Distr.: General
1 April 2016

English only

Economic Commission for Europe

Conference of European Statisticians

Joint Eurostat/UNECE Work Session on Demographic Projections
Geneva, 18-20 April 2016
Item 7 of the provisional agenda
Beyond projections by age and sex

Probabilistic household forecasts for Denmark, Finland, and the Netherlands 2011-2041: Combining the Brass relational method with a Random Walk model

Prepared by Nico Keilman, Department of Economics, University of Oslo

Summary

Probabilistic household forecasts to 2041 are presented for Denmark, Finland, and the Netherlands. Future trends in fertility, mortality and international migration are taken from official population forecasts. Time series of shares of the population in six different household positions are modelled as Random Walks with a Drift. Brass' relational model preserves the age patterns of the household shares. Probabilistic forecasts for households are computed by combining predictive distributions for the household shares with predictive distributions of the populations, specific for age and sex.

If current trends in the three countries continue, we will witness a development towards more and smaller households, often driven by increasing numbers of persons who live alone. We can be quite certain that by 2041, there will be between two and four times as many persons aged 80 and over who live alone when compared with the situation in 2011.

I. Introduction

1. Household forecasts are useful for the planning of welfare provisions, housing supply, and the demand for consumer durables. For instance, elderly persons who live alone are more vulnerable than those who live with a partner. Therefore, household status is an important determinant for the need of formal and informal support and care for the elderly, in addition to health (e.g. Lakdawalla and Philipson 1999; Muller et al. 1999; Lakdawalla et al. 2003; Grundy and Jital 2007). Projections of social welfare spending depend, among others, on the number of lone parents in the future (Jacobsen and Jensen 2014). Falling fertility rates and more frequent divorce and separation in western countries after World War 2 have led to smaller households (OECD 2011); in turn, this has had a strong impact on housing needs (Van Vliet et al. 1985). Another consequence of falling average household size has been an increasing demand for energy, because of economies of scale in households of different sizes (O'Neill and Chen 2002).

2. Deterministic household forecasts have been computed since the 1930s. Nowadays, probabilistic household forecasts are increasingly being accepted as a useful means of quantifying uncertainty; see Bijak et al. (2015) and the review in Section 2. A recent example of a probabilistic household forecast is given by Christiansen and Keilman (2013). They used an approach based on random household shares to compute such forecasts for Denmark and Finland. The household shares represent the proportions of persons who live in a certain household position (living alone, living with marital partner, living in an institution etc.) at a certain point in time, specific for age and sex. Because the future is uncertain, future values of these shares were treated as random variables. The authors obtained expected values of the multivariate predictive distributions for the shares by means of the multi-state household projection model LIPRO, while time series models of the household shares were used to assess the uncertainty in predicted shares, i.e. the variances and covariances of the distributions.

3. The purpose of this paper is to simplify the approach used by Christiansen and Keilman – abbreviated as C&K henceforth. Their method requires transition probabilities between household positions as input data. Transition probabilities of this kind are available from the population registers of Denmark and Finland, but few other countries have similar data for their population as a whole. Panel surveys may be used to compute the required transition probabilities, but one needs a large sample because the probabilities have to be specified for men and women broken down by age. We avoid transition probabilities by formulating multivariate time series models for the shares in each household position. We find that a Random Walk with Drift (RWD) model is a good representation of the data. The typical age patterns of household positions are modelled by means of the Brass Relational method. Predictive distributions of the household shares for future years are obtained from the RWD models. In order to compute a probabilistic forecast for the numbers of persons in certain household positions, simulated predictive distributions for the household shares are combined with simulated predictive distributions of the populations, specific for age and sex. Correlations across ages and between men and women are accounted for. Finally, we use simple accounting rules for computing future numbers of *households* of different types, and the uncertainty therein, based on simulated predictive distributions for the number of *persons* by household position.

4. We apply our model to annual data for Denmark, Finland and the Netherlands. The latter country was added because with only 15 years of data, the series of observations on household shares is relatively short. Hence it is useful to investigate whether the proposed method still gives acceptable results in this situation. The probabilistic household forecasts span the period 2011-2041. Starting point is the population of each country broken down by age, sex, and household position in 2011, based on census data from Eurostat. Future trends in fertility, mortality and international migration are taken from official population forecasts.

II. Earlier work

5. C&K reviewed various methods for household projection and forecasting. Here we give a brief summary.

6. *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 used stochastic simulation and 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 2020 as the product of two other random variables, namely the number of women aged 40 in 2020 and the share of 40-year old women who live as a lone mother in 2020. 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 other words, each year a simulated death rate for a certain age was higher than expected, so were the rates for all other ages, and similarly for birth rates (by mother's age) and immigration numbers. In addition the authors assumed perfect auto-correlation for the random shares: when the simulated shares for one particular year were high, they would also be high in the following years.

7. 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 (2013a, b) computed a probabilistic household forecast for Greater Sidney. Household parameters were modelled as a 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.

8. A problem connected to these probabilistic household forecasts 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. As mentioned in the previous section, C&K used long time series data of observed shares for Denmark and Finland, and formal time series methods to quantify the uncertainty connected to household shares in the future. Expected values of the shares were computed using a multi-state model of household dynamics.

III. Data

9. The starting point of our forecasts is census data on the population of the three countries, broken down by five-year age group, sex, and seven household positions in 2011. 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. Living in a private household, but not in any of the positions described above (OTHR)
7. Living in an institution (INST).

10. Time series of annual data on the number of people in these seven household positions were obtained from the population registers of Denmark, Finland and the Netherlands. The data relate to 1 January of the years 1981-2007 for Denmark, of the years 1988-2009 for Finland, and of the years 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+.

IV. Modelling household shares

a) Household positions

11. The purpose of the modelling exercise is to obtain predictive distributions for the household shares in each country, broken down by sex and five-year age group. Write $V(j,x,s,t,c)$ for the number of people in household position $j=1,2,\ldots,7$ who are in age $x=0,1,\ldots$ and have sex $s=1$ or 2, at time $t=0,1,2,\ldots$ in country $c=1,2,3$. Aggregating over position, we obtain the population $W(x,s,t,c) = \Sigma_j V(j,x,s,t,c)$ of age x and sex s at time t for country c. Household position j has share $\alpha(j,x,s,t,c) = V(j,x,s,t,c)/W(x,s,t,c) = \alpha_j(x,s,t,c)$. The household positions are numbered as follows: CHLD (j=1), SIN0 (j=2), COH (j=3), MAR (j=4), SIN+ (j=5), OTHR (j=6), INST (j=7).

12. For modelling random evolution of the shares, a logit transformation was applied. Building on earlier work (Alho and Keilman 2010; Christiansen and Keilman 2013; Wilson 2013a, 2013b), we have opted for a hierarchy of household positions using a variant of continuing fractions. This led to six types of fractions to be modelled (all specific for age, sex, time, and country). The following fractions were 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.

13. Because of the hierarchy, predicted shares in the logit scale at a higher level are independent of predicted shares at a lower level. The particular sequence 2-6 above is based upon the idea that important shares (numerically, behaviourally) ought to be modelled first, and the less important ones can come last. Hence persons who live together with a partner (points 2 and 3 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 (C&K). Therefore initially they are treated as one group (point 4).

14. Children have been singled out from the beginning, because their shares are assumed to be 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, 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.

15. Temporarily suppressing indices for age, sex, time, and country, the logit transforms of the fractions 2-6 above are

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

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

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

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

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

16. This way, five series were constructed, for each combination of country, sex, and age group.

b) A Random Walk with Drift model

17. Much attention has been given to the age pattern of the shares of each household position. There are many possible modelling strategies for such age patterns. Keilman (2016) gives a brief review, and concludes that a Brass type of relational model works well, at least compared to the more popular Lee-Carter model. The reason is the particular form of the latter model. When the trend in the shares is downward for some ages and upward for other ages, one of the parameters of the Lee-Carter model (namely the age profile b_x), may have both positive and negative values. When in addition the general time trend in the shares is systematically upwards or downwards, the result is a strong distortion in the age pattern of predicted shares; see Lee and Miller (2001) for a discussion of this issue in the context of mortality.

18. 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 with zero expectation and constant variance. 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 thorough discussion.

19. The model is linear in its parameters. Hence, in a first step, we used the method of ordinary least squares (OLS) to estimate the Brass model applied to the age pattern of logit transformed fractions $\xi_k(x,t)$. The standard age pattern $\xi_k^S(x)$ was defined as the average value of $\xi_k(x,t)$, where for each k the average was taken over all years t, for a given combination of age, sex and country. For each k we obtained estimates of parameters a and b that varied over time, between sexes and between countries. In terms of the coefficient of determination (R^2), the fit was excellent in almost all cases, with R^2 values larger than 0.9 - many of them larger than 0.97.

20. In most cases we noticed a gradual increase or decrease over time in the estimates of a and b. To illustrate, the upper panel of Figure 1 shows the annual estimates of the Brass model parameters for Danish women, for k=4 (SIN0 plus INST as a share of the total share of SIN0, SIN+, OTHR, and INST) and k=6 (SIN+ as a share of SIN+ plus OTHR). The latter two cases were selected because they fitted best to Danish data (as judged by a value of R^2 equal to 0.996) and least (R^2=0.912) in terms of a model in which the parameters are assumed to follow a straight line; see model (1) below. Since the standard age pattern is defined as the average pattern for the period, the constant a is close to zero around the middle of the period, while the coefficient b is close to one. In both cases, the constant term increases regularly; for lone mothers, however, the rotation of the age pattern is somewhat irregular. For the case of SIN0 and INST combined the slope of the age profile increases systematically. The lower panel shows estimates for Finnish men, for the cases k=2 (COH plus MAR; R^2 =0.976) and k=6 (SIN+; R^2=0.994).

21. The gradual developments in parameter estimates suggested that a and b could be written as linear functions of time, i.e.

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

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

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

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.

23. Model (2) defines $\xi_k(x,t)$ as a random walk with drift (RWD). The drift $a_k + b_k.\xi_k^S(x)$ consists of a part a_k that is common for all ages, whereas $b_k.\xi_k^S(x)$ is an age-specific part. The term $\xi_k^S(x)$ preserves the age pattern in the random walk increments for each type of fraction k. The innovation variance is $\sigma_k^2 = \mathrm{Var}(d_k(x,t))$.

24. In a second step, parameters a_k and b_k of the model in expression (2) were estimated by OLS (across ages x) assuming an innovation variance independent of age and time. For each type of fraction $\xi_k(x,t)$ (k=2,3,4,5,6), country-specific estimates were very similar, and differences were not significant in most cases. Thus in a third step, the model was re-estimated for all three countries simultaneously. In addition, we did not distinguish by sex: the results from the second step showed small differences between men and women, 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 women's chances of living alone have increased relatively slowly, because their survival chances increased not as rapidly as those of men. Table 1 gives the parameter estimates. 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.

c) Predictions

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

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

26. Hence an h-step ahead forecast is

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

where E[.] denotes expectation, and a_k and b_k have been replaced by their estimated values. The forecast error $F_k(x,T+h)$ equals $\xi_k(x,T+h)$–$\mathrm{E}[\xi_k(x,T+h)]$. Given our assumptions, its variance is

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

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

where $\mathrm{Cov}\left[\hat{a}_k, \hat{b}_k\right]$ denotes the covariance between $\hat{a}_k$ and $\hat{b}_k$.

27. The estimated models (2) were used to extrapolate the logit-transformed fractions $\xi_k(x,t)$ to 2041. Figures 2 gives selected results for observed and predicted shares $\alpha_j(x)$ for Denmark, where the predictions were obtained by back transformation of the logit-transformed fractions $\xi_k(x,T+h)$. We see 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 (aged 80+) 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), resulting from "deinstitutionalization" of elder care that started in the 1970s in Denmark, and a stronger focus on other forms of long-term care such as sheltered housing and home services (Daatland 1997).

d) Variances and correlations

28. The logit-transformed fractions $\xi_k(x,T+h)$= $\xi_k(x,s,T+h,c)$ are assumed to have a multivariate normal distribution, with expected values given by expression (3). Expression (4) specifies their variances. We have no reason 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 (4) by this average.

29. Covariances/correlations remain to be specified. The fractions $\xi_k(x,s,t,c)$ are correlated across ages x, across sexes s, and between countries c. Since each fraction is modelled as a Random Walk with Drift process, it has 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 residuals of model (2).

30. For k=2,3,4,5, and 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 for this group. 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.

31. Following earlier work (Alho and Keilman 2010, C&K) we assumed an AR1 process for the errors in the age dimension. There was little systematic difference in the estimated correlations across ages. Inspecting correlations for different types of fractions (k=2,…,6), we found extremely high estimates age correlation 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. In other words, when the simulated value of ξ_2 is high/low for a particular age group, it is also high/low for all other age groups. For the other types (k=3-6) there was no systematic pattern. We have used the median correlation across ages and types, which is 0.756.

V. Household forecasts

a) 1 *Method*

32. Below we present selected forecast results for the three countries for the years 2021, 2031, and 2041. Starting point was the household structure in 2011 based on census information; see Eurostat (2014). Figure 3 shows the age pyramids of the Netherlands together with the household structures of the population. For the other two countries the graphs are very similar.

33. 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, although cohabitation is frequent among adults under the age of 30. The graph shows a consistent pattern of more young men than young women who live alone. 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 (Peters and Liefbroer 1997, United Nations 2010, US Census Bureau 2014). All this leads to more elderly women than men who live alone.

34. Results of the household forecast are based on 1000 stochastic simulations for the household shares, combined with 1000 simulations for the populations.

35. Both the shares and the populations are for men and women separately, and specific for five-year age groups and for household positions. For example, the number of lone mothers (*j*=5) aged x at 1 January 2021 in country c equals $\hat{\alpha}(5, x, 2{,}2021, c). \widehat{W}(x, 2{,}2021, c)$. The assumption here is that the share $\hat{\alpha}$ˆ and the population number $\widehat{W}$ are independent. Reasons why this assumption is justified are discussed by Alho and Keilman (2010). Note that numbers of cohabiting men and women (COH) are computed independently from each other. Hence there is no provision for maintaining the obvious link between these two. The same is true for household position MAR. This issue is discussed further in Section 6.

36. For the Netherlands we used the results (1000 simulations) of the official probabilistic population forecast published by Statistics Netherlands (Van Duin and Stoeldraijer 2014). For Denmark and Finland, 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 countries of the current paper. For more information about the methodology and assumptions see Alho et al. (2006), Alders et al. (2007), Alho et al. (2008) and the website http://www.stat.fi/ tup/euupe/ .UPE were updated by changing the jump-off years to 2011, and by using expected values for age-specific death rates, birth rates, and net migration numbers from recent population forecasts of the two countries. The remaining assumptions, that is, the variances and co-variances for the mortality rates, fertility rates, and net migration, were left unchanged. The assumption here is that the volatility of fertility, mortality, and migration for the period 2011-2041 in the two countries is

the same as that assumed in the UPE-project for corresponding lead times of 10, 20, and 30 years.

37. To compute numbers of *households* based on *persons* in various household positions, we made the following assumptions.

- Each one-person household, lone father household, or lone mother household corresponds with one person with household position SIN0, SIN+ (men) or 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 equals the number of persons with household position OTHR divided by a fixed ratio, which equals the number of households of type other relative to the number of persons with household position OTHR. For all future years, the ratios were assumed to be 2.05, 3.99, and 5.86, for Denmark, Finland, and the Netherlands, respectively. The latter figures are based on information from the Census of 2011 in each country (Eurostat 2014).

b) 2 Selected results

38. Table 2 shows that predicted developments in important household types in the three countries are as one could expect, given our assumptions. Numbers of one-person households and of cohabiting couples will increase to 2041, whereas there will be fewer married couples. These developments reflect our assumptions of a continuation of historical trends in household shares. Except for Finland, numbers of private households grow faster than population numbers. 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 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 slightly from 2.1 in 2011 to 2.2 in 2041.

39. Now we turn to uncertainty in the predicted predictions. Table 2 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 uncertainty increases with increasing forecast lead time, as one could expect. 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 much more certain than predictions of lone-parent households. A different way of expressing uncertainty is by means of the lower and upper bounds of a prediction interval. The prediction intervals in Table 2 reflect prediction uncertainty in the *absolute* sense. According to the model, chances are 80 per cent that the number of private households in Denmark in 2041 will be between 2.97 and 3.42 million, up from 2.54 million in 2011. A number higher than 3.42 million is not impossible, but chances for that to occur are only ten per cent. Similarly, fewer than 2.97 million households cannot be excluded either - the model predicts that this chance, too, is ten per cent. The prediction intervals become wider when we look further into the

future – this is another expression of increasing uncertainty, as is the case with the CV. Note that in all three countries the prediction interval for one-person households in 2041 is *wider* than that for all private households. In other words, numbers of one-person households are more difficult to predict than all households. The explanation is that random variations in the numbers of households of various types cancel out when these are added together. For instance, shares for COH and MAR are negatively correlated: when the simulated share for MAR is relatively high, the share of COH is often lower than expected, and *vice versa*.

40. How do the results for Denmark and Finland in Table 2 compare with the findings of C&K? Their results are very different from ours. First, whereas Table 2 shows a slight decrease in the number of households in Finland towards the end of the period, C&K find uniformly growing numbers, caused in particular by more one-person households and more married couples. For Denmark, they find that the number of households grows slowly, 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 2.) 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 2 is much larger than that found by C&K. The reasons are not entirely clear, but one explanation is that the latter two authors assumed an estimation variance of the drift estimate in the RWD model equal to the innovation variance divided by the number of observations of the time series minus one. In the current paper we used estimation variance for the drift based on robust standard errors from OLS-regression; cf. expression (4) and Table 1. Although many of the estimates are strongly significant, the error term variances are relatively high (R^2 values are between 0 and 8%).

c) 3 The oldest old

41. Figure 4 shows what we expect for the “oldest old” (persons aged 80 and over) in Finland who live alone, if current trends would continue. This age group is of considerable interest for policy makers. Although the health condition of the oldest old may improve in the years to come, many of these will be in need of formal and informal care. Our simulations predict that in 2041in the three countries, there will be between two and four times as many men and women aged 80 and over who live alone, compared with the numbers in the Censuses of 2011. The growth in these numbers is relatively certain, cf. the increasing lower bound of the 80 per cent prediction interval. The predictions indicate that whereas in 2011 there were between 330 and 420 elderly women who lived alone for every 100 men living alone, by 2041 this sex ratio is likely to be somewhat more balanced, but still very skew: between 240 and 310.

VI. Conclusions and discussion

42. We showed how techniques of data dimension reduction can be used to model and predict patterns of household dynamics. The aim was to simplify the method for probabilistic household forecasting used earlier by Christiansen and Keilman (2013). We have computed probabilistic household forecasts for Denmark, Finland, and the Netherlands, spanning the period 2011-2041. Starting

point was 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 were taken from official population forecasts. Time series of shares of the population in six different household positions were modelled as Random Walks with a Drift. Brass' relational model preserved the age patterns of the household shares. Probabilistic forecasts for households were computed by combining predictive distributions for the household shares with predictive distributions of the populations, specific for age and sex.

43. The results show a continuation of current trends towards more and smaller households, often driven by increasing numbers of persons who live alone. Numbers of households increase faster than population size, which leads to falling average household sizes. A very consistent finding is that more numerous households are easier to predict than households that are less numerous, at least when uncertainty is considered in a relative sense. One can expect a strong growth in the numbers of persons aged 80 and over who live alone in the three countries. A doubling of these numbers towards 2041 is extremely likely. The sex ratio will become slightly less skewed: today there are between three and four elderly women in the three countries who live alone for every elderly man who does so; that ratio is likely to decrease somewhat, to between two and three by 2041.

44. The contribution we hope to make to the literature is that we have shown how one may simplify the data hungry approach of Christiansen and Keilman (2013) by avoiding transition probabilities between household positions. At the same time, a few unresolved issues should be mentioned.

45. First, in Section 4 we 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 are 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.

46. A second issue is that of coherence between men and women. In the observed data for Denmark, Finland, and the Netherland there is a strong 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. When using one random walk model for men and one for women, a practical ad-hoc solution to the problem of coherence between men and women is to adjust predicted numbers of men and women in household positions COH and MAR.

References

Alders, M. (1999). Stochastische huishoudensprognose 1998-2050 [Stochastic household forecast 1998-2050], *Maandstatistiek van de Bevolking* 47(11): 25–34.

Alders, M. (2001). Huishoudensprognose 2000-2050: Veronderstellingen over onzekerheidsmarges [Household forecast 2000-2050: Assumptions on uncertainty intervals], *Maandstatistiek van de Bevolking* 49(8): 14–17.

Alders, M., N. Keilman, and H. Cruijsen (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., M. Alders, H. Cruijsen, N. Keilman, T. Nikander, and D. Quang Pham (2006). New forecast: Population decline postponed in Europe, *Statistical Journal of the United Nations ECE* 12(23): 1-10.

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

Alho, J. and N. Keilman (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.

Bijak, J. , I. Alberts, J. Alho, J. Bryant, T. Buettner, J. Falkingham, J. Forster, P. Gerland, T. King, L. Onorante, N. Keilman, A. O'Hagan, D. Owens, A. Raftery, H. Ševcíková, and P. Smith (2015). Uncertain population forecasting: A case for practical uses, *Journal of Official Statistics* 30(4): 537-544.

Christiansen, S.G. and N. Keilman (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.

Daatland, S.O. (ed.) (1997). De siste årene: Eldreomsorgen i Skandinavia 1960-95 («The final years: Elderly care in Scandinavia 1960-95»). NOVA Report 22/1997. Oslo: Norwegian Social Research. http://www.hioa.no/Om-HiOA/Senter-for-velferds-og-arbeidslivsforskning/NOVA/Publikasjonar/Rapporter/1997/De-siste-aarene

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

Eurostat (2014). http://ec.europa.eu/eurostat/web/population-and-housing-census/census-data/2011-census .

Grundy, E. and M. Jital (2007). Socio-demographic variations in moves to institutional care 1991-2001: a record linkage study from England and Wales, *Age and Ageing* 36(4): 424-430.

Jacobsen, R.H. and S.E.H. Jensen (2014). Future changes in age and household patterns: Some implications for public finances, *International Journal of Forecasting* 30(4): 1110-1119.

Jiang, L. and B. O'Neill (2004). Toward a new model for probabilistic household forecasts, *International Statistical Review* 72(1): 51–64.

Keilman, N. (2016). Household forecasting: Preservation of age patterns. *International Journal of Forecasting* (in press).

Keilman, N. and S. Christiansen (2010). Norwegian elderly less likely to live alone in the future, *European Journal of Population* 26(1): 47-72.

Lakdawalla, D. and T. Philipson (1999). Ageing and the growth of long-term care, NBER Working Paper Series, National Bureau of Economic Research, http://www.nber.org/papers/w6980

Lakdawalla, D., D. P. Goldman, J. Bhattacharya, M. Hurd, G. Joyce, and C. Panis (2003). Forecasting the nursing home population, *Medical Care* 41(1): 8-20.

Muller, C., K. Gnanasekaran, and K. Knapp (1999). *Housing and Living Arrangements for the Elderly: an International Comparison Study*. New York: International Longevity Center.

OECD (2011). Doing better for families. Paris: OECD Publishing.

O'Neill, B. and B. Chen (2002). Demographic determinants of household energy use in the United States, *Population and Development Review* 28 (suppl.): 53–88.

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

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

Scherbov, S. and D. Ediev (2007). Probabilistic household projections based on an extension of headship rates method with application to the case of Russia. Joint Economic Commission for Europe–EUROSTAT Work Session on Demographic Projections, Bucharest, 10-12 October (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.

United Nations Economic Commission for Europe (2006). Recommendations for the 2010 Censuses of Population and Housing, jointly prepared by the United Nations Economic Commission for Europe and the Statistical Office of the European Communities. Geneva: 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 L. Stoeldraijer (2011). Huishoudensprognose 2011-2060: Meer en kleinere huishoudens [Household forecast 2011-2060: More and smaller households]. *Bevolkingstrends* 59(2), 59-67.

Van Duin, C., L. Stoeldraijer, and J. Garssen (2013). Huishoudensprognose 2013-2060: Sterke toename oudere alleenstaanden [Household forecast 2013-2060: Strong increase for elderly living alone]. *Bevolkingstrends* September 2013, 2-26.

Van Duin, C. and L. Stoeldraijer (2014). Bevolkingsprognose 2014-2060: Groei door migratie [Population forecast 2014-2060: Growth due to migration]. *Bevolkingstrends* December 2014, 1-22.

Van Vliet, W., E. Huttman and S. Fava (eds.) (1985). *Housing Needs and Policy Approaches: Trends in Thirteen Countries*. Durham: Duke University Press.

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

Wilson, T. (2013b). The sequential propensity household projection model, *Demographic Research* 28(24), 681e712. http://www.demographic-research.org/volumes/vol28/24/28-24.pdf .

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. Private households and population size. Observed 2011 (Census numbers), and predicted 2021-2041 (averages across 1000 stochastic simulations), in millions. Coefficients of variation (CV) in per cent. Lower and upper bounds of 80 per cent prediction intervals, in millions.

		One person households	Cohabiting couples	Married couples	Lone fathers	Lone mothers	All private households (incl. other private households)	Population size
				Denmark				
2011		0.95	0.30	1.03	0.03	0.15	2.54	5.56
2021	Average	1.19	0.40	1.08	0.03	0.12	2.85	5.83
	CV	9.8	17.1	7.4	48.4	29.0	2.5	1.0
	Interval	[1.04,1.34]	[0.31,0.48]	[0.97,1.18]	[0.009,0.04]	[0.07,0.16]	[2.76,2.93]	[5.75,5.90]
2031	Average	1.39	0.479	1.02	0.02	0.11	3.04	6.10
	CV	13.0	21.3	11.8	52.2	35.2	3.9	2.8
	Interval	[1.16,1.62]	[0.35,0.60]	[0.86,1.17]	[0.01,0.04]	[0.06,0.15]	[2.89,3.20]	[5.88,6.31]
2041	Average	1.54	0.53	0.96	0.02	0.10	3.19	6.32
	CV	15.7	24.8	14.7	59.4	39.4	5.5	5.3
	Interval	[1.23,1.87]	[0.36,0.70]	[0.78,1.14]	[0.01,0.04]	[0.05,0.16]	[2.97,3.42]	[5.92,6.76]
				Finland				
2011		1.04	0.30	0.93	0.03	0.14	2.52	5.38
2021	Average	1.18	0.36	0.93	0.04	0.13	2.69	5.64
	CV	9.8	18.2	8.7	68.7	43.3	3.0	1.5
	Interval	[1.04,1.33]	[0.28,0.44]	[0.82,1.03]	[0.01,0.07]	[0.06,0.20]	[2.58,2.80]	[5.53,5.75]
2031	Average	1.27	0.42	0.87	0.03	0.12	2.76	5.82
	CV	13.6	22.7	13.7	69.4	47.8	5.0	4.0
	Interval	[1.05,1.51]	[0.30,0.55]	[0.71,1.03]	[0.01,0.07]	[0.04,0.20]	[2.57,2.94]	[5.52,6.11]
2041	Average	1.25	0.48	0.82	0.03	0.11	2.72	5.92
	CV	18.3	25.5	17.0	76.5	50.6	7.4	6.8
	Interval	[0.97,1.55]	[0.32,0.64]	[0.64,1.00]	[0.00,0.06]	[0.04,0.19]	[2.47,2.97]	[5.44,6.43]
				Netherlands				
2011		2.71	0.92	3.27	0.09	0.41	7.48	16.7
2021	Average	3.11	0.85	3.51	0.09	0.35	7.96	17.3
	CV	10.0	19.0	6.1	54.7	34.1	2.6	1.0
	Interval	[2.72,3.52]	[0.65,1.05]	[3.22,3.78]	[0.03,0.16]	[0.19,0.50]	[7.69,8.23]	[17.1,17.5]
2031	Average	3.52	0.93	3.42	0.10	0.35	8.38	17.8
	CV	13.0	25.1	9.5	58.3	42.0	3.9	2.7
	Interval	[2.93,4.11]	[0.64,1.23]	[3.00,3.83]	[0.03,0.18]	[0.15,0.54]	[7.96,8.81]	[17.2,18.4]
2041	Average	3.97	0.97	3.29	0.11	0.38	8.78	17.8
	CV	14.9	29.9	11.6	64.8	46.4	5.3	4.3
	Interval	[3.24,4.76]	[0.63,1.35]	[2.80,3.76]	[0.03,0.20]	[0.15,0.61]	[8.18,9.41]	[16.8,18.8]

Table 3. Persons who live in an institution. Observed 2011 (Census numbers), and predictions 2021-2041. Average across 1000 stochastic simulations, in thousands. Coefficient of variation (CV), in per cent. Lower and upper bounds of 80% prediction interval in thousands

	2011	2021	2031	2041
		Denmark		
Average	81.5	69.0	64.2	55.2
CV		21.2	33.2	48.4
Interval		[51.5, 87.7]	[40.6, 93.8]	[28.3, 87.3]
		Finland		
Average	110.7	171.5	287.5	436.1
CV		21.2	29.7	32.7
Interval		[128.2, 223.0]	[184.5, 403.1]	[259.6, 629.0]
		Netherlands		
Average	219.3	224.9	260.8	326.8
CV		21.3	32.4	42.0
Interval		[165.8, 287.2]	[165.8, 376.2]	[180.8, 492.8]

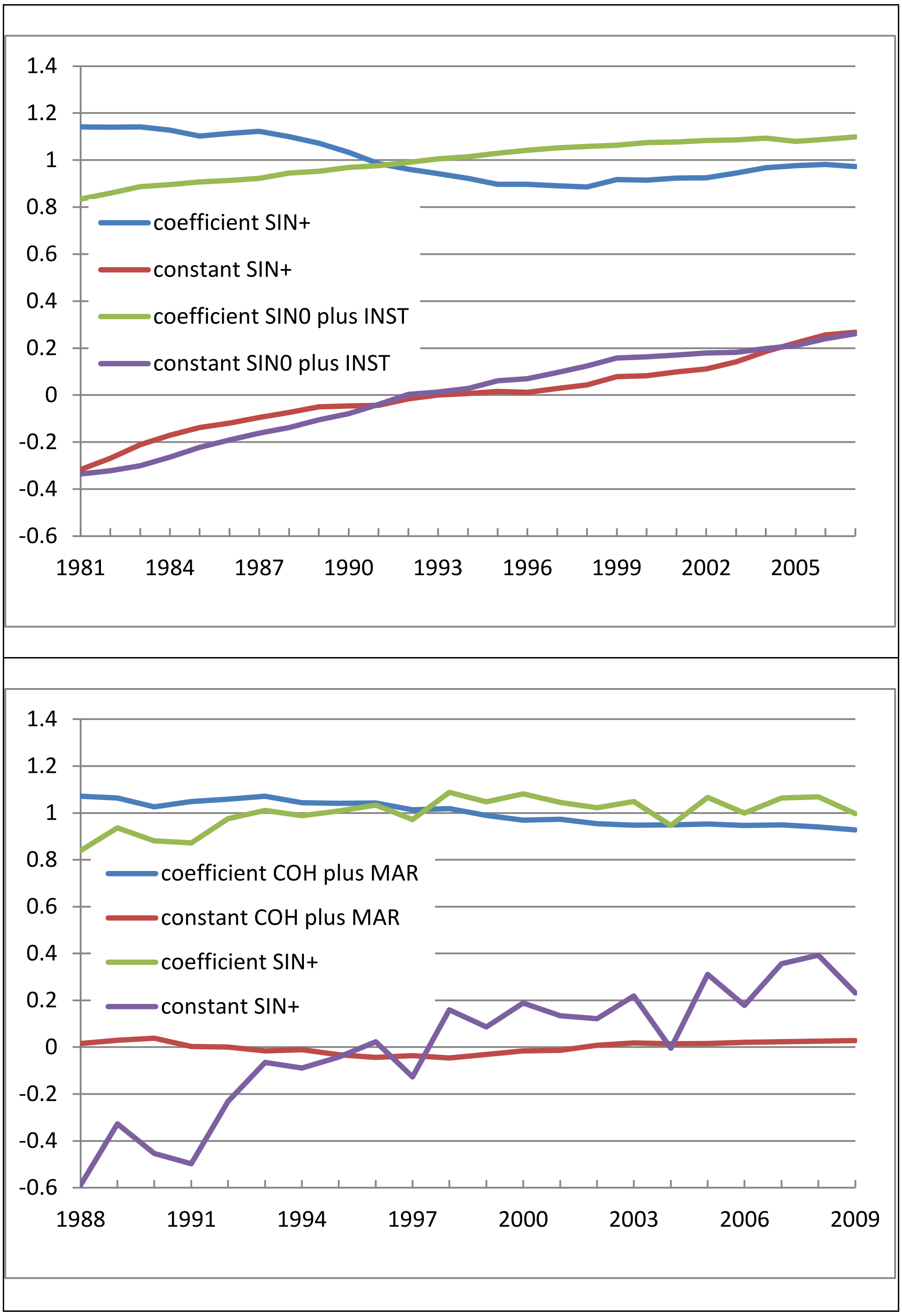

Figure 1. Estimates of Brass model parameters (constant and coefficient) when model is fitted to data from each year separately. Upper panel: data for women in Denmark. Lower panel: data for men in Finland.

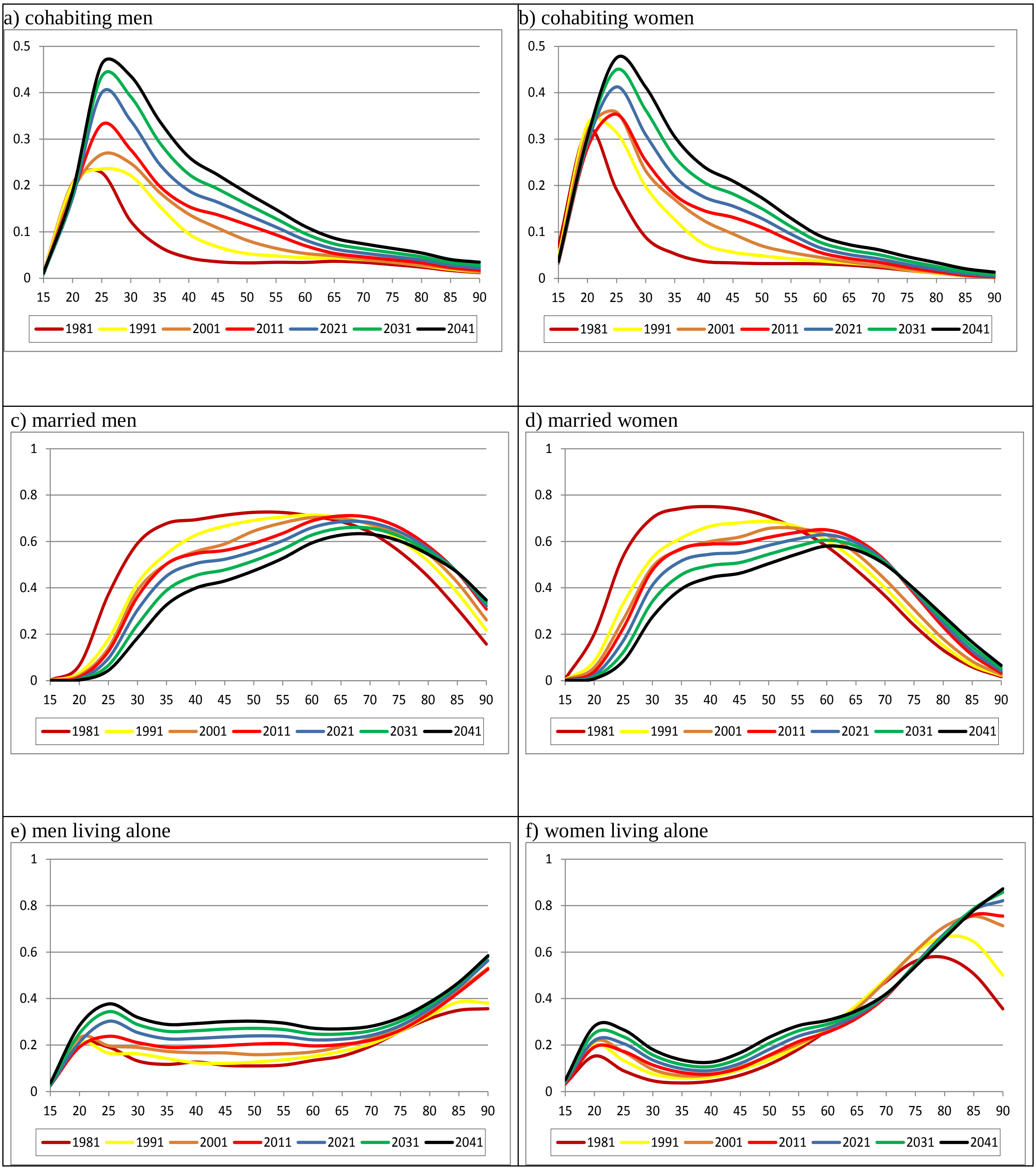

Figure 2. Observed (1981-2011) and predicted (2021-2041) shares of persons in selected household positions, by age, Denmark. Data sources: 1981-2001 register data; 2011 census data, Eurostat 2014; 2021-2041 model extrapolations.

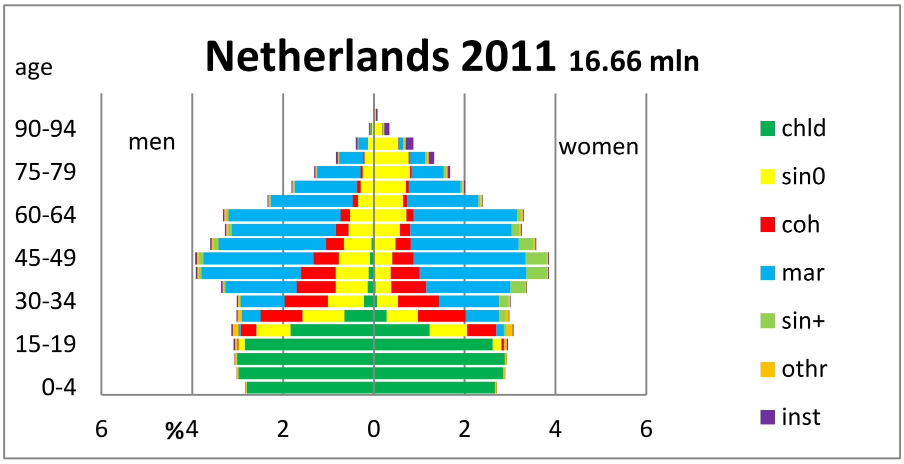

Figure 3. Household structure of the population in three countries. Explanation of legend: “chld”: child living with parent(s); “sin0”: person living alone; “coh”: person living with cohabitee; “mar”: person living with marital spouse; “sin+”: lone parent; “othr”: other private household position; “inst”: person living in institution. Source: Eurostat (2014).

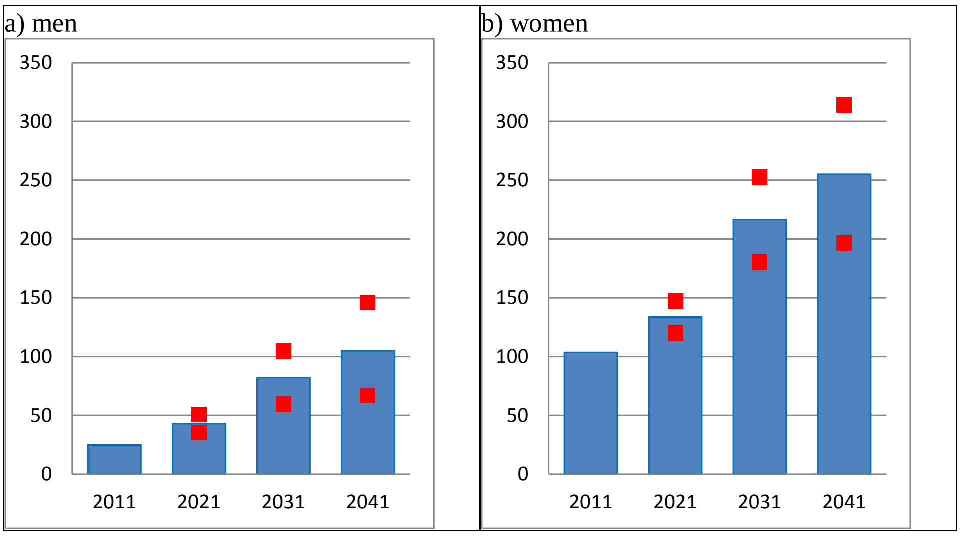

Figure 4. Finland: Men and women aged 80+ who live alone. Observed 2011 (Census numbers), and predicted 2021-2041 (averages across 1000 stochastic simulations), in thousands. Red dots represent upper and lower bounds of 80 per cent prediction intervals, in thousands.