Probabilistic household forecasts for five countries in Europe

Nico Keilman

Department of Economics

University of Oslo

Oslo, Norway

January 2015

Delivery 2.5 of Work Package 2 “Economic Consequences of Ageing”

The MOPACT Coordination Team
The University of Sheffield
Department of Sociological Studies
Northumberland Road
Sheffield, S10 2TU, UK
T: +44 (0)114 222 6458
F: +44 (0)114 276 8125
E: mopact@sheffield.ac.uk

www.mopact.group.shef.ac.uk

MoPAct
Mobilising the Potential of Active Ageing in Europe

Funded by the European Commission's Seventh Framework Programme FP7-SSH-2012-1/No 320333

Abstract

We show how techniques of data dimension reduction can be used to predict patterns of household dynamics in a multi-country context.

Probabilistic household forecasts are presented for Denmark, Finland, Germany, the Netherlands, and Norway, spanning the period 2011-2041. Starting point is the population of each country broken down by age, sex, and household position as reported in the census round of 2011. Future trends in fertility, mortality and international migration are taken from official population forecasts. For changes in household structure we rely on time series of household data.

Long series of household data, in which the population is broken down by household position, age, and sex, are available for Denmark (1981-2007) and Finland (1988-2009) from the population registers in these countries. For the Netherlands the series are rather short (1995-2011). Annual shares of the population by household position, age, and sex for the three time series countries are modeled using an approach that builds on Brass' relational model originally developed to model the age pattern of mortality. We find that the household shares can be modelled as Random Walks with Drifts (RWD), independent of country. The Brass approach preserves the age patterns of the household shares. Future household shares are found by extrapolating the RWD processes. This results in household share forecasts, as well as standard errors of the forecasts. Correlations across ages and between men and women are estimated from model residuals.

No time series data are available for Germany or Norway. For Germany, we use household transition rates borrowed from Denmark and Finland, but adjusted to cohabitation and marriage levels from the German Generation and Gender Survey. For Norway, we have household transition rates for the year 2010 from the population register. Future household patterns for these two countries are computed by using the multistate household model LIPRO, in which the household transition rates are applied to the household pattern from the census. Uncertainty parameters are borrowed from the time series analyses for Denmark, Finland and the Netherlands.

The results show a continuation of current trends towards more and smaller households, often driven by increasing numbers of persons who live alone. The number of households increases faster than population size, which leads to falling average household size. A very consistent finding is that larger households are easier to predict than smaller households, at least when uncertainty is considered in a relative sense.

1. Introduction

Figure 1 shows the age pyramids of five selected European countries together with the household structures of these populations. Data stem from the 2011 round of population censuses; see Eurostat (2014). Population sizes range from a low five million in Norway, to a high 80 million in Germany. Hence the pyramids are plotted in terms of proportions in each population sub-group relative to total population.

Large birth cohorts born during the baby boom period of the 1950s and 1960s are clearly visible. However, the focus of the current paper is on the household structure, and future changes therein. Note the similarities in the five countries. Not surprisingly, most children and adolescents younger than 20 years of age live with one or both parents; see the green bars at the bottom of the pyramids. Among adults, married couples constitute the vast majority (cfr. the blue bars), although cohabitation is frequent among adults under the age of 30 (red bars). Age and sex patterns of those who live alone are strikingly similar across the five countries. The yellow bars show a consistent pattern of more young men who live alone than young women. One explanation is that when a young couple (cohabiting or married) with children breaks up, in many cases the men leaves the household and lives alone for some time, while the woman becomes a lone mother. The sex ratio among one-person households is reversed for the elderly. This is due to three factors: men are often a few years older than women when they form a couple, mortality among (married, cohabiting) men is higher than among women, and after union dissolution, women are less likely to repartner than men (US Census Bureau 2014 p. 131; United Nations 2010 p. 32; Peters and Liefbroer 1997). All this leads to more elderly women who live alone than elderly men.

Data of this kind are important for planning purposes in various sectors of society. Social welfare spending depends on the number of lone parents. Elderly persons who live alone are more vulnerable than those who live with a partner, elderly women in particular. Household status is an important determinant for the need of formal and informal support and care for the elderly, in addition to health. Rising numbers of one-person households in western countries have a strong impact on housing needs. Finally, falling average household size increases the demand for energy, because of economies of scale in energy use in large households.

In addition to information about the current household position of the population, planners need to know how household of various types will evolve over time. Hence it is important to trace possible *future* household dynamics. What patterns of household structure can we expect in the years to come? How certain are we about these developments? In their overview of state-of-the-art knowledge about the dynamics of families and households, Pailhé et al. (2014) argue that the dynamics of family formation have changed in contemporary societies during the past fifty years. Compared to the 1960s, when the nuclear family consisting of a married couple with one or more children was the dominant form, the sequencing of life stages over the life course has become more diverse and more unpredictable. Some life events are experienced by smaller shares of the population; they occur at more diverse ages and for durations that vary more widely. Given the fact that household forecasts are necessary, all this calls for probabilistic forecasts, not deterministic ones.

The aim of this paper is to present probabilistic household forecasts for Denmark, Finland, Germany, the Netherlands, and Norway, spanning the period 2011-2041[1]. Starting point is the population of each country broken down by age, sex, and household position in 2011, as shown in Figure 1. Future trends in fertility, mortality and international migration are taken from official population forecasts. For changes in household structure we rely on time series of household data.

The contribution this paper makes to existing literature is twofold. First, we show how techniques of data dimension reduction can be used to predict patterns of household dynamics in a multi-country context. Second, we show how probabilistic forecasts can be computed for countries with little data on household dynamics (Germany, Norway), by using findings from countries with good data (Denmark, Finland, the Netherlands).

2. Earlier work

The recent paper by Christiansen and Keilman (2013) contains a review of methods for household projection and forecasting. Here we give only a brief summary.

Deterministic household forecasts have a long tradition (e.g. US National Resources Planning Committee, 1938; United Nations, 1973). *Probabilistic* household forecasts were first introduced around the turn of the century by De Beer and Alders; see Alders (1999, 2001) and De Beer and Alders (1999). Alders and De Beer combined a stochastic population forecast with forecasts of random shares. The shares distribute the population probabilistically over six household positions: individuals could live as a child with parents, live alone, live with a partner, as a lone parent or in an institution, or belong to another category. For instance, the authors computed the (random) number of lone mothers aged 40 in 2015 as the product of two other random variables, namely the number of women aged 40 in 2015 and the share of 40-year old women who live as a lone mother in 2015. Expected values for population variables and for the shares for specific household positions were obtained from observed time series, but the statistical distributions that were assumed for the shares were based on intuitive reasoning. Perfect correlations across age and sex were assumed for the mortality rates, fertility rates and migration numbers in the stochastic population forecasts, as well as for the random shares. In addition the authors assumed perfect correlation in the time dimension for the random shares.

Scherbov and Ediev (2007) combined a probabilistic population forecast for the population broken down by age and sex with random headship rates, and applied their method to the case of Russia. A headship rate reflects the proportion of the population that is the head of a private household, for a given combination of age and sex (United Nations, 1973; Jiang and O'Neill, 2004). Scherbov and Ediev based a large part of their uncertainty distributions on intuition. Wilson (2013) computed a probabilistic household forecast for Greater Sidney. Household parameters were modelled as a

[1] These five countries have been selected on two different grounds. First, the data situations in these five are very different. This will give us the possibility to test different strategies for data analysis, in order to accomplish the task. Second, the current project is part of Work Package 2 "Economic consequences of ageing" of the MOPACT project. Scholars with a background from these countries are taking part in that Work Package and the results of the current project will be relevant to them.

random walk. Standard deviations of the random errors were set based on judgement due to the lack of past errors and estimates of living arrangements and households.

An important drawback of the probabilistic household forecasts mentioned here is that uncertainty parameters were largely judgemental. Alho and Keilman (2010) improved on this situation by estimating uncertainty parameters from data. Building on the random share method of De Beer and Alders, they applied their approach to Norwegian data. One important drawback of that work was that the uncertainty assessments were based on limited data, and that simplifying assumptions had to be made. Christiansen and Keilman (2013), in their analysis of future household dynamics in Denmark and Finland, used long time series of observed shares, and formal time series methods to extrapolate them. Keilman and Van Duin (2014) modelled household shares for the Netherlands by means of the Hyndman-Booth-Yasmeen product-ratio variant of the Lee-Carter model.

In the current analysis, we will update the earlier probabilistic household forecasts of Denmark, Finland, and Norway, and compute new forecasts for the Netherlands and Germany. Uncertainty parameters will be derived from residuals that remain after a Brass-type of relational model has been fitted to the household shares in three of the five countries; see sections 4 and 5 below.

3. Data

3.1 Census data

As mentioned earlier, starting point of the forecast is census data on the population of each country broken down by five-year age group, sex, and seven household positions in 2011, as shown in Figure 1. The census data contain the following household positions (household position code in parentheses):

1. Child living with parent(s) (CHLD)

2. Living in one-person household (SIN0)

3. Living in unmarried cohabitation, with or without children (COH)

4. Living with marital spouse, with or without children (MAR)

5. Living as lone parent (SIN+)

6. Other position in private household, for instance member of multiple family household, living with non-family related individuals, homeless (OTHR)

7. Living in an institution (INST).

3.2 Register data

The population registers of Denmark, Finland and the Netherlands provided us with time series of annual data on the number of people in these seven household positions. The data relate to 1 January of the years 1981-2007 for Denmark, of the years 1988-2009 for Finland, and 1996-2010 for the Netherlands. The data comprise the seven household positions listed above, men and women, and ages 0-4, 5-9, 85-89, and 90+.

From the AGHON-project we had data on transitions between household positions, broken down by sex and five-year age groups for Denmark and Finland (Christiansen and Keilman 2013). These data show the number of persons who were in household position k (k=1,…,7) on 1 January of a certain year and in household position j (j=1,…7) on 1 January of the previous year. In this case we had Finnish data for the period 2004–2008 and Danish data for the period 2002–2006.

Time series data of the kind described here are not available for the Netherlands or for Germany. Quite recently, Statistics Norway has established an annual register for family and household data; see Goplen (2014). The data cover the period 2005-2013. Time series analyses of the kind performed for the three countries with longer series will give unreliable results with data for only nine years. The Norwegian data, however, were used to estimate transitions for the year 2010 between household positions for men and women in five-year age groups. For Germany, we use household transition rates borrowed from Denmark and Finland, but adjusted to cohabitation and marriage levels using information from the German Generation and Gender Survey.

4. Modelling household dynamics

4.1 Brief overview of the method

We start by analysing household dynamics in Denmark, Finland, and the Netherlands. The focus in our approach is on the distribution of the population over the seven household positions defined above, given age, sex, time, and country. Each household position corresponds with one share. The shares are different for men and women in different age groups in different countries. Also, they change over time. In order to assess the level of uncertainty in the shares, we analyse time series data on the share for each household position broken down by age and sex in the three countries. First we apply an appropriate transformation of the shares, the purpose of which is to make them stochastically independent across household positions. Next, we apply a Brass type of relational model to the age patterns of transformed shares for each household position, and analyse similarities of model parameters across sex, time, and country. We find that parameter estimates change gradually over time, which means that time can be included as an independent variable. Detrending results in a time series model that is known as Random Walk with Drift (RWD). For each household position, the age pattern of the RWD-increments is preserved by the Brass method. Parameter estimates for the RWD-model are very similar across countries, and hence the model is estimated for the combined data set of the three countries. Except for persons who live alone, parameter estimates show no systematic differences across sexes. Thus sex is not included as an independent variable, except for persons who live alone. The result is a parsimonious description of the data by the Brass Relational Logit Model and a drastic reduction of the dimension of the data set.

Future household shares for the three countries are obtained by extrapolating the RWD-models. This results in household share forecasts, as well as standard errors of the forecasts. Correlations across ages and between men and women are estimated from model residuals. Correlations across household positions are dealt with in a specific manner: see Section 4.2. Using the predicted shares and the estimated standard deviations and correlations, we stochastically simulate 1000 sample paths for the household shares for each age, sex, household position in the three countries.

No time series data are available for Germany or Norway. For Germany, we use household transition rates borrowed from Denmark and Finland, but adjusted to cohabitation and marriage levels based on information from the German Generation and Gender Survey. For Norway, we have household transition rates for the year 2010 from the population register (see Section 3.2). Future household patterns for these two countries are computed by using the deterministic multistate household model LIPRO, in which the household transition rates are applied to the household pattern from the census. This gives us expected values of the household shares. Uncertainty parameters for Germany and Norway are borrowed from the time series analyses for Denmark, Finland and the Netherlands. This way we stochastically simulate 1000 sample paths for the household shares in the two countries.

For each of the five countries, the sample paths are then combined with 1000 simulations from an earlier computed stochastic population forecast that covers the same period. This gives predicted numbers of persons in each household position and each country, as well as predicted numbers of households of corresponding types.

4.2 Predictions of the shares

4.2.1 Denmark, Finland, the Netherlands

We will now explain in further detail each of the steps outlined above. First we assess the uncertainty in predicted household shares in Denmark, Finland, and the Netherlands. Write V(j,x,s,t,c) for the number of people in household position j=1,2, . . .,7 who are in age x=0,1, . . . and have sex s=1 or 2, at time t=0,1,2, . . . in country c=1,2,3 . Aggregating over position, we obtain the population W(x,s,t,c) = Σ_j V(j,x,s,t,c) of age x and sex s at time t for country c. Household position j has share α(j,x,s,t,c) = V(j,x,s,t,c)/W(x,s,t,c) = α_j(x,s,t,c).

We distinguish seven household positions as listed in Section 3.1: CHLD (j=1), SIN0 (j=2), COH (j=3), MAR (j=4), SIN+ (j=5), OTHR (j=6), INST (j=7).

No age restrictions have been imposed on persons who have a certain household position. In particular, children (CHLD) and lone parents (SIN+) can be of any age. In practice, observed or predicted numbers of persons aged 85, say, with positions CHLD or SIN+, will not be interpreted as such, but should be assigned to a different position, for instance to the group of other. Moreover, we have ignored persons aged younger than 15 in the following positions: SIN0, COH, MAR, and SIN+.

For modelling random evolution of the shares, a logit transformation was applied. Building on earlier work (Alho and Keilman 2010; Christiansen and Keilman 2013; see also Wilson 2013), we have opted for a hierarchy of household positions using a variant of continuing fractions. This led to six types of fraction to be modelled (all specific for age, sex, time, and country). By construction, the six fractions as listed below can be interpreted as representing stochastically independent conditional probabilities. Independence is an advantage when we predict the values of these random fractions into the future.

The following fractions are defined, given age, sex, time, and country:

1. The share of CHLD;

2. The relative share of COH and MAR out of the total share of one minus the share of CHLD;

3. The relative share of MAR out of the share of COH and MAR;

4. The relative share of SIN0 and INST out of the total share of SIN0, SIN+, OTHR, and INST;

5. The relative share of SIN0 out of the share of SIN0 and INST;

6. The relative share of SIN+ out of the total share of SIN+ and OTHR.

The particular sequence 1-6 above is based upon the idea that important shares (numerically, behaviourally) have to be modelled first, and those that are less important can come last. Hence persons who live together with a partner (points 3 and 4 above), or alone (points 4 and 5) are given priority. The positions of SIN0 and INST are often difficult to distinguish for elderly persons, due to unclear registration rules for persons who *de facto* live in an institution (Christiansen and Keilman 2013). Therefore initially they are treated as one group (point 4). Children have been singled out from the beginning, because their shares are kept constant over time. The age pattern for this household position shows very little variation: for ages under 15, the shares are almost 100% (some children live in a multi-family household and hence have household position OTHR, a few live in an institution). For ages 15-19 and 20-24 the shares fall rapidly, and they are close to zero for ages beyond 25. Hence any systematic changes over time in the age patterns are difficult to identify. Finally, note that we have selected the household position OTHR as a remainder, which is in agreement with the nature of this position as we have defined it.

In an early stage of the analysis we experimented with an alternative specification of the hierarchy, similar to the one used by Christiansen and Keilman (2013). When applying this approach to data from the Netherlands, we found future age patterns for positions MAR and SIN+ that were difficult to interpret; see Keilman and Van Duin (2014) for details.

Temporarily suppressing indices for age, sex, time, and country, the logit transforms of the fractions 2-5 above are

$\xi_2 = \text{logit}((\alpha_3+\alpha_4)/(1-\alpha_1))$

$\xi_3 = \text{logit}(\alpha_4/(\alpha_3+\alpha_4))$

$\xi_4 = \text{logit}((\alpha_2+\alpha_7)/(\alpha_2+\alpha_5+\alpha_6+\alpha_7))$

$\xi_5 = \text{logit}((\alpha_2)/(\alpha_2+\alpha_7))$

$\xi_6 = \text{logit}(\alpha_5/(\alpha_5+\alpha_6))$

This way, five stochastically independent time series (given age, sex, and country) were constructed. With three countries, two sexes, and 20 age groups, the theoretical number of time series is 600. In practice, we have fewer series, because children younger than 15 years of age can be in household positions CHLD and INST only (in addition to OTHR, which is the reference category).

Much emphasis has been given to the age pattern of the shares of each household position. We have used a Brass type of relational model for the transformed shares. Originally intended to model age-specific survival from birth to age x, the Brass relational model can be written as

$Y(x) = a + b. Y^S(x) + e(x)$,

where Y(x) is the probability of survival from birth to age x in logit transformed form, while $Y^S(x)$ is some standard age pattern of survival, also in logit form. a and b are coefficients to be estimated from the data, and e(x) is an error term. The model is linear in its parameters, and hence one can use ordinary least squares regression to estimate them. Changing the parameter a shifts the age pattern up or down, while b changes its slope. See, e.g. Preston et al. (2001) for a review.

We used ordinary least squares to estimate the Brass Model applied to the age pattern of logit transformed fractions $\xi_k(x)$ (k=2,3,4,5,6) as defined above[2]. The standard age pattern $\xi_k^S(x)$ was defined as the average value of $\xi_k(x)$, where for each k the average was taken over all years t, for a given combination of age, sex and country. Hence for each k we obtained estimates of parameters a and b that varied over time, between sexes and between countries. However, in most cases we noticed a gradual increase or decrease of the estimates of a and b over time. This suggested that a and b could be written as linear functions of time, i.e.

$\xi_k(x,t) = (A_k + a_k.t) + (B_k + b_k.t) . \xi_k^S(x) + e_k(x,t)$.

In order to avoid spurious correlation, we detrended this model by taking first differences, and found

(1) $$\Delta\xi_k(x,t) = a_k + b_k. \xi_k^S(x) + d_k(x,t),$$

where $\Delta\xi_k(x,t) = \xi_k(x,t) - \xi_k(x,t-1)$ and $d_k(x,t) = \Delta e_k(x,t)$ is an error term for which we assume the usual properties.

Model (1) defines $\xi_k(x,t)$ as a random walk with drift (RWD). The drift equals $a_k + b_k. \xi_k^S(x)$, and the innovation variance is $\sigma_k^2 = Var(d_k(x,t))$. The term $\xi_k^S(x)$ preserves the age pattern in the random walk increments for each type of fraction k. Parameters a_k and b_k were estimated by ordinary least squares (across x) assuming an innovation variance independent of age and time.

For each type of fraction $\xi_k(x)$ (k=2,3,4,5,6), differences in estimates between countries were small, and not significant in most cases. Also differences between men and women were small, except for k equal to 4, which reflects the chance of living either alone or in an institution. For women, the estimate of a_4 turned out to be significantly lower (but still positive) than that for men. A possible explanation is that chances of living alone for women have increased relatively slowly, because their survival chances increased not as rapidly as those of men. Table 1 gives the parameter estimates.

---

[2] Note that our model differs from Brass' model: we estimate a model for fractions $\xi_k(x)$ (in logit form) for ages 0-4, 5-9, 10-14, … 90+. The Brass model in its original form is applied to survival <u>from birth to age (-group) x</u>. We prefer not to cumulate over ages, as we want to give equal weight to each age group.

Note that for k=6 (lone parents), both estimates are not significantly different from zero, and hence the process is likely to follow a random walk *without* drift.

Starting from a known value $\xi_k(x,T)$, a future value h years ahead (h = 1,2,…) is

$\xi_k(x,T+h) = \xi_k(x,T) + h.(a_k + b_k.\ \xi_k^S(x)) + d_k(x,T+1) + \ldots + d_k(x,T+h)$.

Hence an h-step ahead forecast can be computed as

$$\text{(2)} \qquad E[\xi_k(x,T+h)] = \xi_k(x,T) + h.(\hat{a}_k + \hat{b}_k.\ \xi_k^S(x)),$$

where a_k and b_k have been replaced by their estimated values.

The forecast error $F_k(x,T+h)$ equals $\xi_k(x,T+h) - E[\xi_k(x,T+h)]$. Given our assumptions, its variance is

$$\mathrm{Var}[F_k(x,T+h)] = \mathrm{Var}\left[\sum_{i=1}^{h} d_k(x,T+i) - h.\left(\hat{a}_k + \hat{b}_k.\ \xi_k^S(x)\right)\right]$$

$$\text{(3)} \qquad = h.\sigma_k^2 + h^2.\mathrm{Var}[\hat{a}_k] + h^2.(\ \xi_k^S(x))^2\mathrm{Var}[\hat{b}_k] - 2.h.\xi_k^S(x).\mathrm{Cov}[\hat{a}_k, \hat{b}_k].$$

During an early phase of the project we experimented with a different approach for preserving the age patterns of the household shares and the fractions, namely the Hyndman-Booth-Yasmeen product-ratio variant of the Lee-Carter model (LC model), adapted to household shares; see Keilman and Van Duin (2014). Originally developed for modelling age-specific mortality rates, the LC model assumes that the logarithm of the rate for age x during year t can be written as an average age pattern a(x) for the whole period for which there are data, plus a time trend k(t). This time trend, however, is not the same at all ages, but is assumed to be age-specific by forming the interaction b(x).k(t). The model is estimated for men and women separately. The Hyndman-Booth-Yasmeen product-ratio variant of this model assumes that the model holds for the (square root of the) product of the death rates of men and women for each combination of age and time, and for the (square root of the) ratio of these two. The advantage of modelling mortality this way is that it preserves the coherence between mortality for men and women. Since coherence between men and women also is important for married couples and cohabiting partners (see below), we attempted to apply this method to household shares for the Netherlands. However, this led to two problems. First, we found that predicted age patterns of the household shares became unrealistic in a number of cases. The product b(x).k(t) changed the age profile too strongly. Second, since expressions for standard errors of estimates of a(x) and b(x) are not known, we assumed that these two parameters were estimated without error. Hence the uncertainty in our forecast predictions was too small, but we do not know by how much.

The estimated models (1) were used to extrapolate the logit-transformed fractions $\xi_k(x,t)$ to 2041. Figures 2-4 give some selected results for observed and predicted shares $\alpha_j(x)$, where the predictions were obtained by back transformation of the logit-transformed fractions $\xi_k(x,T+h)$; see Appendix 1.

Figures 2-4 show a continuation of historical trends, in line with the assumptions. The trends are very similar in the three countries. Cohabitation will become more prevalent, in particular among young adults. For persons aged 60-80, the most dominant position still will be to live with a marriage partner. Among the oldest old we can expect a slight increase in the chances to live with a partner, but a somewhat stronger increase in the chances to live alone. Much of these two time trends is caused by a strong fall in the shares of elderly who live in an institution (not shown here).

4.2.2 Germany and Norway

For Germany and Norway, we predicted household shares for future years (broken down by sex and five-year age group) by means of a model called LIPRO ('LIfestyle PROjections'). The LIPRO model is based on the methodology of multistate demography, but it includes several extensions to solve the particular problems of household modelling. It has been used extensively for household projection and other types of multi state projections; see http://www.nidi.knaw.nl/en/research/al/270101 . For a detailed description of the model and the computer programme, see Van Imhoff and Keilman (1991). Christiansen and Keilman (2013) give a brief summary of the model and apply it to data for Denmark and Finland.

Being based on the methodology of multistate demography, the model starts its predictions from a jump-off population broken down by age, sex, and household position. Then this population is projected forward in time by exposing each population sub-group to a set of household transition rates, death rates, and emigration rates that are dependent on age, sex, and household position. The female part of the population in the age group 15-49 is also exposed to age and household-specific fertility rates. International migration is included in the model as emigration rates and immigration numbers broken down by age, sex, and household position. The result is a projected population structure (by age, sex, and household position) for future years. From this one computes the household shares for each household position (by age and sex).

We modelled the population broken down by five-year age groups (0-4, 5-9, …, 85-89, 90+), sex, and the seven household positions described in Section 3.1. The unit projection interval was five years. Starting point was the population as recorded in the Census of 2011. The projections were carried out for the period 2011-2041, assuming constant rates for household events. Rates for fertility, mortality and migration were calibrated against numbers of births, deaths and migrations taken from official population projections; see below. We applied the exponential version of the model in which intensities are assumed to be constant within the unit interval.

LIPRO projects (aggregates of) individuals, not households. This means that, for example, the number of women who marry during a period will not in general be the same as the number of men who marry during the same period according to the model. To solve this problem, LIPRO employs a consistency algorithm. The paper by Van Imhoff (1992) contains a thorough discussion of this algorithm. In this case the consistency algorithm contains equations that require that equal numbers of men and women marry or enter cohabiting unions in each projection interval. The same applies to the number of men and women experiencing the dissolution of marital and cohabiting unions. When there is a discrepancy between the modelled number of men and of women who experience one of these events, the number is adjusted to the harmonic mean of the inconsistent numbers of men and women.

The consistency algorithm described above assumes that each new couple consists of one male and one female partner. In reality, same sex partnerships are observed as well in the two countries. Census data for Germany show that there are 49490 more men than women who live with a partner, either marital spouse, or registered partner, or cohabitee. This amounts to 0.25 per cent of all men who live with a partner. For Norway, there are 287 more women than men who live with a partner, which is 0.03 per cent of all women with a partner. These small numbers justify the fact that we have omitted same-sex couples from the predictions.

In addition to consistency requirements for union formation and dissolution we have also constrained the capacity of institutions to be constant over time. In practice this was achieved by making the number of persons leaving an institution equal to the number entering an institution in each projection period. As the number of places available in institutions is a result of policy decisions we do not find it reasonable to let the future number of people in institutions be determined purely by transition rates. In addition to the kind of consistency requirements described thus far, the LIPRO program allows the user to set the number of births, deaths, immigrations, and emigrations equal to numbers from an external source. In this case we have chosen to make the total numbers of these events in each projection interval equal to the corresponding numbers from the Eurostat population projection of 2013 for Germany, and Statistics Norway's population projection for 2013 for Norway. For the case of mortality this means that, although initially the death rates are held constant during the 30-year projection period, the consistency algorithm reduces them so as to result in the numbers of deaths from the official population forecast. This implies an increase in the life expectancy. Fertility rates changes over the projection period as implied by the numbers of births, and similarly for emigration and immigration.

Germany

The German Generations and Gender Survey (GGS) gives panel data for the period from March-April 2005 (Wave 1) to September 2008-March 2009 (Wave 2). The survey was held among persons aged 18-79 in private households. The data contain information on household status of 3226 respondents aged who took part in both waves. Of these, only 678 persons changed household status between the two waves. This number is too small to serve as a basis for computing occurrence-exposure rates by sex and five-year age group, even for the most frequent changes. Therefore we borrowed occurrence-exposure rates for changes of household position from Denmark and Finland from the AGHON-project; see Section 3.2. These rates were adjusted such that the numbers of three types of events predicted by LIPRO agreed with information from the GGS. The three events are defined by changes in household status, as follows:

- from "living alone" to "living in a consensual union";
- from "living in consensual union" to any other household position, including "living with marital spouse";
- from "living with marital spouse" to any other household position.

In all three cases the numbers of events predicted by the (average of the) Danish and Finnish rates had to be reduced. This finding suggests that couples in Germany are more traditional than those in Denmark or Finland. Direct comparisons of living arrangements between the populations in the three countries are not known of. However, compared to cohabiting couples in Norway, those in Germany can be typified as more traditional indeed: they consider their living arrangement more often as a prelude to marriage (20 per cent in Germany, 11 per cent in Norway), and less often they consider marriage as an irrelevant option (17 vs. 32 per cent) (Hiekel et al. 2014).

Norway

As mentioned in Section 3.2, we have used data on change in household position (broken down by sex and five-year age group) for the year 2010. These data stem from the register for family and household data held by Statistics Norway.

Results for the two countries

Tables 3 and 4 summarize the results of a multistate life table computation based on the input rates used for the periods 2011-2015 and 2036-2040 for the two countries. The tables show small differences between the countries. In the earlier period, the average man and woman in Germany will spend about one-third of their lifetimes together with a spouse, 11 per cent with a cohabitee, and about one-fifth living alone. The average German woman will give birth to 1.4 children, in three out of four cases when she lives with a husband or a cohabitee. The Norwegian parameters for this early period show a slightly longer part of the life in a consensual union, and considerably fewer years as a married couple. Also, childbearing when living in a consensual union is much more common in Norway than in Germany. These findings are in line with the conclusions by Hiekel et al. (2014). For the later period, the relative distributions over the life time are very similar to those for the early period, but life expectancies are about five years higher. Note that the levels for life expectancy and number of children ever born in these tables may be very different from those published by statistical agencies, because our numbers relate to multi-state stationary populations.

5. Variances and correlations

The logit-transformed fractions ξ_k(x,T+h)= ξ_k(x,s,T+h,c) are assumed to have a multivariate normal distribution, with expected values as given in expression (2). Expression (3) specifies their variances. We have no reasons to assume that the uncertainty in the forecasts differs between countries. Therefore, we computed the average across the three countries of the standard age pattern $\overline{\xi_k^S(x,s)} = \sum_c \xi_k^S(x,s,c)/3$ and replaced ξ_k^S in Expression (3) by this average.

Covariances/correlations remain to be specified. By construction, the fractions ξ_k(x,s,t,c) are independent of household position k, but they are correlated across ages x, across sexes s, and between countries c. Since the fractions are modelled as a Random Walk with Drift process, they have zero autocorrelation. Inter-country correlations may be ignored as long as we present results for the populations of the three countries separately. Correlations across ages and between men and women were estimated from the annual increments $\Delta\xi_k$(x,t) of expression (1).

For k=2,3,4,5,6, we found correlations between sexes equal to 0.626, 0.598, 0.624, 0.891, and 0.065, respectively. Given the low estimate for lone parents (k=6), we have assumed independence between men and women. Reasons for becoming and remaining a lone parent are often very different for men and women. Differences in the estimates for the other groups (k=2-5) are hard to interpret. Therefore we took the median of the four numbers above, which is 0.623.

Following earlier work (Alho and Keilman 2010, Christiansen and Keilman 2012) we assumed an AR1 process for the errors in the age dimension. There were little systematic differences in the estimated correlations across ages. Inspecting age correlations for different types of fractions (k=2,3,4,5,6) we found extremely high correlations for the share of COH plus MAR (k=2; median value across ages equal to 0.982). An intuitive explanation is that the age pattern for living with a partner is very regular. In the simulations described below we have assumed that ages are perfectly correlated for this group. For the other groups (k=3-6) there was no systematic pattern. The median correlation across ages and groups turned out to be 0.756.

6. Illustrative results

Below we present selected simulation results for the five countries for the years 2021, 2031, and 2041. For each country and each year, the results are based on 1000 stochastic simulations for the shares, combined with 1000 simulations for the populations. Both the shares and the populations are for men and women separately, and specific for five-year age groups. For example, the number of lone mothers aged x at 1 January 2021 in country c is found as $\widehat{\alpha}(5, x, 2{,}2021, c). \widehat{W}(x, 2{,}2021, c)$.[3]

The stochastic population forecasts are updates of the results from the Uncertain Population of Europe (UPE) project. The aim of that project was to compute stochastic population forecasts for 18 European countries, including the five countries of the current paper. For more information about the methodology and assumptions see Alho et al. (2006), Alders, Keilman, and Cruijsen (2007), Alho et al. (2008) and the website http://www.stat.fi/ tup/euupe/.

We calculated the stochastic population forecast using the Program for Error Propagation (PEP) developed by Juha Alho. This program takes as its inputs the jump-off population and predicted mortality rates and fertility rates (for women) as well as net migration, all by one-year age groups for all the forecast years. In addition one must specify uncertainty parameters for these rates and the rates' co-variances across time, age, and between the sexes.[4] The program then draws sample values from a standard normal distribution, and transforms them into correlated errors. Adding these errors to the specified rates in the logarithmic scale creates a sample path for the vital rates. This sample path together with the jump-off population is then used to calculate a sample path for the future population, using a cohort component model. The process is repeated to create the number of desired sample paths for the population.

We updated the results from the UPE project by changing the jump-off years to 2011, and using age-specific death rates, birth rates, and net migration numbers taken from recent population forecasts. The remaining assumptions, that is, the variances and co-variances for the mortality rates, fertility rates, and net migration, were kept unchanged. The assumption here is that the volatility of fertility, mortality, and migration for the period 2011-2041 in the five countries is the same as that assumed in the UPE-project for lead times of 10, 20, and 30 years.

[3] This multiplication assumes independence between the share $\widehat{\alpha}$ and the population number $\widehat{W}$. Reasons why this assumption is justified are discussed by Alho and Keilman (2010).

[4] Fertility, mortality, and net migration are assumed to be independent of each other.

When computing the number of households based on the number of persons in various household positions, a number of assumptions were made.

- The numbers of one-person households, lone fathers, and lone mothers equal the numbers of persons with household position SIN0, SIN+ (men) and SIN+ (women), respectively.
- The numbers of cohabiting and married couples equal half the numbers of persons with household positions COH and MAR, respectively.
- The number of other households lone equals the number of persons with household position OTHR divided by the mean household size of other households. Mean household sizes for this type were assumed to be 2.05, 3.99, 2.52, 5.86, and 1.30 for Denmark, Finland, Germany, Netherlands, and Norway, respectively. The latter numbers are based on information from the Census of 2011 in each country (Eurostat 2014). They are assumed to be the same for all three future years.

Table 4 shows that predicted developments in important household types in the five countries are as one could expect, given our assumptions. Numbers of one-person households and of cohabiting couples will increase to 2041, whereas numbers of married couples will fall in all but one country. In Norway, this number will increase slightly. These developments reflect our assumptions of a continuation of historical trends in household shares (for Denmark, Finland, and the Netherlands) and of constant transition probabilities between household positions (Germany and Norway). Except for Finland, numbers of private households grow faster than population numbers – in Germany one may even expect a *falling* population size. As a consequence, the average size of private households will fall. This development is explained by a strong growth of one-person households, by some 40 per cent or more for the period 2011-2041. Finland is an exception: one-person households grow by no more than 20 per cent 20 per cent during the period. This increase is counteracted by a decline in married couples by 12 per cent. As a result, the increase in the total number of households (8 per cent) is less than that of population size (10 per cent), and average household size will increase a little from 2.1 in 2011 to 2.2 in 2041.

More interesting than predicted numbers of household is the uncertainty in those predictions. Table 4 reports the coefficient of variation (CV) for each prediction, defined as the standard deviation across 1000 simulations divided by the average value. Thus the CV is a relative measure of uncertainty. First note that, without exception, uncertainty increases with increasing forecast lead time. Second, relative uncertainty is small for numerous households. Predictions of married couple households and of one-person households are more certain than those of cohabiting couples, and <u>much</u> more certain than predictions of lone-parent households.

How do the results in Table 4 compare with other probabilistic household forecasts? Van Duin and Stoeldraijer (2011) report a probabilistic household forecast for the <u>Netherlands</u> for the period 2011-2060, based on the approach developed by De Beer and Alders (1999) and Alders (1999, 2001). Van Duin and Stoeldraijer predict decreasing numbers of married couple households (to 3 million in 2041; cf. 3.3 million in Table 4) and growing numbers of cohabiting couples (1.2 million in 2041; 1.0 million), one-person households (3.7 million in 2041; 4.0 million), and lone parents (540 000 in 2041; 482 000). The total number of private households is expected to grow to 8.5 million (8.8 million in Table 4), with a 67 per cent prediction interval stretching from 7.9 million to 9.0 million. This would imply a CV of 6 per cent (5 per cent in Table 4). Thus the household trends predicted by Van Duin and Stoeldraijer (2011) are broadly similar to ours, in spite of a methodology that is very different. More detailed results will likely show larger differences.

Christiansen and Keilman (2013) computed probabilistic forecasts for Denmark and Finland based on random shares. Expected values for the shares were computed based on multi-state household predictions made by LIPRO, whereas uncertainty was derived from Random Walk with Drift (RWD) processes for the shares, for each combination of age, sex, and household position in the two countries. The results are very different from ours. First, whereas Table 4 shows a decrease in the number of households in Finland, Christiansen and Keilman find growing numbers, caused in particular by more one-person households and more married couples. For Denmark, they find a slower growth in the number of households, caused by relatively moderate increases in numbers of one-person households and of cohabiting couples. (The trend in married couple households is similar to that in Table 4.) The diverging findings were to be expected, because the models for household shares differ strongly between the two approaches. More interestingly, uncertainty around predicted numbers in Table 4 is much larger than that found by Christiansen and Keilman. The reasons are not entirely clear, but one explanation is that the latter two authors assumed that the estimation variance of the drift estimate in the RWD model equals the innovation variance divided by one minus the number of observations of the time series. In the current project we used estimation variance for the drift based on robust standard errors; cf. expression (3) and Table 1.

7. Conclusions and discussion

In this paper we show how techniques of data dimension reduction can be used to predict patterns of household dynamics in a multi-country context. We compute probabilistic household forecasts for Denmark, Finland, Germany, the Netherlands, and Norway, spanning the period 2011-2041. Starting point is the population of each country broken down by age, sex, and household position as reported in the census round of 2011. Future trends in fertility, mortality and international migration are taken from official population forecasts. For changes in household structure we rely on time series of household data. The results show a continuation of current trends towards more and smaller households, often driven by increasing numbers of persons who live alone. The number of households increases faster than population size, which leads to falling average household size. A very consistent finding is that larger households are easier to predict than smaller households, at least when uncertainty is considered in a relative sense.

In Section 4.2.1 we have defined a random walk with drift (RWD) process for the fractions $\xi_k(x,t)$. The drift equals $a_k + b_k.\ \xi_k^S(x)$, where $\xi_k^S(x)$ is a standard age pattern. This standard is defined period-wise to account for year-to-year changes in the fraction $\xi_k(x,t)$. The term $\xi_k^S(x)$ preserves the age pattern in the random walk increments. *Cohort* effects in the age profiles not accounted for. For example, one could assume that an increasing share of women who cohabit at age 25 in 1995 goes together with larger shares of cohabiting women aged 45 twenty years later. To implement such cohort effects in the Brass relational model would require a standard profile for birth cohorts, in addition to one for periods. We haven't done that, and as a consequence the predicted age profiles for the shares in Figures 2-4 may be wrong.

Another issue is that of coherence between men and women. In the observed data for Denmark, Finland, and the Netherland there is a close correspondence between the numbers of men and women in household types COH and MAR. The numbers are not exactly equal, caused by partnership formation and marriage across international borders, same-sex couples, and errors in the registration. But the numbers are close. This coherence is lost when we predict shares for cohabiting and married men and women separately. Keilman and Van Duin (2014) attempted to preserve this coherence by modelling the shares by means of the Hyndman-Booth-Yasmeen product-ratio variant of the Lee-Carter model, adapted to household shares. However, as reported in Section 4.2.1 this led to unrealistic results. An ad-hoc adjustment of numbers of men and women in household positions COH and MAR is a practical solution, pending the discovery of a more satisfactory solution to this problem of the sexes. Note that the multi-state household projections for Germany and Norway *did* include coherent numbers of men and women who live as a couple, but only in the expected values of the shares. The stochastic simulations of the shares for these two countries did not preserve the coherence.

The populations of the five countries in this study are ageing, similar to those in other countries. With constant shares for persons who live in an institution (INST), this might lead to an enormous increase in the number of persons in such institutions. Indeed, the historical downward trend in shares for household position INST in Denmark, Finland, and the Netherlands is extrapolated by our RWD model. However, in case ageing is faster than the downward trend of these shares, the capacity of elderly institutions may have to be increased. An example is Finland, where the number of persons who live in an institution is predicted to grow from 111000 in 2011 to an expected 436000 thirty years later. In Denmark the downward trend in the shares results in a decrease of the institutionalized from 81500 in 2011 to 55200 in 2041; for the Netherlands the numbers are about constant between 2011 (219000) and 2031 (261000), but they increase to 327000 in 2041. As mentioned in Section 4.2.2, for Germany and Norway we have assumed constant capacity of the institutions. This was made possible by the multi-state household model. For Norway this resulted in 51800 institutionalized in 2041, almost the same number as in 2011 (52400). For Germany, however, the averages across 1000 simulations indicate a strong increase: 631000 in 2011, and 1.35 million in 2041. The reason is that the Eurostat population forecast for Germany involves stronger ageing than our multi-state household projection.

References

Alders, M. (1999) Stochastische huishoudensprognose 1998-2050 (Stochastic household forecast 1998-2050). *Maandstatistiek van de Bevolking* 11: 25–34.

Alders, M. (2001) Huishoudensprognose 2000-2050: Veronderstellingen over onzekerheidsmarges (Household forecast 2000-2050: assumptions on uncertainty intervals). *Maandstatistiek van de Bevolking* 8: 14–17.

Alders, M., Keilman, N., and Cruijsen, H. (2007). Assumptions for long-term stochastic population forecasts in 18 European countries. European Journal of Population 23(1): 33-69. doi:10.1007/s10680-006-9104-4.

Alho, J., Alders, M., Cruijsen, H., Keilman, N., Nikander T., and Quang Pham, D. (2006). New forecast: Population decline postponed in Europe. Statistical Journal of the United Nations ECE12(23): 1-10.

Alho, J., Cruijsen, H., and Keilman, N. (2008). Empirically based specification of forecast uncertainty. In: Alho, J., Hougaard Jensen,S., and Lassila, J. (eds.). Uncertain demographics and fiscal sustainability. Cambridge: Cambridge University Press: 34-54. doi:10.1017/CBO9780511493393.004.

Alho, J. and Keilman, N. (2010) On future household structure. *Journal of the Royal Statistical Society* Series A 173(1): 117-143. doi:10.1111/j.1467-985X.2009.00605.x.

Christiansen, S.G. and Keilman, N. (2013) Probabilistic household forecasts based on register data - the case of Denmark and Finland. *Demographic Research* 28(43). doi: 10.4054/DemRes.2013.28.43.

De Beer, J. and Alders, M. (1999) Probabilistic population and household forecasts for the Netherlands. Geneva: Joint Economic Commission for Europe–EUROSTAT Work Session on Demographic Projections, Perugia, May 3rd–7th (Working paper 45).

Eurostat (2014) http://epp.eurostat.ec.europa.eu/portal/page/portal/population_and_housing_census/census_data/2011_census_ accessed 31 July 2014. Updated link http://ec.europa.eu/eurostat/web/population-and-housing-census/census-data/2011-census

Goplen, E. (2014) Barn i det moderne samboerskap – en bindende eller splittende faktor? En analyse av SSBs registerbaserte samboerdata (Children in modern consensual unions – a binding or a splitting factor? An analysis of Statistics Norway's registerbased data on consensual unions). Master's Thesis Department of Economics, University of Oslo. http://urn.nb.no/URN:NBN:no-45616

Hiekel, N., Liefbroer, A.C., Poortman, A.-R. (2014) Understanding diversity in the meaning of cohabitation across Europe. *European Journal of Population* 30: 391-410. DOI: 10.1007/s10680-014-9321-1.

Jiang, L. and O'Neill, B. (2004). Toward a new model for probabilistic household forecasts. *International Statistical Review* 72: 51–64.

Keilman, N. and Van Duin, C. (2014) Stochastic Household Forecast by Coherent Random Shares Prediction. Memorandum 10/2014, Department of Economics, University of Oslo. http://www.sv.uio.no/econ/english/research/unpublished-works/working-papers/ .

Pailhé, A., Mortelmans, D., Castro, T., Cortina Trilla, C., Digoix, M., Festy, M. , Krapf, S., Kreyenfeld, M., Lyssens-Danneboom, V., Martín-García, T., Rault, W., Thévenon, O., Toulemon, L. (2014) Changes in the life course: A state-of-the-art report. FamiliesAndSocieties Working Paper 6/2014. http://www.familiesandsocieties.eu/?page_id=131

Peters, A., and Liefbroer, A.C. (1997) Beyond marital status: Partner history and well-being in old age. *Journal of Marriage and the Family*, 59 (3), 687-699.

Preston, S., Heuveline, P., Guillot, M. (2001) *Demography: Measuring and Modelling Population Processes*. Oxford: Blackwell Publishers.

Scherbov, S. and Ediev, D. (2007) Probabilistic household projections based on an extension of headship rates method with application to the case of Russia. Geneva: Joint Economic Commission for Europe–EUROSTATWork Session on Demographic Projections, Bucharest, Oct. 10th–12th (Working paper 16).

United Nations (1973). Methods of projecting households and families. New York: United Nations.

United Nations (2010) World Population Ageing 2009. ST/ESA/SER.A/295. New York, United Nations.

US National Resources Planning Committee (1938). The problems of a changing population. Washington DC: Government Printing Office.

United States Census Bureau (2014) 65+ in the United States: 2010. Washington D.C., U.S. Government Printing Office.

Van Duin, C., and Stoeldraijer, L. (2011) Huishoudensprognose 2011-2060: meer en kleinere huishoudens (Household forecast 2011-2060: more and smaller households). *Bevolkingstrends* 59(2), 59-67.

Wilson, T. (2013) Quantifying the uncertainty of regional demographic forecasts. *Applied Geography* 42, 108-115.

Table 1. Parameter estimates for model (1). Student t-values based on robust standard errors

	a_k		b_k		
k	estimate	t-value	estimate	t-value	$Cov(a_k,b_k)$
2	-0.0005697	-0.7	-0.0076073	-7.6	-5.75e-7
3	-0.0432034	-11.8	0.0083405	5.2	-4.88e-6
4 (men)	0.0385686	27.6	-0.0033708	-1.8	-1.48e-6
4 (women)	0.0211024	18.1	0.0040122	4.4	-3.60e-7
5	0.0652686	14.2	-0.0109857	-6.3	-7.16e-6
6	0.0313597	1.1	0.0121778	0.3	0.0010577

Table 2. Percentage of life time spent in various household positions, and percentage of children born in various household positions of the mother; Germany

	CHLD	SIN0	COH	MAR	SIN+	OTHR	INST	All (=100%)
				Germany; input rates 2011-2015				
				%				Years
Men	27	21	11	32	1	8	1	76.8
Women	24	23	11	31	4	6	2	82.6
				%				Children
	0	8	27	54	5	6	0	1.41
				Germany; input rates 2036-2040				
				%				Years
Men	25	23	10	31	1	8	1	81.4
Women	22	25	10	31	4	6	2	87.0
				%				Children
	0	8	27	54	5	6	0	1.52

Table 3. Percentage of life time spent in various household positions, and percentage of children born in various household positions of the mother; Norway

	CHLD	SIN0	COH	MAR	SIN+	OTHR	INST	All (=100%)
				Norway; input rates 2011-2015				
				%				Years
Men	34	19	12	27	2	5	1	76.8
Women	29	20	13	26	6	5	1	82.7
				%				Children
	0	8	38	42	6	5	0	1.53
				Norway; input rates 2036-2040				
				%				Years
Men	32	21	12	27	2	5	1	81.3
Women	27	22	12	27	6	5	1	88.5
				%				Children
	0	8	38	43	5	5	0	1.74

Table 4. Private households and population. 2011 (Census numbers), and 2021-2041 (averages across 1000 stochastic simulations), in millions. In parentheses: Coefficients of variation, in per cent.

	One person households	Cohabiting couples	Married couples	Lone fathers	Lone mothers	All private households (incl. other private households)	Population size
			Denmark				
2011	0.946	0.300	1.031	0.033	0.152	2.541	5.561
2021	1.190 (9.8)	0.398 (17.1)	1.076 (7.4)	0.026 (48.4)	0.117 (29.0)	2.847 (2.5)	5.825 (1.0)
2031	1.391 (13.0)	0.469 (21.3)	1.020 (11.8)	0.024 (52.2)	0.108 (35.2)	3.044 (3.9)	6.096 (2.8)
2041	1.548 (15.7)	0.525 (24.8)	0.962 (14.7)	0.020 (59.4)	0.103 (39.4)	3.188 (5.5)	6.321 (5.3)
			Finland				
2011	1.040	0.298	0.928	0.028	0.141	2.515	5.375
2021	1.183 (9.8)	0.360 (18.2)	0.929 (8.7)	0.036 (68.7)	0.130 (43.3)	2.690 (3.0)	5.639 (1.5)
2031	1.269 (13.6)	0.416 (22.7)	0.871 (13.7)	0.035 (69.4)	0.122 (47.8)	2.757 (5.0)	5.822 (4.0)
2041	1.248 (18.3)	0.475 (25.5)	0.816 (17.0)	0.027 (76.5)	0.113 (50.6)	2.716 (7.4)	5.923 (6.8)
			Germany				
2011	13.5	2.830	17.2	0.423	2.439	36.5	79.7
2021	17.1 (9.9)	3.976 (21.2)	15.5 (7.7)	0.785 (79.7)	1.795 (51.9)	41.2 (3.1)	80.3 (1.8)
2031	18.5 (13.3)	4.213 (28.5)	14.4 (12.1)	1.038 (75.7)	1.691 (56.5)	41.7 (5.2)	79.7 (4.6)
2041	19.2 (16.0)	4.244 (35.1)	13.5 (15.3)	1.088 (81.0)	1.638 (59.8)	41.4 (7.8)	78.1 (8.1)

			Netherlands				
2011	2.708	0.921	3.274	0.086	0.410	7.479	16.7
2021	3.112 (10.0)	0.845 (19.0)	3.506 (6.1)	0.090 (54.7)	0.353 (34.1)	7.962 (2.6)	17.3 (1.0)
2031	3.520 (13.0)	0.925 (25.1)	3.424 (9.5)	0.104 (58.3)	0.351 (42.0)	8.383 (3.9)	17.8 (2.7)
2041	3.971 (14.9)	0.966 (29.9)	3.295 (11.6)	0.106 (64.8)	0.376 (46.4)	8.777 (5.3)	17.8 (4.3)
			Norway				
2011	0.880	0.275	0.848	0.045	0.158	2.286	4.980
2021	0.923 (10.0)	0.327 (20.2)	0.846 (9.2)	0.065 (63.7)	0.158 (46.0)	2.545 (2.6)	5.560 (1.3)
2031	1.082 (13.3)	0.378 (26.7)	0.882 (14.1)	0.087 (64.2)	0.165 (51.0)	2.832 (4.0)	6.123 (3.4)
2041	1.205 (16.2)	0.404 (31.8)	0.911 (16.6)	0.096 (69.6)	0.176 (53.4)	3.037 (5.6)	6.450 (6.1)

Figure 1. Household structure of the population in five countries. Explanation of legend: “chld”: child living with parent(s); “sin0”: person living alone; “coh”: person living with cohabitee; “mar”: person living with marital spouse; “sin+”: lone parent; “othr”: other private household position; “inst”: person living in institution. Source: Eurostat.

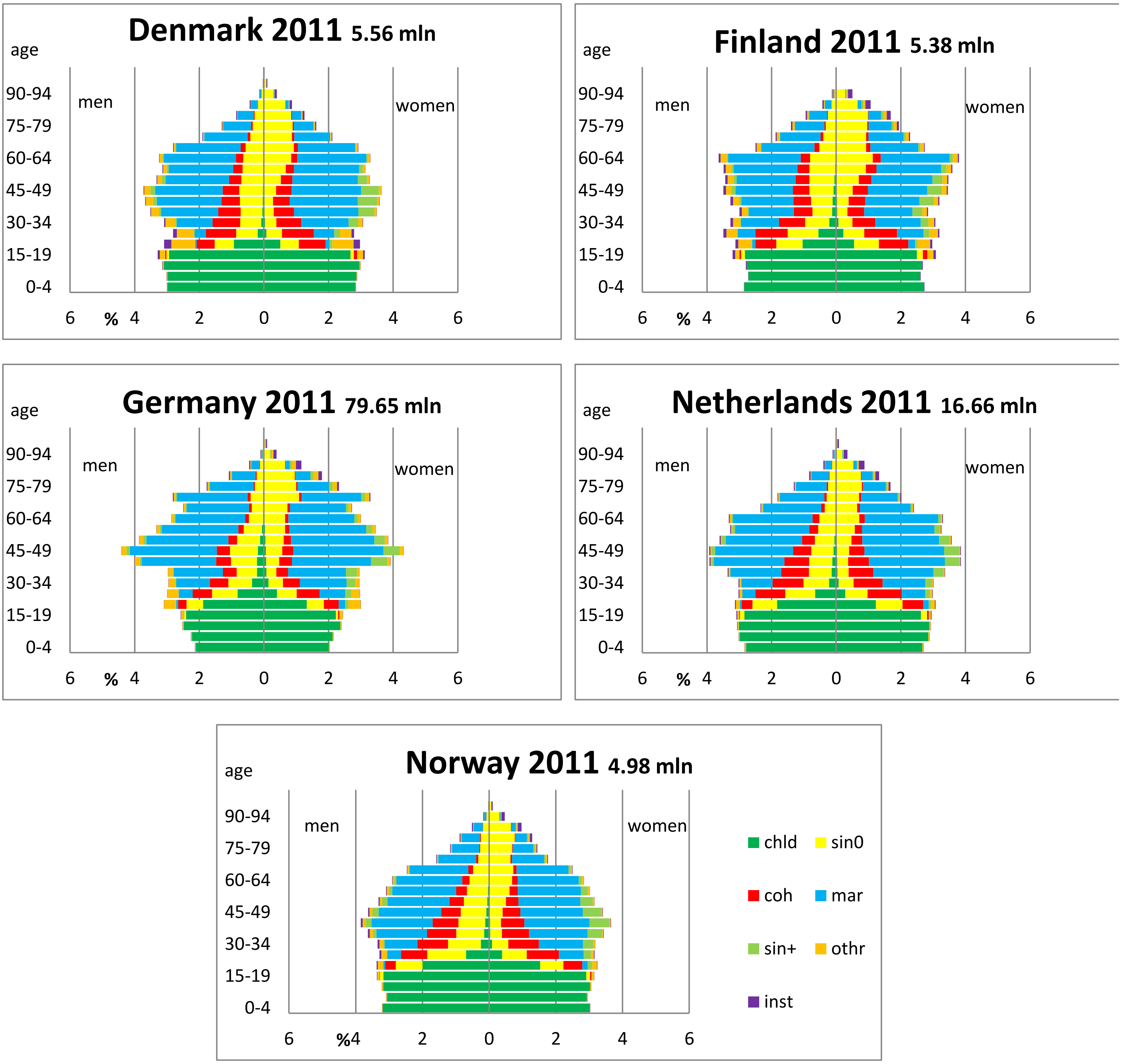


Figure 2. Observed (1981-2011) and predicted (2021-2041) shares of persons in selected household positions, by age, Denmark. Data sources: 1981-2001 register data; 2011 census data; 2021-2041 model extrapolations.

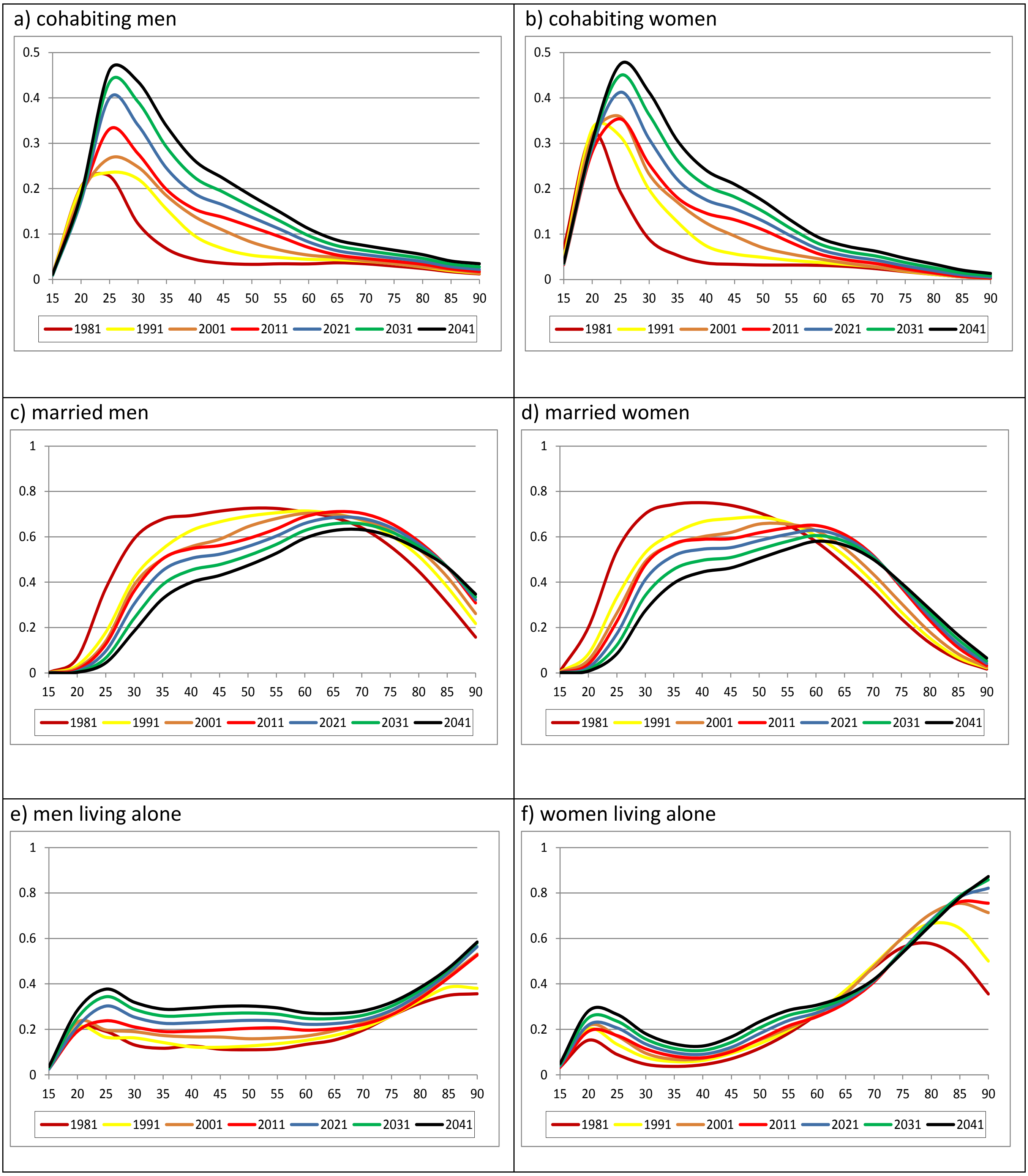


Figure 3. Observed (1991-2011) and predicted (2021-2041) shares of persons in selected household positions, by age, Finland. Data sources: 1991-2001 register data; 2011 census data; 2021-2041 model extrapolations.

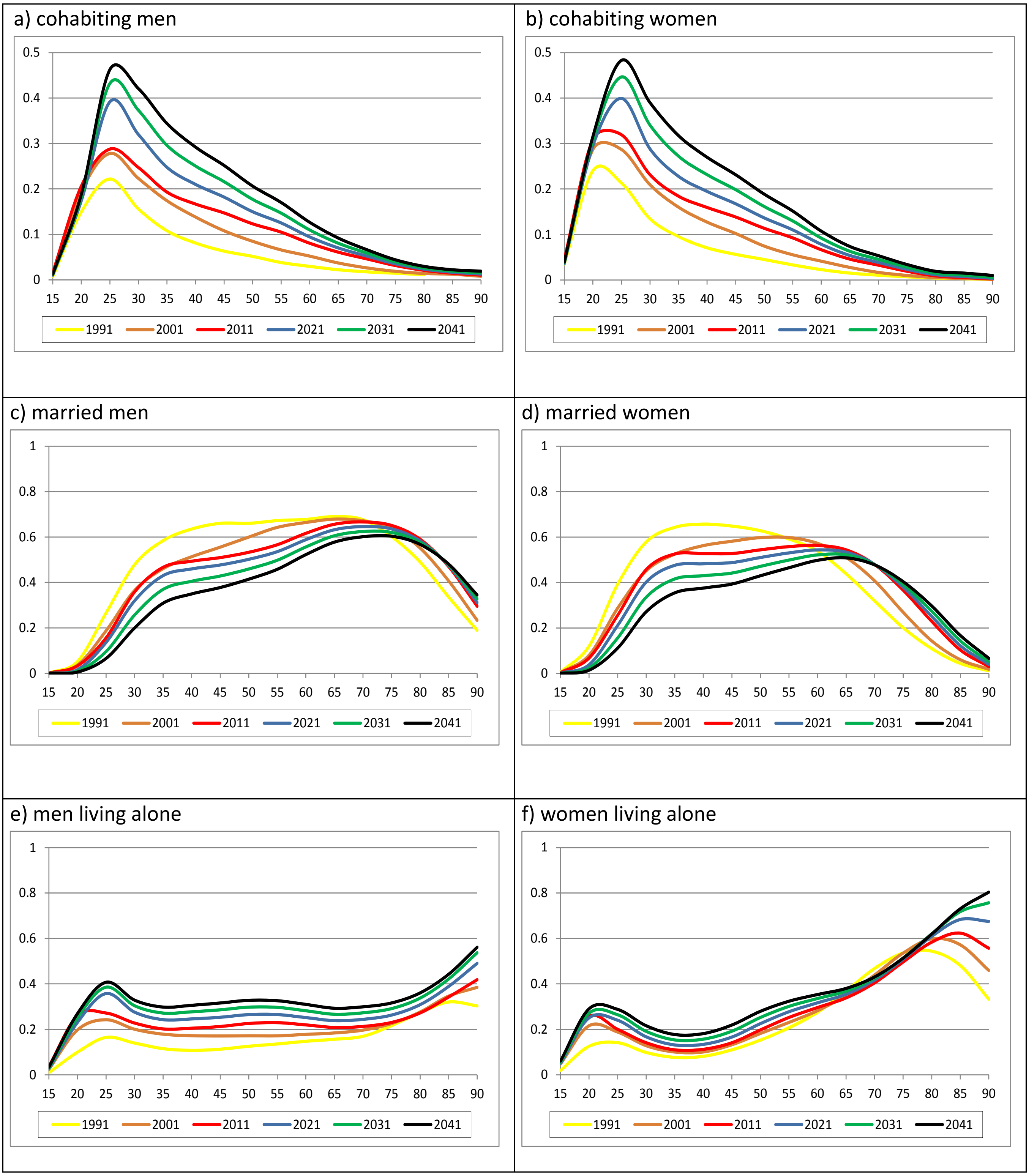
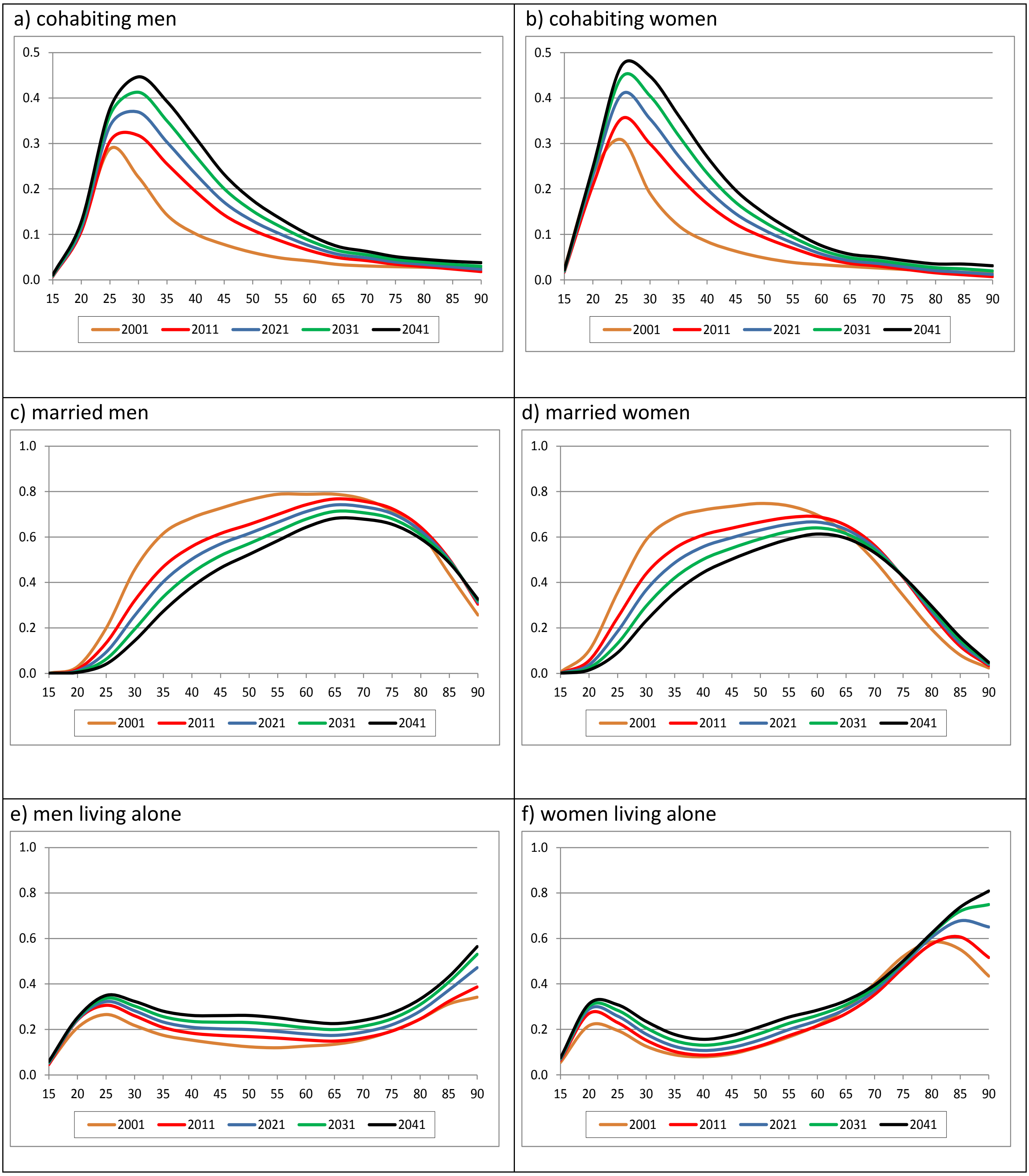


Figure 4. Observed (2001-2011) and predicted (2021-2041) shares of persons in selected household positions, by age, Netherlands. Data sources: 2001 register data; 2011 census data; 2021-2041 model extrapolations.



Appendix 1. Back transformation from ξ to α

In Section 4.2.1 the shares α_j are transformed into fractions ξ_k. In this appendix we outline the back transformation from ξ_k to α_j. We suppress indices for age, sex, time, and country.

Starting point is the set of expressions that transform the shares α_j into fractions ξ_k.

$\xi_2 = \text{logit}((\alpha_3 + \alpha_4)/(1 - \alpha_1))$

$\xi_3 = \text{logit}(\alpha_4/(\alpha_3 + \alpha_4))$

$\xi_4 = \text{logit}((\alpha_2 + \alpha_7)/(\alpha_2 + \alpha_5 + \alpha_6 + \alpha_7))$

$\xi_5 = \text{logit}((\alpha_2)/(\alpha_2 + \alpha_7))$

$\xi_6 = \text{logit}(\alpha_5/(\alpha_5 + \alpha_6))$

There are many equivalent expressions for the α_j written as functions of the ξ_k. One of these is the following set

$\alpha_2 = (1 - \alpha_1)\exp(\xi_4)\exp(\xi_5)/\{(1 + \exp(\xi_2))(1 + \exp(\xi_4))(1 + \exp(\xi_5))\}$

$\alpha_3 = (1 - \alpha_1)\exp(\xi_2)/\{(1 + \exp(\xi_2))(1 + \exp(\xi_3)\}$

$\alpha_4 = \alpha_3\exp(\xi_3)$

$\alpha_6 = (1 - \alpha_1 - \alpha_3 - \alpha_4)/\{(1 + \exp(\xi_4))(1 + \exp(\xi_6)\}$

$\alpha_5 = \alpha_6\exp(\xi_6)$

$\alpha_7 = \alpha_6\exp(\xi_4)(1 + \exp(\xi_6))/(1 + \exp(\xi_5))$

Note that α_1 is independent of ξ_k (k=2,3,…6)