ISSN 0819-2642
ISBN 0 7340 2570 X

THE UNIVERSITY OF MELBOURNE

DEPARTMENT OF ECONOMICS

RESEARCH PAPER NUMBER 914

SEPTEMBER 2004

GLOBAL PROJECTIONS OF HOUSEHOLD NUMBERS USING AGE DETERMINED RATIOS

by

Victor E. Jennings
&
Conrad W. Lloyd-Smith
&
Duncan S. Ironmonger

Department of Economics
The University of Melbourne
Melbourne Victoria 3010
Australia.

GLOBAL PROJECTIONS OF HOUSEHOLD NUMBERS USING AGE DETERMINED RATIOS

Victor E. Jennings, Households Research Unit, The University of Melbourne, Australia
Conrad W. Lloyd-Smith, School of Computer Science and Information Technology, RMIT University, Australia
Duncan S. Ironmonger, Households Research Unit, The University of Melbourne, Australia

A new method based upon age determined population ratios is described and used to estimate household population intensities (households per person). Using an additive and a bounded model household projections are given to 2050 for the world and to 2030 for seven fertility transition subgroups (cohorts) of the countries of the world. Based upon United Nations 2002 Revision data, from an estimated 1.56 billion households at 2000, household growth to 2030 is projected to be an additional 1.1 billion households, whether population increase is 1.3 billion persons under the United Nations low fertility variant or 2.7 billion persons under the high fertility variant. At that date over one third of all households are projected to be Chinese or Indian. By 2050 it is projected that there will be 3.3 billion households with a 95 per cent confidence interval on modelling error only of ± 0.5 billion. This compares with 3.2 billion in the Habitat: Global Report on Human Settlements 1996. The apparent similarity of total household growth under various scenarios conceals a wide range in the growth of household intensities across fertility transition cohorts. It is suggested that models, projections and error be reviewed biennially and that household and population projections be produced jointly.

Web: www.economics.unimelb.edu.au/dept/household/hholdunit.htm

Key words: Household projections, world, age ratios, fertility

Introduction

Households are hybrids of household groups and the dwellings they occupy. Their nature is complex but overall it is contended in this paper that they are driven by two main factors. Stage in the life course represented by the relative proportions of youth, middle persons and elders; and fertility levels, which determine changes in these proportions. These two overriding factors are used in this paper to estimate household intensities (households per person, termed ‘*h*’). The country is the unit for comparison. Age determined ratios and household intensities are the variables used. The central assumption is that all countries track the same *h*-age ratio relationships, as they follow the fertility transition. The work explores the degree to which estimates and confidence intervals on household numbers can be obtained, given limited information. This work is motivated by a need for timely and sufficiently accurate measures of the number of households across the world, despite data limitations in many countries. The procedures adopted in this paper essentially follow on from work described in Jennings, Lloyd-Smith and Ironmonger (1999). The paper is divided into seven sections: Introduction, method, age ratios and the world context, household intensity and age ratios, model evaluation, results, and concluding remarks. Together with references there are two appendices.

Method

This paper focuses on the household group rather than the dwelling. About 97 per cent of the world's population are determined by censuses to be members of one and only one household. A member of a household group is a person who occupies a particular dwelling. This encompasses the single person household under the condition that there are zero common occupancy relationships with others. A household then consists of those factors that relate only to population such as age ratios; those that relate only to the dwelling such as rooms per dwelling; and those factors that are dependent upon the relationships between the two such as persons per dwelling. In this paper only factors related to population are used to calculate household numbers. It turns out that about 87 per cent of the variation in household intensities can be attributed to these population factors. The household projections in this paper are based upon population projections of the United Nations (United Nations 2003) and functions of population developed in this paper.

Past age, sex and household data, which are available for many countries, are used to establish relationships between age ratios and household intensities country by country. Age ratios and household intensities are analogous to dimensionless numbers used in dimensional analysis (Freiberger 1960:240), and allow the development of characteristic features of human populations independently of population size. These features are then used to develop two h-age ratio models. The models are then applied to future population projections produced by the United Nations (2003) to give estimates of household intensities and the number of households. Projections are made of households thirty years ahead for subdivisions of the world's population and fifty years ahead for the world as a whole. Most countries over this period will have age ratios experienced by other countries in the past. Consequently h-age ratio relationships found in the past will also be assumed to apply in the future. Although regular updating of the models is expected as more data comes available it not anticipated that this will make much difference to the projections of households.

For the purposes of this paper population (P) is subdivided into three age classes, young persons aged 0 to 19 years (Y), mature persons aged 20 to 59 years (M) and elders aged 60 years or more (E). Adults (A) are defined here to be persons aged 20 years or more. This subdivision of ages has been found by the authors to provide a near optimum division of the population into three parts for the purposes of household formation analysis. It has been shown (Jennings, Lloyd-Smith and Ironmonger 1999) that the ratio of the number of youth aged 0 to 19 years to the number of women aged 20 to 49 years is also closely associated with total fertility. In this paper Y/M is used as a proxy for fertility. In addition to representing the realised fertility as expressed by the ratio of youth to middle persons, it is also closely related to total fertility for a population. For example, for each of the 192 countries in the world list (United Nations 2002 Revision) the total fertility for the four five-year periods immediately prior to the year 2000 are averaged and denoted by $\overline{TF}$. The regression on $\overline{TF}$ in 2000 yields the following equation.

$$Y/M = 0.138 + 0.201(\overline{TF}) \qquad s = 0.108 \qquad R^2 = 0.92 \qquad (1)$$

where s is the standard error. Similar results, which are not quite as precise, occur with periods related to years earlier than the year 2000. The four five-years periods cover a period 1980 to the year 2000. This is the period over which young persons Y will have been born. The total fertility over-estimates the number of young persons still living at 2000, since some will have died between the ages of 0 and 19 years. For the purposes of this paper however equation (1) is used without the addition of variables related to mortality. For a review of fertility measurement see Preston, Heuveloine, Guillot (2001:92-116) or Hinde (1998:95-106).

The particular advantage of using Y/M as compared to total fertility is that the ratio explicitly relates two key variables in a population which, if either M/P or E/P or either of certain other age ratios is known, completely defines the proportions of Y, M and E. This is so because $Y/P + M/P + E/P = 1$ and thus the age ratios can be mapped in two dimensions, eg Figure 2. It was also suggested in Jennings, Lloyd-Smith and Ironmonger (1999) that age ratios could be used to determine average household size. Since each household in a population can be identified by the age class of one of their members, households per person or household intensities may be considered a function of age ratios. Thus Y/M can provide a link between total fertility levels and household intensities.

The two models used are derived from linear regressions of household intensity on functions of age ratios. One of the models chosen is called the additive model. Using this model the sum of household estimates using age ratios from two or more countries taken separately is equal to the household estimate of the countries when the overall age ratios are used. In this way one can calculate the total number of households worldwide based only on world level age ratios. The other model used is slightly more precise but is only approximately additive. It is called the bounded model. The two models differ in their variables. The additive model has variables h, M/P and Ef/P where h is the household intensity measured as households per person and Ef is the number of female elders. The bounded model has variables h and u where $u = 1/(1 + Y/1.84\,Ef)$. It has the advantage that household intensities can be plotted directly against a single function of age ratios u. It is demonstrated that the models are closely related and yield very similar results.

Household intensity h for a population is taken to be H/P_H where H is the number of households and P_H is the population in households. Examination of censuses shows that for recent data for 88 countries on average $P_H/P = 0.971011$. This figure is used in this paper as the reduction factor on P to give P_H.

The number of households in the 141 countries used for modelling purposes range from 14000 to 348 million. In plotting estimated numbers of households against observed the wide range is simply represented by considering the data at various scale magnifications. This preserves linear relationships. Not all the data can be represented on every graph. It is not appropriate to use logarithms of the number of households since relative error calculated using logarithms is dependent upon population size. For example, consider an observed and estimated value where

$\text{Observed} = 10^{x}$ and $\text{Estimated} = 1.1 \times 10^{x}$ and hence the ratio of Estimated/Observed =1.1

Taking logarithms

log(Observed) $= x$ and log(Estimated) = $0.041 + x$

Now the ratio of log(Estimated)/ log(Observed) $= 1 + 0.041/x$ and this varies according to the value of *x*. Thus a log scale would show a varying relative error across a number of countries of different sizes, even if the observed relative error were constant. In the models used in this paper the residual error term for estimating households is proportional to the population size and hence the observed relative error is constant.

Country as the unit of comparison

'Country" is a defining characteristic of the world's population. It is a subdivision of the world's population that is relatively stable over time. It is subject to changes which mainly arise from its own internal dynamics, and is distinguished from other countries by differences in its pattern of laws and customs, language, geography, economic state, and nature of government. It can be seen as the system unit for the understanding the world. In many ways countries, nations or sovereign states are different from each other. The 'country' defines the boundary of residence for most people. Generally this boundary changes little over time. Although there is migration from country to country this is currently relatively small at about 2.1 to 2.3 per cent per annum of the world population (Slotnik 1998). This would be subject to regular review, as for example when countries form common markets as in Europe.

Household formation is influenced by the laws, customs, climate, geography, economic situation and the fertility levels of a particular country. But household formation is also constrained by the life course common to all. It is also essentially local. The nation may be seen as a summation of many local activities occurring simultaneously within national constraints. Because many of these local changes occur independently the broad changes at the national level are likely to be smooth. This smoothness is also due to inbuilt stabilities. "If it were not for the predictive-supportive impetus of human customs, routines, and habits, human societies would not exist and human existence itself would be in question. Life as we know it demands those stabilities and continuities …." (Rescher 1998:199-200). Thus changes in age ratios with respect to time or other variables may be represented as 'traces', and a point in time may be seen as a snapshot of a near continuous process.

The country is unlikely to disappear as an entity. "No replacement is in sight that can perform the essential functions of establishing an orderly and lawful society, protecting essential freedoms, providing a framework for fruitful economic activity, contributing to effective international cooperation, and providing for the common defence." (Schulz 2004).

Data sources
The data sources used in this paper are from the United Nations and various National Statistical Offices. A listing of sources is given in Appendix 1. The household projections are based upon the data from *World Population Prospects: The 2002 Revision* (United Nations 2003) hereinafter called the 'United Nations 2002 Revision'. In addition to world population summary projections to 2050 the United Nations 2002 Revision has projections for 192 individual countries. It also has annual age-sex estimates for the 192 countries from 1950 to 2000. At regular intervals the United Nations has produced world population projections under various assumptions (United Nations 2001a,b,c; 2002:7-9). The low, medium and high fertility variants are used here. The low-high assumptions have been criticised as logically inconsistent (National Research Council 2000:191-194). However the choice of fertility assumptions does not greatly influence the number of households 30 years ahead, as will be demonstrated.

Since fertility levels are shown to have a major influence on household-population relationships it is desirable to differentiate countries accordingly. In this paper the 192 countries listed in the United Nations 2002 Revision are divided into seven fertility transition cohorts according to the classification of Tables 1.2 to 1.8 in Volume III of the United Nations 2000 Revision (United Nations 2002:26-35). A detailed listing is given in the Appendix 1. In this instance a cohort is that group of countries which has embarked on the fertility transition at about the same time. Each cohort is defined by a range of total fertility levels at a particular time. For reference purposes the equivalent values of Y/M are also shown in Appendix 1. These values are estimated using equation (1). A 41-country data set is also used for exploratory examination of households in relationships to population by age. See Appendix 1 for details.

Other work
Many projections of households are made for individual countries or regions within countries. See Appendix 2 for a partial list. In addition to the number of households in a country and age-sex characteristics the projections usually require at least the number of households of particular types (eg couple families with children). Assumptions about propensities to form various sorts of households are made or probabilities are decided upon empirically. Regional statistics may be used. These forecasts are particularly useful over shorter-term periods such as five years. However in view of increasing fluidity in household arrangements, increasing affluence and range of choices, it would seem that such detailed projections can only be made for the short term.

Extrapolations of the number of households arising from past growth can be applied at the world level but, 'By definition, extrapolation always fails to predict turning points.' (Sherden 1998:143), and turning points need to be allowed for in household projections. Haupt, Oberhofer and Reichsthaler (2003) use a statistical model of household membership rates and apply it to large database to estimate and extrapolate the distribution of household sizes of an autonomous region. The procedure is recommended for application over a time horizon up to thirteen years. They report empirical results demonstrating the robustness of the procedure. Another approach is to relate the number of households to the number of adults aged 20+ years over a number of countries using a regression fit. This relationship can then be used for projections. However the residual error for individual country estimates of the number of households can be large relative to other methods. Some of the problems of determining population forecasts and relating them to households are described in de Beer and Alders (1999). Even in developed countries research is still necessary to improve the methods (van Imhoff 1995:348-350).

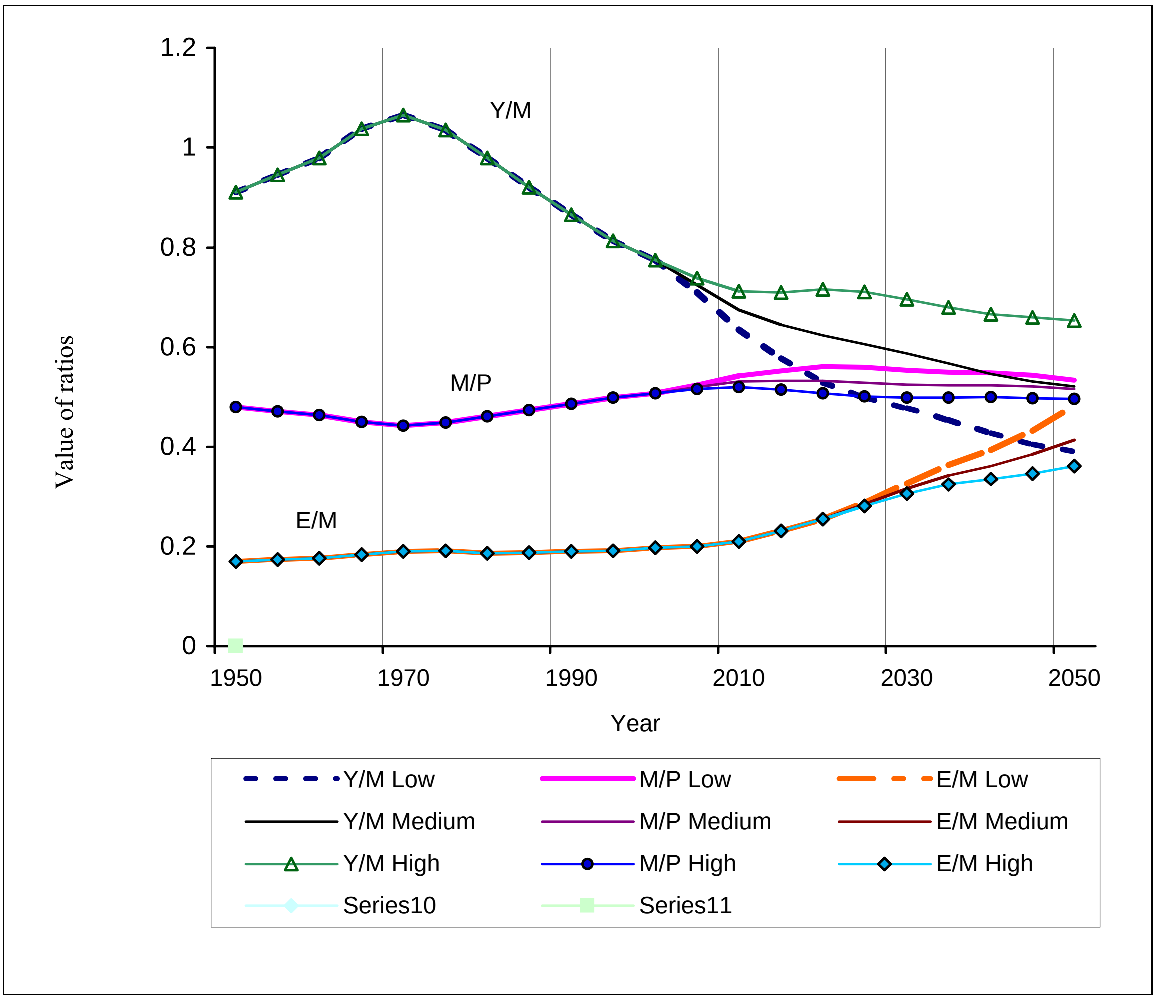
A recent long-term projection of households at the world level was given in the Global Report on Human Settlements 1996 by the United Nations Centre for Human Settlements (Habitat 1996:464-467), hereinafter called Habitat'96. The projections were reported as based upon (a) UNCHS (Habitat), Human Settlements Statistical Database, (b) United Nations, Human Settlements Statistics questionnaire 1992 (Habitat 1996: 467), and (c) the medium fertility projection of the Sex and Age distribution of the World Populations; The 1994 Revision (United Nations 1994. A general review No sensitivity analysis, assessment of error or household size distribution information was provided. Nevertheless the report was a major advance in that it showed a comprehensive long-term projection of households for the world assembled from individual country projections. Details of these projections are included for comparison in Table 3.

Age ratios and the world context

Figure 1 shows the trend in some age ratios when measured at the world level using the United Nations 2002 Revision population data. For the period beyond 2000 the high, medium and low fertility variants for population projections of the United Nations 2002 Revision are used. See Appendix 1 for details. The high fertility variant shows values for Y/M above its medium fertility value and values for E/M and M/P below their medium fertility values. The reverse applies to the low fertility variant. The range for the three variants for M/P is small compared to that for E/M or Y/M.

The outstanding trends are a lowering of the value of Y/M from around 1970, a stability of M/P from around 2015, and a lagged increase in the proportion of E/M. The lower Y/M indicates smaller non-elder households. The counts of age groups M and E can be taken as indicators of the number of non-elder households and elder households respectively. The ratio E/M is therefore an indicator of the ratio of the number of elder households to non-elder households, the 'elderliness' of households. On the other hand Y/M in addition to being an indicator of fertility and youth dependency is also an indicator of the number of youth per non-elder household, the 'youth density' per middle household and its size. Finally M/P which measures the proportion of non-elder persons in the population also indicates the 'middle household intensity' in the population. Taken together the trend is towards a concentration of factors leading to smaller households and hence higher household intensities. The overall trend is towards a one-off transition across the world from a high fertility regime to a low fertility regime. This is of course a reflection of the fertility transition about which much has been written. Similar results are obtained if the geometrical mean of the ratios for individual countries are plotted in order to remove effects of very populous countries such as China and India. These age ratios are also taken to be dependency ratios by others (Eurostat 2002:42-44) although the direction of dependency particularly for the elderly is uncertain.

Figure 1 Estimates from 1950 to 2000 and projections through to 2050 for the world population of the ratio of youth aged 0 –19 years to middle persons 20-59 years (Y/M), elders aged 60 or more years to middle persons (E/M), and middle persons to all persons (M/P)



Source: United Nations 2002 Revision and authors' calculations.

Underlying the general trend shown in Figure 1 is a wide variation in the timing of the fertility transition for individual countries. See for example Table 1 in Caldwell and Caldwell (2001). Some countries such as those in Northern Europe may have reached a lower level stability now, which is projected to occur for world as a whole by 2050. Others have not commenced a movement down in fertility and may stay at the 1970s level shown in Figure 1.

Age ratio structure of the world population – some properties

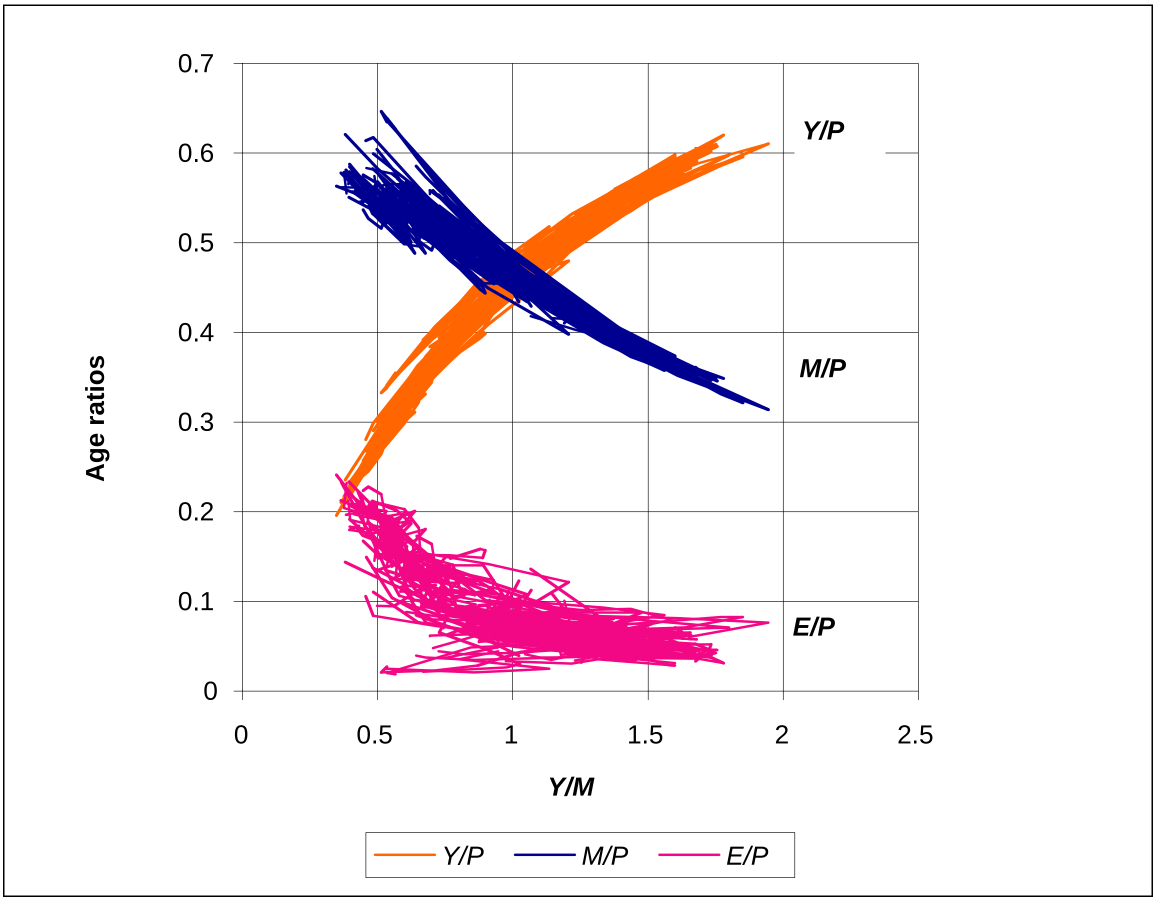
In Figure 1 time was used as the x-axis variable. However various age ratios may be compared directly with Y/M, a measure of fertility, if this is used as the x-axis variable. In Figure 2 the ratios Y/P, M/P and E/P are plotted against Y/M for the 192 countries for which detailed population data is available from the year 1950 to 2000. For each age ratio there is one point taken for each country at 1950, 1960, 1970, 1980, 1990 and 2000, a total of six points for each country. Thus there are 192 'traces' representing 50 year tracking of age ratios of individual countries against Y/M. Most of these traces start at the right and move to the left. Figure 2 is a plan view of a three dimensional graph where time is perpendicular to the surface of the page. It indicates that for a given Y/M there are only a limited range of Y/P, M/P and E/P values that occur for any country. The range of values of Y/M for the 192 countries from 1950 to 2000 is 0.348 (Italy 2000) to 1.946 (Saint Vincent and the Grenadines 1970). Since at the present time countries vary in population size from about one hundred thousand to over a billion the age ratios are clearly not related to population size.

There is an inherent property exhibited in Figure 2: at any point in time for the six age ratios Y/P, M/P, E/P, Y/M, E/M, Y/E (and for certain other combinations) if any two are known, the others can be deduced, since $Y/P + M/P + E/P = 1$. For example if Y/M is determined from measures of fertility and Y/P is also known then the values for the other age ratios can be calculated. This property is quantified by examining Figure 3a and 3b. They are like snapshots of Figure 2 taken at the year 1950 and the year 2000 but are modified as E/P is replaced by Ef/P and u is added. It is convenient to use the ratios M/P, Ef/P and the age ratio function $u = 1/(1 + Y/1.84\,Ef)$ since they are used in determining household intensities for the additive and bounded models. It is calculated later that for a range of data $1.84Ef \approx E$.

The outstanding feature of Figures 3a,3b,3c,3d is the near linearity of the M/P distribution for Y/M greater than about 0.7. This occurs if E/P decreases slowly for these higher values of Y/M. Thus if E/P is a linear function of Y/M it can be shown that $M/P = a + b/(1 + Y/M)$ where a and b are positive constants. At the year 2000 for 122 countries with $Y/M > 0.7$ the regression fit of M/P against $1/(1 + Y/M)$ gives $a = 0.053$ and $b = 0.826$ with $s = 0.006$ and $R^2 = 0.98$. Over this range the regression equation generates a nearly linear curve.

At $Y/M = 1.0$ the Y/P and M/P curves intersect. At this point Y/P and M/P must be less than 0.5 since $E/P \geq 0$, given that $Y/P + M/P + E/P = 1$. From Figure 2 it is seen that the range of values of M/P where $Y/M = 1$ is quite small, from about 0.45 to 0.49. A value of $Y/M = 0$ implies $Y/P = 0$ with high ratios for E/M which would be found for example in retirement communities.

Figure 2 Age ratio structure. Five decade change in age ratios against Y/M for the period 1950 to 2000 for each of 192 countries. General movement is from right to left



Source: United Nations 2002 Revision and authors' calculations.

The function u ranges in value from 0 to 1. For values of Y/M below about 0.5 values of M/P appears to stabilise at around 0.55, see Figure 3b and Figure 8. Taking this value and $u \approx 1/(1+Y/E)$ then $Y/P = (Y/M)\times(M/P) = 0.55(Y/M)$ and $E/P = 1-0.55-0.55(Y/M) = 0.45-0.55(Y/M)$ which gives $u = 1-(0.55/0.45)\times(Y/M)$ i.e. u changes linearly with Y/M. It will be shown that u is linearly related to household intensity, see equation 7a. Hence household intensity is linearly related to Y/M, if M/P is constant.

If $M/P = c + mY/M$ where c and $m > 0$ are constants then
$Y/P = (Y/M)\times(M/P) = (Y/M)\times(c + mY/M)$ which is a quadratic.
The maximum for this quadratic is at $Y/M = -c/2m$
Hence also $E/P = 1 - Y/P - M/P = 1-(1+Y/M)\times(c+mY/M)$
$= (1-c)-(m+c)Y/M - m(Y/M)^2$

which is also a quadratic. The minimum for the E/P quadratic is at $Y/M = -c/2m - 0.5$, i.e. at a value of Y/M always 0.5 less than the Y/M value for the maximum for the Y/P quadratic. This can be seen approximately in Figure 3b. The linear regression fit of M/P on Y/M for the year 2000 (Figure 3b) gives values for $m = -0.18$ and $c = 0.65$ with $R^2 = 0.94$ so that the minimum for the E/P quadratic is then 1.3 and the maximum for the Y/P quadratic is 1.8.

If M/P were a quadratic in Y/M then Y/P and E/P would be cubics in Y/M.

Overall the pattern of age ratio relationships seen in Figure 2 and 3a and 3b is termed the age ratio structure for the population. The structure provides a means for assessing the degree to which the 141 country by two data is representative of the world population. It also provides a context for comparing observed and estimated household intensities. For sections of a country only part of the structure may be observed. For example for the 31 local government areas for Melbourne (Australia) in 1996 only that part of the structure with $Y/M < 0.7$ is seen (Department of Infrastructure: 1998).

Figure 3a Values of Y/P, M/P, u, and Ef/P against Y/M for 192 countries at year 1950

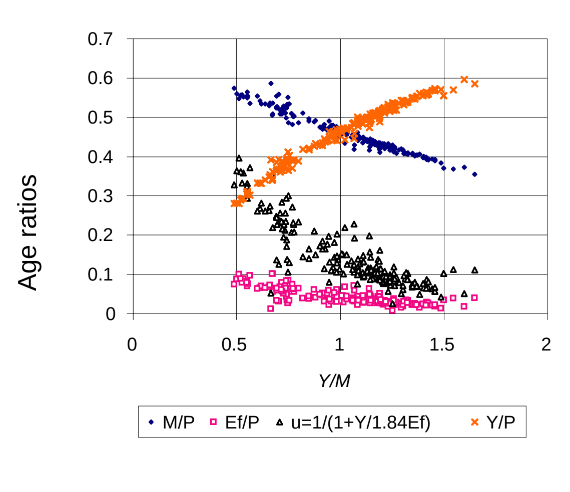


Figure 3b Values of Y/P, M/P, u, and Ef/P against Y/M for 192 countries at year 2000

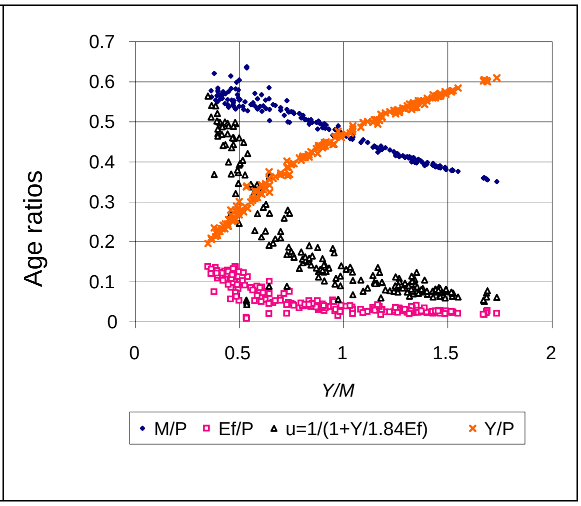


Figure 3c Values of Y/P, M/P, u, and Ef/P against Y/M for 192 countries at year 2030. Low fertility variant

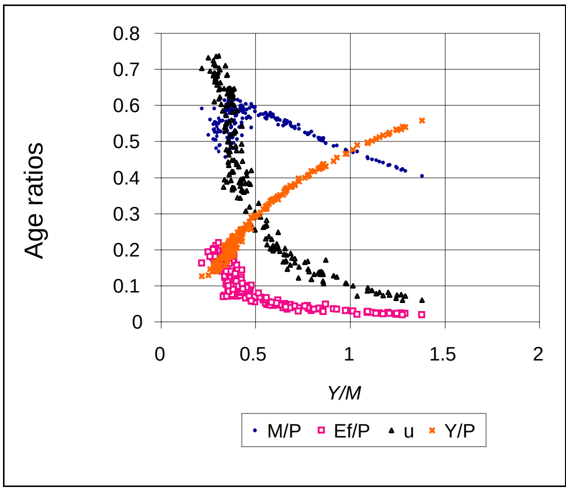
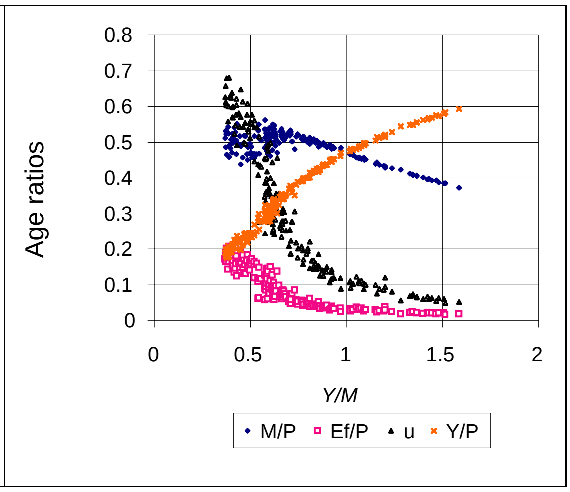


Figure 3d Values of Y/P, M/P, u, and Ef/P against Y/M for 192 countries at year 2030. High fertility variant



Source: United Nations 2002 Revision and authors' calculations.

The main difference between Figure 3a and Figure 3b is that in the latter the curves are extended further to the left because there are lower Y/M values. The elderly component increases with lower Y/M, but so does the proportion of middle persons in the population. In Figure 3a with values of Y/M of about 0.7 the curves are broken. Particular examples of points for u well below trend lines are Guam with coordinates for Y/M and u of (0.67,0.05), São Tomé and Príncipe (0.69,0.14), China–Macao SAR(0.70,0.13), China-Hong Kong SAR(0.75,0.11) and Israel (0.75,0.13). All these countries were countries with small or very small populations at 1950 with perhaps some readjustment still occurring after the Second World War. A detailed investigation of the underlying relationships between Y and the relative proportions of M and E is beyond the scope of this paper.

Figures 3c and 3d are snapshots of projections of age ratios at 2030 for the low and high fertility variants. They show patterns fairly consistent with Figures 3a and 3b. The low fertility case in Figure 3c shows a further movement to the left, while the high fertility case shown in Figure 3d shows a lesser movement to the left. For Y/M greater than 0.5 the distribution of M/P is linear, but for Y/M less than 0.5 it bunches. Ef/P also tends to have a wider spread of age ratio values for Y/M below 0.5.

To what degree is the projected population distribution through to 2030 consistent with the past distribution of the population or with the data used to calibrate the h-age ratio models? Using the age ratio structure of the population comparisons can be made. A regression fit of the distributions of u against powers of Y/M for the already listed 192 countries is made. Table 1 shows the coefficients of the fitted curves for u for the years

1980 and 2000. It also includes the coefficients for the United Nations 2002 Revision high fertility, medium fertility and low fertility projections at 2030. As was shown earlier specifying one of the age ratios together with Y/M is sufficient to specify the others. Since u is linearly related to the bounded model of household intensity, see equation (7a), the comparison also directly relates estimates of household intensities to the age ratio data on which they are based. It is also linearly related to Ef/P. Hence the sensitivity of the household intensity estimates to changes in age ratios can be observed.

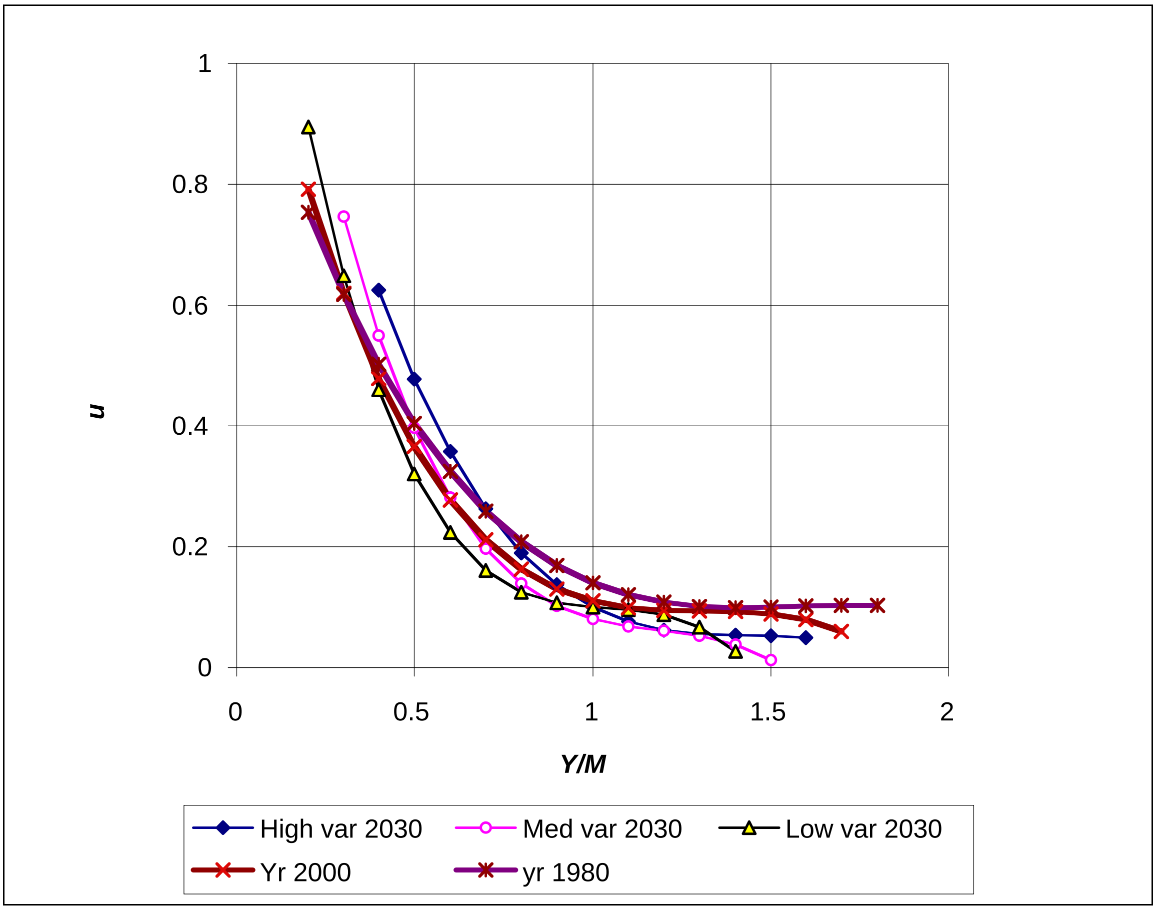
In Figure 4 the values of u against the cubic function of Y/M are plotted. These curves summarise the trends for each of the data sets. The fitted curves for 1980 and 2000 cover approximately the range for the 141 country-by-two data used for calibrating the additive and bounded models and also summarise the trends for the 192 countries for 1980 and 2000. For the range of Y/M from about 0.5 to 1.2 the difference between the 141 country-by-two distribution and the United Nations 2002 Revision projections is not great but increases outside that range. For $Y/M > 1.3$ the number of countries is decreasing over time. However for $Y/M < 0.5$ there is an increasing number of countries. In this case u for the year 2000 is low compared to the United Nations 2002 Revision medium fertility projection, particularly for $Y/M < 0.3$. Since u is equivalent to Ef/P this means that the data used for calibrating the models will give results that lead to low estimates, assuming the United Nations population projections turn out to be accurate. However it is noted that the United Nations 2002 Revision projections are high compared with the population data of the last fifty years as can be seen on the graphs. In addition values of u are very sensitive to Y/M values for $Y/M < 0.3$. The degree to which these factors are significant will be discussed under 'Results'. Where appropriate throughout the paper the coefficient of determination R^2 is adjusted for the number of explanatory variables used.

Table 1 Coefficients for regression of u against powers of Y/M for 1980 and 2000 population data and United Nations 2002 Revision projections for the low, medium and high fertility variants for 2030

	1980 data	2000 data	Low var 2030	Medium var 2030	High var 2030
Intercept	1.09	1.24	1.59	1.64	1.56
Y/M	-1.92	-2.62	-4.23	-3.86	-3.12
$(Y/M)^2$	1.23	2.00	4.03	3.20	2.16
$(Y/M)^3$	-0.26	-0.51	-1.29	-0.90	-0.50
s	0.044	0.054	0.064	0.055	0.050
R^2	0.86	0.87	0.91	0.92	0.93

Source: United Nations 2002 Revision and authors' calculations.

Figure 4 Plot of equations for *u* as a function of *Y/M* for 192 countries at 1980 and 2000 and for the United Nations three fertility variants at 2030



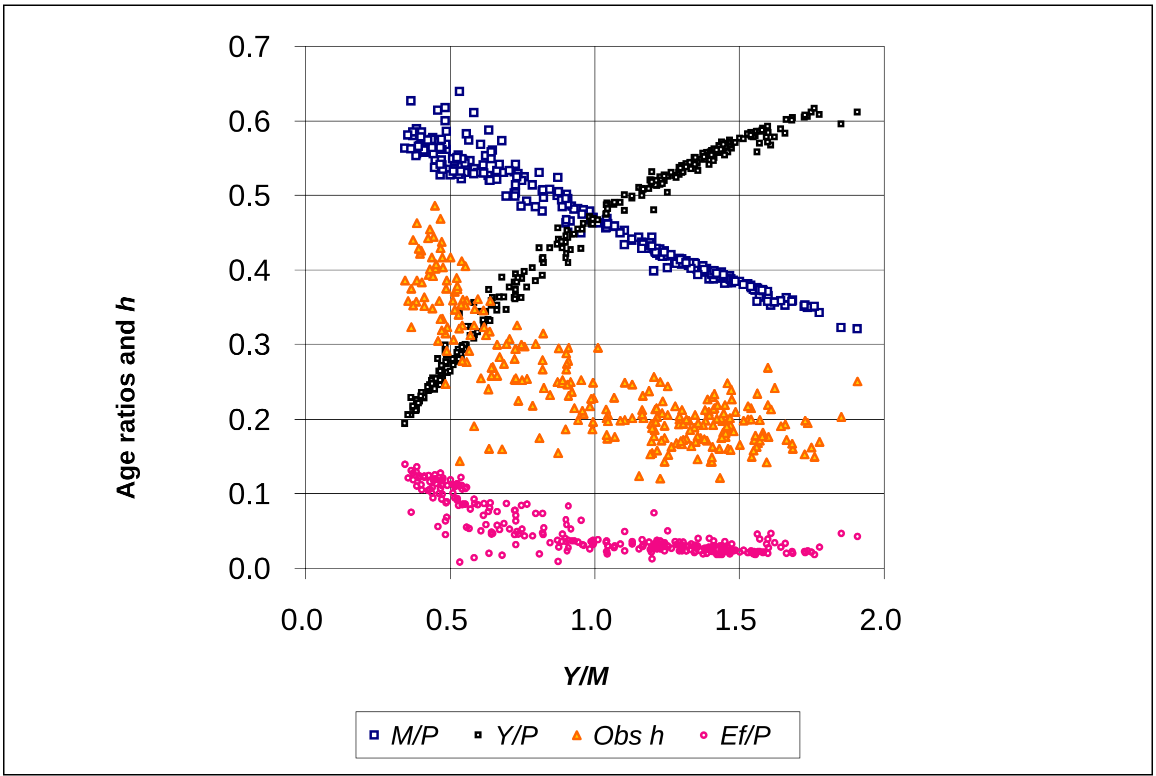
Source: United Nations 2002 Revision and authors' calculations.

Household intensity and age ratios

Data used for modelling

In view of the exploratory nature of this paper the widest possible range of countries is used as a source of data for modelling purposes. There are 141 countries for which both population and household data for at least two censuses are available out of a possible 192 countries. At the year 2000 these countries account for 90.6 per cent of the world's population. These countries are represented by their age ratio structure in Figure 5 and broadly replicate the 196-country age ratio structure at the year 2000 shown in Figure 3b. For example the regression of M/P on Y/M for the data shown in Figure 5 gives $M/P = 0.642 - 0.170Y/M$ with $s = 0.014$ and $R^2 = 0.96$ whereas for the 192 countries represented in Figure 3b it is $M/P = 0.649 - 0.179Y/M$ with $s = 0.017$ and $R^2 = 0.94$. Since values of M/P also define Y/P and E/P the values of M/P can be taken as representative of age ratio structure of the 141 countries. The bias therefore arising from the exclusion of the 51 countries is likely to be small. Household intensities are also plotted in Figure 5. The aim in this paper is to replicate the general trend in observed household intensity data using age ratio models. In Figure 5 since M/P implies E/P ,the ratio Ef/P is shown instead since it is used in the h-age ratio models.

Figure 5 Age ratio structure for the 141 country mean-pairs data



Source: United Nations 2002 Revision and United Nations and others sources and authors'calculations.

The latest available data from each of these 141 countries and the next most recent set about a decade earlier are used to develop two h-age ratio models. Each data point has household numbers, and population numbers by age and sex. The two data points for each country are not independent of each other in that the earlier point gives an indication of the values of the later point. The pair of points may however be considered as independent samples of the random variation that occurs in the data for a particular country. It was decided for modelling purposes to take the mean of the two counts of persons in each of the three age groups, and also of households for each pair, and use these 141 sets of points as data for calibrating the models. This is termed the country mean-pairs data. This introduces a small error due to the non-linearity of the underlying processes determining the data points, but reduces the error due to random variation of the country data. The standard error is very similar for the lower half or the higher half whether the 141 country-by-two data are divided according to country population size or according to household intensity. The higher household intensities in the later period for each country are mostly accounted for by changes in the age ratios rather than time.

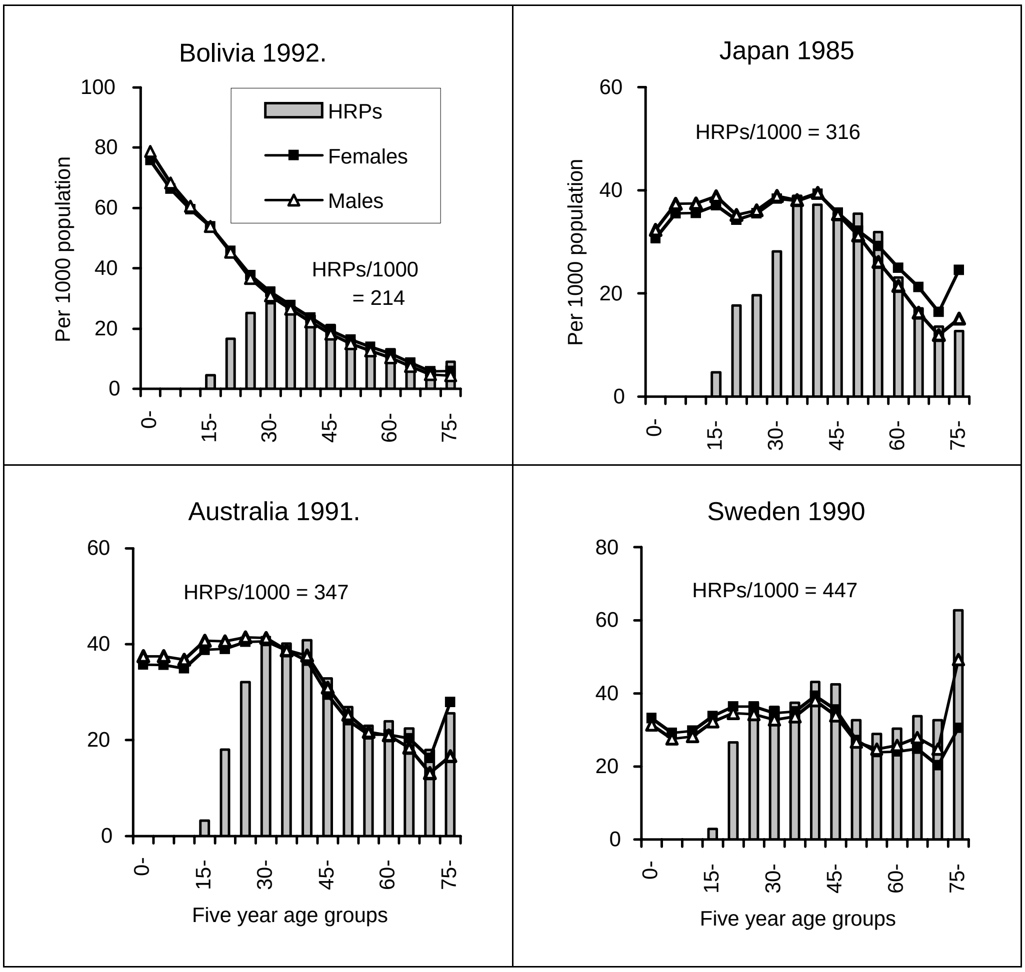
Household formation and numbers of adults

Households may be distinguished by their household reference person (HRP) or household head. For a more extensive definition of the HRP see United Nations (1997b: 107). This person is usually an adult member of a household who is responsible for signing off on the census form. The total number of HRPs is equal to the total number of households (H) in a population. Census data, and general observation confirm clustering by age. That is the age group to which the HRP belongs, is generally close to the age group of at least one other adult member of that household of the same generation, if there is more than one adult of the same generation in the household.

Taking household age as the age of the HRP makes it possible to compare population and households by age strata. A 41-country data set of HRPs and population was used to explore relationships between households and population by age, see Appendix 1 for details of the data set. Figure 6 shows the distributions for four countries, which are selected to represent the range of household intensities. The countries range from Bolivia with low household intensity to Sweden with high household intensity. A comparison is obtained by dividing the household and population counts for each country and age group by the total population P for the country and multiplying by 1000.

Each graph shows two curves close together that connect the ordinates for the number of females and males for each five-year age group. The column graph shows the ordinates for HRPs (households) for each age group. From about the age thirty there are about as many HRPs as there are females or males. There are four distinct phases shown. Up to age 14 years there are no HRPs. Then from age 15 to about age 29 years there is a transition from zero HRPs per age group to HRPs roughly equal to the number of females or males. Billari (2001) in Figure 5 ‘State Distribution by Age’ provides a comprehensive description of factors underlying this age transition. From age 30 years to about age 59 years a stable period follows with about one HRP to every two persons in each age period. Finally there is the 60+ age period where there is slightly more than one HRP per two persons, reflecting the higher proportion of single person households. The upturn at the right hand end is due to the compression of data into a 75+ age category. In reality the number of HRPs and persons per five-year age group continue to decline over these ages. Bolivia contrasts with Sweden in that the ratio of the number of HRPs to half the population is much lower, resulting in a household intensity (households per thousand persons) for Bolivia (214) being about half that for Sweden (447).

Figure 6 Number of males, females and household reference persons (HRPs) in five year age groups per thousand persons in the population



Source: Calculated from the 41-country data set using United Nations and Australian Bureau of Statistics sources, see Appendix.

Note: Each unit on the x-axis indicates five years.. '0-' refers to age 0-5 years, '15-' refers to age 15-19 years and so on.

The outstanding relationship observed is that the number of households is roughly equal to half the adult population. This is independent of the stage in economic development of the country or of population size. It is nearly independent of fertility level. These graphs show period data, but cohort analysis could be used if charts were obtained for consecutive censuses for particular countries. This would help to indicate the degree to which trends independent of age factors influence household intensities. See for example Masnick, Pitkin, and Brennan (1991:163-164). In the case of Australia from 1991 and 1996 there is a slight reduction in estimated household intensities (about 1%) arising from the time component of a linear regression model. For the purposes of this paper the period data are adequate to show the relationships between population and households for the three major age groups, 0 to 19, 20 to 59, and 60+ years for the 41 countries. Assuming independence between the age groups, for the middle group aged 20 to 59 years (middle households designated as H_{20-59}) a regression equation is fitted to obtain $H_{20-59}/M = 0.502 - 0.116\,Y/M$ with a standard error s of 0.0495. This indicates that as the Y/M ratio reduces then the ratio of households to the middle group increases. This may be compared with the growth also in M/P with smaller Y/M as seen in Figure 3a, 3b, 3c, 3d. For the elders, households and persons age 60+ years, the average household intensity is 0.585 with a standard error of 0.0645. There are a small number of HRPs in the 0 to 19 age group and the average household intensity is 0.00631 with a relatively large standard error of 0.0055.

In general only the total number of households H is known, not a division by age. Most countries have values of Y, M, E and H. For comparisons across countries these values are divided by P, the population of a particular country. It is then possible to fit the age ratio data to observed household intensities for the 41 case data. The following regression equation is obtained where $\hat{h}$ is the estimated household intensity

$$\hat{h} = -0.2295 + 0.2362\,Y/P + 0.5002\,M/P + 1.3826\,E/P \qquad R^2 = 0.88 \qquad (2)$$

but this equation over-specifies the variables since $Y + M + E = P$.

Substituting for Y in equation (2) yields

$$\hat{h} = 0.0068 + 0.2640\,M/P + 1.1464\,E/P \qquad s = 0.02796 \qquad R^2 = 0.89 \qquad (3)$$

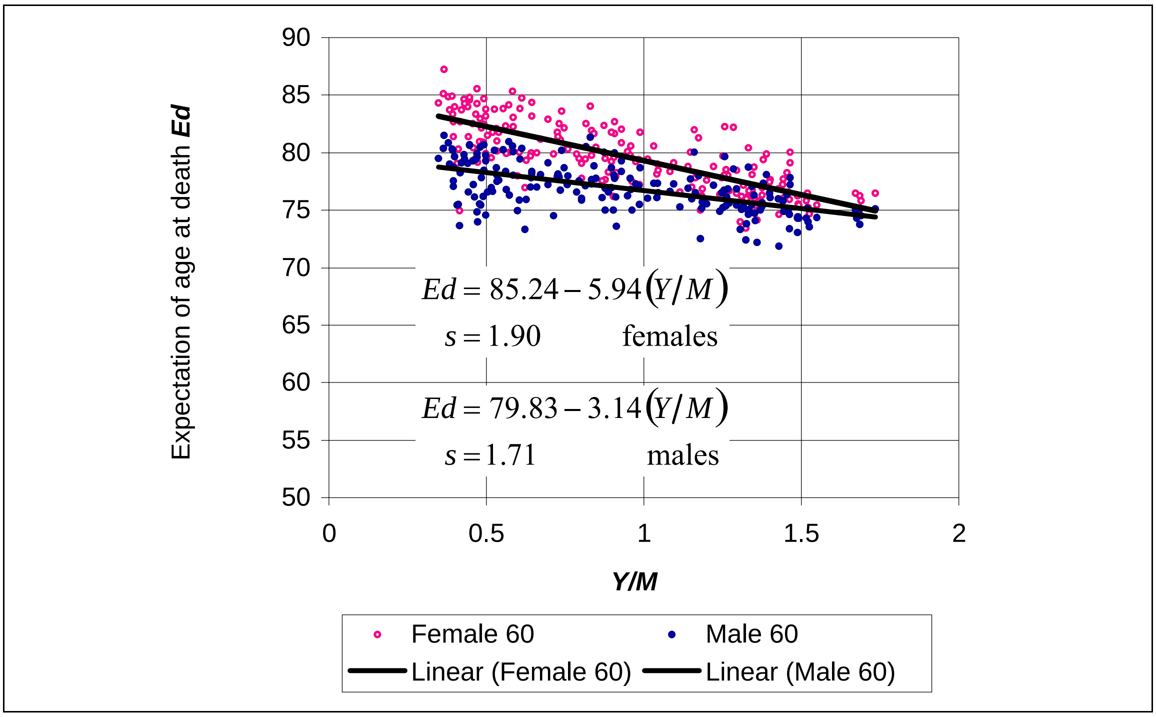
where s is used to denote the standard error.

The data are derived from age clusters and are therefore imprecise. Also the assumption of independence between the household numbers for the three age groups may not be observed. It also appears from the data that the number of older households is more closely related to the number of females age 60+ than the number of 60+ males.

Examination of the much broader data set, the 141 country-by-two data, and past data over the last five decades for large groups of countries shows that replacing E by number of elderly women (Ef) in equation (3) improves the regression fit of the models. This is due mainly to the higher life expectancy at age 60 of females as compared to males and hence more HRPs who are widows.

Figure 7 shows that with decreasing Y/M, the life expectancy at age 60 of females tends to increase at a greater rate than that of males. Expectation of age at death is equal to life expectancy at age 60 years plus 60 years. Since it appears that the number of elder households is more closely related to the number of elder women it was decided to use elder women as the indicator of elder households. This appears contrary to the assertion by de Beer and Alders (1999) that the difference between the number of elder females and elder males will become less in the future, at least in The Netherlands. Here it is taken that the major influence on change in the relative proportions will be the change in the ratio Y/M, rather than time. To date, for the authors, the inclusion of time as a variable in the age ratio regression equations has not improved the fit significantly. This could be regularly reviewed.

Figure 7 Expectation of age at death of person aged 60 years in the period 2000 to 2005, for 192 countries where *Y* = persons aged 0 to19, *M* = persons aged 20 to 59, *Y* and *M* estimated at the year 2000



Source: United Nations 2002 Revision and authors' calculations.

Additive model

The additive model is therefore developed taking $H = bM + cEf$ where Ef is the number of females aged 60 years or more in the population and b and c are parameters to be estimated. For comparisons across a number of countries both sides of the equation are divided by the population of the country P and an intercept term a is added to allow for an unbiased regression fit over the practical range of the age ratios. Thus

$h = a + b(M/P) + c(Ef/P) + e$ where h is the household intensity and e is the residual.

Fitting a regression model to the 141 country mean-pairs data, equation (4) is obtained. See Appendix 1 for details of the data set.

$$\hat{h} = 0.0677 + 0.1752\,M/P + 1.9568\,Ef/P \qquad s = 0.0288 \qquad R^2 = 0.877 \qquad (4)$$

The frequency distribution of fitted less observed household intensities is approximately normally distributed and kurtosis = 0.33 and skewness = 0.02. The upper bound of 2.02

for this equation occurs when $Ef/P = 1$ and is impractical. The range of application of the model is determined by the range of the age ratios of the data. Discussion on extrapolation beyond this range is given later under 'Results'.

By multiplying both sides of the equation by P the additive form is obtained

$$\frac{1}{r}\hat{H} = 0.0677P + 0.1752M + 1.9568Ef \qquad s = 0.0288\,P$$

where *r* is the ratio of the population in households to the total population ie $P_H/P = r = 0.971011$ as described earlier, and $\hat{H}$ is the estimated number of households.

Assuming that households are related to M and E as discussed above then the age ratio structure would indicate an alternative version of equation (3) where E/P is replaced by a quadratic in Y/M. Fitting this regression model to the 141 country mean-pairs gives

$$\hat{h} = 1.702 - 1.721M/P - 0.868Y/M + 0.201(Y/M)^2 \qquad s = 0.0287 \qquad R^2 = 0.878 \quad (6\ )$$

For the purposes of this paper equation (4) is preferred because it is additive and relates directly to Ef. However equation (6) provides additional confirmation of the close relationship between household intensities and the age ratio structure.

Bounded model

In Figure 8 $(M+E)/P$ and M/P are plotted against household intensity *h* for the 141 countries-by-two data. Assume for this analysis that $P_H = P$. For household intensity $h > 0.25$ the ratio M/P is approximately constant at about 0.55. The ratio $(M+E)/P$ is taken to increase approximately linearly from about the point $(0.25, 0.55)$ to a point with coordinates $(h,\ 0.55 + E/P)$. This line has the slope given by $(E/P)/(h - 0.25) = a$ where *a* is a constant so that $h = 0.25 + E/aP$.

Since $M/P = 0.55$ and $(Y+M+E)/P = 1$ then $(Y+E)/P = 0.45$ so that $P = (Y+E)/0.45$. Substituting for P in the equation for *h* and simplifying

$$h = 0.25 + \frac{0.45}{a(1 + Y/E)}$$

Thus it is expected that a relationship of this form should approximately fit the 141 country mean-pairs data. For the reasons given in the section on the additive model E is replaced by $1.84Ef$ where 1.84 is the average of the ratios of E/Ef over the 141 countries-by-two data. Using this approach a regression fit is obtained as follows:

$$\hat{h} = 0.1490 + 0.5310/(1 + Y/1.84\,Ef) \qquad s = 0.0286 \qquad R^2 = 0.878 \qquad (7)$$

or

$$\hat{h} = 0.1490 + 0.5310u \qquad s = 0.0286 \qquad R^2 = 0.878 \qquad (7a)$$

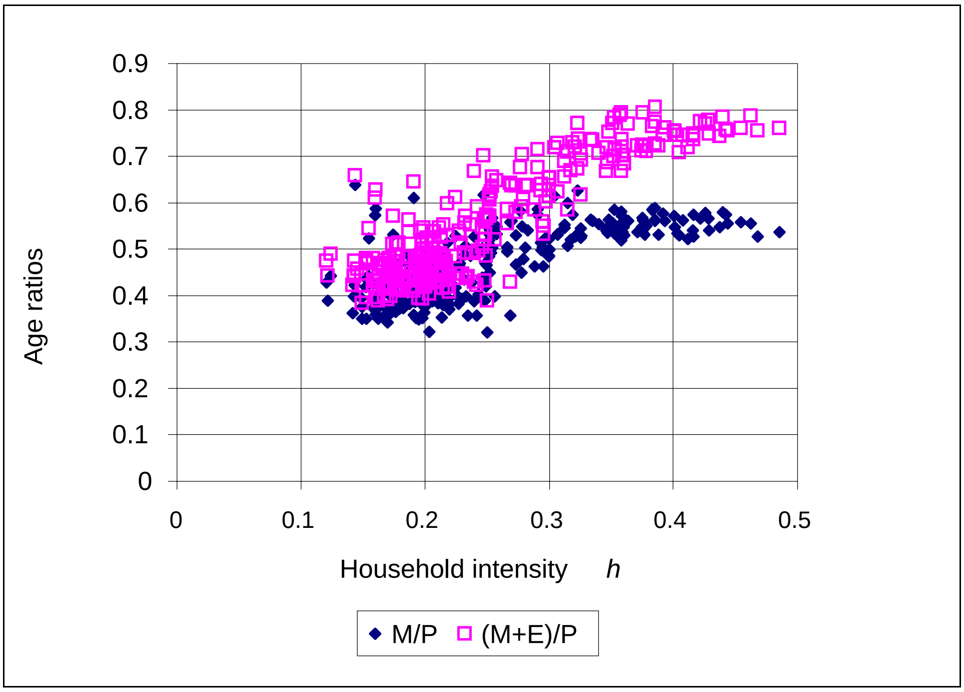
where $u = 1/(1 + Y/1.84\,Ef)$

By multiplying both sides of equation (7a) by P the following form is obtained

$$\frac{1}{r}\hat{H} = 0.1490P + 0.5310uP \qquad s = 0.0286P \tag{8}$$

where r is the ratio of the population in households to the total population i.e. $P_H/P = r = 0.971011$ as described earlier.

Figure 8 Two population ratios against household intensity; the proportion of adults in the population $(M + E)/P$, and the proportion of the middle persons aged 20 to 59 years in the population M/P. Data from 141 countries-by-two



Source: Various United Nations sources, see Appendix 1, and authors' calculations.

Equation (7) is termed the bounded model since it bounded within feasible limits for all positive values of Y/Ef. Maximum value is 0.680 when $Y/Ef = 0$ and there is a lower bound of 0.149 as Y/Ef becomes large. The frequency distribution of fitted less observed household intensities is approximately normally distributed and kurtosis = 0.66 and skewness = 0.02.

Assuming that households are related to M and E as discussed above then the age ratio structure would indicate an alternative version of equation (7) where Ef is replaced by E and a function in Y/M is obtained. Fitting this model to the 141 country mean-pairs gives

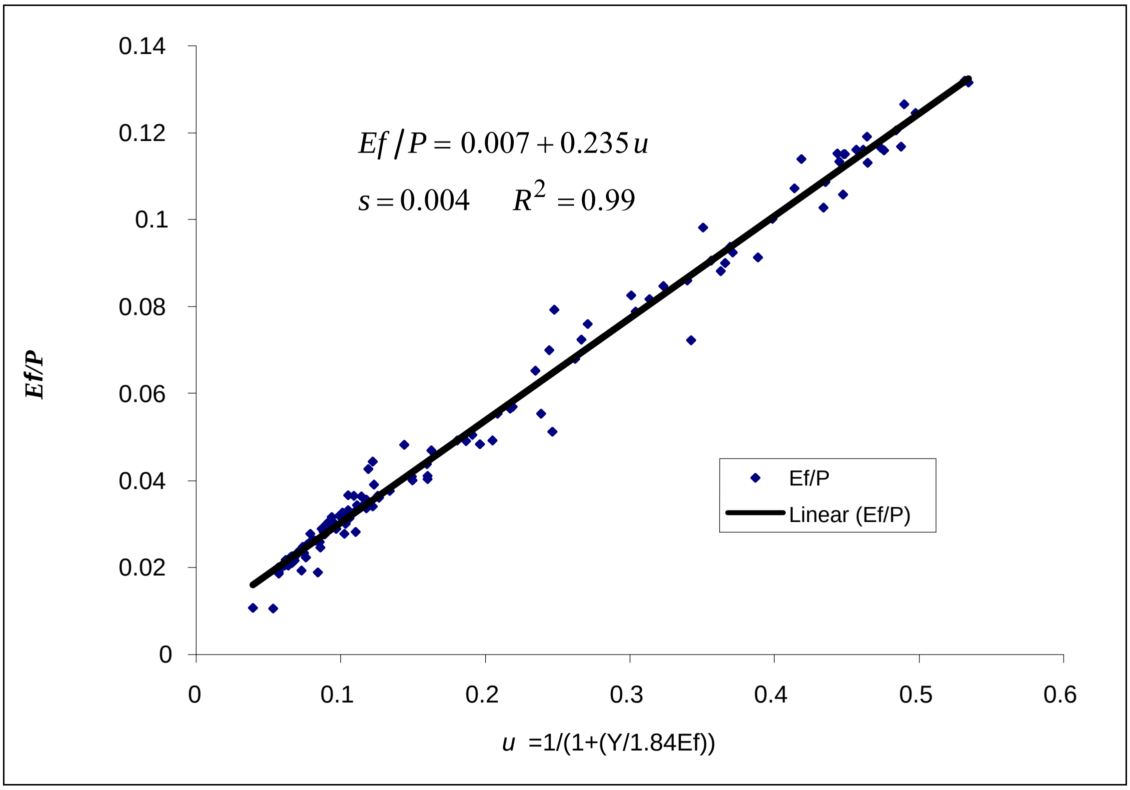
$$\hat{h} = -1.270 + 0.328\,Y/M + 1.978/(1 + 0.7Y/M) \qquad s = 0.377 \qquad R^2 = 0.788 \qquad (9)$$

For the purposes of this paper equation (7) is preferred because it relates directly to Ef and is more precise. However equation (9) like equation (6) also, provides additional confirmation of the close relationship between household intensities and the age ratio structure.

Comparing the two models and the combined model

The bounded model shown in equation (7a) enables the direct plot of h against u, which is not possible for the additive model. The value Ef/P is linearly related to u, see Figure 8, and hence equations (7,7a,8) bear a close empirical relationship to equation (4).

Figure 9 Ef/P **plotted against** $u = 1/(1 + Y/1.84\,Ef)$ **for 141 country mean-pairs data**



Source: United Nations 2002 Revision and authors' calculations.

For example the intercept in equation (4) is small compared to that in equation (7). Since M/P ranges from 0.4 to 0.6 if M/P is assumed constant at 0.5 and this value is substituted in equation (4) then the new intercept value is 0.156, which is similar to that for equation (7). From Figure 9 the remaining variable Ef/P is approximately equal to $0.24/(1 + Y/1.84\,Ef)$ which if substituted in equation (4) gives a value of $0.47/(1 + Y/1.84\,Ef)$ which is similar to that for equation (7). In practice therefore the additive and bounded models give very similar results. Both models indicate that about

87 per cent of the variation in household intensities can be estimated to vary according to age ratios. Using the additive model about half of this variation can be attributed to the ratio Ef/P, and the other half due to variation in M/P.

The two models provide an independent check on each other, and both will be used in unison where possible. That is, the mean of the two estimates of household intensity for a country for a particular period will be used. This mean will be termed the combined model. The United Nations 2002 version has summary world data, which includes a number of very small countries for which separate age data is not provided. In this case the additive model only is used. For the fertility cohorts both models are used. It is possible to adjust the intercept for each model so that for a particular country the estimate of household intensity equals the observed at the year 2000, the jumping off point. However this adjustment, which means a separate equation for each country, does not alter the overall projections significantly, and therefore is not used for world and sub-world level projections.

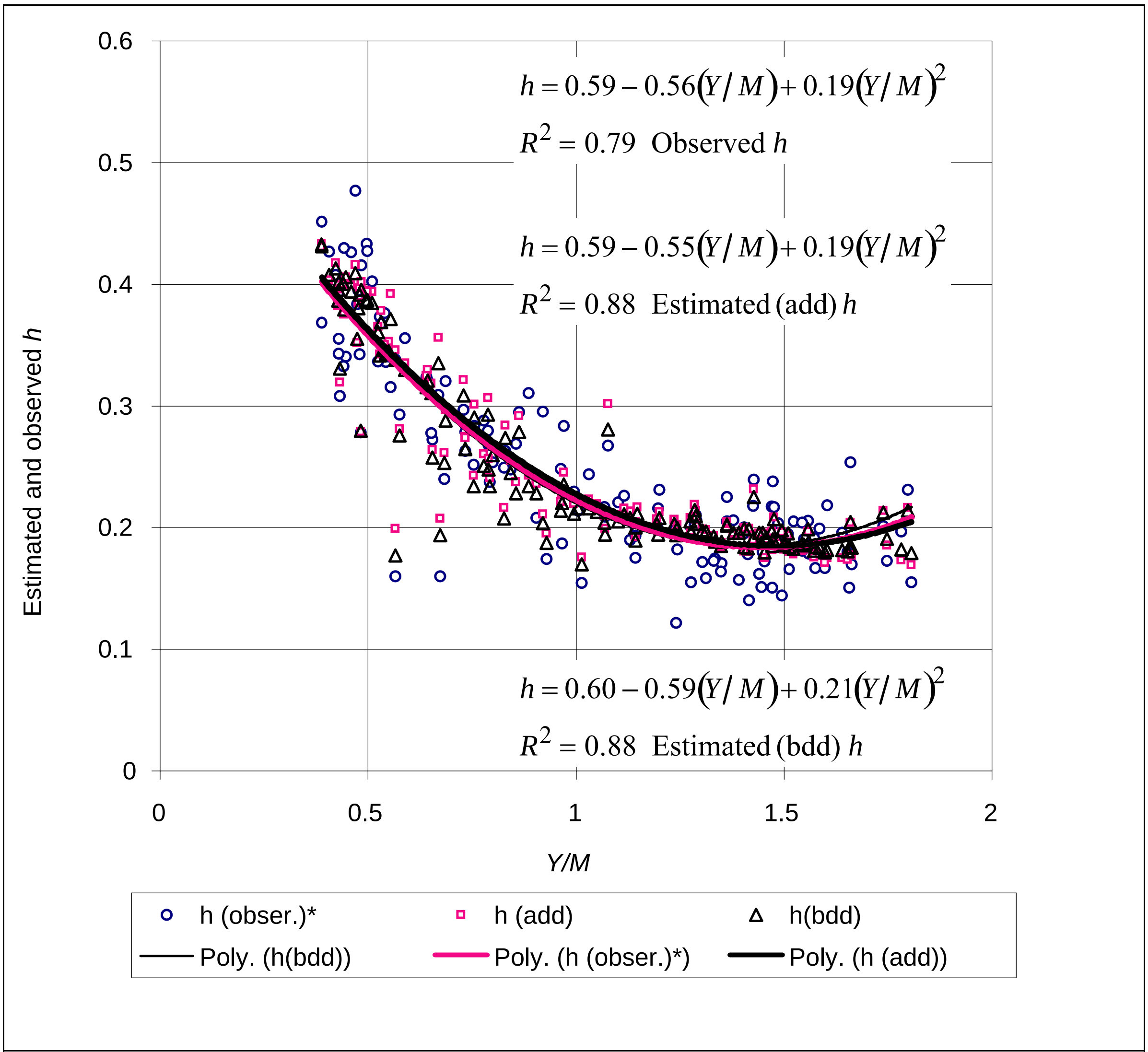
Model evaluation

Household intensities

The additive and bounded models are used to obtain estimates of household intensities for the 141 mean-pairs data. These estimates together with observed household intensities are plotted against Y/M in Figure 10. The equations for the regression fits to the three sets of data are also shown in Figure 10 and the fits are plotted. The curves are very close to each other indicating that the modelling process can replicate trends in the original observed intensities as Y/M changes.

The central assumption in this paper is that all countries will follow along the same household intensity to age ratio path. The age ratio structure shown in Figure 2 is stable, ie for any given value of Y/M there is a unique set of age ratios (subject to residual error). The age ratios for the 141 country mean-pairs data used for modelling household intensities also have this structure. This is demonstrated by the close relationship between the M/P distribution in Figure 5 and that in Figure 3b. For such a set of age ratios there are unique estimates of household intensity obtained, by equations (4) and (7). Thus a trend line based upon these estimates will also be stable. It is projected that many countries in the next 30 years will still have Y/M values within the range of the 141 country mean-pairs data and hence the trends in the model estimates will follow the observed intensities, as seen in Figure 10. A discussion of cases where Y/M is below currently observed values will be given under ‘Results’.

Figure 10 Plot of observed household intensities against Y/M for the 141 country mean-pairs data and also estimates of household intensities using the additive and bounded models, together with regression fits



* h(obser.) represents observed household intensities, h(add) represents additive model estimates of household intensities, h(bdd) represents bounded model estimates of household intensities.

Source: United Nations 2002 Revision, other official sources (see Appendix 1), and authors' calculations.

Estimated household intensities are compared with observed intensities for the 141 countries-by-two data using the combined model, see Figure 11a. The estimated household intensities show a similarity of spread from the observed over a range of observed household intensities from 121 to 486 households per thousand.

The variability seen in Figure 11a represents a 'snapshot' of a continuous sheaf like process shown in Figure 11b. This latter is a plot of estimated against observed household intensities over the last few decades for 99 countries which appear as country 'traces' concentrated along the 45-degree line. It does not follow therefore that individual countries may be able to assume different relationships to the 45-degree line in the short term. This has implications for the way confidence limits may be derived. The approach adopted here is to assume that the variation in the residuals (observed less estimated) as shown in Figure 11a will occur in the longer term, say 20 years or more, for any particular country. Measures of this variation will thus be used for determining confidence intervals and will be discussed later. This assumes perhaps more variation than will occur in practice.

Household numbers

The household intensity data are replotted as household numbers at various scales, see Figure 11c, 11d, 11e and 11f. Applying the combined model to the 141 countries-by-two data gives a ratio of estimates of total households to observed of 1.015. In Figure 10c the highest two values out of the 282 data sets are for China, the next two are for India and the next two are for United States. An ordering of the 282 data sets according to the number of households in each data set shows roughly a ratio scale for household numbers, i.e. if N is the number in one data set then the adjacent data set with more households will have approximately Ne^r households where r is a positive constant. This appears to be related to the power law for size distribution.

The same general pattern applies if all the cases are plotted or if only the lower half by household number are plotted, the lowest quarter, or the lowest one eighth. See Figure 11b, 11c, 11d, 11e and 11f. Most of the extreme cases show estimated households to exceed observed households. In Figure 11c for example estimated households in India in 2001 exceed observed by about 46 million households. This is probably due to the sharing of accommodation by two or more couples as described in the Indian Census (2001). A few outliers are shown in Figures 11d, 11e, 11f. The points representing these countries are shown as empty circles. Overall there are relatively few outliers. These occur partly because the models are unable to represent special living arrangements. For example in Kuwait the male population which includes migrant workers greatly exceeds the females population. There is an overestimate of the number of households since many of these are guest workers do not form households in Kuwait.

Figure 11a Estimated household intensity *h* against observed. 141 countries-by-two (households per thousand persons) using the combined model

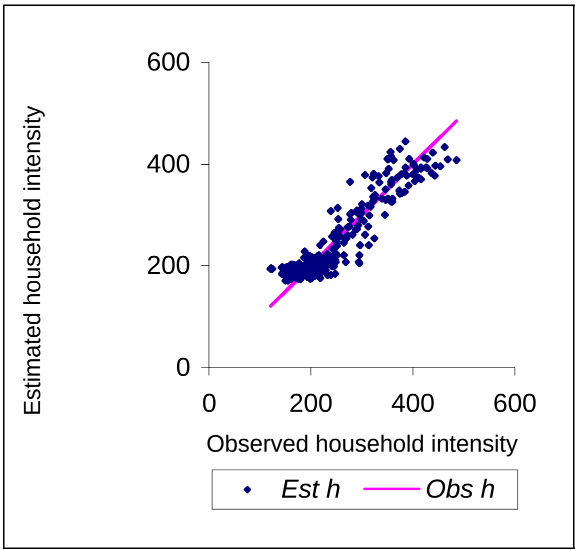


Figure 11b Estimated household intensity *h* against observed. 354 cases forming traces of 99 countries

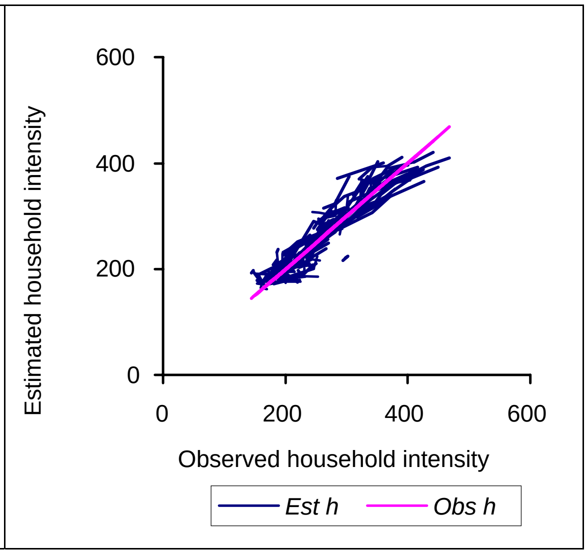


Figure 11c Estimated number of households *H* against observed. 141 countries by two (millions of households)

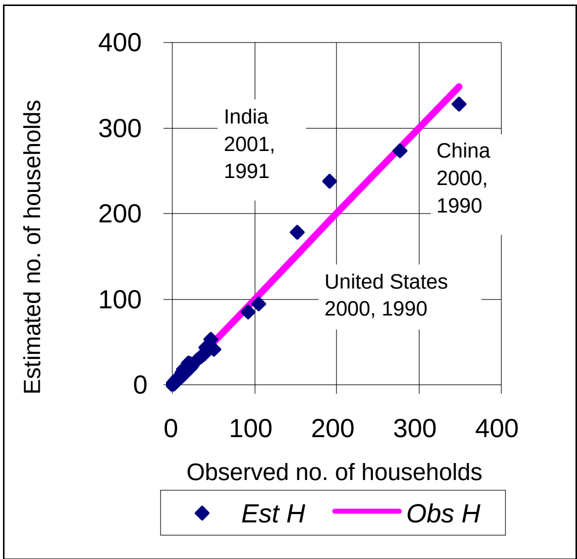


Figure 11d Estimated number of households *H* against observed. 141 data sets out of the 282 with the lowest number of households (millions of households)

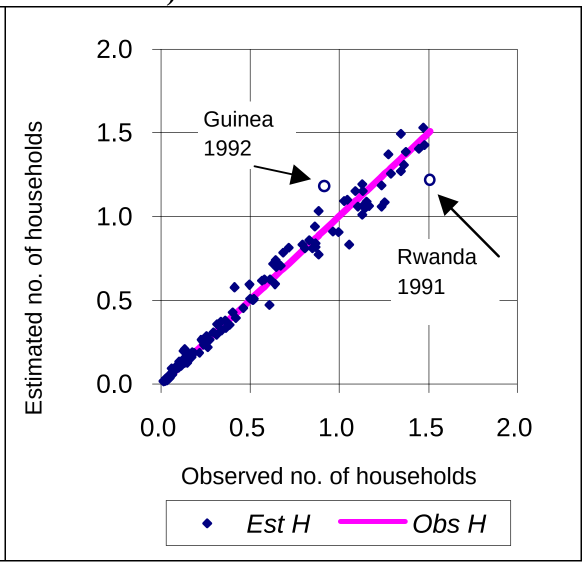
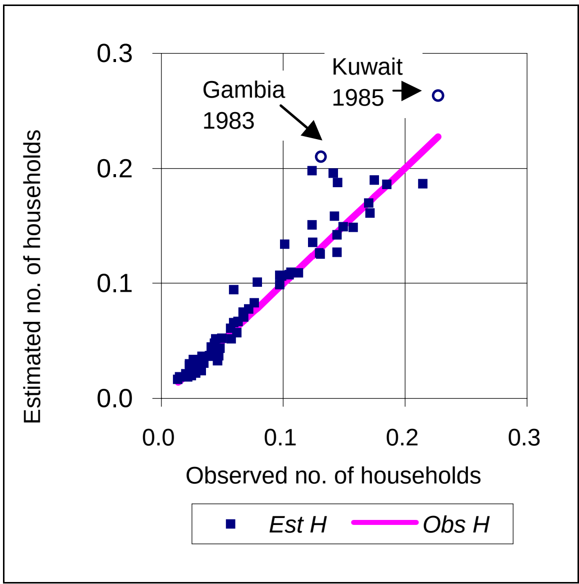
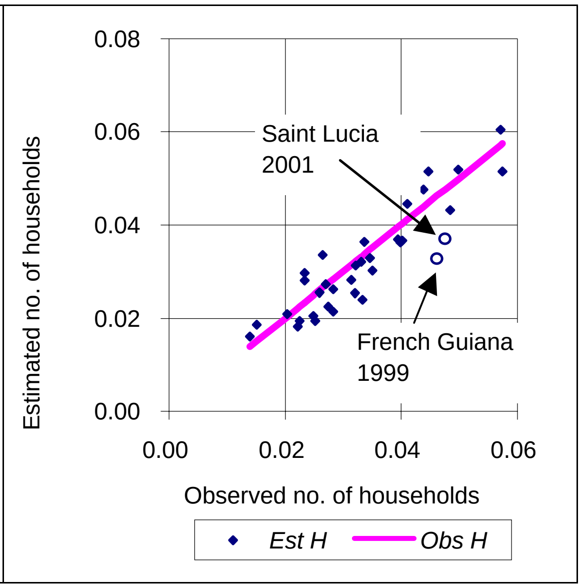


Figure 11e Estimated number of households *H* against observed. 70 data sets with the lowest number of households (millions of households)

Figure 11f Estimated number of households *H* against observed. 35 data sets with the lowest number of households (millions of households)





Source: United Nations 2002 Revision, and various other sources (see Appendix 1), and authors' calculations.

Testing using recalibration

A test of the degree to which past data can be used to indicate current *h* -age ratio relationships is carried out. There are 42 data sets (separate countries) from the 1950s. For each of the data sets for the remaining decades for some countries two data sets are included. There are 83 data sets used in the 1960s, 114 data sets in the 1970s, 148 data sets from the 1980s, 99 data sets from the 1990s and 44 data sets from the 2000s. The two models are recalibrated using each of these sets of data and these recalibrated models are then used to estimate overall household intensities and total household numbers for the 141 country-by-two data. In this way it is possible to determine the error in projections of current household intensities and numbers of households made given only the earlier data. It is assumed that current knowledge of population is available, but not of households.

The results of the calibration test are shown in Table 2. For the 1950s the model estimates are plotted against observed values for household intensities and household numbers and shown in Figure 12a and 12b. A 45-degree line is used for comparison purposes.

Table 2 Comparison of estimated household intensities and household numbers against observed, based on models calibrated using data from the 1950s (1950 to 1959), 1960s, 1970s, 1980s, 1990s; and 1950 to 1999, and the additive and bounded models

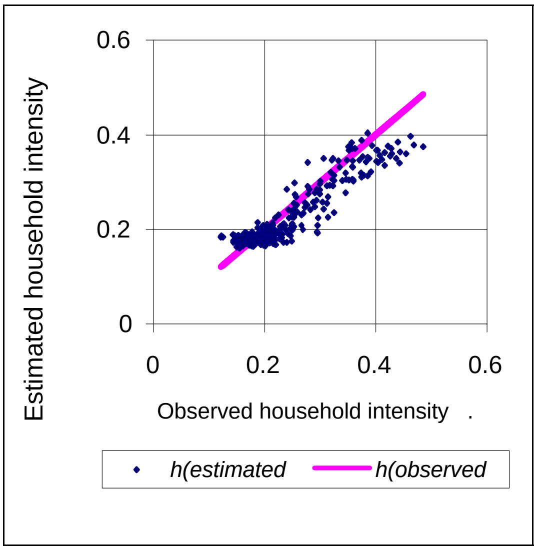
Decade	1950's	1960's	1970's	1980's	1990's	1950-1999	Model
Number of countries	42	83	114	148	99	486	141
Additive model							
Intercept	0.0927	0.0885	0.0522	0.1016	0.0572	0.0751	0.0677
Coeff.of M/P	0.1014	0.1537	0.2373	0.0782	0.1837	0.1526	0.1752
Coeff.of Ef/P	1.824	1.6245	1.4970	1.9425	2.0598	1.8620	1.9568
Std. error	0.0301	0.0286	0.0303	0.0303	0.0335	0.0314	0.0287
Est h/obs h	0.93	0.97	0.96	0.95	0.99	0.97	1.00
Est H/obs H	0.94	0.99	0.98	0.96	1.01	0.98	1.02
Bounded model							
Intercept	0.1430	0.1557	0.1471	0.1425	0.1445	0.1446	0.1490
$1/(1+Y/(1.84Ef))$	0.4869	0.4740	0.4856	0.5174	0.5531	0.5213	0.5310
Std. error	0.0286	0.0287	0.0299	0.0292	0.0335	0.0305	0.0286
Est h/obs h	0.94	0.98	0.96	0.96	1.00	0.98	1.00
Est H/obs H	0.95	0.99	0.97	0.98	1.01	0.99	1.01

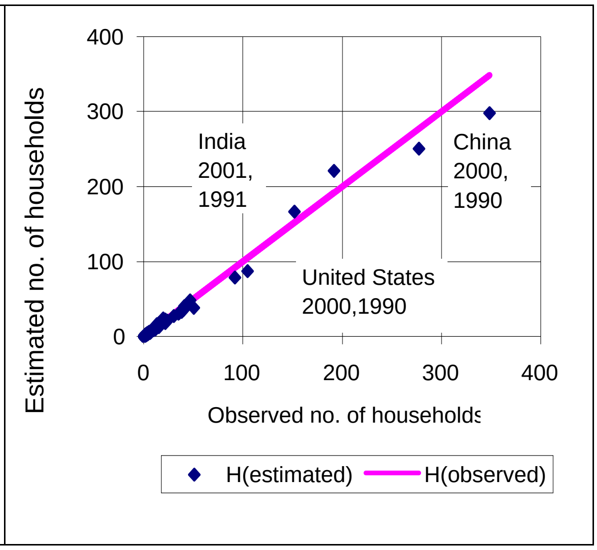
Generally the predictions are reasonably good but are low in some cases. For example the household intensity estimates calculated using the additive model calibrated with the1950s data and plotted in Figure 12a show an underestimation of the higher household intensities. This is possibly because of the low number of countries (42 countries) or because of collection procedures for these early collections. The outstanding characteristic is that the distributions for all the epochs (the 1950s are shown) are similar and show a common pattern. All pairs of graphs (household intensity and number of households) such as Figure 12a and 12b show a consistency of estimated against observed household intensities irrespective of population size. Both models give similar results.

The discrepancies between the observed and estimated households are consistent enough and close enough to the normal distribution to establish 95% confidence intervals using standard techniques. This is taken equal to the estimate for the number of households plus or minus twice times the standard error for the models multiplied by the population in households. This confidence interval is distinguished from the total error, which would include some assessment of the non-sampling error in population estimates, and other factors outside of consideration in this paper. The overall world estimates for households using the additive model are taken to be the additive model estimates of household intensity times the world population in households plus or minus twice times the standard error times the population.

Figure 12a Estimated household intensities against observed. 141 countries-by-two using 1950s calibration of additive model

Figure 12b Estimated number of households (millions) against observed number 141 countries-by-two using 1950s calibration of additive model





Source: United Nations 2002 Revision, and various other sources (see Appendix 1), and authors' calculations.

The influence of China and India on estimates may appear to be considerable since China and India account for about one third of the world's population. However the models used for determining household intensities are not affected significantly whether or not China or India are included since together these countries represent only 2 out of the 141 points used to fit the models.

Residual error for household intensity estimates
Figure 11a shows the residual error after fitting the additive model to the data, and hence the error likely in estimates of projections of household intensities. Using the additive model from equation (4) the 95 per cent confidence interval on modelling error for household intensities is $\hat{h} \pm 0.0576$. This confidence interval is independent of country population size. It is assumed that the population in each country tends to change in unison because of similar laws across the country and much intermingling. Figure 11b indicates that the residual error may persist over several censuses. For individual countries household intensities represented by traces may converge towards the 45-degree line or cross it, or diverge away from it. Knowledge of past history of a trace may

provide some indication of the future trend particularly if one knows why the trend occurred.

Do countries move independently of each other? Figure 2 demonstrates the remarkable structuring of change in age ratios as fertility levels change. The degree of change over time may vary from country to country but not, it appears the age ratio structure. Change of total fertility in the past has exhibited a parallel downwards movement of various countries at similar rates. As a first approach in this paper it is assumed that the confidence interval for the projection is zero now and increases linearly over time to 2050 to that set down in equation (4). That is, the standard error for h used at 2030 will be 0.6 times the standard error used at 2050 or in general will be $(t/50)s$ where t is the number of years projected beyond 2000 and s is the standard error. Further investigation may indicate that the world can be divided into two or more independent parts making a further reduction in the standard error possible. Also standard error could increase at an increasing rate over time.

For a particular country an examination of the residual error over time may show it to be decomposable into a constant plus a smaller residual error, which is increasing over time. The constant then represents a distinct characteristic of the country arising from other factors. The estimate for the country may then be taken to be the global estimate, obtained say from equation (4), plus the constant plus the smaller residual error.

Results

In order to present the 192 country projections more compactly yet still retain the diversity of fertility characteristics the countries are grouped according to the seven United Nations fertility transition cohorts. Table 3 summarizes projections of population and households and the results are graphed in Figures 13 and 14. For the period 2000 to 2030 results are shown for the fertility cohorts. For the period 2030 to 2050 results are shown only at the world level because of greater uncertainties of projections of the fertility cohorts. Over this latter period household projections are increasingly dependent upon fertility assumptions, and there is increasing uncertainty regarding mortality and household to adult propensities. Adults (A) are taken here to be persons aged 20+ years so that $A = M + E$. For comparison the household projections of Habitat'96 are included (Habitat 1996).

Over the next thirty years to 2030 it is projected that there will be between 1 to 1.2 billion additional households formed whether there are 1.3, 2.0, or 2.7 billion extra persons. Over the next fifty years to 2050 there will be 1.6 billion to 2 billion extra households whether there are 1.3, 2.8 or 4.4 billion extra persons.

From equation (5) the 95 per cent confidence interval on modelling error for the world households for the medium fertility case is equal to the estimate of households for the world at the year 2030 of 2655 million $\pm$ twice times standard error times 0.6 times the population, ie

2655 million $\pm \left(2 \times 0.02875 \times 0.6 \times 7{,}894{,}000{,}000 \times 0.971011 = 264 \text{ million}\right)$,

or between 2391 and 2919 million, and for the year 2050 it is
3343 million $\pm(2\times 0.02875\times 1\times 8{,}660{,}000{,}000\times 0.971011 = 484\text{ million})$,
or between 2869 million and 3827 million.

Also shown for the medium fertility variant is a decomposition of household intensity into two components, the proportion of adults in the population $(M+E)/P$, and the ratio households per adult. The world at 2030 is projected to have nearly reached the same adults to population ratio of about 0.7 as the C.7 cohort has now. Assuming $P_H \approx P$ then an adults to population ratio of 0.7 corresponds to a value of Y/P of 0.3. Reading from Figure 3b or 5 this corresponds to a Y/M ratio of about 0.5. At these low values of Y/M household intensities become increasingly sensitive to further reductions in Y/M as seen in Figure 10.

The range of household projections arising from the choice of fertility scenario can be compared with the corresponding range for population using the percentage standardized range. This is 100 times (maximum - minimum) divided by average. At 2030 the standardised range for the population is 16.8 per cent while that for total households is 5.1 per cent. Similar results apply for each of the cohorts with the ratio of the population range to household range varying from 13.4 to 1.7. Globally at the year 2050 the standardised range for the population is 35.9 per cent while that for households is 13.1 per cent. Household projections therefore are still much less sensitive to different fertility variants than are population projections.

The ratio of projections of households at 2030 for the low fertility population variant to that for the medium fertility cohort is about 0.96 for countries in fertility cohort C.1 and C.2, 0.97 for countries in C.3, 0.98 for countries in C.4 , 0.99 for countries in C.5 and C.6, and 1.0 for countries in C. 7. The ratio of projections of households at 2030 for the high fertility population variant to that for the medium fertility cohort is about 1.04 for countries in fertility cohort C.1, C.2 and C.3, 1.02 for countries in C.4 and C.5, 1.01 for countries in C.6, 1.0 for countries in C.7.

The relatively small differences between projections of households for the three fertility variants simplifies comparisons between fertility cohorts up to the year 2030 since data from the medium fertility variant maybe taken to be representative of all variants. Using data from Table 3 a comparison is made between household growth rates and population growth rates for the seven fertility transition cohorts and plotted in Figure 13. For household growth rates the medium fertility variant values only are shown since the other variants are too close to be distinguished. A consistent change in pattern is seen as one moves from cohort C.1 with no fertility transition to cohort C.7 (TF $\leq$ 2.1) with early transition. Except for cohort C.1 household growth rates exceed population growth rates. The disparity between the growth rates for population and households are greatest for the fertility transition cohorts C.5 (2.1 $<$ TF $\leq$3) and C.6 (TF $\leq$ 2.1) and C.7 (TF $\leq$ 2.1).

Table 3 Projected population and number of households grouped by fertility transition cohorts and according to United Nations fertility variants in millions of persons and households; and Habitat '96 projections of households

Cohort		Population in households Millions			Households Millions			Habitat '96[a] Households Millions	Adults /P_H	Hshlds per adult
		Fertility variants			Fertility variants			Fertility variants		
	Year	Low	Medium	High	Low	Medium	High	Medium	Medium	
C.1. Fertility transition not begun TF[c] ≥ 5+	2000	174	174	174	32	32	32	31	0.429	0.425
	2010	232	236	240	42	43	43	42		
	2020	301	314	327	56	57	58	59		
	2030	381	409	436	73	76	79	76	0.472	0.391
C.2. Decline but TF still > 5	2000	566	566	566	108	108	108	93	0.466	0.407
	2010	706	720	734	139	140	141	124		
	2020	846	888	929	178	182	185	167		
	2030	972	1054	1138	223	230	239	210	0.574	0.381
C.3. Decline to 4< TF ≤5	2000	174	174	174	34	34	34	33	0.486	0.402
	2010	203	208	213	43	43	43	43		
	2020	226	240	253	53	54	55	58		
	2030	245	269	295	64	66	68	72	0.627	0.389
C.4. Decline to 3 < TF ≤ 4	2000	1549	1549	1549	341	341	341	297	0.563	0.391
	2010	1754	1797	1841	425	428	430	372		
	2020	1903	2023	2142	533	537	542	449		
	2030	1988	2200	2418	647	658	674	525	0.705	0.425
C.5. Decline to 2.1 < TF ≤ 3	2000	791	791	791	185	185	185	193	0.598	0.392
	2010	877	900	921	231	232	233	241		
	2020	934	995	1050	292	292	293	289		
	2030	962	1066	1168	357	361	367	337	0.739	0.458
C.6. Decline to TF ≤ 2.1	2000	1445	1445	1445	393	393	393	436	0.692	0.393
	2010	1508	1545	1582	472	472	471	550		
	2020	1528	1618	1708	577	572	570	659		
	2030	1496	1644	1798	672	677	686	768	0.789	0.522
C.7. Early transition mainly TF ≤ 2.1	2000	1194	1194	1194	461	461	461	473	0.767	0.503
	2010	1213	1224	1234	504	504	504	519		
	2020	1211	1243	1273	552	550	549	560		
	2030	1193	1251	1308	581	583	585	601	0.812	0.574
Total(Wld.)[b]	2000	5895	5895	5895	1558	1558	1558	1555	0.625	0.423
Total(Wld.)	2010	6495	6632	6765	1848	1857	1866	1892	0.661	0.423
Total(Wld.)	2020	6951	7322	7683	2211	2236	2261	2241	0.688	0.444
Total(Wld.)	2030	7238	7894	8562	2587	2655	2723	2589	0.712	0.472
World	2040	7311	8344	9458	2888	3021	3159	2904	0.735	0.493
World	2050	7194	8660	10325	3132	3343	3569	3220	0.752	0.513

a Numbers for years 2020, 2030, and 2040 are linear interpolations from data for years 2010, 2025 and 2050.
b Additive model only used for calculating the world data from world age data.
c. TF is an abbreviation for total fertility as used in volume 3 of United Nations (2001).

Sources: Calculated using United Nations 2002 Revision, Habitat (1996:464-467): refer Appendix for 141 country-by-two data,192-country data, and list of countries in each fertility transition cohort.

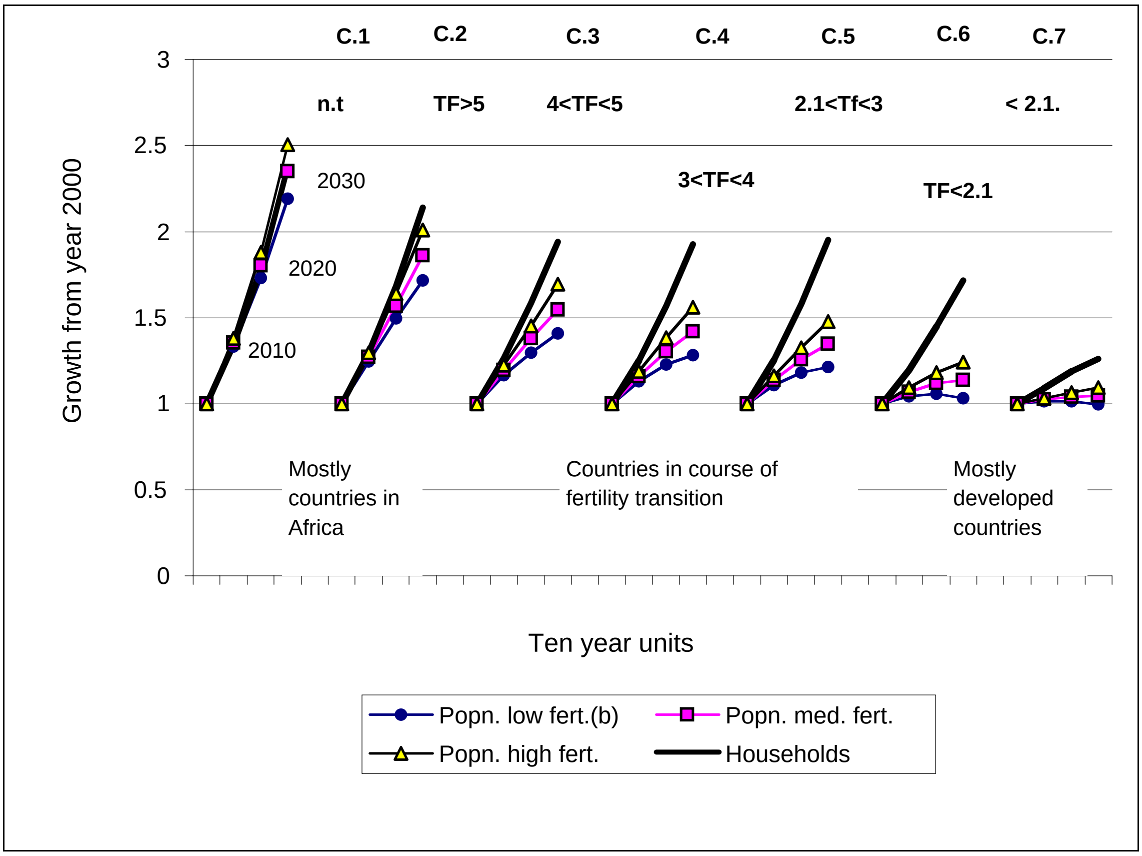
Most of the high purchasing power countries are in the two lowest fertility cohorts C.6 and C.7 while the lowest purchasing power countries are in the two highest fertility cohorts. The ratio of the median purchasing power per middle person (M) between the two groups (The Economist 2004:27) is about twenty to one.
Infrastructure relates particularly to household numbers rather than the population. Figure 13 indicates that the pattern of growth for households and population are different. Therefore it cannot be assumed that national economic policies based on growth in only one of these factors is sufficient.

These results carry through into most individual countries as shown in Figure 14. The annualised growth rates to 2030 are compared for each of the 192 countries listed in the United Nations 2002 Revision for the medium fertility variant. The 45-degree line indicates equal household and population growth rates. Very few observations of household growth rates fall below this line. There are three countries in cohort C.1 (no transition) where this occurs and two countries in cohort C.2 (TF > 5), all from Africa. There are 77 countries where household growth rate exceeds population growth rate by more than one per cent. A similar graph for the low fertility variant version is slightly to the left of the medium fertility case and for the high fertility case is slightly to the right.

For the 192 countries the projected annual growth in households to 2030 for the three fertility variants averaged over countries is approximately 1.9 per cent. The growth in population varies from 0.8 per cent for the low fertility variant, 1.1 per cent for the medium fertility variant, and 1.3 per cent for the high fertility variant. For the medium fertility variant the graphs show that for household growth rate to be zero the population growth rate has to be negative. This also applies to the low fertility variant (not shown). Figure 14 therefore provides a useful check on the feasibility of household projections.

For the world, household intensity increases from 264 households per thousand persons at 2000 to 357 (low fertility), 336 (medium fertility) or 318 (high fertility) by the year 2030. In each case the 95 per cent confidence interval is equal to household intensity ± 35 households per thousand persons. For the year 2050 household intensities are projected to be 435 (low fertility), 386 (medium fertility) and 346(high fertility). In each case the 95 per cent confidence interval is equal to the household intensity ± 58 households per thousand persons. The confidence intervals will be wider for the higher fertility variants, although their mean values will differ by less than 4 per cent for most fertility cohorts.

Figure 13 Comparison of household growth[a] and population growth from 2000 to 2030 for the world seven fertility transition cohorts; C.1 no transition, C.2 decline but total fertility TF still > 5, C.3 decline to 4 <TF ≤ 5, C.4 decline to 3 <TF ≤ 4, C.5 decline to 2.1 <TF ≤ 3, C.6 recent decline to TF ≤ 2.1, C.7 early transition mainly with TF ≤ 2.1



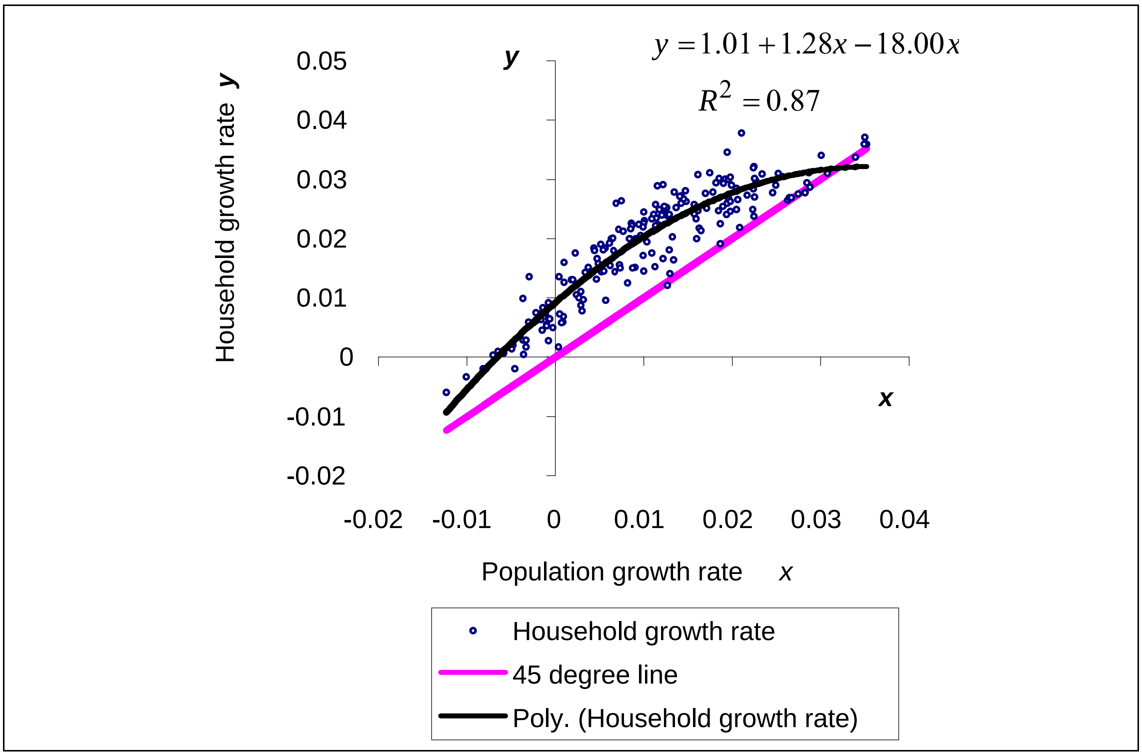
a Only medium fertility variant values for households are shown; low, medium and high fertility values are too close to be distinguished.

b Popn low fert., Popn med fert., Popn high fert., are abbreviations for the population projections of United Nations 2000 Revision using the low fertility variant, medium fertility variant, and high fertility variant respectively.

Source: Calculated using authors' models and 141 country-by-two data described in the Appendix, and based upon United Nations 2002 Revision (United Nations 2003).

By 2030 for the low fertility variant the projections show that about 20 per cent of the world's countries and about 10 per cent of the world's population will live in countries where household intensity exceeds that of any country at 2000, i.e. 486 households per thousand. In the medium fertility variant the corresponding figures are 11 per cent and 7 percent and for the high fertility variant the corresponding figures are 6 per cent and 4 per cent. All these high household intensity countries are members of cohort C.6 (TF ≤ 2.1) or cohort C.7 (early transition mainly TF ≤ 2.1). i.e. mainly developed countries.

Figure 14 Household growth rate per annum (*y*) against population growth rate per annum (*x*) for the United Nations medium fertility variant for 192 countries together with a regression fit. Period 2000 to 2030



Source: United Nations 2002 Revision, and other official sources (see Appendix 1), and authors' calculations.

China and India

The world Habitat'96 projections of households to 2030 are comparable with those given in this paper being 2.5 per cent less. However there are large differences for 2020 and 2030 for both cohort C.4 ($3 < TF \leq 4$) that contains India and cohort C.6 ($TF \leq 2.1$). which contains China. Much of the difference is due to projections for just these two countries. A summary of the projections to 2030 for these countries is given in Table 4.

China and India are at different stages of the fertility transition. China is reaching the later stages and therefore one could expect some slowing down of the rate of growth of household intensity. However India could be moving towards a faster phase of the fertility transition, with household intensity increasing at a greater pace than hitherto.

China

From Table 3 for fertility transition cohort C.6 ($TF \leq 2.1$), the Habitat'96 projection at 2030 is 91 million households higher than the medium scenario projection. This difference is mainly due to a household projection for China that is 91 million higher. The Habitat'96 projections of the Chinese population at 2030 is 103 million more than the United Nations 2002 medium fertility projection. This would account for 44 million more households, using Habitat'96 estimate of household intensity at 2030 (0.431). The remainder would be due to differing projection methods.

Table 4 Comparison of estimates and projections of population and households for China, India, and the World through to 2030, medium fertility variant

		Population (millions)[b]			Households (millions)		
		Census[c]	Habitat' 96[d]	UN' 2002[e]	Census[cf]	Habitat' 96	Model projections
2000	China	1295	1285	1275	348	376	335
2030	China		1554	1451		670	579
2000	China (Wld.%) [a]		21	21	22	24	22
2030	China (Wld.%)		18	17		26	22
2000	India	1009	1022	1017	188	187	224
2030	India		1419	1376		319	421
2000	India (Wld.%)		17	17	12	12	14
2030	India (Wld.%)		17	17		12	16
2000	World		6158	6071		1555	1558
2030	World		8463	8130		2589	2655

a % refers to country estimates and projections as percentages of the United Nations 2000 Revision world estimates and projections.
b Population is the total population, not the population in households.
c Adjustment to 2000 is estimated from Indian Census for 1991.
d Adjustment to 2030 is estimated from Habitat'96 projections for 2010 and 2025.
e UN'2002 refers to United Nations 2002 Revision data.
f Percentages used in this column have as denominator the model projections.

Sources: National Bureau of Statistics China (2001); Office of the Registrar General, India (2001); Habitat'96 (1996); United Nations 1994 and 2000 Revisions; and calculations by the authors.

India

From Table 3 for fertility transition cohort C.4 ($3 < TF \leq 4$). the Habitat'96 projection of households at 2030 is 133 million lower than the medium fertility projection Of this difference, 102 million is due to India and 19 million due to Bangladesh. In the case of India for the period 2000 to 2030 Habitat'96 projections show a growth in household intensity h from 183 households per thousand to 225 households per thousand, an annual growth rate of 0.7 per cent. However the growth rate of adults/person (A/P_H) is also 0.7 per cent, and since $h = (H/A) \times (A/P_H)$ this implies that the growth in households per adult is zero. This seems unlikely since the growth in E/M is 1.6 per cent, i.e. the proportion of the elderly is increasing. In this paper household intensity is estimated to grow from 188 households per thousand to 306 households per thousand, ie at a rate of 1.6 per cent per annum over the thirty years.

The model projections for India show an estimate of households at 2000 of 37 million more than the census figure of 188 million. In the case of India the number of households is sensitive to whether couples move from multiple couple households to single couple households. At 2001 there were about 20 million households with no couples, 135 million with one couple, and 36 million households with two or more couples or 85 million couples in all (Office of the Registrar -General, India 2001). India currently has 28 per cent of the population living in urban areas. If for example urbanisation increases as expected this could lead to a reduction in the number of multiple couple households, and hence an increase in household intensity. It is noted also that household intensity actually dropped from 1990 to the year 2000. At this stage therefore it is not proposed to alter the household estimates on which Table 2 is based.

Extrapolating beyond observed age ratios

The 141 country-by-two data provides a broad representation of existing relationships between population and households and age ratios. However some of the age ratios used in the projections to 2050 are outside those used to calibrate the projection models. This results in some countries having projected household intensities beyond those found in the 141 country-by-two data. This section examines the significance of this effect on the estimation of the number of households.

The age ratios used are M/P, Ef/P together with a function u of the age ratios. Equation (7a) shows $\hat{h}$ as a function of u where $u = 1/(1 + Y/1.84\,Ef)$. The ranges of the age ratios in the 141 country-by-two data used for fitting the models are as follows. For Ef/P the range is 0.01 to 0.14, for M/P it is 0.32 to 0.64 and for u it is 0.04 to 0.57. Maximum observed household intensity data is 0.486 (Sweden 2000) and the minimum is 0.121 (Gambia 1983).

As shown in Figure 1 fertility as expressed by Y/M is generally declining across the world. The age ratio structure is shown in Figure 2 and Figures 3a,3b,3c,3d. With decreasing Y/M the two related variables Ef/P and u increase while M/P appears to be near a peak at about $Y/M = 0.5$. For the projections to 2030 there are a total of 1920 values of age ratios. They consist of 192 values for each age ratio for the year 2000, and values for three fertility levels for three periods 2010, 2020, 2030. The range for Ef/P is 0.02 to 0.22, for M/P it is 0.37 to 0.64, and for u it is 0.05 to 0.74. At the year 2030 out of 576 points there are 121 cases where u exceeds 0.57 and 151 cases where Ef/P exceeds 0.14 . For u this corresponds to an additional two per cent of households to those that would have been obtained if u were 0.57. A similar result applies to Ef/P. Therefore the extrapolations beyond existing known values of variables is not considered to be a significant source of error for projections to 2030.

Most of the countries where projections of age ratios and intensities are beyond the observed values are members of fertility transition cohort C.7 (TF ≤ 2.1) and are developed countries. There are a few countries in fertility transition cohort C.6 (TF ≤ 2.1) also. China, which is a member of C.6, has an Ef/P just exceeding 0.1394 (by 0.0001) in the low fertility scenario. At 2050 there is increased uncertainty as to individual country age ratios and the above analysis is not considered warranted without more detailed investigation of trends in age ratios.

Mortality assumption
It is possible to use the constant mortality variant of the United Nations 2002 Revision to estimate the sensitivity of world household projections to the mortality assumption used in the three fertility variants. To do this the additive model is applied directly to the world population summaries. For the period 2000 to 2030 the assumption of decreasing mortality used in the three fertility variants, as compare to that of constant mortality, has the effect of increasing the annual household growth rate from about 1.55 per cent to 1.79 per cent, and to 2050 from 1.21 per cent to 1.53 per cent. At 2030 for the medium fertility variant it is projected that there will be 2.655 billion households whereas under the constant mortality assumption there would be 2.475 billion. At 2050 the respective figures are 3.343 billion as against 2.841 billion. The differences in growth are less for population than for households.

Stable growth rates
Estimates of household intensities used in this paper are functions of age ratios. Therefore a population with a stable growth rate also has a stable household intensity since the proportions of various ages in the population remain constant. For such stable populations household intensity is estimated to be higher the lower the population growth rate, assuming an equal number of males and females. For example Table S.48 of Hinde (1998:278) is based upon England and Wales 1992 official data. This table gives stable population data. Using this data and applying equation (4), for a stable population growth rate of three per cent the household intensity is 220 households per thousand, but for a stable population with a negative growth rate of one percent it is 490 households per thousand. These results are consistent with the results in Table 3 for the medium fertility variant for cohort C.1 (no transition) and at 2030 for C.7 (early transition). They may be considered to mark the stable end points of the beginning and end of the fertility transition. For further discussion on the criteria for stable populations see Preston, Heuveline and Guillot (2001:147).

Revision of projections
The models may be regarded as linear approximations to non-linear functions that may change in the future. Standardized revision techniques can enable current models to be upgraded, provide information on the nature of the underlying processes and relationships, enable projections to be revised, and error to be reassessed. It is noticeable in Figure 2 and Figures 3b, 3c and 3d that for low Y/M the age ratios tend to scatter. Since countries generally are moving to smaller values of Y/M this points to the necessity of regular review. In this paper it is recommended that a revision take place every two years, following the biennial United Nations Population Revision, and that a published standardized revision technique be adopted. This is particularly so with increasing fluidity of household forms, and because of the intrusion of world events which may temporarily change trends in population scenarios.

Concluding remarks

The world at the country level has a unique age ratio structure. Household intensities are shown to be related to this structure enabling the development of household intensity models. These models account for about 87 per cent of the variation in household intensities. Two models were used in parallel to provide complementary information. Paired data were used to reduce error. The discrepancies between observed and fitted values for household intensity are approximately normally distributed. These discrepancies in household intensities are independent of population size and therefore can be applied to a country of any size population without modification.

The sources of error in model estimates need further investigation. It may be possible to achieve a parsimonious improvement in the models using variables other than age ratios. Since the life course underlies much of this work some modelling of processes under uncertainty may be possible. People change behaviour in unison implying feedback processes, leading to a new level of stability. Since households are hybrids, special constraints need to be considered. Modelling, beyond the simple procedures used here, implies an understanding of the overall dynamics of these processes. Overall approaches may be more rewarding than attempts to identify the many interactions. Household projections are less variable than population projections for a given range of fertility assumptions. It seems therefore that their joint production may improve the predictability of population characteristics. There seems little reason why population and household projections should not be produced jointly at the world level using one or more models.

References

Australian Bureau of Statistics (ABS) 1999. *Household and Family Projections:Australia 1996 to 2021*. Catalogue No 3236.0. Canberra.

Australian Bureau of Statistics (ABS) 2002. *Census of Population and Housing: 1991 and 1996. Age and Sex of the population*. (Provided to authors on CD ROM by ABS). Canberra.

Becker, G. S. 1991 *A Treatise on the Family* : *Enlarged Edition* Cambridge, Mass.: Harvard University Press.

Billari, F.C. 2001. The Analysis of Early Life Courses. *Journal of Population Research* 18(2):119-142.

Building Division, Department of Environment, Housing and Community Development (DEHC) 1977 *A Method for the Projection of Households and Dwelling Completions* 1975-2000 Canberra: Australian Government Publishing Service

Caldwell, J.C. and P. Caldwell 2001. Regional Paths to Fertility Transition. *Journal of Population Research* 18(2):91-117.

Cooper, J., Department of Environment and Planning South Australia 1986 *Methods for Obtaining Household Forecasts* Adelaide South Australian Department of Environment and Planning .

De Beer, J and M. Alders 1999. Probabilistic population and households forecasts for the Netherlands , *Statistics Netherlands, Department of Population; Paper for the European Population Conference EPC99, 30 August –3 September 1999.*

Department of Infrastructure Victoria 1998. *Melbourne in Fact: 1996 Census for Melbourne's Local Government Areas*. Melbourne: Department of Infrastructure.

Department of Infrastructure Victoria 2000. *Victoria in future: Methodology and Assumptions*. Melbourne: Department of Infrastructure.

Eurostat 2002. *European Social Statistics, Demography: Theme 3-Population and social conditions.* 2002 edn. Luxembourg: Office for Official Publications of the European Communities.

Freiberger, W. F. (Ed.) 1960 *The International Dictionary of Applied Mathematics* Princeton, D.Van Nostrand Company Inc.

Gober, P. 1990. The Urban Demographic Landscape: A Geographic Perspective. 232-248 in D. Myers, (Ed.), *Housing Demography.* Wisconsin: The University of Wisconsin Press.

Habitat (United Nations Centre for Human Settlements) 1996. An *Urbanizing World: Global Report on Human Settlements, 1996*. Oxford: Oxford University Press.

Haupt, H., W. Oberhofer and T. Reichsthaler. 2003. A varying-coefficient approach to estimation and extrapolation of household size. *Mathematical Population Studies* 10(4) October-December 2003:249-273.

Hinde, A. 1998. *Demographic Methods*. London: Arnold.

Ironmonger, D.S. and C.W. Lloyd-Smith. 1992. Projections of households and household populations by household size propensities. *Journal of the Australian Population Association* 9(2):153-171.

Jennings, V.E., C.W. Lloyd-Smith and D.S. Ironmonger. 1999. Household size and the Poisson Distribution. *Journal of the Australian Population Association* 16(1/2): 65-82.

Kobrin, F.E. 1976. The fall in household size and the rise of the primary individual in the United States. *Demography* 13(2):127-138.

Masnick, G.S., J.R.Pitkin and J. Brennan. 1990. Cohort Housing Trends in a Local Housing Market: The Case of Southern California. Pp. 157-173 in D. Myers, (Ed.), *Housing Demography.* Wisconsin: The University of Wisconsin Press.

Myers, D. and A. Doyle 1990. Age-Specific Population–per-Household Ratios: Linking Population Age Structure with Housing Characteristics. Pp.109-130 in D. Myers, (ed.), *Housing Demography.* Wisconsin: The University of Wisconsin Press.

National Bureau of Statistics of China 2001. *Communiqué on Major Figures of the 2000 Population Census (No 1) 2002-04-23 10:58:27*. National Bureau of the People's Republic of China, March 28th, 2001.
http://www.stats.gov.cn/english/newrelease/statistics Beijing.
Date accessed: 10 May 2003.

National Research Council, Panel on Population Projections. 2000. *Beyond Six Billion: Forecasting the World's Population*, ed. John Bongaarts and Rodolfo A. Bulatao. Committee on Population, Commission on Behavioural and Social Sciences and Education. Washington DC: National Academy Press.

Office of the Registrar General, India 2001. *Census of India 2001. H-Series Tables on houses, household amenities and assets.* Table S00-014 Created on 17th April 2003 and 5th April 2001. http://.www.censusindia.net/2001/s00-114.html Delhi India. Date accessed: 10 May 2003.

Preston, S.H., P. Heuveloine and M. Guillot 2001 *Demography* Oxford:Blackwell Publishers.
Rescher, N, 1998 *Predicting the Future* Albany: State University of New York Press.
Slotnik, H. 1998. International Migration 1965-96: An Overview. *Population and Development Review* 24(3): 429-468.
Schulz, George 2004 A Changed World *Kissinger Lecture* (February 11, 2004) Washington: The Library of Congress.
Sherden, W. A. 1999 *The Fortune Sellers* New York: John Wiley and Sons.
The Economist, 2004. *Pocket world in Figures* 2004 Edition London: Profile Books Ltd.
United Nations. 1955-1997. *1955, 62, 63, 71, 73, 76, 82, 90, Demographic Year Book.* New York.
United Nations. 1989. *1987 Demographic Year Book*. New York.
United Nations. 1994. *The Sex and Age Distribution of the World Populations: The 1994 Revision.* New York.
United Nations. 1995. *World Population Prospects: The 1994 Revision*. New York.
United Nations. 1997. *1995 Demographic Yearbook.* New York.
United Nations 1999 *Wistat Women's Indicators and Statistics Database Version 4*, ST/ESA/STAT/SER.K/CD/15, New York.
United Nations. 2001a. *World Population Prospects: The 2000 Revision*. Vol.1, *Comprehensive Tables*. New York.
United Nations. 2001b. *World Population Prospects: The 2000 Revision*. Vol.2, *Sex and Age*. New York.
United Nations. 2001c. *World Population Prospects: The 2000 Revision*. Disk 2; Extensive Set ST/ESA/SER.A/205. New York.
United Nations. 2001d. *Living Arrangements of Older Persons: Critical Issues and Policy Responses. Special Issue Nos 42/43 2001 Population Bulletin of the United Nations.* New York.
United Nations. 2002. *World Population Prospects: The 2000 Revision*. Vol. 3, *Analytical Report*. New York.
United Nations. 2003. *World Population Prospects: The 2002 Revision*. Datasets in Excel and PDF Format, ST/ESA/SER.A.225. New York.
Van Imhoff E. 1995. LIPRO: A multistate household projection model. 273-291 in E.van Inhoff E., A. Kuijsten, P. Hooimeijer, and L. van Wissen (eds), *Households Demography and Household Modelling*. New York: Plenum Press.

Recent household numbers were obtained from the World Wide Web sites for the following National Statistical Offices

American Samoa*	Mauritius	South Korea
Argentina	Mexico	Sweden
Chile	Netherlands	South Africa
China	New Zealand	Spain
China – Hong Kong	Norway	Sri Lanka
Czech Republic	Romania	Switzerland
Greece	Peru	Tanzania
Iran	Philippines	Thailand
Ireland	Portugal	Trinidad and Tobago

Israel	Russia	Tunisia
Italy	Poland	United Kingdom
Japan	Singapore	United States
Latvia	Slovakia	Venezuela
Luxembourg	Slovenia	
Malawi	Saint Lucia	

*The data for American Samoa was obtained via the U.S. Census Bureau World Wide Web site, (Census 2000).

Data for Guadeloupe, Guyane, Martinique, Réunion, and France are obtained from *Annuaire Statistique de la France*, Edition 2003 Paris, Statistique Publique.

Data for the following countries is taken from the national statistical yearbooks for those countries: Australia, Germany, New Zealand as well as their World Wide Web sites.

Appendix 1

The world and the 192-country data set

The population data used in these projections comes from the United Nations 2002 Revision (United Nations 2003). There are 192 countries listed separately in the tables. The additive and bounded models are applied to the 192-country data to give the projected number of households for the seven fertility transition cohorts. There are 16 household counts obtained for each of the 192 countries giving a total of 3072 data points. These comprise an estimate of the number of households at 2000, and three projections for each of the years 2010, 2020, 2030, 2040 and 2050, based upon the low, medium and high fertility variants of the United Nations 2002 Revision. These projections are consolidated into cohort projections to 2030. Cohorts are described below.

There is a balance of 36 small countries and areas for which detailed information is not readily available. Each had less than 140000 inhabitants (at the year 2000) and the total population was 1,951,000 persons or about 0.03 per cent of the world population (United Nations 2001c). Habitat'96 estimates of 490000 households for this group give 0.03 per cent of total households for the world (Habitat 1996:464-467). Thus a sufficiently accurate estimate for households for the world can be made without the need to include this small proportion of the population. The overall world projections include this group of countries. At the year 2000 the overall world population estimated by United Nations 2002 Revision is 6,070,581,000 persons; the 192 countries listed separately have an estimated population of 6,069,248,000 persons.

The United Nations data sets are used since they are made comparable between countries, subject to biennial revision (United Nations 2002:1), and are readily accessible. The methodology for the data sets is described in detail in chapters 6 and 7 of United Nations 2000 Revision, Volume III (United Nations 2002). This methodology is subject to some modifications as described in the United Nations 2002 Revision. In Volume III countries are grouped into seven tables according to their stage in the fertility transition that is discussed in Chapter 1. Table 1.1 describes the stages and Table 1.2 to Table 1.8 lists the countries in each stage (United Nations 2002:27-35). In this paper countries are grouped according to these stages in the fertility transition to produce seven cohorts. See below for list. Countries with a + attached are those that are included in the 141 countries-by-two' database. The United Nations reserves 'rate' to 'age-specific fertility rates', which when summed and multiplied by 5/1000 gives 'total fertility'. See for example Table 1.23 in United Nations 2000 Revision, Volume III (United Nations 2002). Also included in the headings for each fertility cohort is an estimate using equation (1) of the equivalent Y/M value using only current total fertility rather than the average of fertilities over previous periods.

Fertility Transition cohorts

C1. *Countries where the fertility transition had not begun by 2000*

(16 number) Angola, Burundi+, Chad, Congo+, Democratic Republic of Congo, Equatorial Guinea, Gabon, Guinea-Bissau+, Liberia, Mali+, Mozambique, Niger+, Sierra Leone+, Somalia, Uganda+, Yemen+.

C2. *Countries whose total fertility was still above 5 children per woman in 1995-2000 (equivalent to Y/M of 1.14).*

(33 number) Afghanistan+, Benin+, Bhutan, Burkina Faso+, Cambodia, Cameroon+, Central African Republic+, Comoros+, Côte d'Ivoire+, Djibouti, Eritrea, Ethiopia, Gambia+, Guinea+, Iraq+, Laos, Madagascar+, Malawi+, Maldives+, Mauritania, Namibia, Nigeria+, Occupied Palestinian Territory, Oman, Pakistan+, Rwanda+, São Tomé and Príncipe+, Saudi Arabia, Senegal, Solomon Islands+, Togo+, United Republic of Tanzania+, Zambia+.

C3. Countries whose total fertility had declined to a level between 4 and 5 children per woman in 1995-2000 (equivalent to Y/M between 0.94 and 1.14).

(22 number) Bolivia+, Botswana+, East Timor, French Guiana+, Ghana+, Guatemala+, Haiti+, Honduras+, Jordan+, Kenya+, Lesotho+, Nepal+, Nicaragua+, Papua New Guinea, Paraguay+, Samoa+, Sudan+, Swaziland, Tonga+, Vanuatu+, Western Sahara, Zimbabwe+,

C4. *Countries that had attained fertility levels between 3 and 4 children per woman in 1995-2000 (equivalent to Y/M between 0.74 and 0.94)*

(23 number) Algeria+, Bangladesh+, Belize+, Cape Verde+, Ecuador+, Egypt+, El Salvador+, Fiji+, Guam+, India+, Iran+, Libya+, Malaysia+, Micronesia (Federated States of), Morocco+, Myanmar, Philippines+, Qatar, South Africa+, Syrian Arab Republic+, Tajikistan, Turkmenistan, United Arab Emirates+.

C5. *Countries that had attained fertility levels below 3 children per but above replacement level of 2.1 in 1995-2000 woman (equivalent to Y/M below 0.74 but above 0.56)*

(31 number) Albania, Bahamas+, Bahrain+, Brazil+, Brunei Darussalam+, Chile+, Colombia+, Costa Rica+, Dominican Republic+, French Polynesia+, Guyana+, Indonesia+, Israel+, Jamaica+, Kuwait+, Kyrgyzstan, Lebanon, Mexico+, Mongolia, New Caledonia+, Panama+, Peru+, Réunion+, Saint Lucia+, Suriname, Tunisia+, Turkey+, United States Virgin Islands+, Uzbekistan, Venezuela+, Viet Nam+.

C6. *Countries that had attained fertility levels at or below replacement level of 2.1 children per woman by 1995-2000 (equivalent to Y/M of 0.56).*

(24 number) Armenia, Azerbaijan, Barbados+, Bosnia & Herzegovina. China+, China, Hong Kong SAR+, China, Macao SAR+, Cuba+, Cyprus+, D.R., Korea, Guadeloupe+, Kazakhstan, Malta+, Martinique+, Mauritius+, Netherlands Antilles+, Puerto Rico+, Republic of Korea+, Saint Vincent and the Grenadines+, Singapore+, Sri Lanka+, TFYR Macedonia, Thailand+, Trinidad and Tobago+.

C7. *The low fertility countries that had undergone the transition early and are generally below 2.1 children per woman (equivalent to Y/M of 0.56).*

(43 number) Argentina+, Australia+, Austria+, Belarus+, Belgium+, Bulgaria+, Canada+, Channel Islands, Croatia, Czech Republic+, Denmark+, Estonia, Finland+, France+, Georgia, Germany+, Greece+, Hungary+, Iceland+, Ireland, Italy+, Japan+, Latvia, Lithuania, Luxembourg+, Netherlands+, New Zealand+, Norway+, Poland+, Portugal+, Republic of Moldova, Romania+, Russia, Serbia and Montenegro, Slovakia+, Slovenia+, Spain+, Sweden+, Switzerland+, Ukraine+, United Kingdom+, United States of America+, Uruguay+,

There were 36 countries where population was less than 140,000 inhabitants and specific age ratio data was not available. These countries are listed below and are taken from United Nations 2000 Revision (United Nations 2001a:26-28).

American Samoa, Andorra, Anguilla, Antigua and Barbuda, Aruba, Bermuda, British Virgin Islands, Cayman Islands, Cook Islands, Dominica, Faeroe Islands, Falkland Islands (Malvinas), Gibraltar, Greenland, Grenada, Holy See, Isle of Man, Kiribati, Liechtenstein, Marshall Islands, Monaco, Montserrat, Nauru, Niue, Northern Mariana Islands, Palau, Pitcairn, San Marino, Seychelles, St Helena, Saint Kitts and Nevis, Saint-Pierre-et-Miquelon, Tokelau, Turks and Caicos Islands, Tuvalu, Wallis and Futuna Islands.

The 41-country data set

This data set enables household representative persons (HRPs) and population to be related by age. HRP data is derived from the 1995 United Nations *Demographic Yearbook*, Table 30 and Table 34 (United Nations 1997) and also the 1987 United Nations *Demographic Yearbook*, Table 33 and Table 37 (United Nations 1989). Population data is taken from the United Nations 2000 Revision and Australian data is from the Australian Bureau of Statistics (2002).

There is a four-way stratification of populations. Firstly, into 41 countries. Secondly, into households or non-households. Thirdly, into households defined by the age of the HRP, the ages being 15-19, 20-24,…, 75+ so that there are 13 five year age classes in all. Finally into households by household size. The 41 countries and year of collection of data are:

Australia 1991, Austria 1981, Bangladesh 1981, Bolivia 1992, Brazil 1980, Bulgaria 1985, Canada 1991, Central African Republic 1988, China, Hong Kong SAR 1991, Comoros 1991, Cuba 1981, Denmark 1991, Finland 1990, France 1990, Germany 1987, Greece 1991, Guadeloupe, 1990, Hungary 1990, Italy 1981, Ireland 1991, Japan 1985, Luxembourg 1991, Macau 1991, Mauritius 1990, Netherlands Antilles 1992, New Zealand 1991, Norway 1990, Pakistan 1981, Philippines 1990, Poland 1988, Portugal 1991, Republic of Korea 1990, Réunion 1982, Romania 1992, Slovenia 1991, Spain 1981, Sweden 1990, Switzerland 1990, United States 1990, Zambia 1990, Belize 1990.

The 141 country-by-two data set of households and population

The calibration of the models is obtained by fitting to an 'averaged' version of 141 country-by-two data as described in the main text. The estimated population for the years 1950 to 2000 is obtained from United Nations 2002 Revision. The estimation of the population in households is described in the Introduction to this paper. The household data is mainly obtained from the United Nations Demographic Yearbooks (United Nations 1955-1997), particularly United Nations (1989; 1997), United Nations Women's Indicators and Statistics Database Version 4, CD-ROM, (United Nations 1999), the official world wide web sites from about 40 national statistics offices listed in the references, and some official national year books. The latest year for which a wide range of both household and population data could be obtained was 2002. These data provide the bases for determining the coefficients in the estimating models, and for measuring error. There are 141 countries where the authors have at least two sets of data, which have both household counts, and population data by age and sex. The later set has a population of 5.2 billion. These pairs are used to provide 282 data sets, called the '141 countries-by-two' data. The pairing of the data points for each country reduces variability arising from differences within the country. The countries in the data set range from those in the late stages of fertility transition such as Sweden to those in the early stages of fertility transition such as Zambia. The data for China includes that for Taiwan.

Appendix 2

Other work

Many projections of households are made for individual countries or regions within countries. For example, household projections are produced for the State of Victoria in Australia through to the year 2021 (Department of Infrastructure Victoria 2000). This State projection is based upon reconciling three methods; the cohort component method, a partly developed household formation method and the housing unit method. Five propensity methods have been described (Cooper, Department of Environment and Planning South Australia 1986). A propensity method that provides bivariate size distribution of households is given in Ironmonger and Lloyd-Smith (1992) and Australian Bureau of Statistics 1999:36). Van Imhoff (1995:273-291) describes 'LIPRO': a multi-state household projection model, which is applied to produce household forecasts by type through to 2050 for the Netherlands. Household reference person (HRP) data is limited in its use in that it necessitates measurement of the age of persons in households. Gompertz and Logistic functions (Building Division, DEHC 1977) have been used. In all these cases however cohort and other forms of data are required and this type of information is not readily available across the world.

Previous work by the authors (Jennings, Lloyd-Smith and Ironmonger 1999) has shown that the Poisson distribution gives a good overall measure of household size distribution when applied at the country level, as well as at the sub-country level. The single parameter used was average household size (persons per household). Thus countries appeared to provide a suitable unit by which to compare household numbers. The results appeared to be independent of the size of population. It was suggested in that paper that age ratios could be used to estimate average household size.

Demographic changes in household formation have been shown to be related to changes in fertility and ageing. In the context of the United States Kobrin (1976:127) observed that falling fertility had decreased the number of very large units (households) while the fall in mortality had increased the proportion of small units by increasing the length of time couples survive after their children are grown. Gober (1990:244) noted the diversity of contemporary households and diversity of life cycles. Nikolai Botev (cited in United Nations 2001d:9) commented upon the links between co-residence, social networks and other family transfers for older people. Myers and Doyle (1990:115) pointed out the advantages of housing-based age ratios. Becker used rational choice theory to discuss various factors affecting families including the number of children. (Becker 1991:137-154).