics faces rapidly changing challenges: rising importance of social statistics and indicators for evidence-based policy making, rising expectations concerning flexibility and data quality, innovation of methodology and IT, availability of new data sources, and at the same time reduced resources and need to avoid over-burdening respondents with too lengthy questionnaires. In response to these challenges and in line with the orientations endorsed by the Directors General of National Statistical Institutes in Wiesbaden in September 2011, Eurostat has been working on the modernisation of social statistics (Clémenceau, 2014). General objectives include the streamlining and integration of social surveys across European Union (EU) countries, which originally were designed independently to target different purposes. The EU modernisation programme of social surveys is part of the portfolio of actions expected to contribute to the overall objectives of the European Statistical System (ESS) Vision 2020. The programme includes actions pushing towards integration of data collections, wider use of innovative data sources, in particular administrative data and possibly Big Data, and improved statistical frames (ESSC, 2015)[1]

In this context, the EU envisages to implement a new legislative architecture for the ESS relating to persons and households, as proposed by the European Commission back in 2009 (Commission of the European Communities, 2009). This proposal for a common framework regulation of the European Parliament and the Council for an Integrated System of Social Surveys is currently being developed. The framework regulation captures topics like the planning stage, identical definitions and standardisation of variables, definition of the statistical population and observation units, specifications of precision requirements, quality reporting. Better harmonization of concepts and variables, common data collection and data processing instruments allow better joint analysis across countries (ESSC, 2015, p. 34). Besides this, two further framework regulations are planned: the first one will cover population statistics including population and housing census and the

[1] The European Statistical System Committee (ESSC) provides professional guidance to the ESS for developing, producing and disseminating European statistics. It is chaired by the Commission (Eurostat) and composed of the representatives of Member States' National Statistical Institutes (NSIs). EEA and EFTA countries' NSIs participate as observers. Observers from the European Central Bank, OECD, etc. may also participate in ESSC meetings.

second one will cover the administratively-based statistics and accounts (ESSC, 2015).

Statistics play an increasingly important role in the policy making process and are regularly used by the Commission, the European Parliament and by (sub-)national governments. Especially evidence-based policy making requires high-quality statistics for assessing policy effects at different stages (ex ante and ex post). The new integrated system of household statistics will lead to significant improvements of quality in European household statistics. However, in order to fully benefit from the possible gains, statistical research has to develop new and emerging methods that enable considerable quality improvements with respect to the new household surveys structure. These methods will surely comprise small area estimation, data matching, microsimulations, and advanced sampling methods as well as new ideas with respect to the use of Big Data[2] or cross-border statistics methods that will be briefly presented in this note.

[2] Big data is a term for data sets that are so large or complex that traditional data processing applications are no longer adequate. Challenges include inter alia analysis, capture, data documentation, search, sharing, storage, transfer, visualisation, querying, updating, etc.

3 From Data to Policy & Policy Evaluation

The European Commission and Member States have been jointly developing EU social indicators for monitoring policies as well as progress towards the EU objectives in the field of social protection and social inclusion since 2001 (cf. Social Protection Committee, 2015, see also Atkinson et al., 2002 and Marlier et al., 2007).

The discussion on *GDP and beyond* [1] has further enhanced the need for EU social indicators for use in the policy-making process. Additionally, there are discussions going on which focus on the way the statistical production process can lead to better (use of) data for better policies. These discussions have to be seen as part of the debate on the new information society, where many sources of data shall help putting together the pieces of a puzzle to get the full picture. Within the past years, and in the context of *Big Data* initiatives, this was mainly seen as technical or computer science problems, where important properties of statistics seemed to be ignorable for reasons of massive data.

Indeed, the future of data use will have to consider many aspects as the InGRID Delphi-survey clearly shows. Essential is the use of data from different sources. This of course includes classical survey data provided by national statistical institutes. Those data are in general of good quality, since they have to comply with the European statistics code of practice (Eurostat, 2011). However, using additional data sources is of growing importance. First of all, administrative and register data are to be mentioned. But just as well, further surveys have to be considered, which are not necessarily produced by national statistical institutes. In fact, web-surveys and Big Data arouse rising interest. If such data can be used to foster the information situation, this immediately raises questions about the quality of output. Is *more* always *better*? This question results directly in one central challenge and research direction, namely the need to judge the quality of statistical outputs based on non-probability sampling (i.e. new modes of data collection and use of samples that are not subject to classical sampling theory) or on Big Data.

Secondly, policy impact assessment attracts increasing attention. Methods of microsimulation are a well established tool for policy evaluation. An example of this is EUROMOD (cf. https://www.euromod.ac.uk/). Modern computer systems and extensive data sources

[1] For example, cf.
http://ec.europa.eu/environment/beyond_gdp/index_en.html or
http://www.insee.fr/fr/publications–et–services/dossiers_web/stiglitz/doc-commission/RAPPORT_anglais.pdf

allow for implementation of in-depth microsimulations. Modern statistical methods are required to exploit this potential.

The third crucial topic stressed by the Delphi-survey concerns special issues of regional support. Within the policy of the Euopean Union, comparison of regions is of major importance. But many regions of interest are located next to national borders, be it EU-internal or even external borders. However, regional indicators are commonly based on country-specific data and concepts. The last section on *Cross-Border Statistics* briefly examines some of the issues that need to be addressed when comparing regions from neighbouring countries.

4 Non-probability Sampling & New data

One of the major issues identified by the respondents of the InGRID Delphi survey (cf. Szekér and Van Gyes, 2015) concerns the limitations of survey data sources currently used. This topic was raised by experts from different areas, namely *poverty and living conditions*, *inclusive growth*, *labour market and precariousness*, *social policy and inclusion*, *working conditions*, and *inequality and welfare state*. Data accuracy is questioned due to limited sample sizes, especially when it comes to subgroups – be they of regional or demographic nature.

Those experts suggest better use of *new* data sources and increasing digitalisation as a remedy. Examples include online surveys and (at least several aspects of) what is referred to as *Big Data*. In fact, most forms of these data sources can – at best – be seen as some form of non-probability sampling, which differs from classical probability sampling in mainly two aspects:

1. Information about the sampling process may be missing.
2. As coverage of the target population of interest is not assured, non-random samples may be subject to poor representativity and selection bias.

For example, both difficulties may be due to self-selection in online surveys. The main reason to use non-probability sampling is lower costs.

There is a vast theoretical framework regarding classical design-based inference in random samples. Following an inductive logic, statistical inference is made by linking the (not entirely known) population of interest to observed sample data in probabilistic terms (for example, cf. Cochran, 1963; Wolter, 2007). In contrast, a single framework that includes all forms of non-probability sampling does not (yet) exist, since non-probability sampling is an extensive field of methods, also including Big Data sources. It is thus difficult, if not impossible, to describe properties which apply to all methods herein.

If data from non-probability sampling are identifiable (i.e. if they can be related to an entity or place in time), they may still be used in statistical inference. Other forms of (Big) Data that are not identifiable, such as social media or twitter data, may be useful for examination and visualisation of social phenomena, but can hardly be generalised for inference.

But even if data are identifiable, making inference from non-probability sampling requires reliance on modelling assumptions. There are two main ways towards adjusting for the non-randomness of these samples, namely **pseudo design-based** and **model-based inference**.

Pseudo design-based approaches use weighting data to handle selectivity. One example hereof is propensity score matching, which goes back to Rosenbaum and Rubin (1983). The unknown sample selection (or volunteering) process is treated as a quasi random process. As randomness comes from sample selection (cf. Neyman, 1934), probabilities of volunteering can be used to compensate for selectivity (for example see Enderle et al., 2013; Valliant and Dever, 2011). However, this approach can only be used if volunteering mechanisms are known or can reasonably be modeled. And again, the whole population must have the chance to be part of the sample. Hence, it solves only the first of the issues mentioned above.

In contrast, model-based inference assumes randomness in the variable of interest itself rather than the sample selection. Inference is made with regard to superpopulation parameters (the parameters generating the population variables) rather than fixed population parameters (cf. Fisher, 1922, Särndal et al., 1978). This can be achieved by e.g. using regression models.

If not only estimation from existing data but also designing surveys is taken into consideration, a further approach consists of combining probability and non-probability samples. This can be useful when costs per case are much lower for the latter, e.g. in convenience sampling, where easily reachable units are selected. If the mean-squared error (MSE) at given costs is – at least for certain covered subgroups – considerably smaller for the non-probability sample, combining estimates from both samples may yield a lower overall MSE than if only a probability sample is used. An example is given in Elliott and Haviland (2007), where a weighted combination of means from probability and non-probability sample is proposed. However, since this weighting is based on the MSE of the estimates, which resembles small-area methods (cf. Rao, 2003), the problem of how to estimate MSEs under non-probability sampling remains.

When it comes to statistical inference from non-probability samples, pseudo design-based approaches like propensity weighting seem promising, if the selection bias does not lead to completely excluding parts of the target population. Model-based inference appears to be a better solution if reasonable assumptions about such subpopulations being excluded can be made. However, such assumptions are necessary in both frameworks. Pseudo design-based approaches need good proxies for the selection process, while model-based inference requires presumptions about how the non-probability sample differs from the population regarding the variable(s) of interest.

When developing surveys, combining probability and non-probability samples might allow for a sensible synthesis. First, design-based approaches like propensity scores and calibration of the non-probability sample can be done using the reference probability sample. Secondly, model-based assumptions concerning the selection bias of the non-probability sample can be evaluated with regard to the reference sample. If per-case costs of probability samples are much higher, and the selection bias of the non-probability sample is small, the MSE might be reduced in this way.

5 Microsimulations

Microsimulation models have become a very important and necessary tool in applied social and policy research. These are efficient instruments to support economic and political decision-making. Microsimulations go back to 1950 already. But rapidly improving computer technologies such as storage expansion, increased Central Processing Unit (CPU) power and memory as well as powerful development tools can help with complex microsimulation models and reduce computational calculating time. The question arises about what enhancements and improvements are feasible and sensible.

The European Commission has organised several conferences underlining the importance of microsimulations to understand the effects of policies, taxes, benefits and reforms. In general microsimulations models are categorised in static, behavioral and dynamic models. The static model like EUROMOD (https://www.euromod.ac.uk/) is an excellent example of a *Tax-benefit microsimulation model covering all EU countries*. This model is often used to analyse and compare the effects of tax and benefit regulations on household incomes across EU countries and for the European Union as a whole without reference to longer run implications and behavioral adaptation. There are also statistical models which consider resonance of individual behaviours (e.g. on labour supply) and can complement static models like EUROMOD (e.g. Peichl et al., 2010).

Finally, when the time dimension is relevant in view of the questions to be answered, dynamic microsimulation models provide a very useful tool. Examples of such models include DREES, DYNAMOD, MOSART, DYNASIM, MIDAS and PENSIM (e.g. Duc et al., 2015, King et al., 1999, Fredriksen, 1998, Orcutt et al., 1976, Hancock et al., 1992). Dynamic microsimulations are developed to explore future changes and long-term consequences of programmes. The time dimension is respected and those models are able to support projections for the trend of economic development under current policies such as pensions, health care and other social welfare programmes (Li et al., 2013). It is also possible to analyse inter-temporal changes with the generated longitudinal data.

The implementation of microsimulation models raises a number of questions related to the data needed for those models, the exchange of knowledge and good practices between model developers and users who are not all IT specialists or statisticians, and the validation of the models. Those questions have become central in the microsimulation research agenda and are now examined in turn, with a special concern for the first topic.

The backbone of every microsimulation is the quality of the basic data set, as this determines the reliability and accuracy of the generated output. For example, statistical inference (hence estimation) is required in most models, in particular in behavioural and dynamic ones (Klevmarken et al., 2007, p. 45). Therefore, the input dataset has to contain all necessary information required by the simulation process at large and those estimations in particular. The goal is a basic dataset with a lot of information and a good sample size (if the whole population at stake is not covered), which of course leads to longer run times of models and requires more computer power.

Administrative data are often used because they contain much information and are available for bigger universe than surveys (see for example Liégeois et al., 2011). Household surveys contain smaller sample sizes and weights for the individuals which may lead to more complex frameworks, especially when considering dynamic models where the units of analysis (households or individuals) are evolving (including mixing) over time (cf. Dekkers and Cumpston, 2012). An alternative procedure is to create synthetic data, which is a very flexible and easy way to produce missing information. However, constructing such data for use in longitudinal analysis is complicated and still in progress (cf. the AMELIA description in Berger et al., 2016).

Yet legal and privacy reasons may make it harder to use administrative data (Li et al., 2013, p. 16). Some microsimulation models already use census data together with data imputed on the basis of information coming from data sources like household surveys, because census data do not contain all the information the models require for achieving qualified output data. In this step, use of model predictions, imputation, matching, and calibration are important methods and have to be used accurately and carefully to secure the quality of the input data.

Also the regional aspect becomes more important for economic and political decisions based on regional indicators and other statistical measurements (Münnich et al., 2013, p. 150). Most sample based microsimulations do not yield the necessary information for regional research tasks. In view of the growing importance of this research, it is essential to investigate the conditions required to perform regionalised microsimulations. The keyword here is Small Area Estimation. Miranti et al. (2015) use spatial microsimulation and small area estimation for analysing regional inequality. Finally the outcome data have to be validated. In general, microsimulation methods are evaluated under the given settings (this is related to model-based evaluation). Little is known on the accuracy of microsimulations in a more general framework, e.g. under consideration of sampling effects or the variance of future predicts. A validation has to include data, coefficient, parameter, programmers, algorithmic, module-specific, multi-module and policy impact validation (Li et al., 2013, p. 29).

Until now, most microsimulation methods originate either from demography and specialised samples (like EU-SILC) or from agent-based methods. Both worlds are still separated. Using specialised examples (currently under investigation), an interaction of both methodologies would be interesting and forward-looking.

All these questions are of high technicity and microsimulation modelling involves developers who do not necessarily belong to the research community. In terms of infrastructure, it might therefore be useful to identify and make available more widely to the community of developers key resources related to microsimulations. These resources involve:

1. the listing of models presently available or under development (together with a necessary nomenclature)

2. the required applications and data

3. various tools needed for quality development, simulation and validation

4. a link to key actors in the field and

5. a platform for exchanging experience, outcomes and questions.

In particular, the synergy between the different actors might be encouraged, e.g. through the creation of a community of developers fostering an exchange of best practices and the analysis of key methodological questions. Stakeholders including users, certainly, should be carefully considered in the various specific steps.

6 Cross-border Statistics

A crucial topic stressed by the Delphi-survey concerns the need for better regional data and indicators. The in-depth analytical comparison of regions is of major importance – not only within a same country but also across countries. Indeed, many regions are located next to national borders, be it EU internal or external borders. However, regional indicators are commonly based on country-specific data and concepts. This section briefly examines some of the issues that need to be addressed when comparing regions from neighbouring countries.

Comparing indicators at regional level is becoming increasingly important both at national and EU levels (see Barca, 2009 and Jouen, 2010 for very useful analyses of the latter). It is one of the emerging and rapidly developing areas of survey statistics and it is strongly connected to small area statistics (Pfeffermann, 2013, Pratesi et al., 2012, Rao and Molina, 2015). Most approaches of small area statistics so far consider *within-country* estimation methods, which certainly facilitate the use of coherent methods and data, and it is important to build cross-border data-set and develop expertise in this complex area. The following example shows the development of a population in two boarder regions at the Swiss-German border (see Figure 6.1) and the Luxembourg-German border (see Figure 6.2).

The demographic forecast between the two countries at the borders yields entirely different figures (Münnich et al., 2016). One would expect at least some compensation in terms of inter-country mobility. However, the models and parameters used for statistics are generally built on country-specific settings which hardly can cover effects across borders. And this leads immediately to the question: can we model border effects solely with country-specific data or do we have to use a common dataset?

Münnich et al. (2015) have started investigating area-level small area models for measuring the “at-risk-of-poverty rate” (EU definition of income poverty) on NUTS 3 and LAU1 regions while including information like proximity to economic centres *behind the border*. The proximity to economic centres, in general, influences developments in the region *behind* the border which may not be directly covered by some indicators. An extension to models in the *Greater Region*, consisting of five regions (Luxembourg, Wallonie, Lorraine, Rhineland-Palatinate, and Saarland) in four countries, is currently under investigation. However, immediately several questions occur:

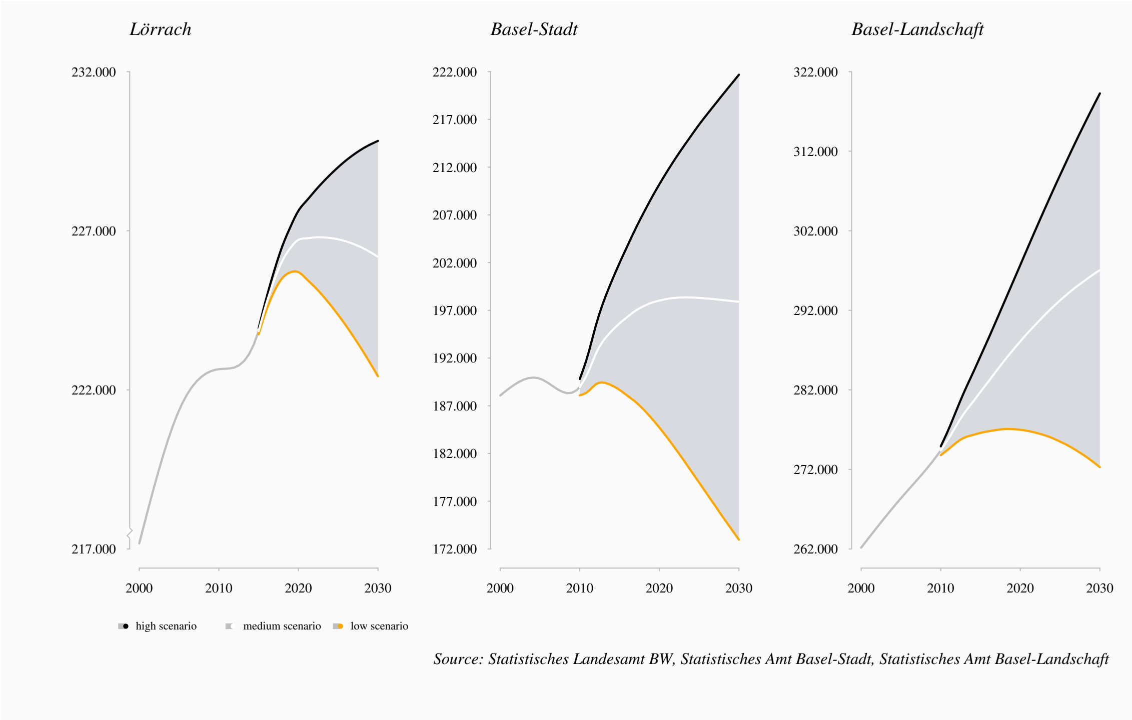

Figure 6.1: Demographic forecast in Lörrach (Germany), Basel (Switzerland) town and Basel Country (Landschaft), 2000-2030.

- Is all the information required for proper modeling available for all regions in the Greater Region?
- What is the impact of possibly different definitions of variables?
- Is the information of a border effect already in the country-specific variables or do we need the commonality of modeling?
- How do the different (sub-)national statistical systems in different countries affect a possible common modeling?

There is no doubt that, in future, reliable regional comparisons will have to consider these different issues. Hence, it is very important to identify areas where (sub-)national figures can be built separately. In all other fields of interest, only the collaboration between the different statistical offices and academic experts may yield reliable and comparable figures at regional level.

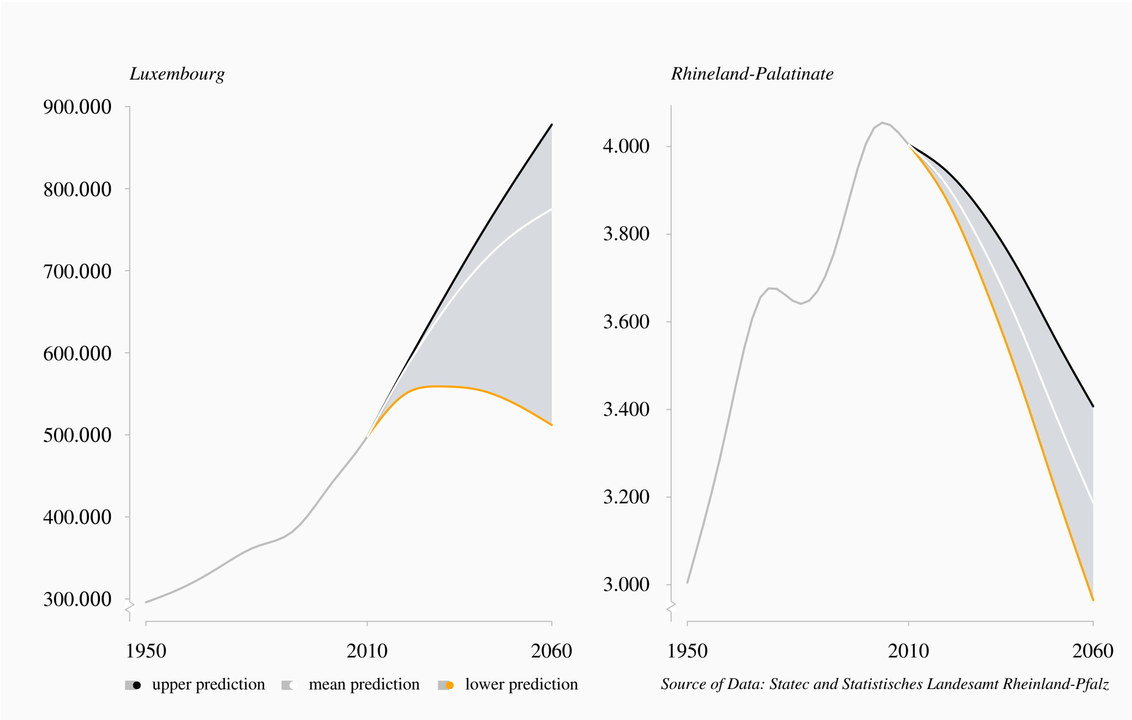

Figure 6.2: Demographic forecast in Luxembourg and Rhineland-Palatinate, 2000-2060.

Bibliography

Atkinson, T., **Cantillon, B.**, **Marlier, E.** and **Nolan, B.** (**2002**): Social indicators: The EU and Social Inclusion. Oxford (u.a.): Oxford Univ. Press.

Barca, F. (**2009**): *An agenda for a reformed cohesion policy. A place-based approach to meeting European Union challenges and expectations. Independent Report prepared at the request of Danuta Hübner, Commissioner for Regional Policy.* http://www.europarl.europa.eu/meetdocs/2009_2014/documents/regi/dv/barca_report_/barca_report_en.pdf.

Berger, Y. G., **Burgard, J. P.**, **Byrne, A.**, **Cernat, A.**, **Giusti, C.**, **Koksel, P.**, **Lenau, S.**, **Marchetti, S.**, **Merkle, H.**, **Münnich, R.**, **Permanyer, I.**, **Pratesi, M.**, **Salvati, N.**, **Shlomo, N.**, **Smith, D.** and **Tzavidis, N.** (**2016**): *Case Studies.* InGRID Deliverable 23.1.

Clémenceau, A. (**2014**): *The LFS review in the context of the Eurostat programme for modernising social micro data collections. 9th Workshop on LFS methodology.* Rome, 15.–16. May 2014.

Cochran, W. G. (**1963**): Sampling techniques. New York (u.a.): Wiley, 2 ed.

Commission of the European Communities (**2009**): *Communication from the Commission to the European Parliament and the Council on the production method of EU statistics: a vision for the next decade.* http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52009DC0404&from=en.

Dekkers, G. and **Cumpston, R.** (**2012**): *On weights in dynamic-ageing microsimulation models.* In: International Journal of Microsimulation, 5 (2), pp. 59–65.

Duc, C., **Housset, F.**, **Lequien, L.** and **Plouhinec, C.** (**2015**): *The Trajectoire microsimulation model: an estimation tool for pension reforms across all schemes.* In: Economie & Statistique.

Elliott, M. N. and **Haviland, A.** (**2007**): *Use of a web-based convenience sample to supplement a probability sample.* In: Survey methodology, 33 (2), pp. 211–5.

Enderle, T., **Münnich, R.** and **Bruch, C.** (**2013**): *On the Impact of Response Patterns on Survey Estimates from Access Panels.* In: Survey Research Methods, 7 (2), pp. 91–101.

ESSC (**2015**): *Item 7 of the agenda, Draft regulation on statistics relating to persons and households, IESS: Integrated European Social Statistics, Work Programme Objective 7.* In: 27th Meeting of the European Statistical System Committee Luxembourg.

Eurostat (**2011**): *European statistics code of practice. For the national and community statistical authorities.* http://ec.europa.eu/eurostat/documents/3859598/5921861/KS-32-11-955-EN.PDF/5fa1ebc6-90bb-43fa-888f-dde032471e15.

Fisher, R. A. (**1922**): *On the mathematical foundations of theoretical statistics.* In: Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 222, pp. 309–368.

Fredriksen, D. (**1998**): Projections of Population, Education, Labour Supply and Public Pension Benefits Analyses with the Dynamic Microsimulation Model MOSART. Statistisk sentralbyrå.

Hancock, R., **Mallender, J.** and **Pudney, S.** (**1992**): *Constructing a computer model for simulating the future distribution of pensioners' incomes for Great Britain.* In: Microsimulation models for public policy analysis: new frontiers, pp. 33–66.

Jouen, M. (**2010**): *A Territorialised Social Agenda to Guide Europe 2020 and the Future EU Cohesion Policy.* In: **Marlier, E.**, **Natali, D.** and **Van Dam, R.** (editors) Europe 2020: Towards a More Social EU?, Peter Lang.

King, A., **Robinson, M.** and **Baekgaard, H.** (**1999**): DYNAMOD-2: An overview. National Centre for Social and Economic Modelling.

Klevmarken, N. A., **Bolin, K.**, **Eklöf, M.**, **Flood, L.**, **Fransson, U.**, **Hallberg, D.**, **Höjgård, S.**, **Lindgren, B.**, **Mitrut, A.** and **Lagergren, M.** (**2007**): *Simulating the future of the Swedish baby-boom generations.* Technical report, Working Paper, Department of Economics, Uppsala University.

Li, J., **O'Donoghue, C.** et al. (**2013**): *A survey of dynamic microsimulation models: uses, model structure and methodology.* In: International Journal of Microsimulation, 6 (2), pp. 3–55.

Liégeois, P., **Berger, F.**, **Islam, N.** and **Wagener, R.** (**2011**): *Cross-validating administrative and survey datasets through microsimulation and the assessment of a tax reform.* In: International Journal of Microsimulation, 4 (1), pp. 1–18.

Marlier, E., **Atkinson, A. B.**, **Cantillon, B.** and **Nolan, B.** (**2007**): The EU and social inclusion: Facing the challenges. Bristol (u.a.): The Policy Press.

Miranti, R., **Cassells, R.**, **Vidyattama, Y.** and **McNamara, J.** (**2015**): *Measuring Small Area Inequality Using Spatial Microsimulation: Lessons Learned from Australia.* In: International Journal of Microsimulation, 8 (2), pp. 152–175.

Münnich, R., **Articus, C.** and **Burgard, J. P.** (**2015**): *Estimating regional ARPRs in a border region.* In: Presentation at the InGRID Satellite Meeting to NTTS 2015.

Münnich, R., **Burgard, J. P.** and **Vogt, M.** (**2013**): *Small Area-Statistik: Methoden und Anwendungen.* In: AStA Wirtschafts-und Sozialstatistisches Archiv, 6 (3-4), pp. 149–191.

Münnich, R., **Kühn, A.** and **Simoes, E.** (**2016**): *Impact of demographic change on health care in a border region.* In: **Hoffmann, L.**, **Hill, J.** and **Müller-Fürstenberger, G.** (editors) Socio-Ecological Transformation Processes in a European Cross-Border Region, Springer.

Neyman, J. (**1934**): *On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection.* In: Journal of the Royal Statistical Society, 97 (4), pp. 558–625.

Orcutt, G. H., **Caldwell, S.** and **Wertheimer, R. F.** (**1976**): Policy exploration through microanalytic simulation. The Urban Insitute.

Peichl, A., **Schneider, H.** and **Siegloch, S.** (**2010**): *Documentation IZAΨMOD: the IZA policy simulation MODel.* In: .

Pfeffermann, D. (**2013**): *New important developments in small area estimation.* In: Statistical Science, 28 (1), pp. 40–68.

Pratesi, M., **Giusti, C.** and **Marchetti, S.** (**2012**): *Small area estimation of poverty indicators.* In: **Davino, C.** and **Fabbris, L.** (editors) Survey data collection and integration, Springer.

Rao, J. N. and **Molina, I.** (**2015**): Small area estimation. John Wiley & Sons.

Rao, J. N. K. (**2003**): Small Area Estimation. London: Wiley.

Rosenbaum, P. R. and **Rubin, D. B.** (**1983**): *Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome.* In: Journal of the Royal Statistical Society. Series B (Methodological), pp. 212–218.

Särndal, C.-E., **Thomsen, I.**, **Hoem, J. M.**, **Lindley, D.**, **Barndorff-Nielsen, O.** and **Dalenius, T.** (**1978**): *Design-based and model-based inference in survey sampling [with discussion and reply].* In: Scandinavian Journal of Statistics, pp. 27–52.

Social Protection Committee (**2015**): *Portfolio of EU Social Indicators for the monitoring of progress towards the EU objectives for social protection and social inclusion – 2015 Update.* Technical report, European Commission.

Szekér, L. and **Van Gyes, G.** (**2015**): *The future policy research agenda of an EU inclusive growth research infrastructure.* Futuring briefing note 1, deliverable 2.1, Leuven, http://inclusivegrowth.be/downloads/output/d2-1-futuring-briefing-note-1.pdf.

Valliant, R. and **Dever, J. A.** (**2011**): *Estimating propensity adjustments for volunteer web surveys.* In: Sociological Methods & Research, 40 (1), pp. 105–137.

Wolter, K. M. (**2007**): Introduction to Variance Estimation. New York: Springer, 2nd ed.

InGRID
Inclusive Growth Research Infrastructure Diffusion

Referring to the EU2020-ambition of Inclusive Growth, the general objectives of InGRID – Inclusive Growth Research Infrastructure Diffusion – are to integrate and to innovate existing, but distributed European social sciences research infrastructures on 'Poverty and Living Conditions' and 'Working Conditions and Vulnerability' by providing transnational data access, organising mutual knowledge exchange activities and improving methods and tools for comparative research. This integration will provide the related European scientific community with new and better opportunities to fulfil its key role in the development of evidence-based European policies for Inclusive Growth. In this regard specific attention is paid to a better measurement of related state policies, to high-performance statistical quality management, and to dissemination/outreach activities with the broader stakeholder community-of-interest, including European politics, civil society and statistical system.

InGRID is supported by the European Union's Seventh Programme for Research, Technological Development and Demonstration under Grant Agreement No 312691.

More detailed information is available on the website: www.inclusivegrowth.be

Co-ordinator

Guy Van Gyes
Monique Ramioul

KU LEUVEN HIVA
RESEARCH INSTITUTE FOR WORK AND SOCIETY

Partners

TÁRKI Social Research Institute Inc. (HU)
Amsterdam Institute for Advanced labour Studies, Universiteit van Amsterdam (NL)
The Swedish Institute for Social Research, Stockholms Universitet (SE)
Fachbereich IV, Wirtschafts- und Sozialstatistik, Universität Trier (DE)
Centre d'Etudis Demogràfics, Campus de la Universitat Autònoma de Barcelona (ES)
Luxembourg Institute of Socio-Economic Research (LU)
Centre for Social Policy, Universiteit Antwerpen (BE)
Institute for Social & Economic Research, University of Essex (UK)
Bremen International Graduate School of Social Sciences, Universität Bremen (DE)
Department of Dynamics of Organisations of Work, Centre d'Etudes de l'Emploi (FR)
The Centre for European Policy Studies (BE)
Dipartimento di Economica e Management, Università di Pisa (IT)
Social Statistics Division, University of Southampton (UK)
Luxembourg Income Study, asbl (LU)
WageIndicator Foundation (NL)
School of Social Sciences, The University of Manchester (UK)

InGRID

Inclusive Growth Research
Infrastructure Diffusion
Contract No 312691

For further information about the
InGRID project, please contact
inclusive.growth@kuleuven.be
www.inclusivegrowth.be
p/a HIVA – Research Institute
for Work and Society
Parkstraat 47 box 5300
3000 Leuven
Belgium