Choosing a suitable projections model for Wales

Background
Recently the Office for National Statistics (ONS) has been responsible for producing sub-national population projections for Wales, under contract from the Welsh Assembly Government (WAG). Since the creation of the 22 local authorities in Wales, there have been difficulties in creating suitable projections at the local authority geography using the current methodology. This paper describes some of these difficulties and compares some alternative software packages to be considered by the projections working group.

Current Methodology
The ONS methodology uses a cohort-component based model and the base mid-year population estimates as the starting point. The population is 'aged on' by one year and assumed local fertility and mortality rates are applied to calculate the number of projected births and deaths. Adjustments are then made about migration into and out of each area. The projections are then constrained to the totals for Wales in the national population projections produced by the National Statistics Centre for Demography (formally the Government Actuary's Department).

Fertility and Mortality
Assumptions about local fertility and mortality by age and sex are based on birth and death registration information for the previous five years. As part of the projection process, for each year projected forward, the local fertility and mortality rates are scaled, so that the national rates and the projected numbers of births and deaths are consistent with the national projections.

Migration
Migration is the most difficult component of population change to estimate. International migration flows are taken from the International Passenger Survey and allocated to local authorities using Census distributions. Data on asylum seekers and visitor switchers are used from the Home Office. The numbers are constrained to the totals for Wales in the national population projections.

Internal migration flows within Wales are derived from GPs patient registers, which provides details on the registrations of the population with GPs down to postcode level. Five years of historic registration data are used to derive assignment matrices at single year of age and sex showing movements to and from each geographical area. The out migration probability by age and sex is derived from this matrix by fitting an assumed age profile model to the data for each area. These age-specific probabilities are then applied to the projected population in each area to provide a number of out-migrants by age and sex. Applying the original assignment matrices produces the assumed destination of these out-migrants. Totalling this up for each destination area in turn gives an assumed level of in-migration by age and sex. The net internal migration is then calculated for each area as the difference between these two. Net internal migration sums to zero across all areas.

Cross-border migration flows are allocated using the GPs patient register. The national projection assumptions for Wales are allocated to the local authorities. These flows are assumed to be constant for each age to each local authority.

Problems with the current methodology
The sub-national population projections are, at best, a hybrid by virtue of being controlled to the national projections. Birth and death rates are scaled, whilst this ensures the total births and deaths are consistent with the national projections, the local rates do not reflect recent, local trends. Scaling will be more significant with small areas such as the 22 local authorities in Wales.

Moreover, they do not take account of local factors that will be key in influencing population growth such as housing developments, or other factors and developments that may influence the demographic profile of the local population.

It is also the case that these projections incorporate a phasing-in of the long-term assumptions. The 'base' year of the population projection starts with the latest year's population estimates; there is then a process of moving fairly smoothly from the base to the start of the long term assumptions, typically after three years. Those initial years of the projections do not, therefore, reflect the underlying, long-term assumptions nor are they forecasts of what is likely to happen. They are simply a means of moving from the base to the start point of the long-term assumptions (and are commonly referred to as the short-term assumptions).

There are issues concerning the adequacy of the underlying data, particularly for international migration (including visitor switchers), and the influence of particular groups – such as students – which can be significant (but not well measured) in particular areas of the country. These are being addressed through the ONS-led Improving Migration and Population (IMPS) project.

Future projections
The sub-national projections working group in Wales met at the end of January to discuss proposals for a change in the approach in producing suitable local authority projections. It was agreed to pursue the possibility of a methodology based on a 'bottom up' approach using local trends for each local authority, on a consistent basis, unconstrained to the national projections. A questionnaire completed by local authority colleagues identified two software packages developed in the UK, Popgroup and the Chelmer model, that local authorities have historically used for their own locally produced projections.

Software packages for population projections
Popgroup
Popgroup is a family of software developed to forecast population, households and the labour force for areas and social groups. It is based on Excel to build on users' existing spreadsheet skills. The software has been used to forecast

national and sub-national populations, as well as social and ethnic groups. It uses standard demographic methods of cohort component projections, household headship rates and economic activity rates. Its flexibility allows integration of official statistics and ancillary data.

Some of the features include:
- Single years of age, males and females separately
- One or more groups, which can be of any type for example:
  1. Districts in a county
  2. Ethnic groups in a district
  3. Sub-areas in a district
- User's separate estimates of special populations can be incorporated (students, armed forces etc)
- Integrates estimates and projections in one dataset
- Constraints of housing, employment or population can be used to control migration.
- Standard age-schedule of rates for each component is the only required dataset
- Flying pyramids show the dynamic evolution of a population
- User's own area and group labels appear in all files and output
- Notes pages for documentation on every file
- Many options allow user to over-ride or adjust the standard with the user's data for:
  1. Counts, and/or rates
  2. Data for individual groups, and/or for the total of all groups
  3. Age-sex detail, and/or data for the total of all persons
  4. Data may be provided explicitly, or the model linearly interpolates between provided values
- Gross migration flows with one or two external areas eg. within and outside the UK
- Summary output measures of fertility, mortality, life expectancy and migration levels
- Security of results with cell and sheet protection
- Facilities to compare forecasts
- Dedicated report and chart generators in addition to full Excel facilities
- All input and output can be copied for use and re-formatting in the user's own documents

Chelmer Population and Housing Model (CPHM)
Chelmer Population and Housing Model (CPHM) is a demographic regional housing model developed by the Population and Housing Research Group (PHRG) as part of a research project supported by the Housing Research Foundation. The model has been purchased by over 50 organisations, including local authorities, the Housebuilders Federation and various consultancy firms.

It is user friendly, and allows the testing of a variety of assumptions, particularly relating to migration and household formation. It produces reports in both text and spreadsheet formats.

In addition to conventional migration-led (total migration and/or international migration) forecasting, CPHM offers a housing-led mode in which the forecasting of population is evaluated in terms of the capacity of an area to accommodate households. This is calculated from the base numbers of dwellings and building/demolition rates. From this, the population growth and migration implications are derived. Within the model framework, calculations may be repeated with various controls applied, such as total populations and total households.

The Chelmer Model is based on cohort survival population projection methodology allowing the projection of natural change (births and deaths) and the projection of migrants into and out of an area.

The natural change of the population (births and deaths) is projected using cohort-survival methodology and national rates drawn from the national projections of fertility and mortality rates, corrected to local registrations of births and deaths for an appropriate period.

Migration into and out of the area is modelled similarly for both internal and international elements. The internal migrants (i.e. within UK) have an age structure taken from the 2001 Census of Population, whilst the international migrant structure uses a combination of Census and ONS population projection data.

Moving forward
The next stage is for the projections working group to decide on Popgroup or the Chelmer Model as a suitable software package. Both software packages use similar methods but the approach in doing the work is different. Provided one is suitable we have ruled out a bespoke model for Wales, but if neither is suitable this approach may be adopted.

Popgroup can be purchased for a one off cost and the projection work completed by the user. The advantage is that at each stage of the process, the user can follow how the underlying raw data and assumptions are being projected forward, making adjustments where necessary. Constraints can be incorporated and modified and variants in the assumptions can be tweaked to produce an alternative set of projections. The disadvantage of this approach is that the work requires a number of staff hours and training although a step by step user guide is available.

The Chelmer Model is such that the projection is run by the PHRG, at a cost, with the user supplying the raw data for the projection. This means very little staff resource is needed and no training required. However work would is still needed in obtaining and preparing the raw data. Understanding the numbers, identifying how the underlying assumptions are being projected forward and tweaking them for variant projections would be more difficult with an external body producing the projections.