

POSTER SESSION ABS19

TUESDAY, JUNE 25th, 16.00-18.00

Veronica Ballerini

Population size estimation via multiple lists: dealing with out-of-target units

During the last years, National Statistics Institutes have been exploring the possibility to produce statistics based on administrative data only. In particular, the interest in populations' size estimation is increasing. In such cases, some issues naturally emerge since the aims of those who collect data and those who use them differ. On the one hand, it is very likely to have out-of-target units in the datasets. On the other hand - and obviously enough, some units that belong to the target population are not observed. In practice, this is the case of incomplete contingency tables whose cells counts consist partially of overestimations. This project's aim is to explore the potential efficiency of data augmentation algorithms which deal with both over and undercoverage issues in a population size estimation problem. Furthermore, we aim to allow the inclusion of important and non-negligible a priori information, in a fully subjective Bayesian approach.

Camille Belmin

Integrating energy access scenarios in probabilistic population projections: A novel approach to address sustainability

One of the biggest challenge humanity has to face is to enhance the standards of living of a growing population in the poorest countries, while remaining in the planetary boundaries. Population is a natural driver of energy consumption and carbon emissions. But energy consumption can also affect population growth in high fertility countries: access to modern energy substitutes human and child labor and allows a re-allocation of men and women's time use, in particular toward education, which is a widely accepted driver of fertility decline. Simplistically, an increase of modern energy consumption in developing countries would have two opposing effects: 1) increased per-capita energy consumption, all else being equal, increases total energy consumption. 2) smaller populations (resulting from reduced population growth), all else being equal, reduce total energy consumption in relation to a business as usual scenario. We present in this poster the double goal of our project (i) understand past relation between energy and fertility and (ii) use this understanding to build a probabilistic projection model of population to assess under which condition the population effect is stronger than the consumption effect. This would constitute a very interesting co-beneficial solution for development and reduction of energy consumption (and eventually carbon emission mitigation). One main challenge to apply the UN's methodology of population projection to our problem is that our population prediction would depend on one predictor variable while the UN's projection do not depend on a specific socio-economic variable.

David Brazel

Intensive longitudinal assessments in twins to investigate the development of adolescent environment and substance use behavior

Smartphones may facilitate the collection of self-report data and passive measures of environmental exposure. Such measures allow new approaches to investigate the interplay between genes and environment in producing complex behavior, including substance use and addiction. We conducted a study

of 670 adolescent twins. At intake, twins were 14-17 years old and subsequently followed for two years with weekly questionnaires and continuous logging of GPS location, from which we extracted twin similarity, time spent at home during the night, and time at school during school hours. Dizygotic twins' locations were less similar than monozygotic twins', and became increasingly dissimilar as the twins aged. This suggests not only that geographical location and related experiences are genetically influenced, but that the genetic influence increases as adolescents age. Through a series of mixed-effects models on age, we found that e-cigarette, alcohol and marijuana use rates increased dramatically after age 18, attributable to genetic and environmental influences. At intake, use of the three substances were correlated with each other and with self-reported parental monitoring at intake. Adolescent change in parental monitoring, time at home, and time at school were not associated with change in substance use. While there are clear developmental changes in all measures, the results suggest that change in one domain are not strongly associated with change in another domain. Adolescents who use substances, for example, tend neither to stay out later or more often than adolescents who drink less.

Silvia Ceausu

Gender-disaggregated migration at subnational scale across three continents

Understanding the dynamics of migration can help improve migrants' wellbeing. Moreover, understanding human mobility helps planning and management of public resources. Gender has been recognized as an important determinant of migration experiences. Women are more vulnerable to exploitation and trafficking, and can suffer inequalities similar to those in other areas of life. Women and men migrants have also different needs and their migration patterns might have different drivers. Here, we construct gender-disaggregated models of migration at subnational level. We also improve on the predictive power of previous models of migration at subnational scale by adding new covariates. We use census data from 40 countries for the period 2005 - 2010 available in the IPUMS International database. The census data include information on where the respondents lived in the previous 5 years. We use Poisson generalized mixed effects models with country as the random effect and covariates at the subnational level as fixed effects to predict number of people moving between any pair of administrative units within a country, implicitly migration flows corresponding to the previous 5 years. We find that the best predictive models for male and female migration are very similar in terms of covariate structure. We cross-validate the models by training them on the migration data of 39 countries and then predicting migration flows for the 40th country in the dataset. We repeated this procedure for all countries in the dataset and we obtained an average R2 between predicted and real data of 0.71 for both male and female migration models. In the next step, we will connect our models to international migration models to predict international migration at subnational level.

Daniel Ciganda

*Micro-level Dynamics Behind the Recuperation of Fertility at Higher Development Levels:
A Computational Modelling Approach*

The recuperation of fertility rates in some of the most advanced economies in the world triggered a reinterpretation of the link between development and fertility among researchers interested in demographic dynamics. The reversal of the long-term negative correlation between fertility rates and a series of development indicators at the macro level has been well documented for female labor force participation, the Human Development Index and GDP per-capita. Recent fertility theories have tried to explain the emergence of the U-shaped pattern in the relationship between development and fertility suggesting that the same set of factors that reshaped women's roles and pushed fertility to low levels, namely increasing education and labor force participation, will drive a re-increase of fertility levels as we move from a male-breadwinner equilibrium to a dual-earner dual-career equilibrium. In this paper we present a computational model of reproductive decision-making that allows us to reproduce the observed pattern of decline and recuperation of fertility in

the context of rather linear increases in the proportions of working women and women with higher education. We exploit the potential of the agent-based approach by modelling aggregate fertility patterns as a result of the interconnected decisions of individuals over the life course and across cohorts. We estimate the parameters of our model from macro level data with the aid of a Gaussian process emulator using data on the evolution of age-specific fertility rates in France and Spain, two countries that exhibit the pattern of decline and recuperation of period fertility but present very different economic, institutional and policy contexts.

Emanuele Del Fava

Integrated Bayesian modeling of migration flows using multiple data sources

For a better understanding of the causes and consequences of international population migration, we need to overcome the limitations of the various data sources employed by countries to measure the migration, e.g., censuses, population registers, and surveys.

Population registers can capture migration data in the form of relocations (changes of address), conditional on being followed by a stay of a minimum duration that is country-specific. Possible issues that may hinder the comparison between countries include under-registration, differences in the minimum duration, and accuracy of the collection system.

On the other hand, censuses and surveys typically inquire about transitions, namely, differences in the country of residence between the time of the interview and one year (or five years) before. These data, particularly those from surveys, may suffer statistical problems of precision and bias, coverage, and non-response, which all can eventually lead to underestimation.

Previous research has shown how Bayesian modeling can be used to harmonize different data sources and estimate migration rates. These models allow to integrate various data sources on migration and other country-specific covariates that may be predictive for the migration rates; while adjusting for differences in the measurements (under-reporting, different timing, accuracy) and ascertaining the uncertainty around the migration estimates and the model parameters.

In this work, we aim to extend previous research on Bayesian modeling of European migration flows with a twofold purpose. First, we develop a model that integrates multiple data sources on migration (e.g., population registers and surveys); second, we address the specificities and the limitations of each data source to estimate the number of migrations per year between pairs of countries under a duration of 12 months. We present here some preliminary work, which consists of the modeling framework and an application to data from a selection of European countries.

(Joint work with Emilio Zagheni)

Martin Klesment

Educational heterogamy in late 19th century Tartu, Estonia

BACKGROUND. In historical perspective, the transition from pre-industrial system to modern societies is associated with increasing social status heterogamy. As individual's acquired characteristics became more important for partner selection than inherited class status, the importance of status homogamy declined and marrying outside one's own social group became more frequent.

OBJECTIVE. We investigate educational heterogamy in a 19th century town at the eastern border of the Hajnal line. We ask whether marriage of unequally educated partners is related to status exchange with regard to other characteristics of the partners. Ethnic background, origin, and age are considered as characteristics that may associate with sorting into educationally hypergamous unions.

METHODS. The analysis uses census data from 1897, including both married and non-married population. We estimate the effects of age difference, origin heterogamy, and ethnic heterogamy on the probability of educational hypergamy. The model takes into account women's selectivity into marriage.

RESULTS. The results indicate a positive relationship between educational hypergamy and marrying outside own ethnic or origin group. Also, age difference between the spouses is positively associated with educational hypergamy. The results suggest that educational sorting into marriage was not random with respect to other potential criteria of partner selection.

CONCLUSIONS. Our study provides new evidence about the late 19th century marriage market, more specifically about the role of formal education as a factor in partner selection. It is often assumed that people prefer to marry someone wealthy but with a similar background. We show that educationally upward marriages in the late 19th century Tartu implied a higher likelihood of ethnocultural and origin heterogamy. This can be interpreted as evidence in support of the status exchange theory.

(Joint work with Hannaliis Jaadla and Mark Gortfelder)

Luca La Rocca

Inference on the Derrida parameter of a Random Boolean Network

This poster presents an analysis of genome-wide mRNA expression data from a battery of gene deletion experiments on the yeast *S. cerevisiae*, using a well-known model of genetic regulatory networks and focusing on the parameter that determines its dynamical regime.

References:

Villani, M., La Rocca, L., Kauffman, S. A. & Serra, R. (2018). Dynamical criticality in gene regulatory networks. *Complexity*, 5980636, 1-14.

Stefan Leknes

Estimating age-specific fertility rates for small areas using Norwegian register data

There is large demand for small area estimates of demographic components. They are important in understanding demographic change, in constructing regional populations, and for planning purposes. Reliable estimates are challenging to construct, as the units of interest are scarcely populated. In this application we will use Norwegian register data to create age-specific fertility rates by the use of empirical Bayes methods. This method exploits support from aggregate areas, and keep relevant info from the local level.

Rosa Maria Lipsi

Regional Population Estimation with Italian Administrative Data: preliminary results of the Bryant and Graham approach

The European Statistical Program provides the framework for the development, production, and harmonization of European statistics. One aim of the program is to encourage Countries to extend the use of administrative data to produce official statistics in all statistical fields, especially for population estimation. Next Italian Census, expected in 2021, will be based on data deriving from population register, instead of the traditional one. This register will be built combining different data sources. For this purpose, many statistical

agencies proposed and evaluated new methods, as the Bryant and Graham approach (2013a, 2013b, 2015). In their approach the Bayesian model is composed by an observation model and a system model. Both have a hierarchical Poisson-gamma structure. The system model is a complete description of the 'demographic account' (Rees 1979), that contains counts of births, deaths, migrations, and population, all disaggregated by age, sex, region, and time. The observation model describes the relationship between the demographic account and the observed data. The posterior distribution of the parameters is obtained by simulation using Markov chain Monte Carlo methods.

In this study, the Bryant and Graham model has been applied to Italian data. Preliminary analysis are restricted to the estimation at the level of the cell count of the population, the newborn and deaths, by using official statistics and comparing the estimate with a single administrative data source (Lista Anagrafica Comunale - LAC). Each cell count is identified by sex, 19 age groups and, 22 regions from 2006 to 2014 years.

First results highlight the model's ability to simultaneously detect the peculiarities of demographic components (e.g. the fertility pattern by age or region) and the potential use of Bayesian methods in the production of official population statistics.

(Joint work with Simona Toti and Sara Giavante)

References:

Bryant, J. R. and P. J. Graham (2013a). "Bayesian Demographic Accounts: Subnational Population Estimation Using Multiple Data Sources", *Bayesian Analysis* 8: 591–622.

Bryant, J. R. and P. J. Graham (2015). "A Bayesian Approach to Population Estimation with Administrative Data". *The Journal of Official Statistics*, Vol. 31, No. 3, 2015, pp. 475–487.

Bryant, J. and P. Graham. (2013b). "A Bayesian Method for Deriving Population Statistics from Multiple Imperfect Data Sources". Paper presented at the World Statistics Congress, August 25–30, Hong Kong.

Rees, P. (1979). "Regional Population Project Models and Accounting Methods". *Journal of the Royal Statistical Society. Series A (General)*, 142(2): 223–255. 592.

Andrea Nigri

Life expectancy: a realistic forecasting model

Starting from demographic analysis reasons, other than actuarial and financial implications, mortality forecasting continues to represent a real challenge driving researchers and practitioners to find a more plausible way in order to portray the future shapes. In this context, the aim of our work is to produce an accurate forecast life expectancy using a deep learning approach based on the Recurrent Neural Network with Long Short-Term Memory (LSTM) architecture. This approach has not yet been considered in the literature concerning the investigation of life expectancy evolution. As general result, we observe that LSTM network shows better predictive accuracy in all forecasts tried on several countries worldwide.

Francesco Rampazzo

Following a Trail of Breadcrumbs: a Study of Migration through Digital Traces

Measuring International migration is challenging. The lack of timely and comprehensive data about migrants, and varying measures and definitions used by countries are a barrier to understanding international migration. In this paper, we complement traditional data sources with digital traces, i.e. social media data.

We use the Integrated Model of European Migration to combine the data from the Labour Force Survey and Facebook Advertising Platform to study the number of European migrants in the UK, aiming to produce estimates of European migrants closer to their true stock number. The model used in the analysis provides a framework which assesses the limitations of the datasets in terms of the definition of migrants used; the bias and the accuracy are also considered to create an appropriate prior distribution, which could adjust these data issues. The model is divided into a migration theory based model, and a measurement error model. The estimates produced in the model suggest that there are more European migrants than suggested by the official estimates. For the ten most numerous nationalities, we produced estimates by 5 years age group and sex. We discussed the advantages and limitations of this approach, and we suggest how we can complement even more data sources in this framework.

Budi Setiawan

Implementation of the Bayesian model in population projection and its implications for changes in Indonesia population structure 2015-2045

In the latest population projection, the national statistical office of Indonesia still uses the component method to produce deterministic projections for the population of Indonesia and the provincial level. Therefore, this study aims to provide probabilistic projections of the population of Indonesia and the provincial level. Probabilistic projections for total fertility rate (TFR), life expectancy at birth for males and females, net migration, and population totals using Bayesian Hierarchical Modeling.

Sandile Simelane

TBA

Astri Syse

Prediction intervals around deterministic projections

Norway's national population projections are currently deterministic and scenario-based (Syse et al 2018). We do, however, provide 15 alternatives thus ensuring that the components do not have to vary in similar manners. Our most used alternative is entitled 'main alternative', and uses the medium scenario for fertility, life expectancy and immigration (MMM). Other much used alternatives are high national growth (HHH), low national growth (LLL), low fertility (LMM), high immigration (MMH), strong aging (LHL) and weak aging (HLH). The age profile of the future population varies depending on the combination of the alternatives for the various components. All information about our projections may be found at www.ssb.no/en/befolkning/statistikker/folkfram.

As of today, mortality (life expectancy) and immigration scenarios result from models. For mortality, we use a stochastic framework (Lee-Carter, product-ratio version). For immigration, we use parameters that estimate future immigration based on income differences, unemployment differences, population changes in sending areas and network effects (Cappelen et al 2015). This model is not used in a stochastic framework. For fertility, we use expert judgement to create a high, low and medium scenario.

For the 2020-projections, we will provide prediction intervals (67, 80 and 95 per cent) around the deterministic medium scenario, using simulations. For future projections, we hope to provide also a full-scale stochastic projection – in addition to the formally acknowledged deterministic projections.

Our mortality projections usually do quite well. However, we are currently attempting to incorporate cohort effects into our model using the StMoMo-method (Villegas et al 2018). As of today, we are not able to fully replicate the published results, and we are working to figure out what might be the cause for the differences we observe. Provided our programming is not too faulty, it seems that possible cohort effects might be rather

weak in Norway, particularly for later time periods. How to forecast future cohort effects, if they are to be incorporated in our formal projections, is currently being debated.

References:

Cappelen Å., T. Skjerpen and M. Tønnessen (2015). Forecasting Immigration in Official Population Projections Using an Econometric Model. *International Migration Review* (49) 4: 945–80.

Syse, A., S. Leknes, S. A. Løkken and M. Tønnessen (2018). Norway's 2018 population projections – Main results, methods and assumptions. Reports 2018/22, Statistics Norway. www.ssb.no/en/befolkning/artikler-og-publikasjoner/attachment/354133?ts=1643ab3eaf8

Villegas A., P. Millosovich and V.K. Kaishev (2018). StMoMo: An R Package for Stochastic Mortality Modeling. <https://cran.r-project.org/web/packages/StMoMo/vignettes/StMoMoVignette.pdf>

Kathryn Weny

Eliminating FGM – Is the 2030 agenda target within reach?

In 2015, UN member states committed to eliminate Female Genital Mutilation (FGM) by 2030 within the Sustainable Development Agenda. To reach this goal, interventions need to be targeted and guided by the best available evidence. To date, however, estimates of the number of girls and women affected by FGM and their trends over time and geographic space have been limited by the availability, specificity and quality of population-level data. We presented new estimates based on all publicly available nationally representative surveys collected since the 1990s that contain both information on FGM status and on the age at which FGM occurred. Using survival analysis, we have generated estimates of FGM risk by single year of age for all countries with available data, and for rural and urban areas separately.

However, to further detect areas with little decline in FGM as well as refined projections on FGM trends, better models are needed in order to provide reliable estimates and support programmatic activities to eliminate Female Genital Mutilation.

Kathryn Weny

Sexual and reproductive health and rights in sudden onset crises

To date, UN agencies including UNFPA often have to make rough assumptions about the size and profile of populations affected by humanitarian settings. High quality and up-to-date data on population by age, sex, and geography as well as key socio-economic indicators are often not available. Consequently, the UN humanitarian system is forced to operate under a lot of uncertainty, which affects the accuracy of procurement decisions, fund raising and deployment of resources.

In the case of the Minimum Initial Service Package (MISP) that primarily serves to address Sexual and Reproductive Health needs in sudden onset crises, the UN and its partners use the best available data from humanitarian data sources, nationally representative household surveys or modelled estimates provided by the UN or partner agencies. Given the variation in humanitarian contexts, complexity of population data landscapes and the short response time available for humanitarian response, data preparedness and prepositioning are critical. UNFPA is looking at innovative solutions and ideas to improve the current situation and to improve estimates in poor data settings.

Sainan Zhang

TBA