



Accuracy of Local Authority Population Forecasts Produced by a New Minimal Data Model: A Case Study of England

Philip Rees¹ · Tom Wilson²

Received: 11 April 2023 / Accepted: 21 October 2023
© The Author(s) 2023

Abstract

The preparation of forecasts for small and local area populations involves many challenges. Standard cohort-component models are problematic because of small numbers, which make estimation of rates unreliable. Because of this, the Synthetic Migration Population Projection (SYMPOPP) model was designed to forecast local populations without need for detailed area-specific information. This model had been used successfully for small area forecasts in Australia. The objective of the paper is to assess its performance when applied to local areas in England. The model uses a bi-regional structure based on a movement population account. Sub-models of total population change are employed to control future change. Fertility, mortality and migration rates are borrowed from national statistics, constrained to small area indicators. The model uses an Excel workbook with VBA routines and is relatively easy and quick to use. Model inputs were calibrated for 2006–2011 and used to forecast for 2011–2021. Results were tested against the census-based 2021 mid-year populations. A new error statistic, Age Structure Error, was used to evaluate Basic and Refined model versions against official projections. The two versions of SYMPOPP posted lower errors. The simple models had fewer areas with errors of 10% or more (12.3–12.6%) compared with the official projections (14.5% of areas). Investigation revealed that these errors occurred in local authorities with high military, student, prison, or ethnic minority populations, influenced by factors not captured in a projection model for the general population.

Keywords Population forecasts · Simple models · Cohort-component models · Age Structure Errors · Forecast performance

✉ Philip Rees
p.h.rees@leeds.ac.uk

Tom Wilson
wilson.t1@unimelb.edu.au

¹ School of Geography, University of Leeds, Leeds LS2 9JT, UK

² Melbourne School of Population and Global Health, The University of Melbourne, 207 Bouverie St, Melbourne, VIC 3010, Australia

Introduction

Population forecasts for local authorities inform a variety of planning, budgetary, policy formulation, market assessment, and service delivery activities. They help assess future demand for schools, hospitals, housing, public transport, aged care services, and water supply and sewerage infrastructure, among other uses. Important investment decisions involving millions of pounds/dollars/euros are made, at least partly, based on such forecasts. These decisions can result in the provision of essential new services, buildings, or infrastructure for local areas. Sometimes decisions are instead made to close or relocate services or to sell off or demolish buildings or infrastructure. It is important therefore that population forecasts are as accurate, timely, and informative as possible.

Preparing regional and local scale population forecasts by age and sex which perform well on all these criteria is challenging. Previous research demonstrates that forecast errors increase the further into the future forecasts extend, the more volatile an area's population growth rate, and the smaller an area's population (Rayer & Smith, 2010; Tayman, 2011; Wilson et al., 2018). The preparation of forecasts for hundreds or thousands of small or local areas involves considerable amounts of work. This includes substantial (often expensive) data purchases; much data preparation, estimation, smoothing, and adjustment; complex assumption-setting; creating or running demographic models; checking of data outputs; liaising with stakeholders; and communicating forecast results to users. A multiregional cohort-component model with single year of age detail is a complex, data-intensive, and time-consuming macro-scale model (Rees et al., 2017) to implement. Dynamic spatial micro-simulation is even more challenging (Marois & Bélanger, 2015). The whole forecast production process can take many months to complete and requires much staff time.

Modelling approaches which can reduce the data inputs, time, costs, and complexity of the forecasting process offer many benefits, especially if they do not adversely affect accuracy. Hajnal's (1955, p. 321) comment is still relevant today: "[forecasting] should involve less computation and more cogitation than has generally been applied". Forecasters should spend time on analysing and understanding demographic trends, and their likely future trajectories, rather than on preparing detailed input data and calculating complex forecasts.

Several models with low data input requirements may be used to produce age-sex forecasts, including various forms of the Hamilton–Perry model (Hamilton & Perry, 1962; Wilson & Grossman, 2022a), net migration cohort-component models, and directional (inward and outward) migration cohort-component models (Smith, 1986; Smith et al., 2013; Wilson, 2016). Often these models are implemented with 5-year age groups and forecast intervals, thereby reducing the data inputs, complexity, and time required relative to single year of age models. A key advantage of a cohort-component model with directional, rather than net, migration is the simple adjustment of age-sex migration to either a total population constraint or a total net migration assumption. Inward and outward migration age profiles are relatively stable in shape and can easily be adjusted up or down, while a net migration age profile is much harder to adjust because it tends to change

shape as inward and outward migration levels change. In addition, directional migration models better represent demographic reality because they incorporate people moving in and out of areas, rather than non-existent 'net migrants'. In terms of accuracy, directional migration cohort-component models are often slightly more accurate than other types of cohort-component model, and constraining to independently forecast total populations has been shown to improve the accuracy of age-sex forecasts further (Reinhold & Thomsen, 2015; Wilson, 2016). However, even a bi-regional cohort-component model, in which internal migration is modelled between each local area and the rest of the country, still requires a considerable amount of input data and data preparation.

A projection model has been proposed which needs little input data and data preparation (Wilson, 2022). The synthetic migration population projection (SYMPOPP) model does not require the user to supply any age-sex migration data, any age-sex mortality data, or any fertility data, even though it is a cohort-component model incorporating births, deaths, and directional migration flows by age and sex. The model has been operationalised in an Excel/VBA workbook. The main data inputs are population estimates by 5-year age group and sex for the jump-off (launch) year and for 5 years earlier, local area total forecast populations, and a national forecast by age and sex which acts as a constraint. Synthetic migration flows are calculated from a model migration rate age schedule (included in the programme) and base period-cohort population change; fertility rates are estimated indirectly from the jump-off year population age structure; and mortality rates are derived from life expectancy at birth assumptions and a national mortality surface consisting of past and forecast life tables (included in the programme). Details of the methods are provided in "Methods" section. SYMPOPP has been tested on Australian local and small area populations and found to produce age-sex population forecasts with a respectable level of accuracy, and with greater accuracy than those generated by the Hamilton–Perry model (Wilson, 2022; Wilson & Grossman, 2022a). However, it has not yet been tested outside Australia, or on larger subnational populations.

The aim of this paper is to present an application of the new programme to Lower Tier Local Authority (LTLA) populations in England, focussing primarily on forecast error. There are 309 districts in the study, with a mean population of 183,000 in 2021. Retrospective population 'forecasts' were prepared from 2011 to 2021 and error measured by comparing against official 2021 mid-year population estimates (ONS, 2022), based on 2021 census data. Comparisons are also made with 2011-based subnational population projections for England published by the Office for National Statistics (ONS, 2014) as well as forecasts produced from a simple Hamilton–Perry model. The synthetic migration model, software for implementation and evaluation measures are described in the "Methods" section. The "Results" section specifies the input data and forecast assumptions used in the England case study. The "Discussion" section presents the main results, including the accuracy, usefulness, strengths, and weaknesses of the model, comments on the geography of errors and suggests further possible applications of the model in England. The "Conclusions" section summarises the contribution of the paper to the field of population forecasting.

Methods

What are Local Area and Small Area Populations?

Local areas are subnational spatial units associated with an executive authority and council of representatives. To test the SYMPOPP model, we used LTLAs, harmonised on boundaries used in the 2021 census. Fortunately, the harmonisation was a very simple process because it only required the merger of a few districts to obtain the 2021 census geography. Small areas are subnational units often but not always nested within local authorities. Within cities, small areas are small in population and territorial extent; in rural territory they have small populations but occupy large areas. The first application of the SYMPOPP model in Australia used small areas (Wilson, 2022), whereas the application in this paper uses local areas.

Overview of Cohort-Component Models for Local and Small Area Projection

Wilson (2022) proposed a new demographic projection model that simplifies the forecasting of populations for local and small areas when detailed component data are either not available or difficult to estimate. The component which causes most issues for estimation is migration, both internal migration within a country (out-migration and in-migration) and international migration to and from the rest of the world (emigration and immigration). Table 1 presents the numbers of cells in three cohort-component models used in projecting the populations of SA3 areas in Australia (Wilson, 2022) and Lower Tier Local Authorities (LTLAs) England (this paper) and the average number of migrations per array cell in three alternative cohort-component CC) models: the multiregional (MRCC), the bi-regional model (BRCC), and the synthetic migration model (SYMPOPP).

The number of migrations per cell in the MRCC model is tiny: 3.3 in the case of SA3 areas in Australia and 1.7 for LTLAs in England. Most cells are empty. Migration rates cannot be reliably estimated and are unsuitable for setting forecasting assumptions. The model can be made feasible by shrinkage so that migrations are out-flows and inflows are between each area and other areas combined. This is the Bi-Regional Cohort Component (BRCC) model (Rogers, 1976; Wilson & Bell, 2004) used in projecting local populations by ethnicity in the UK (Rees et al., 2017). An alternative approach is to estimate migration rates by age and sex using origin rates, destination rates, and an interaction term in an iterative proportional fitting model (Van Imhoff et al. 1998) or a log-linear model (Rogers et al. 2011).

The Movement Population Account and Projection Model

A population account is a consistent representation of population change and its components (births, deaths, migrations) from which the necessary demographic rates can be estimated, for use in projection. However, the format of the demographic account and derived projection model depends on the migration measures

Table 1 Flow cells and migrations per cell, for three cohort-component models, 2011–2016

Dimensions	Australia: SA3 areas	England: LTLA areas	Internal migration	International migration
Origins (O)	328	309		
Destinations (D)	328	309		
Age groups (A)	17	17		
Sexes (S)	2	2		
No. of migration flows			Formulae	
Model: MRCC	3,669,008	3,256,860	ODAS-OAS	OAS + DAS
Model: BRCC	33,456	31,518	OAS + DAS	OAS + DAS
Model: SYMPOPP	11,152	10,506	OAS	
Migrations per flow				
Model: MRCC	3.3	1.7		
Model: BRCC	363.9	180.1		
Model: SYMPOPP	1091.6	540.4		

Formulae follow the scheme used in Van Imhoff et al. (1998). For example, *ODAS* number of origins (O) × number of destinations (D) × number of age groups (A) × number of sexes (S)

MRCC Multi-Regional Cohort Component model, which includes regional mortality, fertility, and international migration components by region and internal migration between regions, by age and sex; *BRCC* Bi-Regional Cohort Component model, which includes regional mortality, fertility, and international migration components by region and internal migration for a set of regions and the rest of the country by age and sex; *SYMPOPP* Synthetic Migration Population Projection model, which includes regional mortality and fertility and total migration (international and internal) between a region and the rest of the country and rest of the world, combined

Sources Official estimated migration flows from ABS (Australia) and ONS (England)

used (Rees & Lomax, 2019). The main measures used are firstly migration as changes between location at the start and end of a time interval and secondly events of crossing an area boundary to effect a change of usual residence. Depending on author, these two types are labelled “migrants” and “migrations” (Courgeau, 1973) or “transitions” and “movements” (Rees, 1984; Rees & Willekens, 1986). The sources that collect transition data are the census or a survey, using a question about residence at a previous point in time. The sources that tabulate movement data are full or partial official registers which record all changes of residence across boundaries. Rees (1985) shows that both data types can be used in a population forecasting model if the model is consistent with the data type. The migration variables used in the SYMPOPP model, net migration, a migration turnover measure, and the profile of migration rates by age, can be used in a movement account and projection model.

In Section S2 of the Supplementary Online Materials (SOM) document, the structure of a bi-regional population account is set out. Bi-regional projection models have been shown to produce results close to the more detailed multiregional projection model (Rogers, 1976; Wilson & Bell, 2004). The bi-regional model still requires information on total migration flows to and from local or small areas, disaggregated by age and sex, which may suffer from the problems described earlier. The synthetic migration projection model is designed to be used without detailed migration flow data. The relationship between bi-regional migration and synthetic migration accounts is described

in the SOM, Section S2. Preliminary base period total migration is computed by multiplying local or small area populations-at-risk by model migration rates. A schedule of migration rates by age may be “borrowed” from national regional data or if these are not available from other similar countries (Rees, 2023). The resulting total migration for an area is split into inward and outward migration using net migration computed as a residual by subtracting natural change from population change for each area. To these, in- and out-migration estimates are added mortality rates schedules by age to estimate local or small area deaths. Estimates of mortality rates for local areas may be obtained directly, for example, if local or small area life tables have been computed or indirectly estimated from national life tables corrected to match data on deaths for local or small areas. Fertility rates by age of mother for local or small areas may be available from the national statistical agency or estimated indirectly from population age-sex structure, adjusted using more reliable national or regional fertility rates.

The Synthetic Migration Population Projection Model

The SYMPOPP model combines cohort-component modelling using an estimate of local area net migration and a schedule of migration by age to estimate outward and inward migration. The model minimises the input data needed, time, and expense. The model has a bi-regional structure: one region is the local area of interest; the other region combines all other subnational areas plus all other countries. In previous applications of a bi-regional model for UK subnational projections (Rees et al., 2011, 2017), three regions were distinguished: local area, rest of the country, and rest of the world.

The term ‘model’ covers the full set of projection computations: the cohort-component model, the simplified methods for preparing fertility, mortality, and migration inputs, and constraining results to independent forecasts of local total populations. The SYMPOPP model calculations are explained and then methods for preparing input data are described.

The SYMPOPP model includes area-to-area migration in a simple way. The model combines immigration with internal in-migration as inward migration and emigration plus internal out-migration as outward migration. The result is a bi-regional model comprising the region of interest and the rest of the world. The design of projection models incorporating migration depends on the type of migration available, either residential movement in a time interval or migrants who transition between two points in time. Each type is associated with its own account of population change which includes births and deaths and each can be used to project the population (Rees, 1984, 1985). The SYMPOPP model is based on demographic movement accounts and uses 5-year age groups and 5-year time intervals. The population accounting equation that links starting and ending population in a time interval is

$$P_{s,a+5}^i(t+5) = P_{s,a}^i(t) - D_{s,a \rightarrow a+5}^i(t, t+5) - O_{s,a \rightarrow a+5}^i(t, t+5) + I_{s,a \rightarrow a+5}^i(t, t+5) \quad (1)$$

where P stands for population, D for deaths, O for outward migration, I for inward migration, i represents a local area, s represents sex, t represents point in time, $t, t+5$ refers to the projection interval between t and 5 years after t , a stands for age

group, and $a \rightarrow a + 5$ is the change in a cohort’s age from a to $a + 5$ during the t to $t + 5$ projection interval.¹ The model is based on the period-cohort parallelogram space in a Lexis diagram (Fig. 1). The first period-cohort, which represents events that occur between birth and the first age group, has a triangular form. The final period-cohort tracks the population from age group 80 and over at the start of a 5-year time interval to age group 85 and over at the end.

Births replace the start population in Eq. (1) in the newborn infant period-cohort. Births are projected by multiplying age-specific fertility rates by age-specific female populations-at-risk, with the sex ratio at birth used to convert total births into boy and girl babies. Deaths and outward migration are projected through multiplying the area’s population-at-risk by occurrence/exposure rates. Inward migration is, however, treated as a flow because the population-at-risk (of the rest of the world) is very large and challenging to forecast. The projection equation for sex s , period-cohort $a \rightarrow a + 5$, and local area i is

$$P^i_{s,a+5}(t + 5) = \frac{\left(1 - \frac{5}{2}d^i_{s,a \rightarrow a+5} - \frac{5}{2}o^i_{s,a \rightarrow a+5}\right)}{\left(1 + \frac{5}{2}d^i_{s,a \rightarrow a+5} + \frac{5}{2}o^i_{s,a \rightarrow a+5}\right)} P^i_{s,a}(t) + \frac{1}{\left(1 + \frac{5}{2}d^i_{s,a \rightarrow a+5} + \frac{5}{2}o^i_{s,a \rightarrow a+5}\right)} I^i_{s,a \rightarrow a+5} \tag{2}$$

where d is the death rate and o is the outward migration rate. The code for implementing the model adopts a ‘deconstructed’ approach in which each component is calculated separately in an iterative computation. The code is relatively simple and transparent and allows constraints to be applied easily. Deaths and outward migration are forecast as

$$D^i_{s,a \rightarrow a+5}(t, t + 5) = d^i_{s,a \rightarrow a+5} \frac{5}{2} \left(P^i_{s,a}(t) + P^i_{s,a+5}(t + 5)\right) \tag{3}$$

$$O^i_{s,a \rightarrow a+5}(t, t + 5) = o^i_{s,a \rightarrow a+5} \frac{5}{2} \left(P^i_{s,a}(t) + P^i_{s,a+5}(t + 5)\right) \tag{4}$$

Start-of-interval populations substitute for end-of-interval populations in the initial iteration. End-of-interval populations are updated in successive iterations. This iterative approach is adopted so that initial migration flows can be adjusted to an independent total population forecast.

The total population constraint is adjusted in three steps. First, the net migration estimate needed to obtain the total population constraint is computed using the demographic accounting relationship:

$$N^i = P^i(t + 5) - P^i(t) - B^i + D^i \tag{5}$$

where N refers to total net migration, P represents total population, B stands for total births, and D indicates total deaths. Second, the inward migration which is

¹ SOM Table S3 defines all variables, subscripts, and rates used in the paper.

Age Group a to a+4	Age a				Period-Cohort a→a+5
85+	85	Last Period-Cohort	Last Period-Cohort	Last Period-Cohort	80+→85+
80+	80				75-79→80+
	20				
15-19	15				10-14→15-19
10-14	10				5-9→10-14
5-9	5	Period-Cohort	Period-Cohort	Period-Cohort	0-4→5-9
0-4	0	New-Born Period-Cohort	New-Born Period-Cohort	New-Born Period-Cohort	Birth→0-4
Time Mid-Year		2006	2011	2016	2021
			Jump Off Population		Test Population
Period		2006-2011	2011-2016	2016-2021	
		Calibration Period	Forecasting Period	Forecasting Period	

Fig. 1 Lexis diagram illustrating the concepts and notation used in SYMPOPP

consistent with the required net migration is computed as outward migration plus the required net migration:

$$I^i = O^i + N^i \tag{6}$$

Table 2 Data required to calculate the base period population accounts

Variable	Summary
Populations	Local area population estimates by sex and 5-year age group for the start and end of the base period
Births	Either recorded local area base period births by sex (if available) Or model age-specific fertility rates (e.g. national rates) to estimate births indirectly
Deaths	Either recorded base period deaths by sex and period-cohort (if available) Or a national life table for the period, or a mortality surface of ${}_nL_x$ values, to estimate deaths indirectly
Net migration	No input additional data are required (net migration by sex and period-cohort for the base period is calculated as an accounting residual)
Outward and inward migration	Synthetic base period outward and inward migration flows by sex and period-cohort are estimated from three data inputs: (i) model migration rates by sex and period-cohort, (ii) a crude migration turnover rate (inward + outward migration divided by the total population-at-risk) which can be area-specific or universal, and (iii) local area net migration by sex and period-cohort (just calculated)

Third, inward migration by sex and period-cohort is multiplied by the ratio of the required inward migration to the sum of preliminary inward migration over sex and period-cohort:

$$I_{s,a \rightarrow a+5}^i [2] = I_{s,a \rightarrow a+5}^i \frac{I^i}{\sum_s \sum_a I_{s,a \rightarrow a+5}^i}, \quad (7)$$

where [2] refers to an adjusted value. The iterative computations continue, updating the end-of-interval populations using Eq. (1) and forecast births, deaths, and migrations until convergence is achieved.

Both outward and inward migration could be adjusted to fit the net migration value needed. However, if net migration assumptions incorporated in the total population forecasts were substantially lower than in the base period, this would result in substantial increases in outward migration and large reductions in inward migration. For local areas, outward migration rates in some ages are very high. This approach risks increasing outward migration and might result in negative populations. For these reasons, only inward migration is adjusted.

Data Preparation for the Base Period Population Accounts

The outward migration rates and preliminary inward migration flows for the forecasts are derived from population accounts for the base period which is the 5-year period leading to the launch point of the forecasts. The data preparation described here is consistent with population accounting principles (Rees & Wilson, 1977; Rees & Willekens, 1986 (illustrated in SOM Tables S1 to S5) and with the data adjustment approach of Simpson and Snowling (2011). Data inputs needed to assemble the population accounts are listed in Table 2. At a minimum, local fertility, mortality, or migration data are not required. Local area population

estimates by sex and 5-year age group for the launch year and 5 years earlier are needed, together with national fertility, mortality, and model migration rates. Wilson (2022) tested the SYMPOPP model for 328 SA3 areas in Australia. In the "Results" section, we describe the data preparation for a further test of the model using 309 local authority areas in England, which have, on average, about twice the population of the SA3 areas.

Error Measures for Assessing the SYMPOPP Forecasts

The Absolute Percentage Errors (APE) for total populations and age-sex group forecasts were measured using

$$APE = \frac{|F - T|}{T} 100 \quad (8)$$

where F represents the forecast population and T represents the estimate of the population for the mid-point of the test year. We report median, 90th percentile, and mean values of APE.

A single alternative absolute percentage error measure for total population was used to summarise forecast errors across age-sex groups for a local population. The conventional APE for total population fails to assess the quality of a forecast by age and sex. The APE could be close to zero, suggesting an accurate forecast, when there are offsetting age-sex errors. For example, there could be substantial over-forecasts of younger ages and severe under-forecasts of older ages. To bypass this effect, an alternative measure, Age Structure Error (ASE), was computed by summing absolute errors for each age-sex group and then dividing by the total population:

$$ASE = \frac{\sum_s \sum_a |F_{s,a} - T_{s,a}|}{T} 100 \quad (9)$$

In graphic terms, the numerator is the absolute difference in area between the forecast population pyramid and the actual population pyramid. ASE is effectively a weighted Mean Absolute Percentage Error and gives results equivalent to the Mean Absolute Deviation/mean error measure designed by Kolassa and Schutz (2007). However, the ASE is calculated differently to accommodate age-sex populations of zero, which occur in small populations. When all age-sex forecast errors take the same sign (all positive or all negative), the alternative ASE will be the same as the conventional APE for the whole population. The ASE is equivalent to the mean APE for all sex-age groups, when the weights for averaging are derived from the Test population (see SOM, Section S7). If there is a mix of over- and under-forecasts across age-sex groups, then ASE will be greater, indicating a less successful forecast of the age-sex structure. Other ways of measuring age-sex structure error could have been used, such as a simple mean of all age-sex-specific Absolute Percentage Errors. We chose the ASE to retain consistency with the APEs for total populations. ASEs are used in the "Discussion" section to assess the accuracy of local area forecasts.

A Model for Forecasting the Total Population of Local Areas

A vital ingredient in the SYMPOPP model is a forecast of the total population of all local or small areas, which acts as a top-down constraint on the detailed sex and age forecast. We chose a Linear/Exponential model which applies linear extrapolation if base period growth is positive and an exponential model where base period growth is negative. This is to avoid forecasting negative populations. The simple linear (LIN) model is

$$P_i(t + 5) = P_i(t) + G_i, \quad (10)$$

where $P_i(t)$ is the launch year population for small area i , G_i is the growth in population over the base period, and $P_i(t + 5)$ is the forecast population 5 years ahead. For growing areas, the LIN model forecasts a gradually declining growth rate, avoiding runaway growth. However, for rapidly declining areas, it can result in negative projected populations. When this occurs, an exponential (EXP) model is used instead:

$$P_i(t + 5) = P_i(t)e^{r_i}, \quad (11)$$

where e^{r_i} is e (base of natural logarithms) raised to the power r_i , which is the rate of growth over the 5-year period for local area i . The exponential function reduces growth to lower and lower levels but never becomes negative. The extrapolative forecasts of total local area populations were constrained to the forecasts of the 2010-base national population projection for England.

Data for the England Case Study

The input data described here refer to the components needed to fill the bi-regional account for local areas in England from which the demographic rates are computed. These are assumed to be constant at the start of the forecast but as they are controlled to the independent forecast of total population, the rates will be adjusted.

To prepare model inputs of mortality, fertility, and migration rates, we use open access data from official data repositories and web sites. More precise inputs could have been obtained by commissioning tables from the Office for National Statistics, but this option was not pursued because the SYMPOPP model does not require detailed data inputs. Table 3 lists the data from open access sources we used to prepare model inputs.

Populations

The Office for National Statistics (ONS) has prepared local area (LTLA) population estimates by single year of age for mid-years from 2001 to 2021. SOM, Section S6 provides details of how these were reconciled with census populations and converted to a standard geography. For this case study, population estimates for 2006, 2011, and 2021 by five-year age groups are selected. The 2006 and 2011 populations

Table 3 Data and sources used in the England Case Study

Variable	Description
Attributes	Geography: 309 Lower Tier Local Authorities (LTLAs) for 2021 Years for populations: 2006, 2011 (calibration), 2011, 2021 (forecasting) Periods for components: 2006–2011 (calibration), 2011–2021 (forecasting) Sex: Males, Females, Age groups: 0–4, 5–9, 10–14, ..., 80–84, 85+
Populations	Local authorities: Population estimates by sex and 5-year age groups Base period: 2006, 2011, Forecast period: 2011, 2021 Source: Nomis (2023a), ONS (2023b)
Births	Base period: 2006–2011, Forecasting period: 2011–2021 Basic model: TFRs and ASFRs are estimated indirectly Refined model: Local authorities, total live births, total fertility rates England, Age-specific fertility rates, ages 15–44, sex ratio for new-borns Source: ONS (2023a)
Deaths	Base period: 2006–2011, Forecasting period: 2011–2021 Basic model: England life expectancies for all local authorities Refined model: Local authorities, Total deaths, Life expectancies Source: ONS (2023c, 2023d) Basic and Refined models: Model life table surfaces (stationary population variable), Source Wilson (2018, 2020)
Net migration	Base period: 2006–2011, Forecasting period: 2016–2021 Source: Computed as a residual by the SYMPOPP model
Out-migration	Base period: 2006–2011, Forecasting period: 2011–2021 Definition: Internal out-migration plus international emigration Basic model and Refined model (three option used): <ul style="list-style-type: none"> • SA3 Migration age schedule from Wilson (2022) • England, migration age schedules computed from 2011 census migration and population data for England from Nomis (2023b) and migration indicators from ONS (2021) with Sex differences in migration rates borrowed from Wilson (2022) • Basic standard migration schedule from Rogers and Castro (1981)
In-migration	Base period: 2006–2011, Forecasting period: 2011–2021 Definition: Internal in-migration plus international immigration Computed by the SYMPOPP subtracting out-migration from net migration

are used to calibrate the base period account, while the 2021 population serves as check on the accuracy of our projections from the jump-off year 2011 to the census year 2021.

ONS uses a cohort-component model drawing on vital statistics data from the registers of births and deaths. Estimates of internal migration between local areas are drawn from the NHS patient database, which records patient address changes. International migration is estimated from the migration information in the International Passenger Survey (IPS) and administrative records of entries and exits maintained by the Home Office, together with models for estimating international migration to and from local areas. In 2020, the IPS was discontinued because of the Covid-19 pandemic and data on visas and other administrative sources were substituted. ONS has recently produced mid-year population estimates based largely on administrative data sources for 2011 to 2022 (ONS, 2023e). The key feature of these new estimates is the use of administrative data for international migration, replacing

the IPS survey. However, these have not yet been classified as official statistics, so the estimates produced by the demographic “roll forward” method are used in this paper.

The ONS mid-year population estimate methodology makes separate estimates for sub-populations, such as students, military personnel, and prisoners, using administrative data. The 2021 population estimates were produced using results for the 2021 Census (March 21) “rolled forward” to 30 June, using components of change for the intervening period of three months and ten days. ONS re-estimated previous populations and components of change using the LTLA geography of the 2021 Census.

Fertility Rates

For both the Basic and Refined forecasts, we used the xTFR method of Hauer and Schmertmann (2020), which indirectly estimates base period total fertility rates from the age structure of the population, as described in Wilson (2022). Base period local area fertility rates were calculated by constraining model age-specific fertility rates to estimated local TFRs:

$$f_{+,a}^i = f_{+,a}^{model} \{TFR^i / TFR^{model}\} \quad (12)$$

Births are allocated to male and female infants using the sex ratio at birth for England.

Mortality Rates

In the Basic projection, we assume that all areas have the same life expectancy (at birth) as England as a whole and derive mortality rates using one vector of life table stationary populations. In the Refined projection, we employ ONS estimates of base period local area life expectancies and link local life expectancy projections to those at the national level by assuming local-national life expectancy differences remain constant. Age-specific death rates are calculated by selecting a vector of stationary populations corresponding to the life expectancy of the local area. The data used to estimate mortality rates for local areas in England are as follows: total deaths for each year 2011–2021, published local area life expectancies at birth, and a library of stationary populations by age and sex based on a library of modelled life tables that enable life expectancies to be associated with period-cohort mortality rates (see Table 3 for sources). A detailed description is given in Wilson (2018).

The first step is to average the local life expectancies published by ONS over the 2006–2011 base period. The second step is to select the modelled life table that matches local life expectancy from the ${}_nL_x$ mortality surface embedded in SYM-POPP. The resource provides the stationary populations² by age subject to the

² Equivalent to the life years lived between exact ages.

mortality rates of the model life table which has life expectancies closest to those reported for local area i . Period-cohort mortality rates are defined as

$$d_{s,a \rightarrow a+5}^i = \frac{L_{s,a}^{k(i,s)} - L_{s,a+5}^{k(i,s)}}{\frac{5}{2} \left(L_{s,a}^{k(i,s)} + L_{s,a+5}^{k(i,s)} \right)} \quad (13)$$

where $k(i,s)$ refers to the model life table interpolated to exactly match to the life expectancy of local area i and sex s .

Outward Migration Rates

Because the SYMPOPP model uses a measure of migration turnover and migration by age profiles, they are less affected by the distinction between the types of migration (transition or movement) than would be conventional multi- or bi-regional cohort survival models, where it is necessary to match type of migration measure with the format of the projection model (Rees, 1985). Schedules of migration rates by age and sex are needed for the SYMPOPP model.

These are referred to as either migration age schedules (MAS) or model migration schedules (MMS) (Rees, 2023). MAS are based on tables of migration classified by age and matching tables of population-at-risk. For areas with large populations, the migration flow data produce reasonably accurate rates. The migration tables for areas with small populations have zero entries or very small values. In those situations, the out-migration tables for local or small areas can be summed or a national table can be used (Wilson, 2022).

Alternatively, rather than use the “raw” rate numbers directly, a model is fitted to the empirical schedule. The results are model migration schedules (MMS). Where no empirical data are available on migration by age, MMS can be borrowed from other studies such as Wilson (2022) for Australian SA3 areas or for inter-regional migration. Rogers and Castro (1981, p. 42) derive basic standard MMS parameters by averaging the inter-regional migration parameters in 17 countries that collaborated in the IIASA Migration and Settlement Study, using their simplest model. This has a childhood slope, a labour/higher education peak with a steep rise to the peak followed by decline to older ages. Intermediate between MAS and MMS are schedules of empirical rates smoothed using a suitable procedure (Bernard & Bell, 2015).

Figure 2 plots the MAS used for LTLAs in England. This MAS uses 1-year migrant transition data from the 2011 UK census combined with migration (movement) data from the NHS register for mid-2011 to mid-2016. The method, explained in SOM, Section S5, converts the MAS from a transition to a movement basis, using the assumption that migrants make just one migration per year. Figure 3 plots the census-based MMS for SA3 areas in Australia (Wilson, 2022), which are used in the England case study to assess the gains from using “native” data rather than “borrowed” data. This MMS uses retrospective 5-year migration data (transitions) from the 2016 Australian census. Figure 4 uses the basic standard parameters of Rogers and Castro that can be borrowed in the absence of data

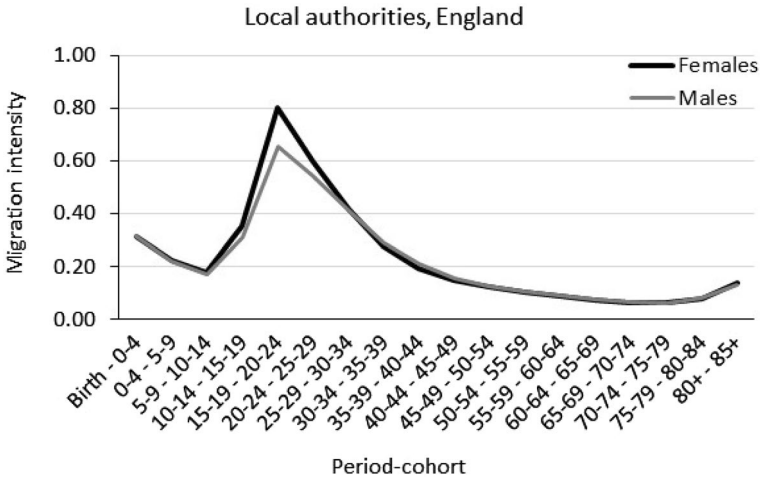


Fig. 2 Migration between Lower Tier Local Authorities, England, 2011–2016. *Source* Computed by the authors using migration tables from the 2011 Census and associated with the migration components used in the ONS 2011–2016 mid-year population estimates for local authorities in England

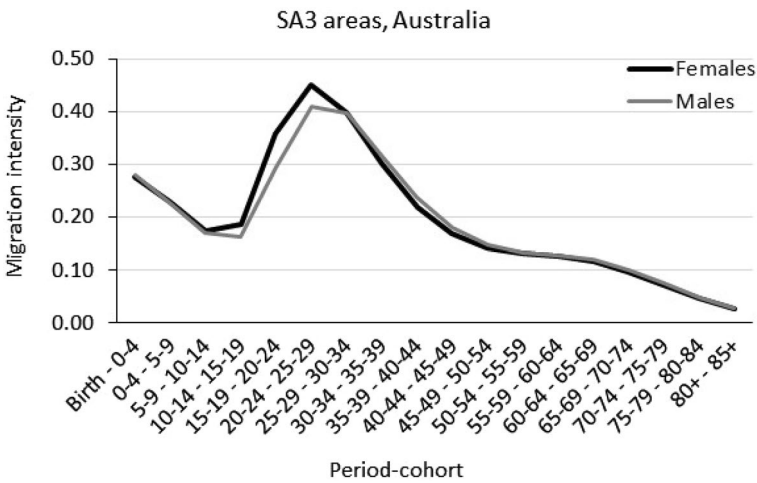


Fig. 3 Migration between SA3 areas, Australia, 2011–2016. *Source* Computed by the authors from migration data from the 2016 Census (Wilson, 2022)

for a country of interest. The sex-age-specific migration rates associated with the alternative MAS/MMS schedules are provided in the SOM Section S4.

Although the shapes of the three plots are very similar, the levels of inter-area migration are different reflecting the combination of migration measure differences and scale effects. Figure 2 uses estimated 5-year movement data over a 5-year period for local areas. Figure 3 uses transition data for a 5-year interval for

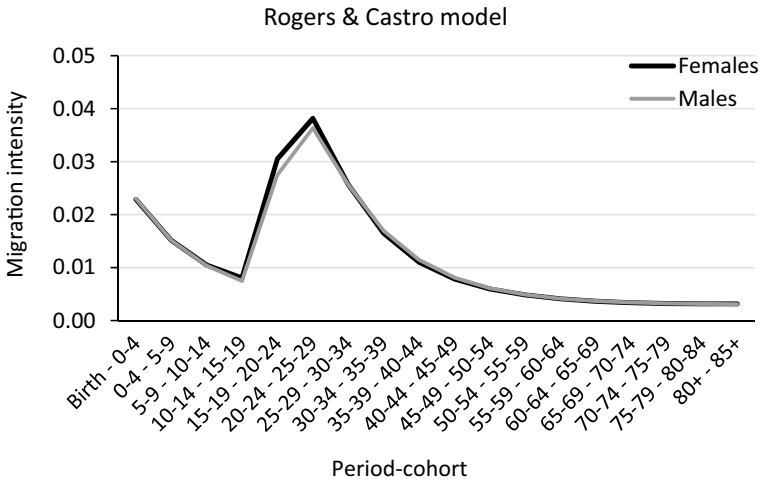


Fig. 4 Basic standard model migration schedule, for migration between regions in 17 countries. *Source* Computed by the authors using the model developed by Rogers and Castro (1981) and the average parameters reported in Table 17, p.42

Table 4 Alternative population forecasts for local authorities in England

Projection	Description	Migration age schedule variants
Basic	SYMPOPP model using component estimation methods in Wilson (2022) for SA3 areas in Australia	Australia MMS (SA3 areas)
		England MAS (LTLAs)
		Rogers-Castro MMS (Regions)
Refined	SYMPOPP model using component estimation method described in "Results" section of the paper	Australia MMS (SA3 areas)
		England MAS (LTLAs)
		Rogers-Castro MMS (Regions)
ONS	Multi-Regional Cohort Component model described in ONS (2016)	Migration Rates by Age, averaged over 5-year period (LTLAs)

Sources Basic, from Wilson (2022); Refined, from Section "Results"; ONS, from ONS (2014, 2016)

small areas. Figure 4 uses a mixture of transition and movement data from both censuses and registers for large subnational regions.

Results

The Alternative Forecasts

The key features of the alternative forecasts are presented in Table 4. Two sets of inputs are used: the Basic set and the Refined set. To test how well the SYMPOPP model performs in forecasting populations for local areas, we compare outcomes

Table 5 Local authority total population forecast errors, 2021

Error measure	Linear/exponential (%)	ONS projections (%)
25th percentile APE	1.3	1.3
MedAPE	2.8	3.1
75th Percentile APE	4.7	5.1
MAPE	3.4	3.8
% areas < 5% APE	78.3	72.5
% areas 5–10% APE	19.1	22.7
% areas > 10% APE	2.6	4.9

Source Authors' calculations

against the June 30th 2021 mid-year population estimates of local areas, which are closely linked to the March 27th, 2021 Census enumerated populations. To the two SYMPOPP projections, we add a third, the subnational projections for England produced by ONS (2014, 2016), which use a multiregional cohort-component model, in which the problem of sparsity in the inter-region migration matrix is partially overcome by pooling rates from five years of data.

A further set of forecasts is listed in the rightmost column of Table 4, in which the three MAS/MMS estimates are used as migration variants in the Basic and Refined forecasts. The research question is whether the choice of MAS/MMS makes a difference.

Constraining Population Forecasts

Although the main focus of the study was the errors in local authority population forecasts by age and sex, these forecasts are affected by errors in the local authority total population forecast and national population constraints, so it is useful to examine the level of errors. The national projection of the total population of England (after applying a jump-off year adjustment to the 2010-based projections) was 1.7% too high by 2021. Age-specific population forecasts were higher at the youngest and oldest ages, and a mixture of higher and lower in the young and middle adult ages. This was due to fertility declining below the assumed level and life expectancy not increasing as much as anticipated. As a result of these offsetting errors, the national Age Structure Error was fairly high at 4.7%.

Errors in local authority total population forecasts produced by the Linear/ Exponential model for 2021 are summarised in Table 5. The table includes a comparison with total population forecast errors from the ONS 2010-based projections for 2021. Average errors, as measured by the Median Absolute Percentage Error (MedAPE) and Mean Absolute Percentage Error (MAPE), were fairly similar for the Linear/ Exponential and ONS forecasts, with marginally lower errors obtained from the Linear/Exponential model. Interestingly, the errors of the two sets of local authority total population forecasts are only moderately positively correlated ($r=0.54$).

Table 6 Local authority population forecast Age Structure Errors, 2021

Error measure	Synthetic model: basic (%)	Synthetic model: refined (%)	ONS projections (%)	Hamilton–Perry (%)
25th percentile Age Structure Error	4.9	4.8	5.0	5.2
Median Age Structure Error	5.9	5.8	6.2	6.3
75th percentile Age Structure Error	8.0	7.6	7.8	8.2
Mean Age Structure Error	6.8	6.8	7.3	7.3
% areas < 5% Age Structure Error	28.2	32.4	24.9	19.7
% areas 5–10% Age Structure Error	59.5	55.3	60.5	67.6
% areas > 10% Age Structure Error	12.3	12.3	14.6	12.6

See Section "[Methods](#)" for the definition of the Age Structure Error

Source Authors' calculations

If errors below 5% are regarded as good quality forecasts, then most local authority populations were forecast well. The Linear/Exponential model produced forecasts less than 5% error for 78.3% of local authorities, while the ONS projections did so for 72.5%. If errors of 10% or more reflect a poor-quality forecast, then only 8 local authorities (2.6%) were forecast poorly by the Linear/Exponential model and only 15 (4.9%) by the ONS projections. London Borough populations proved particularly difficult to forecast well, which is in part due to recent population declines related to COVID and Brexit (GLA, 2022). For the Linear/Exponential model, 5 of the 8 poor forecasts were in London, while for the ONS projections, 10 out of 15 poor forecasts occurred in London.

Local Authority Forecasts of Populations by Age and Sex

The Age Structure Error summarises how well the age-sex structures of local authority populations were forecast (Table 6). For the Basic set of forecasts created by the synthetic migration model, the median Age Structure Error for 2021 was 5.9%, while for the Refined set of forecasts, it was only marginally better at 5.8%. Age Structure Errors across individual districts were also similar in the Basic and Refined sets of synthetic model forecasts. Nonetheless, the use of area-specific migration turnover and life expectancy assumptions in the Refined forecasts resulted in a modest increase in the share of local authorities with good quality age-sex forecasts (where the Age Structure Error was under 5%) from 28.2% to 32.4%. As would be expected, there is a positive association between the errors of the Linear/Exponential forecasts of total population and error in forecasting the population age-sex structure. The correlation between the total population APE and Age Structure Error was $r=0.72$ for both Basic and Refined forecasts. As Table 6 shows, Age Structure Errors were marginally higher in the ONS projections and Hamilton–Perry forecasts.

While the Age Structure Error provides a single number measure of age-sex structure errors, it is also useful to examine the age patterns of error given that many uses of population forecasts are limited to specific age groups. Median APEs for male

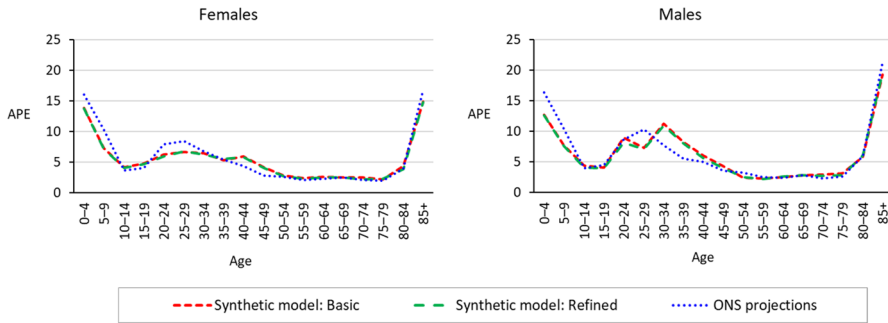


Fig. 5 Median APes of age-specific population forecasts, 2021. *Source* Authors' calculations

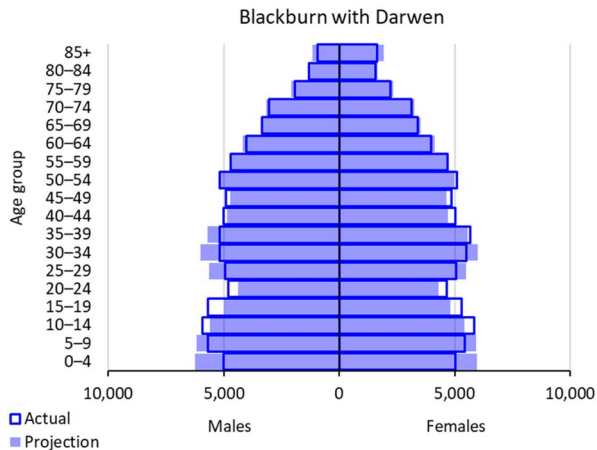
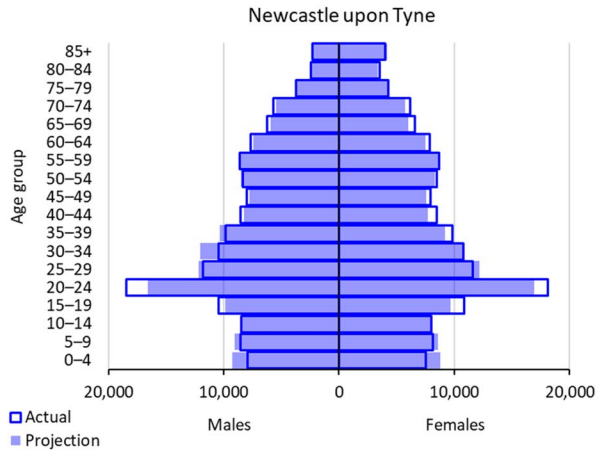
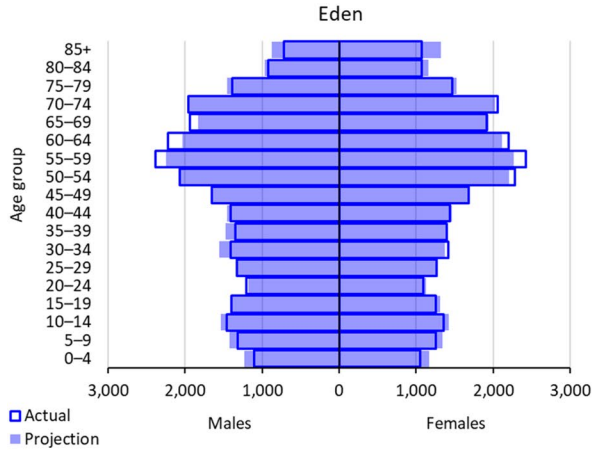
and female age-specific population forecasts for 2021 are shown in Fig. 5. Errors are highest at the youngest and oldest ages, a finding in common with other studies (e.g. Smith & Tayman, 2003), and moderately high at the peak migration age groups in the young adult ages. Births and young infant populations suffer large forecast errors due to difficulties in forecasting fertility rates and the size of the childbearing age population. In the young adult ages, the volatility of migration is generally a major contributor to population forecast errors. At the oldest ages, populations experience more noise in demographic rates due to their smaller size. Differences between the median errors of two sets of synthetic model forecasts are tiny.

The ONS projections experienced slightly higher errors than the synthetic model forecasts in the young childhood ages, higher errors in some young adult ages, but lower errors in the 40–49 age range for females and 30–49 for males. At middle and older ages, the synthetic model and ONS projections had very similar median errors. Differences in median errors between males and females are most noticeable in the 30–39 age range and at age 85+. The age and sex pattern of error from the Hamilton–Perry model (not shown) were very similar to that of the synthetic migration model, although with marginally higher errors at some ages.

To give a visual representation of the magnitude of forecast errors by age and sex, Fig. 6 shows the forecast and actual populations in 2021 for three local authority districts. These examples have Age Structure Errors at the 25th percentile, median, and 75th percentile of the error distribution from the Refined synthetic model forecasts. The differences in error are not easy to discern from the graphs due to the narrow distribution of Age Structure Error. The Age Structure Error is 4.8% for Eden (25th percentile), 5.8% for Newcastle upon Tyne (median), and 7.6% for Blackburn with Darwen (75th percentile).

How do the errors vary across local authorities? Fig. 7 maps the percentage Age Structure Error using three classes: low errors are less than 5%, medium errors range from 5 to 10%, and high errors are those above 10%. Table 6 indicates that the Refined SYMPOPP model run generates one third in the lowest category, half in the middle category, and one eighth in the high category. We focus here on those local authorities with the highest error, which are named on the map. The LTLAs fall into four types: (1) those known to house large military bases, (2) those where

Fig. 6 Population forecasts by age and sex of three local authority districts, 2021, created by SYMPOPP. *Note* These forecasts experienced Age Structure Errors at the 25th percentile (top graph) median (middle), and 75th percentile (bottom graph) of the error distribution. *Source* Authors' calculations using the Refined forecasts from SYMPOPP



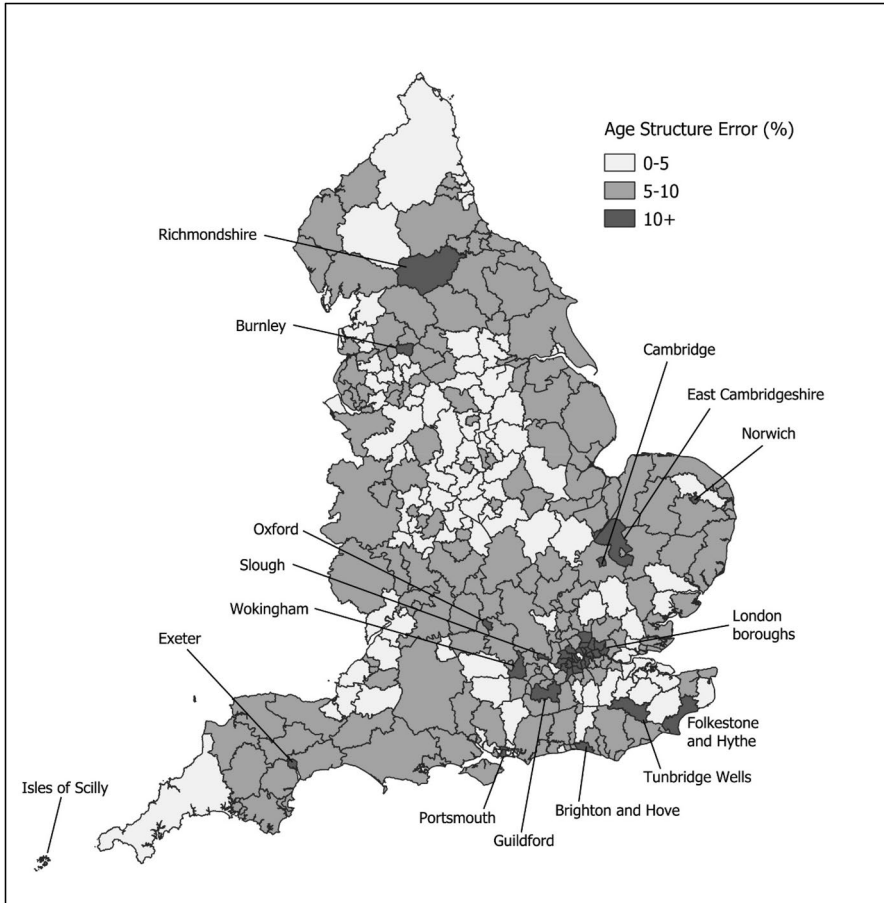


Fig. 7 Map of Age Structure Errors for local authorities in England, 2021 projected populations. *Source* Data generated by SYMPOP, boundaries from ONS Geography and map produced using QGIS

university students make up an important share of the total population, (3) those with ethnically diverse populations which tend to be younger because of high fertility rates and past immigration waves that increased the fertile age population, and (4) a residual category.

LTLAs with large military populations include the following. Richmondshire is the location for the British Army's largest army base. East Cambridgeshire is where both the Royal Air Force and United States Airforce have bases. Gosport is closely linked with His Majesty's Naval Base (HMNB) Portsmouth. Wokingham is home to the Royal Electrical and Mechanical Engineers (REME), a unit of the British Army.

Estimating and forecasting military populations in local authorities are difficult because recruitment is sex-selective (more men than women) and retirement happens at earlier ages than in the general population. Precise statistics are highly sensitive and are released to ONS under strict confidentiality conditions. ONS (2023e)

identifies errors for Richmondshire in the Admin-Based Population Estimates (ABPE) for 2021 and 2022 equivalent to those from the SYMPOPP model. Recruitment to and exit from the military are determined by difficult-to-forecast Ministry of Defence budgets and plans.

Other local authorities with high errors include medium-sized cities with elite universities. These include Cambridge, Oxford, Exeter, Brighton, Norwich, and Guildford. Student populations are difficult to monitor because they are in transition between domicile with their parents and at their university. In censuses, their usual residence can be recorded at either location. Between censuses, the main instrument for monitoring internal migration in the UK is the demographic spine of the National Health Register. Students are usually required to register with the student health practice associated with their university. However, when they migrate away after completing their studies, they are slow at re-registering, delaying until treatment for ill-health is needed. Both ONS and academic researchers use data from the Universities and Colleges Admissions Service (UCAS, <https://www.ucas.com/>) to help estimate migrations on entry to higher education. Information on exits is more problematic and rule-based assumptions are used (Wu et al., 2008). These difficulties in measuring the internal migration of domestic students also affect international migration by overseas students.

The map also places a majority of London Boroughs (LBs) in the high error category. LBs may have military and student populations, and immigrant-origin populations (identified as ethnic groups in UK censuses) with a demographic structure and dynamics different from the host population. Their populations are refreshed by new immigrations which have not declined since the Brexit referendum vote but rather changed origins, with non-EU citizens replacing EU citizens who have emigrated. The fertility levels of non-EU immigrants from South Asia and Africa have been higher than the national average, though diminishing through the generations (Harrison et al., 2023). Some local authorities outside London, such as Burnley and Slough, have also received major streams of South Asian migrants. Immigrant status also affects internal migration behaviour with South Asian families preferring their daughters not to migrate from home on entry to tertiary education (Bagguley & Hussain, 2007).

Further Communal Establishment populations whose demographic behaviour diverges from that captured in the SYMPOPP model include Prison Populations, People living in Sheltered Accommodation, the residents of Old People and Nursing Homes, and retirement migrants. The SYMPOPP model was designed to be simple, parsimonious in its data demands, and for rapid implementation, so the special dynamics of these institutional populations are not well represented.

Births, Deaths, and Net Migration

Errors in forecasting births, deaths, and net migration over the 2011–2021 period are expressed as the APE in total population which *would* have occurred if only the component in question was forecast, while the other two components were observed (or perfectly forecast) values. This is a simple approach which assumes independence between

Table 7 Contribution of component errors to total population APE, 2021

Median total population APE with	Synthetic model: basic (%)	Synthetic model: refined (%)	ONS pro- jections (%)
Forecast births	1.2	1.2	1.3
Forecast deaths	0.7	0.7	0.7
Forecast net migration	2.4	2.4	3.6

Source Authors' calculations

Table 8 Age Structure Errors for migration age structure variants

Migration age schedule	Median Age Structure Error (%)	
	Basic version	Refined version
Rogers-Castro	5.96	5.89
Australia	5.92	5.84
England	5.85	5.84

births, deaths, and net migration, but it has the advantage of being easily interpretable: it is the error in total population contributed by one component of change. Note, however, that there are often offsetting component errors (e.g. forecast births too high and net migration too low), so the three component APEs often sum to more than the APE calculated directly for total population.

Table 7 presents median APEs of total district populations from births, deaths, and net migration being forecast inaccurately. If only births had been forecast inaccurately, the median total district population APE in 2021 would have been 1.2%. If only deaths had been forecast inaccurately, it would have been 0.7%. For net migration, the median error was 2.4%, indicating a greater contribution to error from this component of change. Equivalent errors in the ONS projections are similar for deaths and a little higher for births and net migration.

Results for Migration Age Schedule Variants

To test alternative specifications of the migration–age relationship, six further forecasts using SYMPOPP were run. The median age structures errors are reported in Table 8. As before, the errors were reduced a little in the Refined version compared with the Basic. Comparing MAS specifications, the Rogers-Castro standard MMS performed worse than the Australian MMS or England MAS profiles. The England specification performed better than the borrowed Australian profile in the Basic version of SYMPOPP, though the errors were equivalent in the Refined version.

Discussion

The aim of the paper was to test how well the SYMPOPP model, with its minimum data requirements, performed in forecasting the populations of local authorities in England over one ten-year forecast period (2011–2021) using the demographic rates established in the calibration period (2006–2011). The 2021 forecast populations of alternative model specifications were compared with the Census-based mid-2021 population estimates. Did the forecasts and the estimates match? The answer was mostly yes. The SYMPOPP Refined forecasts had the highest percentage of low errors and combined low and middle errors, although the Basic forecasts had the same percentage of high errors. Both forecasts from the SYMPOPP model performed a better than the 2012-based ONS Sub-National Population Projection.

Extension to More Time Periods

These findings might have been more robust if we had evaluated the forecasts over more time periods. For example, 2001–2006 could have been used as a calibration period with 2006–2021 as the forecast horizon. An additional improvement would have been to use a mid-2016 set of estimates reconciled with the 2021 population outcomes. ONS (2023f) have analysed the differences between the rolled forward local population estimates not using 2021 Census data against the 2021 estimates using census data. However, a re-based set of mid-year population estimates for 2012 through 2020 has yet to be produced. Future work could include a comparison of SYMPOPP results against the revised estimates series as well as the ONS forecasts.

Extension to Single Years

The SYMPOPP model is based on five-year intervals for time and age. However, many local authorities have school age populations that do not fit to quinquennial age groups. A useful addition to the model would be a routine to interpolate single of age and single year of time populations.

Future Mortality Assumptions

Forecast changes in mortality rates might be introduced after an analysis of trends rather assuming that those estimated in the calibration period continue. For England and its local authorities, key issues are whether life expectancies will improve in the future and whether the gap between best performing local authorities and worst will narrow. In the 2011 to 2021 period, life expectancy did not improve as expected in the long run assumption used in official forecasts (The Economist, 2023). In forecasts of future mortality assumptions in England, ONS (2023g) employed a model (Dodd et al., 2021) which gives too low a weight to recent experience, predicting a return to steadily improving life expectancy. Given the UK's experience of continuing excess deaths in 2022 and 2023 that seems unlikely. The UK economy is

forecast to perform poorly in the 2023–2028 period, because of the effects of Brexit. (OBR, 2023; Singh, 2023). This will result in a shortfall in National Health Service funding meaning little chance of life expectancy recovery to pre-2010 trends.

Comments on the Geography of Errors

In the "Results" section, we commented on reasons for high errors in some local authorities. The rigour of these interpretations could be tested through a regression analysis of errors against variables capturing the special population effects, other socioeconomic factors, and official classifications of local authorities.

Opportunities for Using the SYMPOPP Model

The model has been tested using local areas responsible for administration. A first opportunity would be to develop sub-local authority forecasts. Extensive use was made of small areas, within local authorities of middle (level) super output areas (MSOAs) during the Covid-19 pandemic (2020–2022) to report cases of infection, deaths, and vaccinations. Many analyses of the pandemic at small area scale relied on 2011 Census data, which were radically out of date, particularly for fast growing ethnic minority groups. Small area forecasts often rely on published housing plans for forecasting new households. Unfortunately, these plans provide poor forecasts of actual numbers of houses built, being usually under-fulfilled. A recent test of SYMPOPP on small areas with populations of just a few thousand in Tasmania proved successful, suggesting that application to small areas in other countries is likely to work well (Wilson & Grossman, 2022b). A second opportunity would be to extend the forecasts to sub-populations such as ethnicity. Rees et al. (2017) carried out 2011-based ethnic population projections for local government areas in the UK which were complex and detailed. The model used 12 ethnic groups and involved modelling of components for which ethnic information was sparse. This would be useful preparation for the design and execution of a new 2021-based round of ethnic group projections. A third opportunity would be to extend tests of the SYMPOPP model to other countries, particularly those with accurate population registers.

Conclusions

The key finding of the paper is that a simplified population forecasting model for small populations developed in Australia proved generally effective when applied to local authority districts in England. The simplest version in terms of data requirements worked well and a refined version did only marginally better. Both versions of SYMPOPP performed slightly better than the official multistate cohort-component model. In terms of total population forecasts, 78% of local areas were forecast within 5% error after 10 years, with the median error being 2.8%. For age-sex-specific populations, the Refined version of SYMPOPP generated forecasts 10 years ahead with Age Structure Errors below 5% for 32% of areas and below 10% for 88% of areas,

with the median error being 5.8%. The 'success' or otherwise of these levels of error depends on the exact uses of the forecasts. However, focus group discussions suggest that many users of population forecasts regard errors of 3–5% as acceptable for populations of large areas, while 5–10% error is viewed as reasonable for small populations (Wilson & Shalley, 2019).

Compared to other local area population forecast studies, the SYMPOPP forecast errors for English local authorities are reasonably competitive. For example, in the original application of the model to SA3 areas in Australia, median APEs for total population were 4–6% after 10 years, while the Age Structure Errors were 7–10%. In forecasting the populations of US counties with various implementations of the Hamilton–Perry model, Hauer (2019) obtained Median APEs for total populations of about 5% for 10 years ahead, with median APEs for age-specific population forecasts being about 8%. Tayman et al. (2021) produced forecasts for US counties 10 years ahead with a modified Hamilton–Perry approach. They applied total population constraints from extrapolative models and adjusted county Cohort Change Ratios in line with those from a population forecast for states. The Mean Absolute Percentage Errors for total county populations were 6%, while the Mean Absolute Percentage Error across 10-year age groups was about 8%. Of course, the comparisons of errors here are not perfect due to variations in population sizes and characteristics between studies, but they at least give an approximate indication of forecast performance.

However, there is still room for improvement. There were local areas where errors of more than 10% in the Age Structure Error occurred, mainly because of the concentration of special populations linked to institutions. The UK's Office of National Statistics has long recognised the same issues affecting preparation of population estimates between census years. Novel attempts to base those estimates on a variety of administrative sources have not escaped the same difficulties (ONS, 2023e). Perhaps populations living in communal establishments need new methods for forecasting because despite their small size (circa 2% of the total population), they have impact in local areas where they constitute a larger share of the population.

In the "Discussion" section, we made suggestions about further applications of the SYMPOPP model for areas within local authorities where small numbers mean direct measurement of demographic rates is unreliable. The model may also be useful when forecasting sub-populations such as immigrant or ethnic groups when accurate statistics on component rates at local scale are not available. The authors would also welcome applications to countries beyond the two so far implemented. The usefulness of this work is that it reduces the burden on local areas without skilled demographers to run their own forecasts. Instead, the SYMPOPP code, built using Excel/VBA, could be employed.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11113-023-09839-2>.

Acknowledgements The authors are grateful to Gareth Rees (Undo Ltd, Cambridge) for help in producing the map of local Age Structure Errors and Stephen Clark (University of Leeds) for converting the map into image format. Stephen Clark and Irina Grossman (University of Melbourne) read a draft of the paper and provided helpful comments.

Author Contributions PR and TW contributed equally to the research.

Funding This research was supported by the Australian Research Council Discovery Projects scheme (Project DP200101480).

Data and Code Availability Projection input and results data with SYMPOPP code in Excel Workbooks, and an Excel Workbook on Estimation of Migration Rates by Age are available from the authors. They will be deposited on GitHub. A copy of SYMPOPP with illustrative data can be downloaded from <https://doi.org/10.6084/m9.figshare.19372784.v1>.

Declarations

Conflict of interest The authors declare that they have no conflicts of interest.

Ethical Approval None was sought. The data are all aggregate (non-disclosive) populations provided by Statistical Agencies under Open Data Licences.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bagguley, P. & Hussain, Y. (2007). The role of higher education in providing opportunities for South Asian women. The Policy Press (for the Joseph Rowntree Foundation). Retrieved 16 March, 2023, from <https://www.jrf.org.uk/report/role-higher-education-providing-opportunities-south-asian-women>
- Bernard, A., & Bell, M. (2015). Smoothing internal migration age profiles for comparative research. *Demographic Research*, 32(33), 915–948.
- Courgeau, D. (1973). Migrants et migrations. *Population (french Edition)*, 28(1), 95–129.
- Dodd, E., Forster, J., Bijak, J., & Smith, P. (2021). Stochastic modelling and projection of mortality improvements using a hybrid parametric/semi-parametric age–period–cohort model. *Scandinavian Actuarial Journal*, 2021(2), 134–155. <https://doi.org/10.1080/03461238.2020.1815238>
- GLA. (2022). Population change in London during the Covid-19 pandemic: Annex—Assessment of updated evidence. London: Greater London Authority. Retrieved 13 March, 2023, Retrieved 18 March, 2023, from <https://data.london.gov.uk/dataset/population-change-in-london-during-the-pandemic>
- Hajnal, J. (1955). The prospects of population forecasts. *Journal of the American Statistical Association*, 50(270), 309–322. <https://doi.org/10.1080/01621459.1955.10501267>
- Hamilton, C. H., & Perry, J. (1962). A short method for projecting population by age from one decennial census to another. *Social Forces*, 41(2), 163–170. <https://doi.org/10.2307/2573607>
- Harrison, J., Keenan, K., Sullivan, F., & Kulu, H. (2023). Union formation and fertility amongst immigrants from Pakistan and their descendants in the United Kingdom: A multichannel sequence analysis. *Demographic Research*, 48(10), 271–320. <https://doi.org/10.4054/DemRes.2023.48.1>
- Hauer, M. E. (2019). Population projections for U.S. counties by age, sex, and race controlled to shared socioeconomic pathway. *Scientific Data*, 6, 190005. <https://doi.org/10.1038/sdata.2019.5>
- Hauer, M. E., & Schmertmann, C. P. (2020). Population pyramids yield accurate estimates of total fertility rates. *Demography*, 57(1), 221–241. <https://doi.org/10.1007/s13524-019-00842-x>

- Kolassa, S., & Schutz, W. (2007). Advantages of the MAD/mean ratio over the MAPE. *Foresight*, 6, 40–43.
- Marois, G., & Bélanger, A. (2015). Analyzing the impact of urban planning on population distribution in the Montreal metropolitan area using a small-area microsimulation projection model. *Population and Environment*, 37(2), 131–156.
- Nomis. (2023a). Population estimates—Local authority based by five year age band—Nomis—Official Census and Labour Market Statistics. Office for National Statistics. Retrieved 13 March, 2023, from <https://www.nomisweb.co.uk/datasets/pestnew>
- Nomis. (2023b). 2011 Census. Table MU01BUK_ALL—Origin and destination of migrants by age (grouped). Office for National Statistics. Retrieved 13 March, 2023, from https://www.nomisweb.co.uk/census/2011/mu01buk_all
- OBR. (2023). Economic and fiscal outlook—March 2023. Office for Budget Responsibility. Retrieved 16 March, 2023, from <https://obr.uk/publications/economic-and-fiscal-outlook-march-2023/>
- ONS. (2014). Sub-national population projections for England: 2012-based. Office for National Statistics. Retrieved 13 March, 2023, from <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationprojections/bulletins/subnationalpopulationprojectionsforengland/2014-05-29>
- ONS. (2016). 2012-based Subnational Population Projections: Methodology. Office for National Statistics. Office for National Statistics. Retrieved 13 March, 2023, from <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationprojections/methodologies/2012basedsubnationalpopulationprojectionsmethodology>
- ONS. (2021). Dataset: Local area migration indicators, UK (Discontinued after 2020). Office for National Statistics. Retrieved 13 March, 2023, from <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/migrationwithinthek/datasets/localareamigrationindicatorsunitedkingdom>
- ONS. (2022). Bulletin. Population estimates for the UK, England, Wales and Northern Ireland: National and subnational mid-year estimates for the UK and its constituent countries by administrative area, age and sex. Retrieved 18 March, 2023, from <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2021>
- ONS. (2023a). Births in England and Wales: Summary tables 3 or worksheet 3. Office for National Statistics. Retrieved 13 March, 2023, from <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/datasets/birthsummarytables>
- ONS. (2023b). Datasets. Estimates of the population for the UK, England, Wales, Scotland and Northern Ireland. Office for National Statistics. Retrieved 13 March, 2023, from <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalescotlandandnorthernireland>
- ONS. (2023c). National life tables—Life expectancy in the UK. Office for National Statistics. Retrieved 13 March, 2023, from <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/bulletins/nationallifetablesunitedkingdom/2018to2020>
- ONS. (2023d). Life expectancy for local areas of the UK: Between 2001 to 2003 and 2018 to 2020. Office for National Statistics. Retrieved 13 March, 2023, from <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandlifeexpectancies/bulletins/lifeexpectancyforlocalareasoftheuk/between2001to2003and2018to2020>
- ONS. (2023e). Admin-based population estimates for local authorities in England and Wales, 2011 to 2022. Office for National Statistics. Retrieved 13 March, 2023, from <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/articles/adminbasedpopulationestimates/provisionalestimatesforlocalauthoritiesinenglandandwales2011to2022>
- ONS. (2023f). Reconciliation of mid-year population estimates with Census 2021 at local authority level: Exploration of the differences between the official 2021 Census-based mid-year estimates and the 2021 rolled-forward mid-year estimates at local authority level. Office for National Statistics. Retrieved 13 March, 2023, from <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/articles/reconciliationofmidyearpopulationestimateswithcensus2021englandandwales/2023-02-28>
- ONS. (2023g). Prospective new method for setting mortality assumptions for national population projections, UK: January 2023. Office for National Statistics. Retrieved 13 March, 2023, from <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationprojections/>

- articles/prospective新方法forsettingmortalityassumptionsforationalpopulationprojectionsuk/january2023
- Rayer, S., & Smith, S. K. (2010). Factors affecting the accuracy of subcounty population forecasts. *Journal of Planning Education and Research*, 30(2), 147–161. <https://doi.org/10.1177/0739456X10380056>
- Rees, P. (1984). Spatial population analysis using movement data and accounting methods: Theory, models, the 'MOVE' program and examples. *Working Paper 404*, School of Geography, University of Leeds. Retrieved 13 March, 2023, from https://www.researchgate.net/publication/344645571_Rees_1984_Spatial_population_analysis_and_the_MOVE_program?channel=doi&linkId=5f86b258a6fdccfd7b5fcc18&showFulltext=true
- Rees, P. (1985). Does it really matter which migration data you use in a population model? In P. White & G. Van der Knaap (Eds.), *Contemporary studies of migration* (pp. 55–77). Geobooks.
- Rees, P. (2023). The relationship between migration intensity and age. In S. Torres & A. Hunter (Eds.), *Handbook on migration and ageing* (pp. 138–150). Edward Elgar Publishing.
- Rees, P., & Lomax, N. (2019). Ravenstein revisited: The analysis of migration, then and now. *Comparative Population Studies*, 44(2019), 351–412.
- Rees, P., & Willekens, F. (1986). Data and accounts. In A. Rogers & F. Willekens (Eds.), *Migration and settlement: A multiregional comparative study* (pp. 19–58). Reidel.
- Rees, P. & Wilson, A. G. (1977). *Spatial population analysis*. Edward Arnold.
- Rees, P., Wohland, P., Norman, P., & Boden, P. (2011). A local analysis of ethnic group population trends and projections for the UK. *Journal of Population Research*, 28(2–3), 149–184. <https://doi.org/10.1007/s12546-011-9047-4>
- Rees, P., Wohland, P., Norman, P., Lomax, N., & Clark, S. (2017). Population projections by ethnicity: Challenges and a solution for the United Kingdom. In D. Swanson (Ed.), *Frontiers in applied demography* (pp. 19–58). Springer.
- Reinhold, M., & Thomsen, S. L. (2015). Subnational population projections by age: An evaluation of combined forecast techniques. *Population Research and Policy Review*, 34(4), 593–613. <https://doi.org/10.1007/s11113-015-9362-0>
- Rogers, A. (1976). Shrinking large-scale population-projection models by aggregation and decomposition. *Environment and Planning A*, 8(5), 515–541. <https://doi.org/10.1068/a080515>
- Rogers, A., & Castro, L. (1981). Model migration schedules. Research Report RR-81-30, International Institute for Applied Systems Analysis. <https://pure.iiasa.ac.at/id/eprint/1543/>
- Rogers, A., Little, J., & Raymer, J. (2011). *The indirect estimation of migration*. Springer.
- Simpson, L., & Snowling, H. (2011). Estimation of local demographic variation in a flexible framework for population projections. *Journal of Population Research*, 28(2), 109–127. <https://doi.org/10.1007/s12546-011-9060-7>
- Singh, A. (2023). Budget 2023: Living standards won't return to pre-Covid levels in next five years, damning OBR verdict warns (inews.co.uk). *inews*, 15 March. Retrieved 16 March, 2023, from https://inews.co.uk/news/budget-2023-obr-verdict-covid-2211509?ico=more_from_this_author
- Smith, S. K. (1986). Accounting for migration in cohort-component projections of state and local populations. *Demography*, 23(1), 127–135. <https://doi.org/10.2307/2061413>
- Smith, S. K., & Tayman, J. (2003). An evaluation of population projections by age. *Demography*, 40, 741–757. <https://doi.org/10.1353/dem.2003.0041>
- Smith, S. K., Tayman, J., & Swanson, D. A. (2013). *A practitioner's guide to state and local population projections*. Springer.
- Tayman, J. (2011). Assessing uncertainty in small area forecasts: State of the practice and implementation strategy. *Population Research and Policy Review*, 30(5), 781–800. <https://doi.org/10.1007/s11113-011-9210-9>
- Tayman, J., Swanson, D. A., & Baker, J. (2021). Using synthetic adjustments and controlling to improve county population forecasts from the Hamilton-Perry method. *Population Research and Policy Review*, 40(1), 355–1383. <https://doi.org/10.1007/s11113-021-09646-7>
- The Economist. (2023). Life expectancy: Nasty, British and short. *The Economist*, March 11th, 19–21.
- Van Imhoff, E., Van der Gaag, N., Van Wissen, L., & Rees, P. (1998). The selection of internal migration models for European regions. *International Journal of Population Geography*, 3(2), 137–159. [https://doi.org/10.1002/\(SICI\)1099-1220\(199706\)3:2%3c137::AID-IJPG63%3e3.0.CO;2-R](https://doi.org/10.1002/(SICI)1099-1220(199706)3:2%3c137::AID-IJPG63%3e3.0.CO;2-R)
- Wilson, T. (2016). Evaluation of alternative cohort-component models for local area population forecasts. *Population Research Policy Review*, 35(2), 241–261. <https://doi.org/10.1007/s11113-015-9380-y>
- Wilson, T. (2018). Evaluation of simple methods for regional mortality forecasts. *Genus*, 74(14), 1–22.

- Wilson, T. (2020). Mortality surface illustration. Retrieved 13 March, 2023, from <https://doi.org/10.6084/m9.figshare.17088539.v1>
- Wilson, T. (2022). Preparing local area population forecasts using a bi-regional cohort component model without the need for local migration data. *Demographic Research*, 46(32), 919–956. <https://doi.org/10.4054/DemRes.2022.46.32>
- Wilson, T., & Bell, M. (2004). Comparative empirical evaluations of internal migration models in sub-national population projections. *Journal of Population Research*, 21(2), 127–160. <https://doi.org/10.1007/BF03031895>
- Wilson, T., Brokensha, H., Rowe, F., & Simpson, L. (2018). Insights from the evaluation of past local area population forecasts. *Population Research Policy Review*, 37(1), 137–155. <https://doi.org/10.1007/s11113-017-9450-4>
- Wilson, T., & Grossman, I. (2022a). Evaluating alternative implementations of the Hamilton-Perry model for small area population forecasts: The case of Australia. *Spatial Demography*, 10, 1–31. <https://doi.org/10.1007/s40980-021-00103-9>
- Wilson, T., & Grossman, I. (2022b). Evaluation of a new simplified population projection model: A case study of local government area projections in Tasmania. *Australasian Journal of Regional Studies*, 28(1), 55–73.
- Wilson, T., & Shalley, F. (2019). Subnational population forecasts: Do users want to know about uncertainty? *Demographic Research*, 41(13), 367–392. <https://doi.org/10.4054/DemRes.2019.41.13>
- Wu, B., Birkin, M., & Rees, P. (2008). A spatial microsimulation model with student agents. *Computers, Environment and Urban Systems*, 32(6), 440–453. <https://doi.org/10.1016/j.compenvurbsys.2008.09.013>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.