Independent review of population projection methodology of the Greater London Authority

Report for the Greater London Authority

Jason Hilton, Jakub Bijak and Jonathan J Forster

ESRC Centre for Population Change, University of Southampton.

Southampton, 20 October 2016

Acknowledgments

This work has been prepared for and was funded by the Greater London Authority (GLA), under the contract No. 17329. All the views and interpretations in this report are those of the authors only, and they do not necessarily reflect the views of the GLA.

Contents

Independent review of population projection methodology of the Greater London Authority

Executive Summary

- The projection model of the Greater London Authority (GLA) largely follows the methodological- state of the art in demographic projections, except for a few aspects listed below. The model is multi-regional, with local populations changing through births, deaths, international migration, and origin-destination-specific internal migration.
- The data and estimates in the model come from the Office for National Statistics (ONS) and are usually adequate, given a lack of population register in the UK, except for some issues with projected fertility data. The inputs originally come from the vital registration system, survey estimates (international migration) and patient flows (internal migration).
- The method of estimating the demographic rates is sound, although the baseline data for individual local authorities can be characterized by very large random volatility. For a revised version of the model, we recommend smoothing some of this variability first, so that random noise does not produce spurious results in the projections.
- The methodology offers some flexibility with respect to assumption setting: the various components can be constrained to national population projections of the ONS. For migration assumptions, different time periods can be used for the fitting of trends to be extrapolated. As a result, some of these assumptions may differ from those of the ONS.
- A comparison of the model results with the ONS National Population Projections (NPP) for England shows that the fully constrained variant of the GLA projection almost perfectly aligns with the NPP, with slight discrepancies due to different data used for constraining. Larger differences observed in a previous version of the model were a result of larger cross-border flows to other constituent nations in the GLA model and missing data in migration constraints.
- There are some minor issues with how output life tables are produced, which require further attention. The projections would also benefit from a detailed documentation, and detailing the mathematical specification of the model, although the computer code is very neatly presented and well understandable for specialists with knowledge of R.
- Over a longer horizon, further development of the model could focus on a framework for integrated smoothing of data irregularities and forecasting of the trends into the future, ideally coupled with some formal measurement of uncertainty. The pioneering work at Statistics New Zealand offers a great starting point in that regard.

1. Introduction and Background

1.1. Overview

This report reviews the Greater London Authority demographic projection model (henceforth: "the GLA model") with the aim of verifying the model's technical validity, assessing the reasonableness of the assumptions made, and highlighting areas for possible improvement.

The report was authored by members of the Demographic Estimation and Forecasting team at the ESRC Centre for Population Change, University of Southampton. The review was conducted through thorough examination of the model code and the supporting documentation, and by the analysis of selected model outputs. Some of the issues identified in the preliminary version of the report have already been amended, and these instances are clearly indicated in the text.

The review is focused primarily focused on the methodology implemented by the GLA, not those elements of the model based on work by third parties (such as the Office of National Statistics or the Departments for Communities and Local Government), who produce many of the inputs to the model. However, a discussion of the input data is also provided to facilitate discussion of the model and to understand potential sources of uncertainty.

1.2. Structure

The report is divided into five sections. After the current Introduction, in Section 2 we examine the sources of the input data required for running the model, and, where relevant, discuss the adjustments made to such data by the GLA. In Section 3 the methodology underlying the GLA model is described, with focus on how the model projects each demographic component, and how these are combined in the overall projection model in order to obtain estimates of future population. In Section 4 some outputs from the projection model are presented and compared to the outputs from the Office of National Statistics (ONS) National Population Projections. Finally, in the concluding Section 5 the findings of the report are summarised and suggestions for model improvements are offered.

2. Review and assessment of available demographic data

The Greater London Authority model relies on a number of pre-existing datasets produced by the Office of National Statistics. The GLA have made a small number of adjustments and additions to these datasets, which are described below where applicable.

2.1. Components of Change Time Series for Mid-Year Estimates

The Components of Change is the main dataset underpinning the projections provides information at a local authority level about the numbers of births, deaths, and international and internal migratory moves between each mid-year point, so between 1 July 2001 and 30 June 2015 (Office for National Statistics 2016d). The data are utilised in the GLA model to provide the basis for most of the local authority-level rates in the model.

Necessarily given the range of the components included, the data are based on information from a wide range of sources. The dataset can be seen as a by-product of the production of the Office of National Statistics Mid-Year Estimates, and detailed information about the data can be found in the methodological notes for this series (Office for National Statistics 2016b). Taking the population numbers estimated by the censuses as a basis, mid-year estimates for successive years are made by "ageing-on" the population from the previous year – that is, for each local authority, adding births, and updating each age group to reflect the passage of time and changes due to any deaths and migrations. The dataset enumerates all such changes applied between each mid-year point, and additionally includes ad-hoc adjustments made by the ONS and changes to 'special populations', which include serving members of the Armed Forces and the prison population. The sources for these data are described below:

- Deaths in the dataset are based on vital registration data, which provides information on age, sex, and the area of usual residence. This information is generally of high quality, given the legal requirement to register deaths when they occur.
- Births are similarly based on vital registration data and are broken down by sex. The age of the mother at birth is not recorded in this dataset.
- International migration is provided by year and single years of age for each local authority. These data are less reliable, as they are is based on estimates from the International Passenger Survey, a limited survey of entries and exits at UK borders. These estimates are then divided into streams based on the types of migrants, and a wide variety of administrative sources are used to distribute these streams to local authorities. Age and sex distributions are obtained by clustering local authorities based on demographic characteristics, and applying distributions based on the cluster average, to reduce noise.
- Internal migration data are also provided by age and sex. Gross inflows and outflows are provided for each Local Authority, but are not cross-classified by origin and destination. These estimates are based on data from the Higher Education Statistics Authority (HESA), the Patient Register Data Service (PRDS), and the National Health Service Central Register (NHSCR). These data are of reasonable quality, but have well known difficulties in registering movements of young adults and in particular male students, who are mobile and tend not to register with doctors.

Data in this series have been subject to revisions after the 2011 population census (Office for National Statistics 2012). The total count of the resident population resulting from this census differed from the "aged-on" mid-year estimate based on the 2001 census by around 460,000 people. Therefore, the historical series of components of changed had to be updated to reflect the new census-based mid-year estimate for 2011. The majority of the difference is believed to result from under-counts of immigration from 'A8' countries joining the European Union in 2004, and from an underestimate of net flows from the Republic of Ireland early in the decade, and estimates of these factors were obtained by the ONS. However, the net effect of identifiable adjustments was to add more young males than were required to make up the differences with the census based estimate, and also to leave a large proportion of the total difference unattributed. Thus, an unattributable adjustment that boosted the numbers of females while reducing the numbers of young males had to be applied uniformly across the inter-census years 2001–2011 (Ibid).

Because of concerns about the suitability of the methods used by the ONS to conduct this correction, the Greater London Authority use a different method for adjusting the back-series for the components of change for the London boroughs only. In this particular case, a natural demographic change model (without migration) has been estimated for 2002–2011, and the differences between the output of this model and the census-based estimates for 2011 have been attributed to migration.

Overall, the Components of Change dataset combines the best available local authority-level data on demographic change, and has the advantage of also being the data used by the ONS to carry out their projections. This simplifies the problem of attempting to harmonise GLA and ONS projections, and understanding where sources of differences may come from.

2.2. National and Sub-National Population Projections

The ONS National Population Projections (NPP) give age- and sex-specific projections of future population at a national level, based on assumptions about the future trends of the demographic components informed by the past data and expert opinion. Variants are provided corresponding to combinations of principal, high and low scenarios for each component of change. Sub-national Population Projections (SNPP), which are constrained to sum to the national total across all local authorities, are also produced. The methodology is not discussed further here, but is provided in Office for National Statistics (2016c). As part of the SNPP, data on births by single year of age of mother are provided at the local authority level, offering crucial detail on sub-national fertility differences.

2.3. Internal Migration Dataset

Internal migration estimates by age and sex including cross classification by origin and destination are available as a separate dataset from the Office for National Statistics (2016a), overcoming the shortcomings of the Components of Change dataset, which does not provide this breakdown. As before, these estimates are based on the HESA, PRDS and NHSCR datasets. In the GLA projections, only the data for the latest year are used.

3. GLA projection methodology

3.1. Model Summary

In this section, the working of the GLA model is described in some more detail. However, a top-level overview is first offered, to provide the necessary context. Overall, the model aims to project population in a manner broadly consistent with the ONS methodology for Sub-National Population Projections (SNPP), and works with age- and sex-specific data for single year of age, with a top, open age-group of 90+. The model includes four major components: fertility, mortality, internal migration and international migration. The geographies included consist of the 326 English local authorities. plus Scotland, Wales, and Northern Ireland, with the latter three each treated as single, monolithic entities. This set of areas are referred to Local Authorities (LAs) throughout as shorthand for the complete set of geographical areas. The population of interest is the usually resident population, thus discounting temporary or seasonal migrants and visitors to the UK who stay for less than 12 months.

The 'natural change' components of the model (mortality and fertility) function by projecting forward sub-national patterns from the ONS data according to national trends drawn from the National Population Projections. The migration components assume that the future rates will continue at the average of the past n years, where the value of n can be set by the projection user. These elements are combined using the well-known cohort-component projection model (see e.g. Preston et al. 2000), so that the population at a given mid-year point can be updated according to projected births, deaths and migrations in order to obtain an updated population at the following mid-year.

The model is implemented in an open-source environment (R), and the code conforms to the tidy data paradigm (Wickham 2014), so that all data are stored in the same format, with each row of data representing a single observation. This minimises potential sources of error, and increases code readability through the use of 'verb' functions that describe what transformation is being applied and to what data. As a result, the code is generally robust, easily comprehensible and transparent. Additionally, the ability to store parameter settings through a GUI aids reproducibility of results.

The order of calculation for the cohort-component demographic projection model is as follows:

1. Mortality
2. Fertility
3. Infant Mortality
4. International emigration
5. International immigration
6. Internal migration

3.2. Mortality

Historical mortality

To obtain the history of sub-national mortality for projection purposes, deaths and population counts (exposure) are taken from the Mid-Year Components of Change estimates for each local authority. Death probabilities are then calculated for each combination of age, sex and local authority:

$$q^s_{x+0.5,i}(t,t+1) = \frac{D^s_{x,i}(t,t+1)}{N^s_{x,t,i}},$$

with x referring to age at last birthday, t to time, i to geographical unit (LA), and s to sex. To calculate infant mortality, historical births are used as the denominator.

It has to be noted that the deaths $D_{x,i}(t,t+1)$ in the Mid-Year Components of Change dataset, which are in the numerator in the above equation, refer to the number of those aged x *at time t* who died before time $t+1$. This corresponds to the period-cohort Lexis parallelogram coloured red in Figure 1 below (Office for National Statistics 2016b). As the population $N_{x-1,t,i}$ refers to those aged $[x,x+1)$ at time t (red line), this approximates the respective life-table probability, $q_{x+0.5}$.

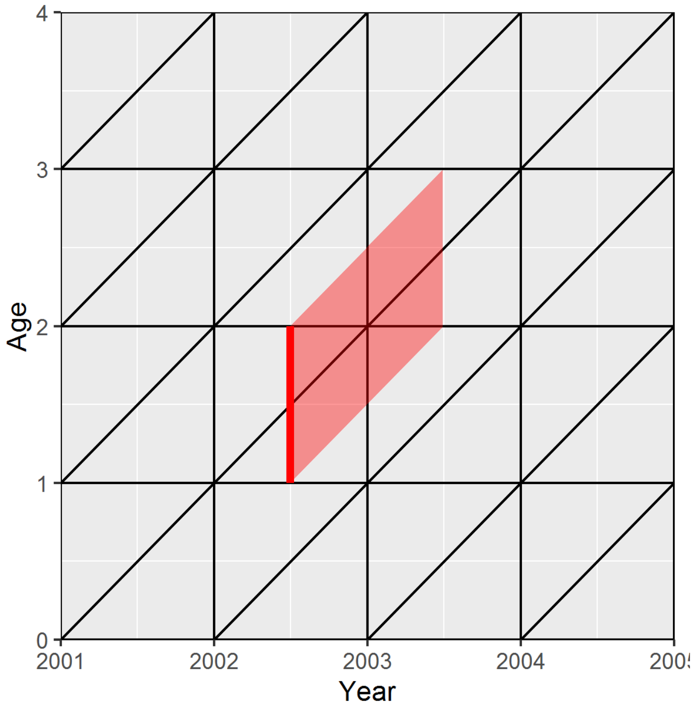

Figure 1: Lexis diagram of deaths

There seems to be some confusion in the code between rates and probabilities when it comes to mortality measures. Keilman (2000) provides a description of the problem of confusing these two concepts; in summary, these quantities are conceptually very different. A probability describes must lie in the interval [0,1], while a rate can potentially take any non-negative real number. In the projection loop, the quantities calculated are correctly treated as probabilities; the initial population between age x and $x + 1$ is multiplied by $q_{x+0.5}$ to obtain the number of deaths, which is then subtracted from the initial population to obtain survivors.

However when life tables for output files are computed, the quantities are treated as though they were classical occurrence-exposure rates (m_x in the life-table convention), and converted to another q_x^* by using the standard formula below:

$$q_x^* = \frac{q_x}{1+(1-a_x)q_x}.$$

This suggests an underestimation of the true q_x and an overestimation of true life expectancy. This terminology is also evident in ONS work, where the quantities q_x are also described as 'rates' in some documentation. Still, this issue does not affect the results of the underlying populations, which are correct, but only some output, such as life expectancies.

Baseline probabilities

The jump-off probability $\hat{q}_{x+0.5,i}$, that forms the basis of the projections, is calculated by either averaging or regressing against time the previous n years of probabilities *for each age, sex and LA combination separately*, in order to estimate the first year of rates. This results in rather noisy estimates (Figures 2 and 3), as there is no smoothing or borrowing of strength from the national or regional levels.

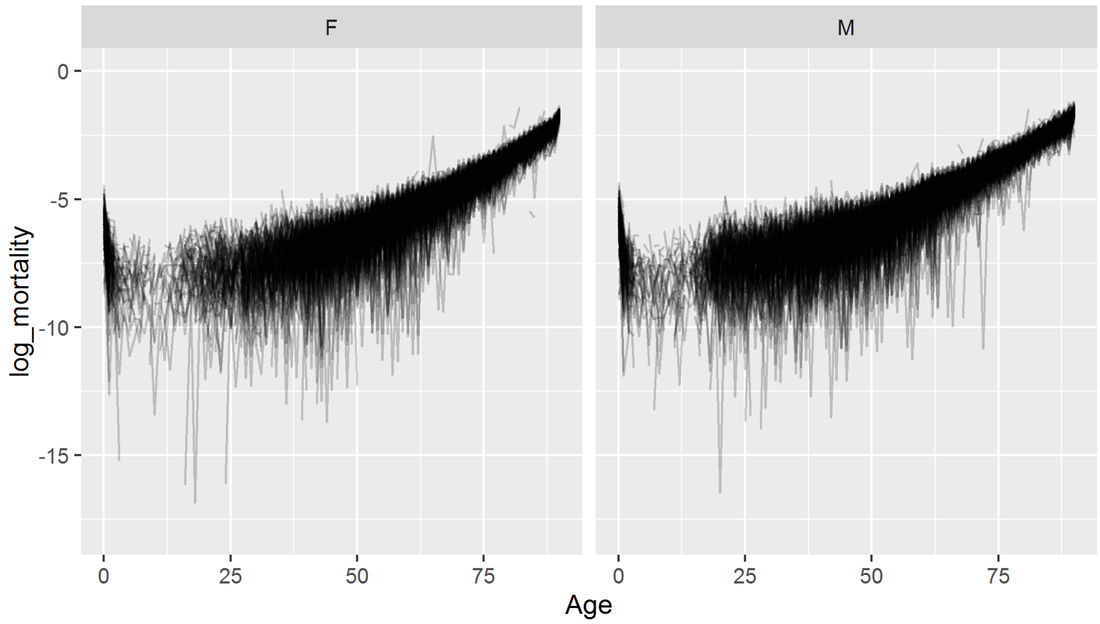

Figure 2: Estimated Log Probabilities of Death, 2015. Data source: ONS

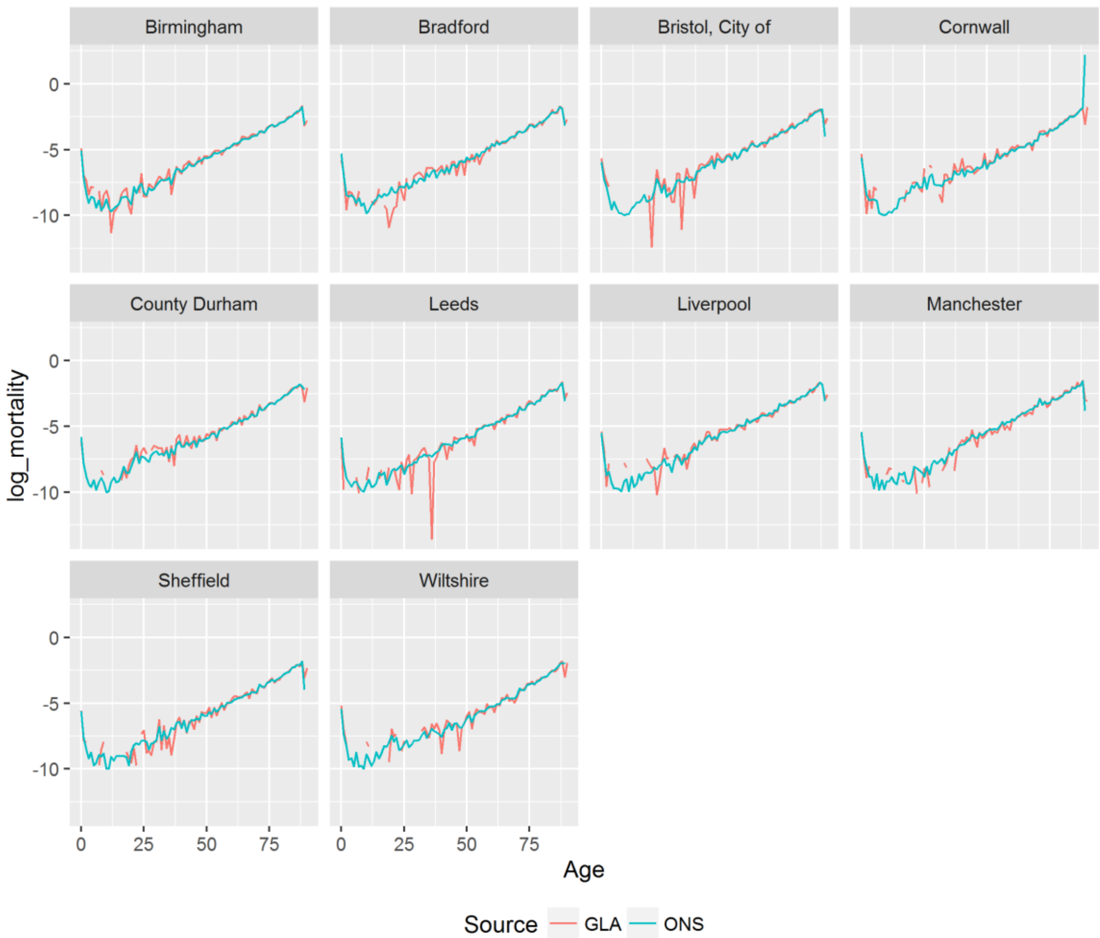

Figure 3: Comparison of GLA and ONS Log probability projections for 2018 for the largest 10 English LAs by Population. Data source: ONS

This problem is evident, when the estimates are compared with the equivalent values from the ONS NPP, even for the largest districts, where statistical variability should be less of an issue. The ONS method consists in calculating an average differential between local and national mortality rates over the past five years (presumably separately for each age and sex group also, although this is not clear from the documentation), and then using these differentials to scale down national rates for the first projection year. The ONS also replace zero rates with national equivalents. These elements together provide for a smoother estimates of mortality, as can be seen from the implied log death probabilities from ONS and GLA forecasts (Figure 3). Most of this roughness occurs at young ages were deaths are rare in any case, however, and so may not cause large differences in final population totals.

Future Probabilities

Using the NPP, which provide national projections of future death probabilities $q_{x+0.5}$, annual improvement factors δ by single year of age *x* are calculated for all projection years t. We drop the sex-specific notation for simplicity and allow the superscript to denote the source of projection. Thus, for $x \leq 90$, we have:

$$\delta_{x,t} = \frac{q_{x+0.5}^{NPP}(t, t+1) - q_{x+0.5}^{NPP}(t-1, t)}{q_{x+0.5}^{NPP}(t-1, t)}$$

To get the local authority specific probabilities q, the jump off probabilities $\hat{q}_{x+0.5,i}$ are simply repeatedly multiplied by the values of δ calculated from the trends.

$$q_{x+0.5,i}^{GLA}(t+h, t+h+1) = \hat{q}_{x+0.5,i} \prod_{k=1}^{h} \delta_{x,t+k}$$

These probabilities can then be used as the basis for calculating survivorship in the cohort-component model.

Constraints

Projected deaths by local authorities can be constrained to match the National Population Projections for either 2012 or 2014. Total national deaths implied by the GLA model for each year, sex and age combination are calculated by summing projected deaths over LAs. Scaling factors are then simply the ratio of these totals to equivalents in the NPP, and are applied to all LAs. Where NPP information is unavailable (beyond 2037, for example), scaling factors are set to 1 by default.

3.3. Fertility

Fertility Structure

The first step towards obtaining estimates of future births in the GLA model is to estimate the sub-national structure of fertility. 'Structure' here implies patterns of fertility rates by age for each LA. Births for a baseline mid-year to mid-year period by mother's single year of age and LA (from the SNPP) form the numerator for the calculation of these rates. The at-risk population is calculated as the average of the mid-year population from the current year and the previous year (from MYEs), correctly accounting for changes in number of potential mothers over the period. The age-specific fertility rates, *f*, are thus calculated as:

$$f_{x,i}(t, t+1) = \frac{b_{x,i}(t,t+1)}{0.5*(N_{x,i,t}+N_{x,i,t+1})}.$$

As with mortality, no smoothing or borrowing of strength is attempted. Fertility is not quite as rare as young-age mortality over most of the age range, so rates are not as noisy as the estimates of death probabilities in the mortality model, but they could still benefit from some modelling that would account for smoothness over age or correlations across areas. Figure 4 below shows estimates of age-specific fertility rates for all local authorities.

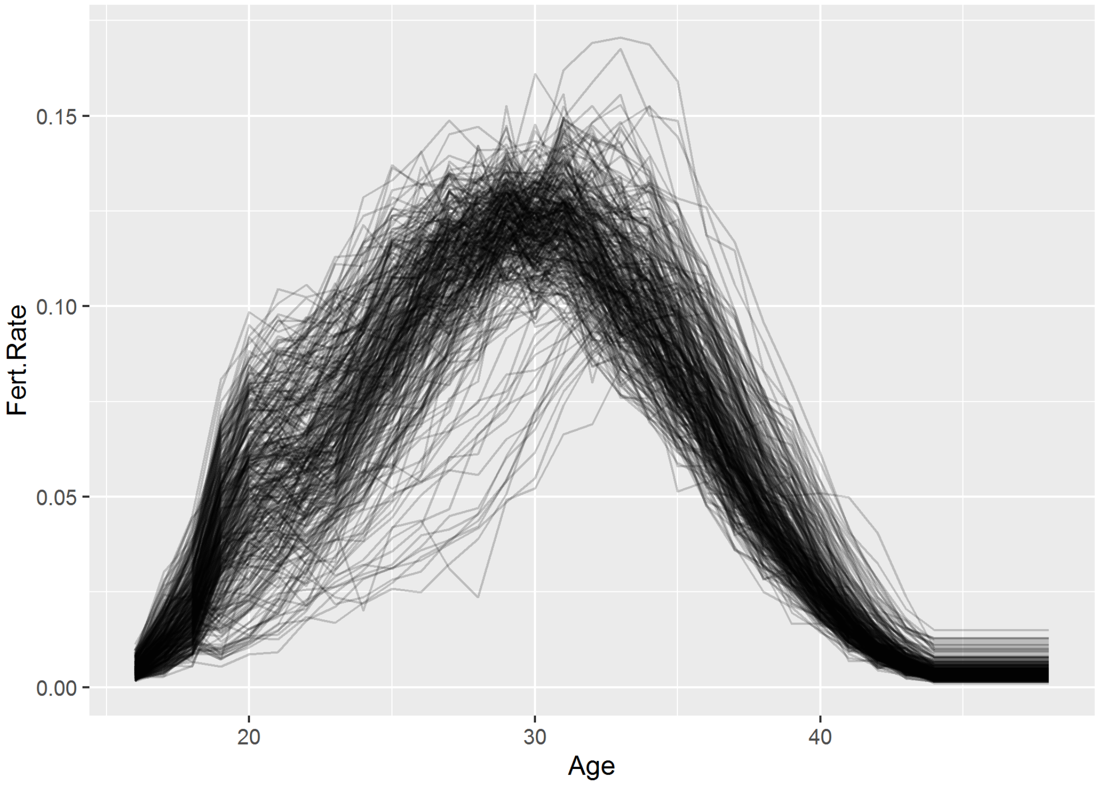

Figure 4: Fertility Rates by Local Authorities. Data source: ONS

Scaling

Births by local authority, implied by this procedure do not necessarily match those from the Mid-Year Estimates (MYEs) for the most recent year, presumably because of differences in the data sources, and also because the first projection year is not necessarily the same as the year for which data is available at the LA level. Some scaling is therefore necessary.

The actual number of total births to scale against is taken from the MYEs for the most recent year where LA-specific data are available. The inferred number of births is calculated using the structured rates above and the correct population at risk (again, the average of population for this year and that preceding it). This is summed over age for each local authority. The actual number of births is then divided by these implied numbers to get a single scaling factor for each LA, which is then used to scale the structured fertility rates to meet the LA totals.

Future Trend

For the future years, the proportionate yearly changes in age-specific fertility rates from the NPP are calculated thus:

$$\delta_f = \frac{f_{x,t}^{NPP} - f_{x,t-1}^{NPP}}{f_{x,t-1}^{NPP}}$$

The user can choose which ONS variant projection is the source for these changes, which are subsequently used to multiply the baseline age-specific fertility rates for each Local Authority in order to get projections for subsequent years.

In the preliminary version of the model, there was a data processing error: the input file provided with the model appeared to show that fertility rates were constant for every variant at each year, so that δ_f was always zero. This has been now corrected, and the assumptions are in line with the 2014 NPP fertility projections: relatively stable over the projection horizon, but with some initial movement towards an asymptote being present.

Projection in the Cohort Component Model

The numbers of births are obtained by taking an average of the projected fertility rates applied to the initial and survived populations. Note that 'survival' here only reflects mortality, not migration processes, so emigrants for that year are still counted fully in the population at risk, while immigrants are ignored. This may potential lead to biases in counts of births, but these biases are likely to be small.

Constraint

Total numbers of births can also be constrained to those from the NPP in 2014 (or 2012). This is done in the same way as for mortality, by working out the ratios of the sum of the LA-level projected births to the national totals by age and sex, and using these factors to scale up the projections. As scaling data are only available for some regions, scaling factors are thus simply set to 1 elsewhere. Similarly, projections against which to scale are given only up to 2037, so beyond this scaling factors are assumed to be equal to 1. This can potentially lead to a step change after the constraints are lifted, so it is advised that constrained projections are terminated after 2037.

3.4. International migration

Description

International migration data from the Mid-Year Components of Change dataset are used directly to obtain estimates of inward and outward migration. For immigration, raw counts are used by age and area. For emigration, probabilities are used, dividing emigration counts between mid-years by the mid-year population count for each cell. Baseline age-specific rates or counts are established by averaging separately for each LA over the past n years, where n is a parameter to be set – five or twelve years are typically used. A multiplicative dampening parameter is available for both immigration and emigration, to allow for an assumption that international migration will be lower (or higher) in the future than it has been in the historical series. International migration is assumed to remain at the base average for all future years. Net migration is then added to the survivors for each LA.

Constraint

International immigration and emigration estimates for individual Government Office Regions are available up until 2037. Constraints can be applied in a similar way as for fertility and mortality, excepting for the use of a higher level of regional aggregation.

3.5. Internal migration

The ONS internal migration estimates of counts of movements from LA to LA by age are first restricted to the latest n years, as was the case for international migration. Then, these estimates are divided by relevant populations from the sending LA to get flow probabilities between each directed pair (dyad) by year. The average over these years is calculated, and this probability is then used for all projection years. Outward and inward migration for each LA-LA tie is calculated for each projection year. Then, for each LA, the total in- and out-migration counts are respectively added to and subtracted from the surviving population.

Differences in internal migration relative to the ONS SNPP are mostly due to differences in the averaging period. Figure 5 below shows the time-series of gross internal flows into and out of London, together with the averages used by the ONS and GLA. The GLA prefer to average over a longer period, as the five-year period used by the ONS includes a potentially exceptional reduction of outflows which seems to be linked to the economic downturn post-2008. The increase in migratory outflows in more recent years appears to at least make the use of a longer time series plausible and better justified.

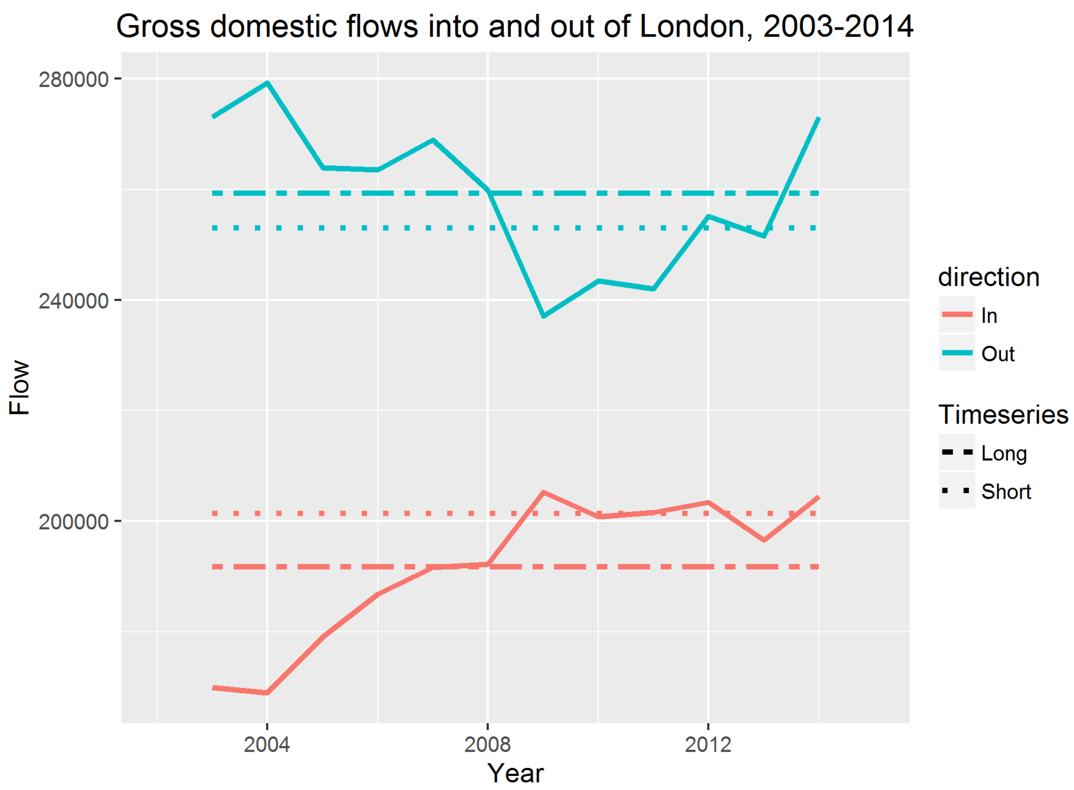

Figure 5: Domestic flows into and out of London, with averages using different time-series lengths. Data source: ONS

4. Empirical comparison to ONS projections

In order to compare the empirical results with the equivalent ONS projections, the GLA model was run while constraining fertility, mortality and international migration to sum to equivalents from the National Population Projections, while leaving internal migration unconstrained. In the initial version of the projections, from examining population trends over the projection period for England only, relatively small differences remained between the two projected totals. The majority of these differences were due to larger cross-border outflows to the rest of the UK in the GLA model; in the ONS model, these flows stayed relatively constant in absolute terms, while with the GLA model they grew, presumably in line with rising populations.

The remainder of the difference was small, and could be attributed mainly to international migration (Figure 6). Although this component was constrained to the National Population Projections, the constraint file contained some flows which were not attributed to any region. As a result, such flows were not included in the constraining procedure, which involves calculation of regional-level scaling factors. Although the annual difference was very small, the effect accumulated over successive years, as shown in Figure 6.

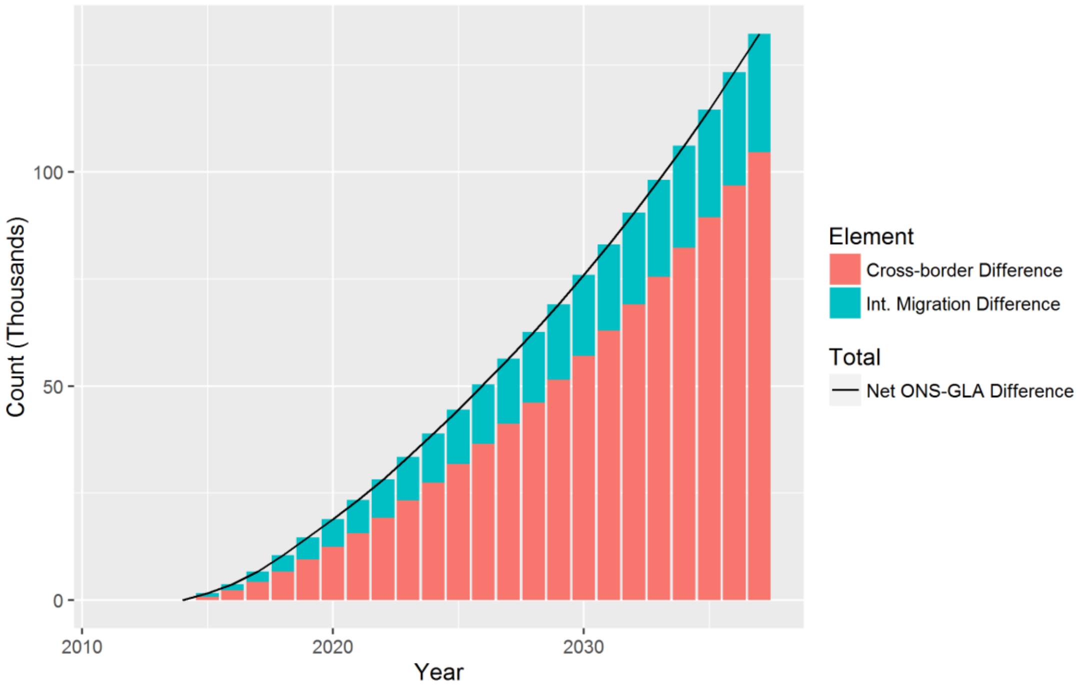

Figure 6: Decomposition of the difference between ONS and initial GLA projections

In the most recent version of the projections, the issues regarding cross-border flows to the rest of the UK and the international migration constraints have been resolved, with another constraint being added for flows between the UK constituent countries. As a result, there is now a near-perfect alignment between the constrained GLA projection and the ONS NPP. The only residual difference is due to the underlying data used for constraining rather than being a result of the model itself. In particular, the migration outflows are slightly larger in the GLA constraint file than in the ONS NPP, which is the reason for the discrepancy.

5. Comparison to the state of the art and other methods

The GLA model is robust, well-implemented and adequate for the purpose to which it has been put. With regard to the model specification, in particular, it follows the state of the art of demographic projections at a sub-national level, that is, multi-regional models, which link different areas explicitly through origin-destination-specific migration flows, embedded within the cohort-component model of population renewal (see Rogers 1975; and Rogers and Philipov 1979). This is especially important, as sub-optimal model specification, for example based on net or gross flows, can bias the projection results (Raymer, Abel, and Rogers 2012). Over the recent decades, the multi-regional approach has become dominant across Europe, including in the UK (Kupiszewski and Kupiszewska 2003; Wohland et al. 2010). A fairly recent methodological summary is available in Swanson and Tayman (2012).

Nevertheless, in comparison to the most cutting-edge developments in academic literature and official statistical practice on population estimation and forecasting, the GLA model could be improved in several important respects. Firstly, as discussed before, demographic rates are not smoothed to attempt to negate the effect of random volatility, and neither is strength borrowed across regions. Secondly, no attempt is made to account for uncertainty in future predictions, except through the use of variant projections from the ONS National Population Projections. These issues are problematic because there is little indication of how likely such scenarios are to come about, and the nature of the variants produced is somewhat arbitrary, rather than being based on past data variability.

There exist models that take care both of smoothing the patterns and structures, and of embedding the projections in a probabilistic framework. Many of these approaches have been pioneered in New Zealand, both in academic literature (Cameron and Poot 2010; Bryant and Graham 2013; 2015), and in official statistical practice. Importantly from the point of view of the GLA model and its programming environment, Statistics New Zealand (2016) has just made available R libraries for sub-national probabilistic population estimates and forecasts based on multiple sources of data.

In the academic literature, there have also been some other methodological suggestions, such as averaging or combining forecasts yielded by different models (see, e.g., Reinhold and Thomsen 2015; or Wilson 2016). Another methodological aspect that can be formally brought into the GLA framework for sub-national projections is related to developing methods for evaluating the projection performance *ex post* in a formal way (see Bell and Skinner 1992; Wilson and Bell 2004). Having such a formal and programmed evaluation mechanism in place would enable spotting departures from projection trends early on, which in turn would facilitate timely revisions of projection assumptions.

6. Conclusions and recommendations

Overall, the GLA model utilises the best data sources available together with a trusted projection method to obtain credible estimates of future population. The assumptions are realistic, and sometimes more robust than those implemented as a part of the standard ONS sub-national population projections, for example with respect to how long data series are used for estimating trends in specific variables. Still, a few specific improvements to the current version of the GLA projection methodology can be suggested. In particular, in the immediate future, we recommend to:

R1. Correct the calculations of life-tables for outputs files. At present, the quantities which are already probabilities of dying (q_x) seem to be treated as rates, leading to incorrect quantities being calculated as life expectancies.

R2. Produce less noisy estimates for rates and probabilities at the local level. This is particularly an issue for mortality. Ideally, a hierarchical statistical model which 'borrows strength' across areas by considering similarities in patterns of rates can be considered (Gelman et al. 2014), or one that smooths the data over age (Wiśniowski et al. 2015; Bijak et al. 2015). Failing this, parametric models can be used. A short-term solution would be to adopt the ONS sub-national population projection method for mortality, average the rates over the previous five years to obtain a scaling factor, and use this to scale-down national rates, replacing zero rates with national equivalents.

R3. Produce thorough documentation, describing the model formally in demographic notation, which would be helpful for professional users not directly familiar with R to quickly understand the methodological basis and the workings of the model. Besides, the use of defensive programming, such as through the `R` package `assertr`, might allow data-frames to be automatically checked when running the model to ensure that input data contain the expected content. This will safeguard against future problems with updated datasets going unrecognised.

In a longer term, we recommend developing a framework for ex post forecast evaluations, and exploring the feasibility of using probabilistic methods in order to properly assess the uncertainty of the outcomes. These long-term recommendations notwithstanding, and except for the minor limitations listed above, which should be easy to fix, we consider the GLA model in its present form to be developed in accordance with the methodological demographic state of the art, well thought through, using the best information available, and thus fit for purpose.

References

Bell, Martin, and Jim Skinner. 1992. “Forecast accuracy of Australian subnational population projections.” *Journal of Population Research* 9 (2): 207–35.

Bijak, Jakub, Erengul Dodd, Jonathon J. Forster, and Peter W. F. Smith. 2015. “English Life Tables No. 17 Methodology.” Office of National Statistics.

Bryant, John R., and Patrick J. Graham. 2013. “Bayesian demographic accounts: Subnational population estimation using multiple data sources.” *Bayesian Analysis* 8 (3): 591–622.

———. 2015. “A Bayesian Approach to Population Estimation with Administrative Data.” *Journal of Official Statistics* 31 (3): 475–87.

Cameron, Michael P., and Jacques Poot. 2010. “A Stochastic Sub-national Population Projection Methodology with an Application to the Waikato Region of New Zealand.” PSC Discussion Papers. Hamilton, New Zealand: Population Studies Centre, University of Waikato.

Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. 2014. *Bayesian Data Analysis*. Third Edition. Abingdon: CRC Press.

Keilman, Nico. 2000. “Review of Andrew Hinde: Demographic Methods.” *European Journal of Population* 16 (2): 187–89.

Kupiszewski, Marek, and Dorota Kupiszewska. 2003. “Internal Migration Component in Subnational Population Projections in Member States of the European Union.” CEFMR Working Paper 2/2003. Warsaw: Central European Forum for Migration Research.

Office for National Statistics. 2012. “Methods used to revise the national population estimates for mid-2002 to mid-2010.” Office for National Statistics. http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/method-quality/specific/population-and-migration/population-statistics-research-unit--psru-/methods-used-to-revise-the-subnational-population-estimates-for-mid.

———. 2016a. “Internal migration, England and Wales: year ending June 2015.” June 2016. Office for National Statistics. http://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/migrationwithintheuk/bulletins/internalmigrationbylocalauthoritiesinenglandandwales/2015-06-25.

———. 2016b. “Methodology Guide for Mid-2015 UK Population Estimates (England and Wales).” June. Office for National Statistics. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/methodologies/methodologyguideformid2015ukpopulationestimatesenglandandwalesjune2016.

———. 2016c. “Methodology used to produce the 2014-based subnational population projections for England.” Office for National Statistics. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationprojections/methodologies/methodologyusedtoproducethesubnationalpopulationprojectionsforengland.

———. 2016d. “Population estimates and components of population change Detailed time series 2001 to 2015.” Office for National Statistics. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland.

Preston, Samuel, Patrick Heuveline, and Michel Guillot. 2000. *Demography: Measuring and Modeling Population Processes*. New York: Wiley.

Raymer, James, Guy J. Abel, and Andrei Rogers. 2012. “Does specification matter? Experiments with simple multiregional probabilistic population projections.” *Environment and Planning A* 44 (11): 2664–86.

Reinhold, Mario, and Stephan L. Thomsen. 2015. “Subnational Population Projections by Age: An Evaluation of Combined Forecast Techniques.” *Population Research and Policy Review* 34 (4). Springer Netherlands: 593–613.

Rogers, Andrei. 1975. *Introductions to multiregional mathematical demography*. New York: Wiley.

Rogers, Andrei, and Dimiter Philipov. 1979. “Multiregional Methods for Subnational Population Projections.” IIASA Working Paper 70. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Statistics New Zealand. 2016. “Demographic Estimation and Forecasting R Package.” https://github.com/statisticsnz/R.

Swanson, David A., and Jeff Tayman. 2012. *Subnational Population Estimates*. Dordrecht: Springer.

Wickham, Hadley. 2014. “Tidy Data.” *Journal of Statistical Software* 59 (1): 1–23.

Wilson, Tom. 2016. “Does Averaging Yield More Accurate Local and Regional Population Forecasts?” *Applied Spatial Analysis and Policy*. Online first, DOI: 10.1007/s12061-016-9194-2.

Wilson, Tom, and Martin Bell. 2004. “Comparative empirical evaluations of internal migration models in subnational population projections.” *Journal of Population Research* 21 (2): 127–60.

Wiśniowski, Arkadiusz, Peter W. F. Smith, Jakub Bijak, James Raymer, and Jonathan J. Forster. 2015. “Bayesian Population Forecasting: Extending the Lee-Carter Method.” *Demography* 52: 1035–59.

Wohland, Pia, Phil Rees, Paul Norman, Peter Boden, and Martyna Jasinska. 2010. “Ethnic Population Projections for the UK and Local Areas, 2001-2051.” Working Paper 10/02. Leeds: School of Geography, University of Leeds.

■

ESRC Centre for Population Change

The ESRC Centre for Population Change (CPC) is a joint initiative between the Universities of Southampton, St Andrews, Edinburgh, Stirling, Strathclyde, in partnership with the Office for National Statistics (ONS) and the National Records of Scotland (NRS). The Centre is funded by the Economic and Social Research Council (ESRC) grant numbers RES-625-28-0001 and ES/K007394/1.

This report publishes independent research, not funded directly from the core grant of the Centre. The views and opinions expressed by authors do not necessarily reflect those of the CPC, ESRC, ONS or NRS.

Website | Email | Twitter | Facebook | Mendeley