Innovative data

Strand organizer: Jason Hilton (University of Southampton)

Innovations in modelling processes & mechanisms – Monday 9 September, 4.45pm

Infectious disease modelling with demography

Signe Møgelmose[1,2], Karel Neels[2], Niel Hens[1,2]

[1]Hasselt University, Belgium, [2]University of Antwerp, Belgium

The demographic structures of populations influence the age-specific contact patterns which are relevant for the spread of infectious diseases. Households especially play a central role in the transmission process of many infectious diseases, due to the higher frequency and intimacy of social contacts among people living together. Nevertheless, it is not fully understood how demographic changes such as population ageing affect the long-term dynamics of disease transmission and infection control. In order to investigate this, we are modifying an existing individual-based model (IBM) and applying it to the Belgian population. The IBM consists of two parts: a demographic model and a disease transmission model. Access to Belgian census and register data makes it possible to obtain a higher level of detail and realism regarding the population composition in the demographic model. The disease transmission model is modified to a household model, which allows for a distinction between household and population-level transmission. Using this modified IBM, the Belgian population is simulated from 2001 to 2050 with influenza-like epidemics of varying total final size. The use of census and register data improves the representation of the initial population. During the simulation period, population ageing takes place. This is the result of increasing life expectancy and below-replacement fertility, which indirectly leads to an increasing share of small households. We show the impact of these demographic changes on disease transmission dynamics at the household and population levels. Moreover, we demonstrate how these effects change for different levels of vaccination coverage.

Email: signe.mogelmose@uhasselt.be

How does information spread affect migration routes? Insights from simulation modelling

Martin Hinsch, Jakub Bijak

University of Southampton

Migration is characterized by complex dynamics and unpredictability, which are especially acute for forced migrations. Among their key features, one important aspect is related to the emergence and changes of migrant routes. In the context of the 2015–16 European asylum migration, it has been posited that information spread has significant impact on the paths the migrants may choose in order to reach their chosen destinations. We explore the possible mechanism by a means of an agent-based simulation model, whereby agents (migrants) navigate through an artificial landscape, exchanging information about the features of different locations in terms of their attractiveness and how difficult they are to reach. The migrant agents in the model are heterogeneous and vary with respect to how much they are willing to explore the landscape and what resources they have at their disposal. Additionally, simulations differ with respect to how the agents follow predefined plans or rely on the information obtained from others. These features, as well as those of the different locations forming a part of the landscape, are described by using free model parameters. In this talk, we will present the model and its key substantive findings, and use meta-modelling techniques based on Gaussian process emulators to assess the model uncertainty and sensitivity of various possible outcomes, such as tendency to concentrate in certain locations,

average time taken to find a way through the landscape, and variations in the input parameters. The paper concludes by offering practical implications for research and policy.

Email: j.bijak@soton.ac.uk

How reducing differentials in education and labour force participation could lessen workforce decline in the EU28

Guillaume Marois[1], Patrick Sabourin[1], Alain Bélanger[2]

[1]International Institute for Applied Systems Analysis, Austria, [2]Institut national de la recherche scientifique, Canada

The objective of this paper is to investigate how improvement in the educational attainment and labour force participation of disadvantaged subpopulations could impact the shape of the future labour force in the European Union. We used a microsimulation model called CEPAM-Mic to project the labour force population of EU28 countries. Compared with traditional projection models, CEPAM-Mic incorporates heterogeneity among different groups and allows the development of alternative scenarios where the educational attainment and labour force participation of disadvantaged groups are comparable to the better advantaged groups. Removing inequalities in educational attainment and labour force participation between disadvantaged groups and more advantaged ones drastically changes the future prospects for labour force size and the labour force dependency ratio (LFDR) in the EU. The expected decline in labour force size is reduced by 54%, while the expected increase in the labour force dependency ratio is narrowed by 70%.

Email: marois@iiasa.ac.at

Innovative statistical models for estimation & forecasting – Tuesday 10 September, 11.00am

Estimating bilateral migration flows from net migration

Guy J Abel[1,2], Peter W F Smith[3]

[1]Asian Demographic Research Institute, Shanghai University, China, [2]International Institute for Applied Systems Analysis, Austria, [3]University of Southampton

Bilateral migration data, summarizing the number of people migrating between each origin and destination, provide a clearer understanding of migration patterns than summary measures such as net migration. However, bilateral migration flow data are commonly unavailable, not up to date, or provide conflicting accounts of population change when compared with changes implied in demographic data. In this paper we introduce an Iterative Proportional Fitting Procedure (IPFP) to update bilateral migration data to match known net migration totals. The IPFP can also be applied in missing data situations to generate synthetic bilateral migration flows to match known net migration totals where no reported bilateral data are available. We illustrate both of these situations using real-world data. The resulting estimates of bilateral migration flows are demographically consistent with changes in reported population totals, births and deaths over the period and provide a detailed depiction of contemporary migration patterns.

Email: guy.abel@shu.edu.cn

Grouped multivariate functional time series method: An application to subnational mortality forecasting

Han Lin Shang, Yang Yang

Australian National University, Australia

Forecasting age-specific mortality rates at the national and subnational levels plays an important role in developing social and economic policies. However, producing forecasts at the finest subnational levels separately will not only fail to utilize correlation among subpopulations to improve forecast accuracy, but also the resulting independent forecasts may not add up to those at the national level. To address these issues, we extend the grouped univariate functional time series to grouped multivariate functional time series forecasting methods. Dynamic functional principal component analysis is first conducted for a group of correlated populations for the purpose of dimension reduction. Point and interval forecasts are then produced at subnational levels, before being aggregated towards the national total in a grouped structure. The advantages of the proposed approach over conventional mortality forecasting methods include improved point and interval forecasts, as well as reconciliation consistency of forecasts across various disaggregation factors within a group. We demonstrate the superiority of our approach via a study of the one-step-ahead to 15-step-ahead point and interval forecast accuracies of regional Japanese age-specific mortality rates.

Email: hanlin.shang@anu.edu.au

Comparing fertility forecasting methods

Jason Hilton, Erengul Dodd, Jonathan J Forster, Peter W F Smith, Jakub Bijak

University of Southampton

Forecasting fertility is vital for a wide range of planning needs, in local and national governments and in the private sector, for instance for the prediction of future demand for housing and various goods. However, fertility is a dynamic phenomenon which is intrinsically intertwined to many other economic and social processes such as the labour market, cultural norms, female emancipation and the housing market. Predicting it is therefore difficult. This paper examines the performance of a variety of different fertility methods according to how well they predict fertility across a range of contexts and periods. This paper builds upon the work of Bohk-Ewald et al. (2018), which also examines the performance of forecasting methods, but is concerned primarily with the errors in completed cohort fertility, rather than the underlying age-specific fertility rates. Existing work (Hilton et al. 2019) has compared the performance of a newly developed Bayesian parametric mixture model to the best-performing models identified by Bohk-Ewald et al. (2018) over a small number of countries, showing that such mixture models are competitive in terms of Root Mean Squared Error. The code for this analysis is available at https://github.com/jasonhilton/fert_mix . This paper extends this analysis to additional countries and methods.

Email: j.d.hilton@soton.ac.uk

Forecasting life expectancy: New evidence from a parsimonious CH survival function

Albert K. C. Tsui[1], Peter Cayton[2], Kin Yip Ho[3]

[1]National University of Singapore, Singapore, [2]University of the Philippines, The Philippines, [3]Australian National University, Australia

We modify the recently proposed CH survival function by Wong and Tsui (2015) to forecast life expectancy at various ages by reducing the number of parameters of the function. Our modification preserves the key features of the 'youth-to-

adulthood' and 'old-to-oldest-old' components of the original CH function, and produces comparable estimates of life expectancy at various ages for the US population. Five component series of parameter estimates generated from the modified function are fitted to multiple time series models, which yield more efficient estimates of parameters. Residual-based bootstrapping is employed to obtain the 95% confidence interval for forecasts of life expectancy at various ages. It is found that the simulated confidence bands for the forecasts are consistently narrower than the corresponding bands from the original CH function.

Email: ecsatsui@nus.edu.sg

Innovative data & methods – Tuesday 10 September, 4.45pm

Probabilistic methods for combining traditional & social media bilateral migration data

Dilek Yildiz[1], Arkadiusz Wisniowski[2], Guy Abel[3]

[1]Wittgenstein Centre (IIASA, VID/ÖAW, WU), Austria, [2]University of Manchester; [3]Asian Demographic Research Institute, China

Official migration statistics are developed and published by national offices of statistics and collated by international organizations. These statistics are based on rigorous internationally harmonized principles, but they come with a considerable time lag. New data sources offer opportunities to complement traditional sources for migration statistics. In particular, the availability of high quantities of individual geo-located data from social media has opened new opportunities. In this research, we develop probabilistic methods to combine traditional and social media bilateral migration data to estimate timelier and potentially more accurate migration statistics. Bayesian methods offer a powerful mechanism for combining data sources. Previously, models have been developed solely for combining traditional migration data sources using the prior models for measurement parameters. We adapt the basic methodologies of these former models to combine migration data from both traditional and new data sources derived from social media.

Email: dilek.yildiz@oeaw.ac.at

Using digital traces to measure the European migrant population in the UK by combining the Labour Force Survey and Facebook advertising data

Francesco Rampazzo, Jakub Bijak, Agnese Vitali, Ingmar Weber, Emilio Zagheni

University of Southampton, Max Planck Institute for Demographic Research

Measuring international migration is challenging. The lack of timely and comprehensive data about migrants, and the varying measures and definitions used by countries are a barrier to understanding international migration. In this paper, we complement traditional data sources with social media data. We use the Integrated Model of European Migration to combine the data from the Labour Force Survey and the 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 five-year age group and sex. We discuss the advantages and limitations of this approach, and we suggest how we can include even more data sources in this framework.

Email: f.rampazzo@soton.ac.uk

Leveraging mobile phones to attain sustainable development

Ridhi Kashyap[1], Valentina Rotondi[2], Luca Pesando[3], Simone Spinelli[2], Francesco Billari[2]

[1]University of Oxford, [2]Bocconi University, Italy, [3]University of Pennsylvania, US

For billions of people across the globe, mobile phones enable cheap and effective communication, as well as access to information and vital services on health, education, and the economy. Drawing on context-specific evidence highlighting the benefits of the digital revolution for social development outcomes, this study is the first to grow this evidence to a global scale by providing causal empirical support for the idea that mobile phones are a vehicle for sustainable development. We do so by assembling a wealth of publicly available macro- and micro-level data, exploring a wide range of demographic and social development outcomes, and leveraging a combination of methodological approaches, enabling us to unravel causal relationships. Country-level analyses covering 200+ countries over the past three decades reveal that mobile phone access is linked to reductions in gender inequality, improvements in contraceptive uptake, and reductions in maternal and child mortality levels. Individual-level analyses drawing on detailed geospatial information from sub-Saharan Africa further show that women who own a mobile phone are better informed about where to access sexual and reproductive health and are better able to make independent decisions. Payoffs are larger among the least developed countries or geographical clusters, pointing to the importance for policymakers around the globe of closing digital divides as a way to foster sustainable development. Overall, our findings suggest that boosting mobile phone access and coverage, and ensuring equitable access to these technologies, particularly among women, can be a powerful instrument for attaining empowerment-related development goals, in an ultimate effort to reduce poverty.

Email: ridhi.kashyap@nuffield.ox.ac.uk

What can administrative data tell us about the household population?

Sabrina Rowlatt

Office for National Statistics

The Office for National Statistics (ONS) is transforming the way we produce population and household statistics, with the ambition to put administrative data at the core of our evidence by 2020. Decision-makers (for example central and local government) increasingly need better local data on the size and characteristics of their populations to build better services, such as schools and housing. Greater use of administrative data has the potential to produce better-quality household statistics at lower levels of geography and including multivariate analysis. The Administrative Data Census Project within ONS has previously published progress towards producing household estimates derived from administrative data, including household income, using the concept of an ‘occupied address’. Progress has since been made on household composition and on identifying how individuals are related within an address. This presentation will provide an update on this research and demonstrate the potential benefits that administrative data has, as well as the impact of this transformation, compared with the existing survey-based method.

Email: sabrina.rowlatt@ons.gov.uk