DCLG 2003-based household projections: Methodology and Sources of Data Provided by Professor Dave King 5th October 2006

The Method

The household projections are compiled by applying projected household membership rates to a projection of the private household population disagreggated by age, sex and marital/cohabitational status and summing the resulting projections of household representatives.

The household projections are compiled through a six stage process:

a. the resident population is projected by the Government Actuarys Department (GAD) at national level and by the Office for National Statistics (ONS) for sub national areas of England. The latest household projections utilise the 2003-based national population projections. The subnational population projections are also 2003-based.

b. the marital status of the population is projected by GAD. This projection is made for England and Wales combined. It projects the population by legal (de jure) marital status (within sex and age) cross analysed by whether or not cohabiting. The proportions cohabiting in 2003 were estimated by ONS and projected for future years by GAD. The marital status projection is at national level only. Estimates of marital status in future years at subnational level are made by applying national/local differentials in marital status from the 2001 census to projected marital status factors.

c. the institutional population is projected by DCLG from the 2001 census (evaluated against information from the Department of Health about residents in residential care homes). At ages under 75, the number of residents in institutions is kept constant at 2001 levels in each sex/age/marital status category. The marital status classification groups individuals into de jure categories only, as members of the institutional population do not cohabit by definition (resident staff with self contained accommodation are not part of the institutional population). In the 75 - 79, 80 - 84 and 85 and over age bands, the proportion of the population in each age/gender/marital status category resident in communal establishments is assumed to remain constant.

d. the institutional population is subtracted from the total resident population projection to leave the private household projection, analysed by sex, age and marital status (cross classified by cohabitation status) in the years required for household projections.

e. within each age/sex/marital cum cohabitational status category, household membership rates are projected from historical data derived from censuses and (at national level) Labour Force Survey (LFS) data. The projected household membership rates are then multiplied by the appropriate private household population projection.

f. projections for sub national areas are initially made independently of the national projections, but are subsequently adjusted for consistency with the national projection.

Similarly, projections for sub regional areas are adjusted for consistency with the regional projections. These adjustment processes are termed "regional controlling".

Data Sources

The data sources used for projecting household membership rates are the 2001 Census (commissioned table CT598; 100 per cent), special analyses of 10 per cent samples of the 1971, 1981 and 1991 Censuses; the ONS Longitudinal Study samples from the 1971 and 1981 Censuses and the Labour Force Survey (LFS) from 2002 to 2004. The Labour Force Survey is considered the best available source of data about household membership rates after the 2001 Census.

Some adjustments to these data were implemented on advice from the Peer Review Group, which advised on the development of the projections. Some adjustments were made to data used in previously published projections. The 2001 Census enumerates students as being usually resident at term-time address, whereas previous Censuses enumerated them on the basis of parental address. At the subnational level the 1971, 1981 and 1991 Census data was adjusted, where possible, to allow for the impact that the different treatment of students would have on marital composition and household membership rates in those earlier years, to achieve consistency in the time series for projection purposes.

Some adjustments were also made to the household membership rates derived from the 2001 Census. The first adjustment was to adjust for differences between the Census and the 2001 mid-year estimates: population (ONS) and marital status (GAD). This amounted to a 187,000 adjustment to the population made by ONS attributable primarily to people deemed to have been missing from within households in the Census count (rather than being household representatives or persons in communal establishments). About 160,000 of these were males aged 25-34. The large majority of the 187,000 are assigned to the single non-cohabiting marital status group, and this has the effect of moderating their 2001 household representative rates in those cases. In addition there was further minor revision of 2001 household membership rates, based on reassigning a limited number of representatives in the commissioned table aged under 15 on advice from ONS, as a result of follow-up examination of those cases.

The Labour Force Survey has been used to extend the time series of household representative rates beyond the most recent available Census year. Compared with the Census 100 and 10 per cent samples, the Labour Force Survey samples are small, less than 1 per cent of the population. To minimise the effect of any systematic bias, the LFS data for 2002 to 2004 have been adjusted to reflect the discrepancies between Census and LFS data in 2001. Because of the smaller sample used in the LFS, LFS household representative rates are calculated only by age and sex. Fully disaggregated household membership rates for 2002 to 2004 are produced by projecting Census data and then controlling the results to be consistent with the LFS based age/sex membership rates. These household membership rates are then used in conjunction with the Census data for national projections of household representative rates.

Projection Method for Household Representative Rates

Household representative rates in future years are projected from time trends estimated from the census based values for 1971, 1981, 1991and 2001 and then the values from the Labour Force Survey. A modified version of the "life cycle" method used in previous projection rounds is used. This "life cycle" method makes used of the fact that the household membership rates in a particular cohort (defined by age) will vary smoothly with time. Furthermore, the changes in a particular rate in a given cohort will be the result of effects such as leaving the parental home, marriage and mortality, which are strongly dependent on the stage of the cohorts life cycle. It follows that, for each household membership status and marital status, plots of the cohorts membership rates against its age will have shapes which are recognisably similar for all cohorts. The projection method models this characteristic development of membership rates within cohorts as life cycle curves, which are assumed to undergo only trend changes in shape from cohort to cohort.

In practice, the household membership rates are modelled in terms of cohort birth date and cohort age by a two dimensional curve fitting method which makes a maximum likelihood fit to the available membership rate data. This method automatically weights data points according to sample size so that, for example, rates calculated for a small population group or from the LFS based data sets (which have comparatively small sampling fractions) are not given undue weight in determining the membership rate projections. In addition, explicit weights are applied for other purposes. An exponential weighting system is used to give greater weight to more recent data and further variations in the weights are incorporated to allow for uncertainties arising from definitional modifications and the errors introduced by estimation procedures such as those used in forming LFS based data sets.

An adjustment to the modelling was implemented triggered by advice from the Peer Review Group, which advised on the development of the projections. The modification to the "life cycle" method which was introduced in the 2003-based projections keeps to the spirit of the original methodology and was enabled by advances in computational efficiency and triggered by evidence from the 2001 Census and LFS that the previous formulation ran the risk of producing projections which overstated household representative rates in age groups under 30. It is in the nature of cohort life cycles that the strongest influences on a particular cohort will be the past experiences of that cohort together with the experiences of comparable cohorts (in the same age/gender group) previously. In the formulation of the original methodology used to determine the “life cycle”-derived curves, this influence was modelled but in order to do so with the computational efficiency necessary at the time, data from all older age groups was allowed to influence the trajectories of younger generations. The modification used in the 2003-based projections introduced substantially greater computation in order to allow curves to be constructed relevant to each cohort in turn, such that the curve relevant to a particular cohort reflects the past experiences of that cohort (and younger cohorts) together with the experiences of comparable cohorts (in the same age/gender groups)

previously up to that particular age only. This has had the effect of dampening projected growth of household representative rates at younger adult ages, particularly those under the age of 30.

Projection by time trend is appropriate where the observed changes are the result of numerous influences whose effects cannot be estimated separately. At the subnational level in particular, with only four census data points, the possible effects of economic variables such as real income, unemployment rates and mortgage rates cannot be distinguished one from another by statistical means, or together be distinguished from the effects of social changes. Upward trends in household membership rates and headship rates have been very long established. Rising trends are evident for the 1960s and 1950, as well as the 1980s and 1970s. Nothing is in view across most age/gender groups that could reasonably be expected to cause changes of trend in this respect, though of course the projected rising trends in individual household membership rates slow down as the maximum limit of 100 per cent is approached. Medium term and longer term trends are the basis of the projections of households and of the underlying projections of population and marital status. Such fluctuations are normally outside the scope of the projections, which can depict only medium term trends.

In the specific case of single non-cohabiting females, the Peer Review Group considered that the uplift in the household representative rates that was projected for 25-39 age groups was disproportionately greater than that for single non-cohabiting males in the same age groups. Consequently the growth in their projected aggregate[1] household representative rates was constrained proportionate to the growth in the equivalent male rates.

Regional Controlling

Separate household projections are made independently for each projection area. This gives rise to inconsistencies, in the sense that the projected number of households in a given area is not normally equal to the sum of the household projections for its constituent sub areas. This problem stems from the non linearity of the household membership rate model and from the use of LFS-based data for England and Wales. A similar situation arises regarding private household population projections, originating in the division of the age/sex groups by marital status. In addition, the non-linearity of the projection method is such that unadjusted the projections will not normally have equal numbers of male and female cohabiters, or of husbands and wives in married couples.

The purpose of the regional controlling procedure is to adjust the household projections so that all of these consistency conditions are achieved. At the same time, it must maintain the age/sex resident population projections made by GAD and ONS. To accomplish this task, a "top down" procedure is used. The separate projections for England and for Wales are first modified to agree with those for the combined area of England and Wales. At the next stage, the projections for the English Government Office Regions are calibrated to the controlled projections for England. This procedure is

[1] In this context, “aggregate” refers to all household representatives, irrespective of household type

continued down the "tree", the controlled projections for each area being used to calibrate the projections of its sub areas. The 2003-based projections use the same generalised scaling methodology as used in the 1996-based projections, although the latter used Standard Statistical Regions, rather than Government Office Regions, as the regional tier for controlling purposes.