Probabilistic household forecasts based on register data
-the case of Denmark and Finland

Solveig Christiansen
Nico Keilman

1. Introduction

Our aim is to compute probabilistic household forecasts for Denmark and Finland, using register data. Household forecasts are useful for planning housing supply, energy use, and the demand for consumer durables (e.g. King 1999, Muller et al. 1999, O'Neill and Chen 2002). For the elderly, the household position also has an effect on their demand for places in nursing homes (e.g. Lakdawalla and Philipson 1999; Lakdawalla et al. 2003; Grundy and Jital 2007).

Traditionally, household forecasts have been computed by models that, roughly speaking, can be divided in two groups: household headship rate models, and household transition models (Van Imhoff et al., 1995). Compared to headship rate models, that are static in nature, transition models have the advantage that they explicitly describe the dynamics of the household composition of the population.

Both types of models are widely used for computing deterministic forecasts. A projection of the number of households of a certain type in a given year in the future is computed as one number (or just a few numbers; see Section 2). Such a deterministic forecast, however, does not give an accurate view of forecast uncertainty. The future is inherently uncertain, and hence probabilistic methods have to be used. Alho and Keilman (2010) have recently developed a method for computing probabilistic household forecasts. They applied their method to Norwegian data. One important drawback of their application is that the uncertainty assessments were based on limited data, and simplifying assumptions had to be made (see section 2).

The purpose of this paper is to improve on the approach of Alho and Keilman by taking advantage of high quality data from the population registers and housing registers of Denmark and Finland. Both countries have register data covering the whole populations dating back to the 1980s. The registers contain information about persons in every dwelling, including all flats in apartment blocks, each having its own unique address (Lind 2008; Niemi 2011). We have constructed time series for household parameters and analysed the prediction errors in those time series. This allowed us to assess the expected errors in the household forecasts for the two countries.

We forecast, with a 30 year horizon, the number of people occupying the following household positions: dependent child, living with a spouse, living in a consensual union, living alone, lone mother or father, and living in other private household. In addition, the elderly can live in an institutional household. Our household forecasts for Denmark and Finland form part of the so-called AGHON project (Ageing Households and the Nordic welfare model (http://www.etla.fi/eng/tutkimushaku.php?type=hanke&id=106). The aim of this project is to combine statistical analysis of household types with economic analysis of population ageing in Denmark and Finland. Probabilistic household forecasts, which describe the developments of different household types and quantify the uncertainty in these descriptions, are used jointly with computable general equilibrium models and partial models describing household behaviour under uncertainty.

Following this introduction, the paper is divided into five sections. We give a brief overview of earlier work in the field of household forecasting in Section 2. Section 3 describes the methods used to forecast household shares and the population. In Section 4 we present the data employed in this paper. Section 5 gives some selected results from our household forecasts. Finally, section 6 summarises and draws some conclusions.

2. Overview of earlier work

As mentioned in the introduction, our model is similar to that used by Alho and Keilman (2010). This so-called random share model can be characterized as a probabilistic and dynamic macro model that projects households of various types, as well as the population broken down by age, sex, and various household positions. Below we will briefly sketch the most important features of our model, as compared with other approaches of household forecasting. Extensive literature reviews of household projection models have been published by Jiang and O'Neill (2004), Bell et al. (1995), and Arminger and Galler (1991). Another useful reference is Van Imhoff et al. (1995).

Probabilistic household projection models are of fairly recent date, compared to deterministic models. De Beer and Alders were the first to publish a probabilistic household forecast. See Alders (1999, 2001) and De Beer and Alders (1999). They combined a probabilistic population forecast with random shares that distributed 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 variable for the number of lone mothers aged 40 years in 2015 as the product of two other random variables, namely the number of women aged 40 years in 2015 and the share of those women who live as a lone mother. 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 both were based on intuitive reasoning, and correlations across age, sex, and time were disregarded.

Scherbov and Ediev (2007) combined a probabilistic population forecast for the population broken down by age and sex with random headship rates. In demography, a headship rate reflects the proportion of the population that is the head of the household, for a given combination of age and sex (United Nations, 1973; Jiang and O'Neill, 2004); see below. Like De Beer and Alders, Scherbov and Ediev based a large part of their uncertainty distributions on intuition. In contrast, our contribution is to show how uncertainty in the forecast of the shares that distribute the population over several household positions can be modelled as a stochastic process, the parameters of which can be derived from time series models estimated from population register data.

In our view, probabilistic forecast models are more appropriate for computing forecasts than deterministic forecast models. There are many possible future household developments for a given population, but some of these are more likely than others. As opposed to a deterministic forecast, which predicts only one number (or perhaps just a few; see below) for a certain year, a probabilistic forecast tells us how likely it is that future household numbers will be within a certain range. Information of this kind allows policy makers, planners and other forecast users in the fields of housing, energy, social security etc. to take appropriate decisions, because some household variables are more difficult to predict, and hence more uncertain, than others. It also guides them once actual developments start to deviate from the most likely path. New actions or updated plans are unnecessary as long as developments are likely to remain close to the expected future. Deterministic forecasts traditionally deal with forecast uncertainty by formulating alternative scenarios, usually in terms of a high and a low trajectory for some key input parameter, in addition to a most likely trajectory (Jiang and O'Neill, 2006). The drawback is that uncertainty is not quantified, and hence the user does not know how likely it is that the high trajectory will materialize, instead of the most likely trajectory. Moreover, the results are not plausible from a statistical point of view, as they implicitly assume perfect correlation across age, time, and type of household (Lee, 1999; Alho et al., 2008).

Our household model is a dynamic one, as opposed to static household models. Dynamic household models (also labelled as household transition models) deal explicitly with household events. A household event is defined as a change in household position an individual experiences during a brief time interval. For instance, a person who lives as dependent child with his or her parents and starts to live with a cohabitee, experiences the event of home leaving. A lone mother whose last child leaves home becomes a one person household. Dynamic household models were first developed in the 1980s, when existing multistate demographic models were applied to household analysis (Kuijsten and Vossen 1988). A prominent example of the group of dynamic household models is the LIPRO ("LIfe style PROjections") model (Van Imhoff and Keilman 1991) which is based on the methodology of multistate demography but includes several extensions to solve the particular problems of household modelling. At present it is used by Statistics Netherlands for their official household forecasts (Van Duin and Harmsen 2009) and by The Office of National Statistics for their marital status projections for England & Wales (http://www.ons.gov.uk/ons/rel/npp/marital-status-population-projections-for-england---wales/2008-based-marital-status-projections/marital-status-projections---2008-based.pdf) . Other dynamic models, which demand less detailed data, have been employed elsewhere (e.g. ProFamy; see Zeng et al. (2007)). In the current forecasts we have used the computer program developed for LIPRO (version 4.0; see http://www.nidi.knaw.nl/Pages/NID/24/841.bGFuZz1VSw.html) to compute the expected values for our random household shares.

The advantage of dynamic household models, as opposed to static models, is that they explicitly model household events. At the same time their data demands are relatively high. Most of the static models are of the headship rate type. Headship rate models compute future numbers of households by combining an independent forecast of the population (broken down by age, sex, and often also by marital status) with future values for the proportions of household heads in the population (specific of age, sex, etc.). These models have a long tradition in demography (US National Resources Planning Committee, 1938; United Nations, 1973; Keilman et al., 1988). Because of their modest data demands they are more often used than dynamic models (e.g. Jiang and O'Neill 2004), in spite of the fact that processes of household change remain a black box.

A final distinction is that between microsimulation models and macrosimulation models. Microsimulation household models (Wachter 1987; Galler 1988; Fredriksen 1998) take the individual as the unit of analysis, and attach a number of characteristics to each person: age, sex, survival status, number of children, household position etc. Pointers indicate which individuals live together in a given the same household. The model updates these characteristics (except for those that are fixed, such as sex) for each individual by means of random draws from assumed probability distributions for events such as death, the birth of a(n additional) child, change in household position, etc. In this sense the microsimulation model is a probabilistic model, but it only captures Poisson uncertainty. The Poisson rates that determine the distributions (death rates, birth rates, rates for household events) are non-random. For this reason microsimulation models are less-well suited to reflect forecast uncertainty, as in reality the rates tend to change over time in an often unpredictable way. The advantage of microsimulation models is that they are very well suited to map complex household, family, and kin structures (Jiang and O'Neill 2004). But the data requirements are large, because the model is applied to a file with information about individual persons. A recent attempt to combine microsimulation and macrosimulation has resulted in the so-called MicMac model http://www.nidi.nl/smartsite.dws?id=24930&ver=&ch=NID&lang=UK).

The model in this paper extends the work of Alho and Keilman (2010) for Norway, who estimated their household transition rates from panel data from around 5000 households. Mortality rates, however, were estimated based on marital status data from the population register of Norway, together with a number of simplifying assumptions. A few other transition rates had to be borrowed from a deterministic dynamic household forecast for Norway published by Keilman and Brunborg (1995).

A major advantage of having register data is that we do not have to rely on small population samples when calculating household transition rates. Having transition data for the total population and for many years, means that we can get quite reliable estimates also for infrequent transitions.
A further merit of the register data we have in this case is the relatively long time series containing the population in different household positions. These series can be used to estimate the uncertainty in the future distribution of the population across household positions. This is an improvement on the Alho and Keilman (2010) study in which uncertainty parameters were based on the empirical errors in the predicted household shares from an earlier Norwegian household forecast.

Using register data, it is also clear that all the data are compiled using the same definition. When household data are taken from different sources, different definitions may have been used. For instance, one part of the data may have been based on a household-dwelling definition, where all those who live at the same address are member of the same private household. Other data sources may have used the housekeeping definition where only those who take meals together and use common household facilities form a household. The first definition includes lodgers as part of the household of the landlord, whereas the second does not. Thus numbers of one-person households will show substantial differences, whether one takes the household-dwelling definition or the housekeeping definition of a private household. The same is true for numbers of large households.

3. Methods

3.1 Brief overview of our approach

We begin by computing deterministic household forecasts with a 30-year horizon for Finland and Denmark. We have set jump-off years to 2007 for Denmark and 2009 for Finland, which were the latest years for which we have reliable data. In 2008 there was a change in some definitions in Denmark, which makes the data from the years 2008 and later difficult to compare to earlier data. The results of interest of the latter forecasts are the distributions of the population over several household positions. Each household position corresponds with one share. These shares are different for men and women in different age groups. 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. The time series models predict, among others, the likelihood that a certain share will be different from its expected value by a certain amount. Also, the data enable us to estimate the correlations of the shares across ages and between the sexes. Correlations across household positions are dealt with in a specific manner; see Section 3.4. Using the shares computed in the deterministic population forecast and the estimated standard deviations and correlations, we

simulate 3000 sample paths for the household shares for each age and sex. These paths are then combined with 3000 simulations from an earlier computed stochastic population forecast that covers the same period. This gives the predicted number of persons in each household position.
We will now explain in further detail each of the steps outlined above.

3.2 Deterministic household forecast

*T*he population is divided into categories defined by sex, 5-year age groups up to 95+, and seven different household positions. These household positions are:

1. CHLD – dependent child living with one or both parents (up to 25 years of age).
2. SIN0 – person living in a one-person household.
3. SIN+ – single mother or father (aged 15-75).
4. COH – living in a consensual union with or without dependent children.
5. MAR – living with a spouse with or without dependent children.
6. OTHR – living in a private household, but not in any of the positions described above.
7. INST – living in an institution for the elderly (from 70 years of age).

These categories refer to living arrangement and not marital status. For example, the category MAR does not include all those who are married, but only those who are currently living with a spouse. An example of a person belonging to the group OTHR is someone living in a multiple family household. Persons who live in households where they have no parent-child relationship, and are not married or cohabiting with any of the other members of the household also belong to this category. In addition, those who in the data were coded as children although they are 25 and older, coded as lone parent and aged 75 and over, and those aged under 70 who are living in institutions were assigned the household position OTHR.
To compute the deterministic household forecast we use the macro simulation model and corresponding computer program LIPRO. We will here give a rough sketch of the LIPRO model. For a detailed description of the model and the computer programme, see Van Imhoff and Keilman (1991).

We start out with a jump-off population broken down by age, sex and the seven household positions described above. This population is then projected forward five years at a time by exposing it to 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 population at time t+1 can then be calculated using the standard demographic bookkeeping equation.

$$V_{t+1} = P_t V_t + Q_t I_t$$

where

V_t - is a column vector of the population broken down by age, sex and household position at time t.

I_t - is a column vector of immigrants who have arrived between time t and time t+1.

P_t and Q_t are square matrices containing transition probabilities determined by the rate matrix M_t which contains age, sex and household position specific rates.

The period *(t, t+1)* is 5 years.

As discussed above, the LIPRO model is based on the projection 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. For a thorough discussion of this algorithm see Van Imhoff (1992). 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. We here employ the harmonic mean version of the consistency algorithm. This means that when there is a discrepancy between the numbers of men and of women experiencing one of these events, the initial numbers are adjusted to the harmonic mean of the initial numbers of men and the modelled number of women experiencing this event. We have chosen also to require constant capacity of institutions, which means that the number leaving an institution must be 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, it is also possible to set the number of births, deaths, immigrations, and emigrations equal to numbers from an external source. In this case we have chosen to set the total number of these events in each projection interval equal to the numbers from Statistics Denmark's population projection 2010 for Denmark and Statistics Finland's population projection for 2009 for Finland. 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.

3.3 Stochastic population forecast

The 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 Denmark and Finland. 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/ .

We calculate the stochastic population forecast using the Program for Error Propagation (PEP) developed by Juha Alho. This programme takes as its inputs the jump-off population, and predicted mortality rates, 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.[1] The programme then draws sample values from a standard normal distribution, and transforms them to 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 year to 2007 for Denmark and 2009 for Finland, and using age specific death rates, birth rates, and net migration numbers taken from Statistics Denmark's population projection of 2010 for Denmark and that of 2009 for Finland. The remaining assumptions, that is, the variances and co-variances for the mortality rates, fertility rates and net migration, were kept unchanged. We simulated 3000 paths for the future population.

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

3.4 Analysis of time series data

In order to asses the level of uncertainty in the household shares we modelled time series for the period 1988-2009 for Finland and 1982-2007 for Denmark. Following earlier work on Norwegian data, see Alho and Keilman (2010), we have opted for a tree-like structure.

This led us to model six types of fractions (all specific for age, sex and time):

(1) the total share of MAR and SIN0;

(2) the relative share of MAR out of MAR and SIN0;

(3) the relative share of COH out of the total share of COH, CHLD, SIN+, OTHR, and INST;

(4) the relative share of CHLD out of the total share of CHLD, OTHR, SIN+, and INST;

(5) the relative share of SIN+ out of the total share of SIN+, OTHR, and INST;

(6) the total share of INST out of the total share of INST and OTHR.

We number the household positions as CHLD j=1, SIN0 j=2, COH j=3, MAR j=4, SIN+ j=5, OTHR j=6, INST j=7. Write V(j, x, s, t) for the number of people in household position j=1,2, . . . who are in age x=0,1, . . . and sex s, at time t=0,1,2, ... Aggregating over position, we obtain the population W(x, s, t)= Σ_jV(j, x, s, t) of age x and sex s at time t. The share of household position j is α(j, x, s, t)=V(j, x, s, t)/W(x, s, t) = α_j (x, s, t). The six fractions defined above are restricted to the interval [0,1]. Therefore, we applied logit transformations to the above mentioned fractions. Temporarily suppressing indices for age, sex and time, this gives:

ξ_1= logit($\alpha_2 + \alpha_4$) = log(($\alpha_2 + \alpha_4$)/(1- $\alpha_2 - \alpha_4$))

ξ_2 = logit(α_4/(($\alpha_2 + \alpha_4$) = log (α_4/ α_2)

ξ_3 = logit(α_3/(($\alpha_1 + \alpha_3 + \alpha_5 + \alpha_6 + \alpha_7$) = log ($\alpha_3$/ ($\alpha_1 + \alpha_5 + \alpha_6 + \alpha_7$))

ξ_4 = logit(α_1/(($\alpha_1 + \alpha_5 + \alpha_6 + \alpha_7$) = log ($\alpha_1$/ ($\alpha_5 + \alpha_6 + \alpha_7$))

ξ_5 = logit(α_5/(($\alpha_5 + \alpha_6 + \alpha_7$) = log ($\alpha_5$/ ($\alpha_6 + \alpha_7$))

ξ_6 = logit(α_7/(($\alpha_6 + \alpha_7$) = log (α_7/ α_6)

We now have, by construction, six statistically independent time series, all of them specific for age and sex.

We conducted tests to see whether there were signs of autocorrelation in the data. This was indeed the case for quite a few of the time series for the first three fractions in both countries, and also for fraction 5 in Denmark. Therefore we experimented with different versions of ARIMA models. All in all we detected autocorrelation in a little less than half the time series for both Finland and Denmark. In the majority of cases, an ARIMA (1,1,0) model, $\xi_k(t) = c_k + \xi_k(t-1) + \phi(\xi_k(t-1) - \xi_k(t-2)) + e_k(t)$, where c_k is a constant and e_k is an error term, gave a good fit although in a few cases models including a moving average part fitted even better.

For each of the time series, we also estimated a random walk with a drift model (RWD model), $\xi_k(t) = \xi_k(t-1) + D_k + e_k(t)$ where D_k is a deterministic drift and e_k is an error term. In the cases where autocorrelation had been detected, we compared the residual standard deviations estimated from the RWD model and the ARIMA model that gave the best fit. Although the RWD model did overestimate the residual standard deviation compared to the more refined model, the differences between the two were generally small. Striving for parsimony, we therefore decided to employ the RWD model throughout. This means that for a few household positions and age groups, our prediction intervals for the household shares are wider than strictly necessary. In this sense our assessment of uncertainty is a bit conservative.

The resulting standard deviations are generally larger for the youngest and oldest age groups than for the middle aged. They are also generally smallest for the shares for fractions 1 and 2 although this is not always the case for young adults, see Figure 1.

FIGURE 1

We estimated the correlation between the sexes to be 0.46 for Denmark and 0.53 for Finland, assuming independence of age and household position. Based on the work on Norwegian data by Alho and Keilman (2010) , we assumed an AR(1) model, $e_k(x+1,s,t) = \beta e_k(x,s,t) + u_k(x,s,t)$, $|\beta|<1$, for the correlation across age groups, assuming independence of sex and household position. Here e refers to the errors from the random walk with drift models; x=age, s=sex, t=time, whereas=1, …, 6 refers to the six fractions defined above. The first-order autocorrelations β was therefore estimated as the empirical correlation between residuals $\hat{e}_k(x+1,s,t)$ and $\hat{e}_k(x,s,t)$. The estimated median values for the correlations were 0.63 for Denmark and 0.29 for Finland.

3.5 Simulation of household shares

We took 3000 draws, from a t-distribution. We assumed that the errors $e_k(t)$ of RWD model for the fractions $\xi_k(t)$, k = 1,...6 have a normal distribution, with expected value zero, and standard deviation estimated from that model. The 1-α level prediction interval [L(h),U(h)] for $e_k(T+h)$ is of the form

$$e_k(T) + \hat{D}_k.h \pm t_{T-2}(1 - ½\alpha)\hat{\sigma}_k\sqrt{\frac{h^2}{T-1} + h} \ ,$$

where T is the number of observations in the RWD model, $\hat{D}_k$ is the estimated drift, $t_{T-2}(1-½\alpha)$ is the (1-½α) quantile of a t-distribution with T-2 degrees of freedom, and $\hat{\sigma}_k$ is the estimated residual standard deviation of the RWD model.

The terms h and $h^2/(T-1)$ under the square root account for innovation variance and for estimation error in the drift, respectively, while the t-distribution accounts for estimation error in the innovation variance.

Assuming standard deviation and correlation between the sexes and across age groups as estimated from the time series analysis, these are used to create correlated errors, for each sex and age group. These errors are then added to the point predictions from the deterministic household projection, which have been transformed into the same type of logit fractions as described in the previous section.

We then transformed the predicted shares ξ_k in the logit scale back to shares α_k in the original scale, for each time t and both sexes, according to:

$\alpha_2 = \exp(\xi_1)/[(1+\exp(\xi_1))(1+\exp(\xi_2))]$

$\alpha_4 = \alpha_2\exp(\xi_2)$

$\alpha_3 = (1- \alpha_2- \alpha_4)\exp(\xi_3)/(1+\exp(\xi_3))$

$\alpha1 = (1- \alpha_2- \alpha_3\ \alpha_4)\exp(\xi_4)/(1+\exp(\xi_4))$

$\alpha_5 = (1- \alpha1- \alpha_2- \alpha_3 - \alpha_4)\exp(\xi_5)/(1+\exp(\xi_5))$

$\alpha_7 = (1- \alpha1- \alpha_2- \alpha_3 - \alpha_4- \alpha_5)\exp(\xi_6)/(1+\exp(\xi_6))$

$\alpha_6 = 1- \alpha1- \alpha_2- \alpha_3 - \alpha_4- \alpha_5- \alpha_7$

This way we obtained 3000 sample paths for each of these shares, specific for age and sex. Finally, we multiplied each of these sample paths for the household shares with one of the simulations from the stochastic population forecast. This then results in 3000 sample paths for the number of people in each household position.

Implicit in this multiplication is the assumption that the household shares and the population numbers are independent random variables. This assumption is difficult to check empirically, but we have reasons to believe that it is a reasonable one. A possible dependence is that between the number of elderly persons (which is determined by mortality) and the share of one-person households in that age bracket. Often, when one of two partners in an elderly couple dies, the surviving partner becomes a one-person household. The implied correlation is likely small, because it refers to a second-order effect, namely the *difference* between *mortality* of men and women who live in a couple.

4. Data and assumptions

As mentioned above, we have used data from population registers compiled by Statistics Denmark and Statistics Finland for January 1st of the years 1987-2008 and 1982-2007, respectively[2].

We have also used data on transitions between household positions, broken down by sex and 5-year age groups. 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. The household transition data were used to compute one-year transition probabilities. We decided to use averages over the period 2004-2008 for Finland and 2002-2006 for Denmark so as to avoid erratic patterns for infrequent transitions. The probabilities of entry into single fatherhood in Finland seemed too high, and were therefore set to 20% of the corresponding numbers for women (but it was set to zero for men aged 10-14). The Finnish birth rate for single mothers in the age group 15-19 also seemed unrealistically high and was adjusted downwards to the Danish rate. In addition, for both countries, the rates for entry to single parenthood after age 70 were set to zero, and the rates for going from single parent to other private households were set to 1. The same applies to dependent children after the age of 25.

Numbers of deaths, emigrants, and immigrants decomposed by age, sex, and household position, as well as births broken down by age and household position of the mother, were available for the same years as the rest of the transition data in the Danish case, whereas in the Finnish case they were only available for the year 2008. To avoid irregular patterns in Finnish age-specific death probabilities , the married, cohabiting, and single parents were combined into one group, and those living alone and those living in other positions in private households were gathered into another group. Similarly, the married, cohabiting, and single parents were grouped together when computing emigration probabilities.

Many of the age patterns for the transition probabilities are qualitatively the same for men and women and also between the two countries, although the magnitudes vary. As an example of the age

[2] For more information on the Danish data see http://www.dst.dk/HomeUK/Guide/documentation/Varedeklarationer/emnegruppe/emne.aspx?sysrid=00076

For more information on the Finnish data see http://www.stat.fi/meta/til/perh_en.html patterns, Figures 2-5 show some of the one year transition probabilities for Finnish women for the period 2004-2008[3].

Among the general features observed for both sexes and in both countries are:

- young people who on their own are likely to enter into cohabitation (Figure 2);
- those cohabiting in their 20s and 30s have high marriage probabilities (Figure 3);
- living with a spouse is a stable position except at the end of the life course when experiencing the death of the spouse or entry into an institution is common (Figure 4);
- for all age groups the cohabiting experience higher probabilities of switching to single household position than do those living with a spouse (Figures 3 and 4);
- young single parents often start a cohabiting relationship (Figure 5). When they are in their fifties, they have an elevated chance of living alone, because their (last) child leaves the household;

Figures 6 and 7 show the probabilities of entering and exiting an institution for men and women in Finland. We see that the probability of leaving an institution is highest for those living alone and lowest for the cohabiting. This might have to do with the use of short term placements. The probability of entering an institution is highest for the cohabiting and lowest for the cohabiting for men and those living in other private households for women.

As described above, what we have computed from the transition data are transition probabilities. What we need as input to our household projection model are, however, occurrence-exposure rates. Under a constant intensity assumption, the probability matrix P_t is an exponential function of the rates matrix M_t. Thus to find the occurrence-exposure rates in M_t we need to compute the logarithm of P_t, defined as a power series. The power series, however, does not always converge; see Van Imhoff and Keilman (1991, p. 77) for details. Hence we assume that the occurrence-exposure rate for a certain household event is equal to the one-year transition probability for the corresponding change in household position. This introduces a small error in the rates. Under the assumption used, a Taylor series expansion shows that the probability matrix P and the rate matrix M are related as $P = I - M + ½M^2 - 1/6M^3 + \dots$, where I the identity matrix. Most rates are in the order of magnitude of a few percent or less. Mortality at high ages is an exception, where rates up to 30 per cent are found. Thus for mortality, we computed rates from numbers of deaths and exposure times assuming that

[3] The age groups on the y-axis refers to age 1st January 2004.

there are no disturbing events in the particular population group defined by age, sex, and household position.

Getting numbers for the institutional population in Denmark was difficult. A law was passed in 1987 which abolished the building of nursing homes from January 1st 1988. The existing nursing homes were to be phased out gradually[4]. These were then to be replaced by nursing apartments which offer the same level of care, but where the residents all have their own apartment with bathroom and a small kitchen. The nursing apartments are not considered institutions in the legal sense. Although residents of these apartments are needs-tested, they are considered tenants, which involves a different set of rights and responsibilities compared to persons who live in an institution. As the nursing apartments are not considered institutions those living there are not registered as living in an institution in the household register. The way the residents are registered can vary among municipalities.

For those living in nursing homes we have detailed information about numbers and transitions, broken down by age and sex. The data we have about the population living in nursing apartments are the numbers in the age groups 67-74, 75-79, 84-89 and 90+. We assume that the distribution across age and sex is the same in the nursing apartment population as in the nursing home population, which numbered about 10000 and 30000, respectively, in 2007. In order to get an estimate of the number of persons living in institutions in the jump-off population, we therefore, adjusted the distribution of residents in nursing apartments to fit into our age group classification and divided the residential population between the sexes using the age and sex distribution of the nursing home population in 2007. To accommodate the increase in the institutional population, the numbers of elderly living alone were adjusted downwards. Although, as noted above, the registration of the those living in nursing apartments varies among municipalities we have reason to believe that the majority are registered as living alone. In the years when extra funding was given for the conversion and replacement of nursing homes, and we therefore witnessed a steep decrease in the share living in nursing homes this was mirrored by a sharp increase in the proportion living alone. The same is not the case for the share living with a partner.

As the Danish transition rates into institutions only reflected those moving to nursing homes, we decided to use the transition rates into institutions from the Finnish data in the Danish forecast.

---

[4] This was to happen through conversion to or replacement by nursing apartments.

Multistate life tables based on the first projection interval, which is 2009-2013 and 2007-2011 for Finland and Denmark, respectively, give a summary view of the input rates for this period (Tables 1 and 2). Table 1 shows that the Finns spend a little more than a quarter of their lives living as a child, a third living with a spouse, 11% cohabiting and around 20% living alone. The Danes spend a somewhat larger fraction of their lives living as a child and a little less living with a spouse (Table 2). Based on this life table, the average Fin is more likely to be married than the average Dane; cf. below. In both countries the majority of children are born by mothers who live with a spouse, although the difference between births by married and cohabiting women is smaller in Denmark than in Finland.

In principle the rates are held constant throughout the projection period, however there are small changes due to consistency requirements; cf. Section 3.2. In Section 5.2 an alternative to holding the rates constant, based on trend extrapolation of the rates, will be discussed briefly.

5. Results

5.1 Main outcomes

We will now discuss selected results concerning households of different types and have a closer look at the results for a few age groups. Additional results are available from the first author upon request.

Some of the tables and figures in this section refer to the number of persons in each household position which is what we get directly from multiplying the sample paths as described in section 3.5. In addition, we also have computed sample paths for the number of private households of each type, as this is what is important for many of the planning needs. The numbers of married and cohabiting households equal half the numbers of married and cohabiting persons. The number of other private households was estimated by dividing the population living in such households by 4.65 which was the mean size in Finland at the jump-off point. The same number was used for Denmark. Adding on the numbers of people living alone and single parents, this gives 3000 paths for the number of private households. Mean household size is also an interesting measure, and this was computed as the size of the population in private households divided by the number of private households.

We will firstly have a closer look at the expected development in the number private households of each type. Tables 3 and 4 show these expected the numbers, the lower and upper bounds of the 80%

prediction intervals, as well as the coefficients of variation (CV), for Denmark and Finland, respectively.

TABLE 3 and 4.

When we look at the growth in the numbers of private households of various types during the 30 year period, the strongest increase is expected for the number of one-person households: by 31% and 50% in Denmark and Finland, respectively. We also notice quite a large increase in the number of households consisting of a cohabiting couple. On the other hand, the number of “Other private household” in both countries, and the number of married couple households in Denmark will decrease slightly. Overall, we expect an increase in the number of Danish private households by 13%, from 2.5 to 2.8 million. For Finland we expect a growth by 27%, from 2.5 to 3.1 million. Married couple households become less important, numerically speaking: from 40% to 34% of all private housholds in Denmark and from 38% to 33% in Finland. The fraction of single person households, on the other hand, is expected to increase from 38% to 44% in Denmark and from 41% to 49% in Finland. It is virtually impossible that there will be fewer private households by 2037/2039: looking at the 3000 draws only 1% of the Danish and none of the Finnish imply a smaller number of households in the final year than in the initial year. The corresponding number for married couple households is a staggering 67% for Denmark but only 0.7% for Finland. The probability for a decrease in single person households is 2% in Denmark, whereas in Finland none of the draws imply a reduction. All in all we expect a decrease in the average household size from 2.16 to 2.13 (80% prediction interval 2.01-2.28) in Denmark and from 2.14 to 1.89 (80% prediction interval 1.79-1.99) in Finland during the period.

We see that there is largest relative uncertainty, as reflected in the CVs, concerning the household types “Other private household” and “Lone parents”. The number of married couple households is easier to predict, as judged by the CV. The Danish predictions are more uncertain than the Finnish numbers. This is due to two reasons. 1. The Danish RWD models show somewhat larger residual standard deviations than the Finnish models. 2. Danish population numbers are somewhat more uncertain, especially among the elderly. For instance, 30 years ahead the CV for Danish men aged 95-99 is 0.83 compared to 0.62 for Finnish men. Likewise, for women the Danish CV is 0.61 and the Finnish 0.49. Note that forecasts for the total number of private households are more certain (CV-values after 30 years of 4.8% and 4.1% for Denmark and Finland, respectively) than forecasts for each of the specific household types (CV-values ranging from 4.8 to 29.1%). This is due to aggregation: some of the specific household types move in opposite directions. Hence their sum is easier to predict than the elements.

Note also that prediction uncertainty (still judged by the CV) increases more steeply during the first two decades than during the last decade of the forecast period. The reason that uncertainty stabilizes towards the end of the projection period is to be found in the transformation of the shares from the logit scale (with linearly increasing prediction intervals and unbounded predicted values) back to the original scale (with predicted values limited between zero and one).

With a few exceptions[5], the coefficients of variation in Tables 3 and 4 are smaller than corresponding CVs for Norway from the article by Alho and Keilman (2010). Thanks to the high quality register data we were able to fit more realistic times series models (RWD) than Alho and Keilman: due to the paucity of their data they estimated very simple Random Walk models. If the real process is random walk with drift, a random walk model will result in too large estimates of the residual standard deviation.

While CVs reflect relative uncertainty, absolute uncertainty can be analysed by inspecting the width of the prediction intervals. The upper and lower bounds of the 80% prediction intervals in Tables 3 and 4 show that there is largest absolute uncertainty regarding the number of single person households in both Denmark and Finland. This reflects the fact that they are the most numerous household type. On the other hand, because of their small numbers, single parents have some of the smallest absolute uncertainties.

We will now have a closer look at the expected development in the number of people in each household category, focusing on selected age groups. Tables 5 and 6 contain the CVs for the number of people in different household positions for the age groups 20-24, 50-54 and 80-84, separately for each sex, for Denmark and Finland, respectively. The relative uncertainty is generally largest for the youngest age group. A notable exception is the group of young adults who live in consensual union, those in Denmark in particular. Although residual standard deviations for young adults are higher than those for middle-aged adults (cf. Figure 1 for the example of Finnish men), the large numbers of cohabiting young adults reduce their relative uncertainty. For the youngest two age groups (20-24 and 50-54) the greatest relative uncertainty concerns single parents. For the oldest age group there

---

[5] Exceptions are Danish results for lone parents in the first period into the forecast, lone parents, married and cohabiting couples in the second period, and single person households in the final period of the forecast.

is a large amount of uncertainty concerning cohabiting, the number living in nursing homes and the number living in other private households. For the youngest age group there is generally least uncertainty regarding the number of cohabiting and those living alone, whereas for the middle aged and elderly the most certain are the married and those living alone. In general, when there are many persons in a particular household position, this category is easier to predict than a less numerous one.

TABLES 5 and 6

The box-and-whisker plots in Figures 8 and 9 display the shares in the household types married, cohabiting and single person households in the age groups 20-24, 50-54 and 80-84, for Denmark in 2037 and Finland in 2039, respectively. These plots give the usual first and third quantiles as well the median, and outliers among the 3000 sample paths. We see how the absolute uncertainty reflects the size of the sub-populations. Note for example the small absolute uncertainty for young married men in Denmark, a group with a high relative uncertainty (see Table 5) but which comprises only a few thousand persons.

FIGURES 8 and 9

5.2 Changing rates

As mentioned in Section 4, the input rates for the deterministic household forecast are held constant throughout the projection period, except for adjustments to satisfy internal and external consistency requirements. We tried to improve on this approach detecting a possible time trend in the rates. We assumed a linear trend in (the logit of) the rates and extrapolated these rates linearly. This meant that there were varying rates for each 5 year period of the projection. Using these types of rates did however in some cases lead to eruptive results. An example is the share of cohabiting among young (20-35) women in Denmark. Using varying rates led to a sharp increase in the share of these women from 2009 to 2019. The share then fell quite significantly from 2019 to 2029, and thereafter increased to about the same level as in 2019. In our opinion these results were implausible. For the majority of other household positions using varying rates did not have much effect on the results and we therefore decided to stick to constant rates throughout the projection period. Loosely speaking, when rates are constant over time, this corresponds to shares that have constant (upward or downward) slopes.

5.3 RWD extrapolations

We also experimented with expected values for the shares computed from direct extrapolations of the random walk with drift models (transformed back from the logit scale to the original scale). This was done in order to directly take account of the trends in the shares. This approach did, however, in some cases lead to implausible results. For example it gave results for Finland in 2037 where only around 60% in the age group 15-19 lived with their parents, and hardly any in the age group 20-24. In Denmark it all but extinguished the share of elderly living in other private households. Compared to the LIPRO findings, the results from this method suggest a much stronger substitution of marriage for cohabitation for the young and middle aged. An additional methodological drawback of this approach is that we cannot take advantage of the internal and external consistency requirements built into the LIPRO model.

6. Conclusion

Given the need for planning based on household structure, spanning from public income and expenditure to the demand for consumer durables, this article has investigated the future household structure in Denmark and Finland with a 30 year horizon. Predictive distributions have been computed for households of several types and for persons in various household positions, including the institutionalized. We have used the random share approach developed by Alho and Keilman (2010), and tried to improve on their results by taking advantage of high quality data from Danish and Finnish population and housing registers. As was done in their article we combined a probabilistic forecast for the share of people in each household position, broken down by age and sex, with simulations from a stochastic population forecast covering the same period. This then gives a probabilistic household forecast for the number of people in each household position.

Our results show an expected further increase in the number of private households, from 2.5 to 2.8 million (80% prediction interval 2.6-3.0 million) in Denmark and from 2.5 to 3.1 million (80% prediction interval 3.0-3.3 million) in Finland. Taken together with an increase in population size, this means a decrease in the mean household size from 2.16 to 2.07 persons per private household in Denmark and from 2.14 to 1.97 p/ph in Finland. We find a further reduction in the share of married couple households and a growing importance of one person households. The largest coefficients of variation are for lone parent households and “other private household”, and smallest for married couple households. The single person household, on the other hand, displays the largest absolute uncertainty reflecting the fact they are most numerous household type.

The fact that we could use register data had several advantages, compared to the data of Alho and Keilman (2010). First, we could estimate all the transition probabilities without making approximations from data based on marital status and small sample surveys. Hence we obtained reliable rate estimates even for household events that occur quite seldom. Second, using register data implies that the same definitions (of households, families, etc.) ha been used throughout. Third, the data, spanning more than 20 years in both countries, could be used to construct time series models of household shares. We could then analyse the empirical prediction errors in these time series models to derive estimates for the uncertainty in the predicted household shares. This is a clear improvement on the Alho and Keilman (2010) approach where the "uncertainty parameters were estimated from observed errors of an old household forecast against subsequent censuses". The better data is reflected in the fact that, when it comes to household numbers, compared to the Norwegian results, the vast majority of the coefficients of variation are smaller, given household position and number of years into the forecast.

Thus an important new insight based on our analysis is that households become easier to predict when household data from administrative registers are available for at least two decades. There are many reasons why administrative registers should get more emphasis in data collection for statistical purposes. An important one is that a traditional population census, based on questionnaires to be filled out by individuals, has become extremely costly to undertake. As an alternative, many countries consider a change away from a traditional census to a register based census. Countries such as Denmark, Finland, Norway, the Netherlands, and Sweden have shown how those registers can be used. The registers of Finland and the Netherlands have excellent household information, the quality of household data from Danish register is good (information on elderly institutions is not reliable), while Norwegian household data are problematic, due to problems in the dwelling register in that country. Statistical agencies should prioritize improving the quality of existing registers, and developing administrative registers in countries where they do not yet exist.

Acknowledgements
The research reported in this paper was carried out as part of the AGHON-project ("Ageing households and the Nordic welfare model"); see http://www.etla.fi/eng/tutkimushaku.php?type=hanke&id=106 . Our work was supported by grant nr. 2135-08-0109XXX from NORDCORP. We are grateful for many useful suggestions from Juha Alho and other AGHON participants during this project, and for the help of Juha Alho, Sirkku Uljas, Rasmus

Høibjerg Jacobsen, and Stephanie Koefoed Rebbe in obtaining the Finnish and Danish data. We also wish to thank Thomas Michael Nielsen and Steffen Hougaard at Statistics Denmark for answering questions about the Danish data.

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., N. Keilman and H. Cruijsen (2007) Assumptions for long-term stochastic population forecasts in 18 European countries, European Journal of Population, 23:33-69

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 ECE12, 2006, 23:1-10

Alho, J., H.Cruijsen and N. Keilman (2008) Empirically based specification of forecast uncertainty. In Uncertain Demographics and Fiscal Sustainability, J. Alho, S. Hougaard Jensen, and J. Lassila (eds.), pp. 34-54. Cambridge: Cambridge University Press

Alho, J. and N. Keilman (2010) On future household structure. Journal of the Royal Statistical Society Series A 173(1), 117-143

Arminger, G. and H. Galler (1991) Demografische relevante Modellrechnungen: Simulations- und Analyseverfahren auf der Basis empirischer Erhebungen. Materialien zur Bevölkerungswissenschaft Heft 72, Wiesbaden: Bundersinstitut für Bevölkerungsforschung.

Bell, M., J. Cooper, and M. Les (1995) Household and family models: A review. Canberra: Commonwealth Department of Housing and Regional Development.

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, May 3rd–7th, working paper 45. Geneva: Economic Commisson for Europe.

Fredriksen, D. (1998) Projections of Population, Education, Labour Supply and Public Pension Benefits: Analyses with the Dynamic Microsimulation Model MOSART. Social and Economic Studies 101. Oslo: Statistics Norway.

Galler, H. (1988) Microsimulation of household formation and dissolution. In Modelling Household Formation and Dissolution, Eds. N. Keilman, A. Kuijsten and A. Vossen, pp. 139–159. Oxford, UK: Clarendon Press.

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, 424-430.

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

Jiang, L. and O'Neill, B. (2006) Impacts of demographic events on US household change. Interim Report IR-06-030. Laxenburg: International Institute for Applied Systems Analysis.

Keilman, N., A. Kuijsten, and A. Vossen (eds) (1988) Modelling Household Formation and Dissolution. Oxford: Clarendon Press.

Keilman, N. and H. Brunborg (1995) Household projections for Norway 1990-2020. Reports 95/21. Oslo: Statistics Norway.

King, D. (1999) Official household projections in England: methodology, usage and sensitivity tests. Joint Economic Commission for Europe–EUROSTAT Work Session on Demographic Projections, Perugia, May 3rd–7th, working paper 47. Geneva: Economic Commission for Europe.

Kuijsten, A. & A. Vossen (1988). Introduction. In Modelling Household Formation and Dissolution, Eds. N. Keilman, A. Kuijsten and A. Vossen, pp. 3–12. Oxford, UK: Clarendon Press.

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, 8-20.

Lee, R. (1999) Probabilistic approaches to population forecasting. Population and Development Review 24 (supplement), 156–190.

Lind, M. (2008) Draft description of addressing in Denmark- Overview of legislation, authority, address registries and address components, Danish Enterprise and Construction Authority.

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

Niemi, T. (2011) The Finnish address system, Population Register Centre, Finland.

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

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–EUROSTATWork Session on Demographic Projections, Bucharest, Oct. 10th–12th, working paper 16. Geneva: Economic Commission for Europe.

United Nations (1973) Methods of Projecting Households and Families. New York: United Nations.

US National Resources Planning Committee (1938) The Problems of a Changing Population. Washington DC: Government Printing Office.

Van Duin, C. and C. Harmsen (2009) Een nieuw model voor de CBS huishoudensprognose ("A new model for the CBS household forecasts"). Bevolkingstrends, statistisch kwartaalblad over de demografie van Nederland 57 (3): 20-42.

Van Imhoff, E. and N. Keilman (1991) LIPRO 2.0: an Application of a Dynamic Demographic Projection Model to Household Structure in The Netherlands. Amsterdam: Swets and Zeitlinger.

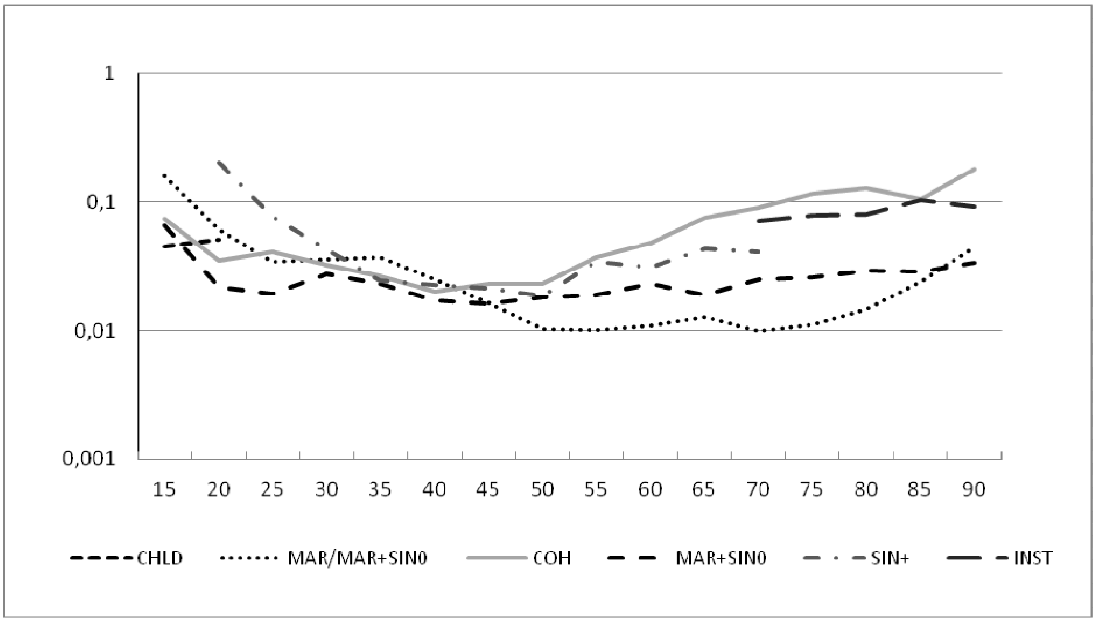
Van Imhoff, E. (1992) A general characterization of consistency algorithms in multidimensional demographic projection models. Population Studies, 46 (1): 159-169.

Van Imhoff, E., A. Kuijsten, P. Hooimeijer, and L. van Wissen (1995) Household Demography and Household Modelling. New York: Plenum Press.

Wachter, K. (1987). Microsimulation of household cycles. In Family Demography: Methods and their Application, Eds. J. Bongaarts, T.K. Burch and K.W.Wachter, pp. 215–227. Oxford, UK: Clarendon Press.

Zeng, Y., K. Land, Z. Wang and D. Gu (2007) U.S. family household momentum and dynamics: an extension and application of the ProFamy method. Population Research and Policy Review, 25, 1–41.

Figure 1 Residual standard deviation of random walk with drift. Finnish men



Note: The categories refer to the six fractions defined in Section 3.4.

Figure 2 One-year transition probabilities. Women who live alone, 2004-2008, Finland

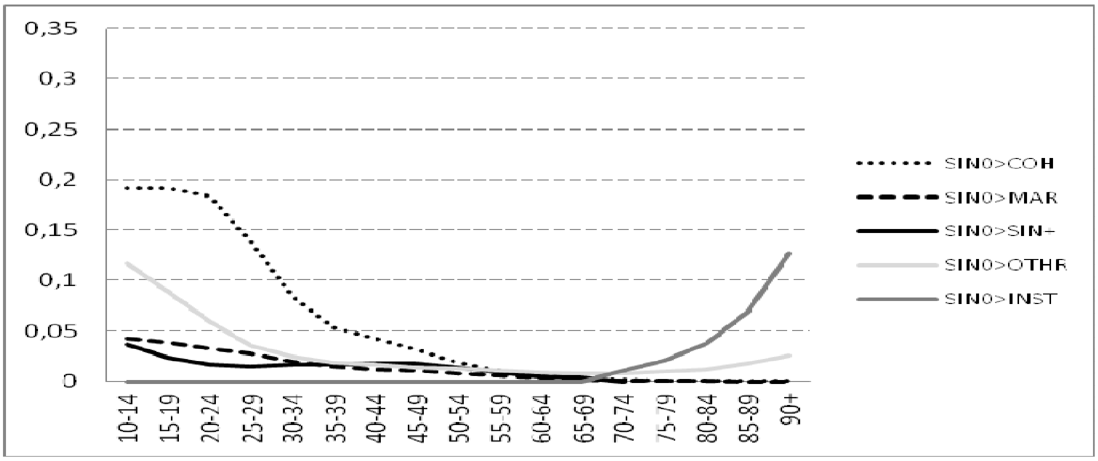


Figure 3 One-year transition probabilities. Cohabiting women, 2004-2008, Finland

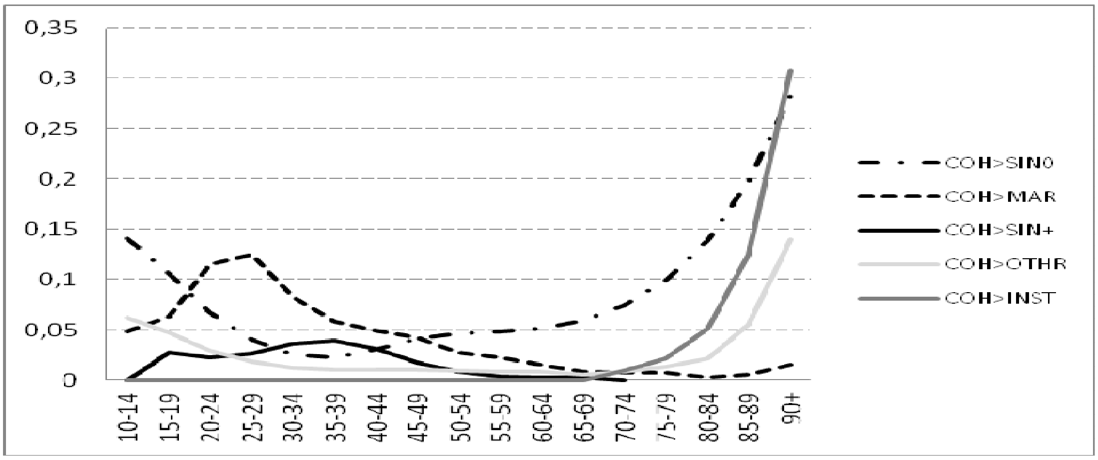


Figure 4 One-year transition probabilities. Married women, 2004-2008, Finland

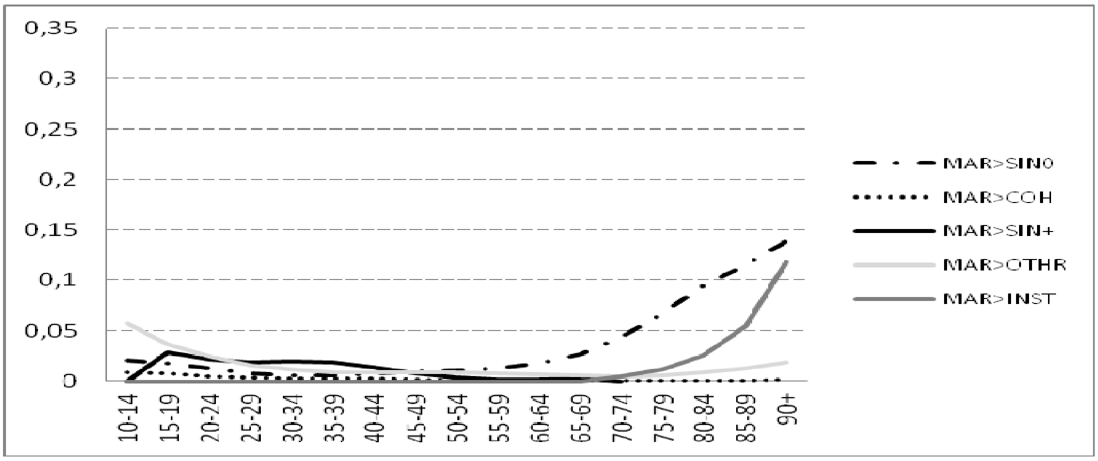


Figure 5 One-year transition probabilities. Lone mothers, 2004-2008, Finland

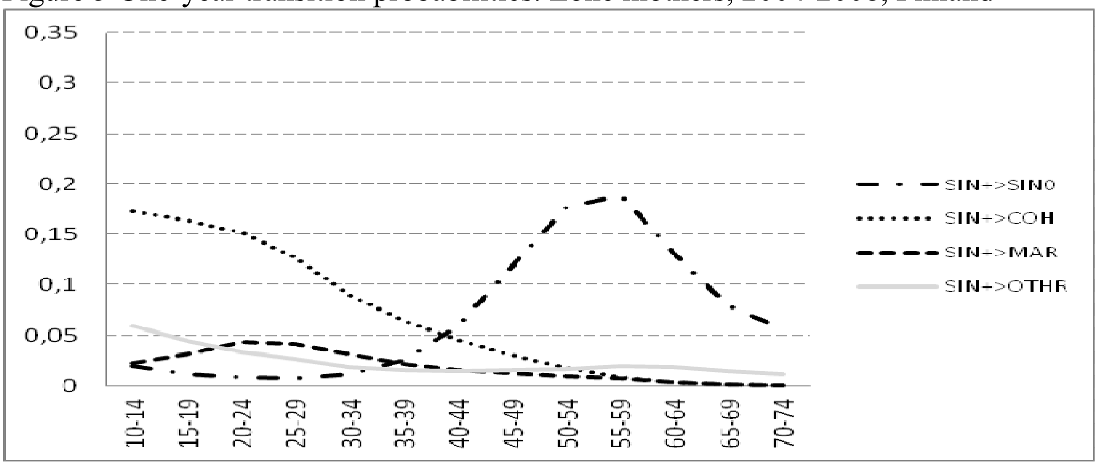


Figure 6 One-year transition probabilities. People entering an institution, 2004-2008, Finland

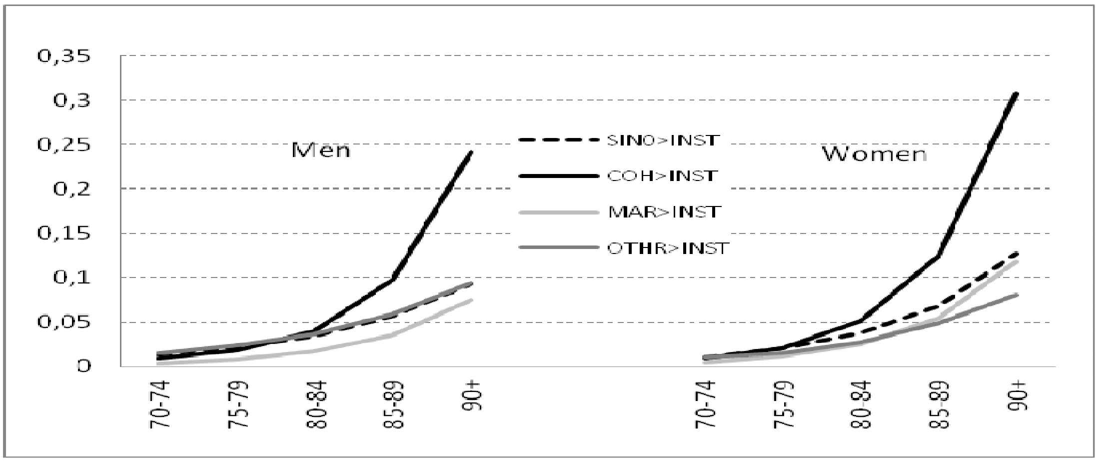
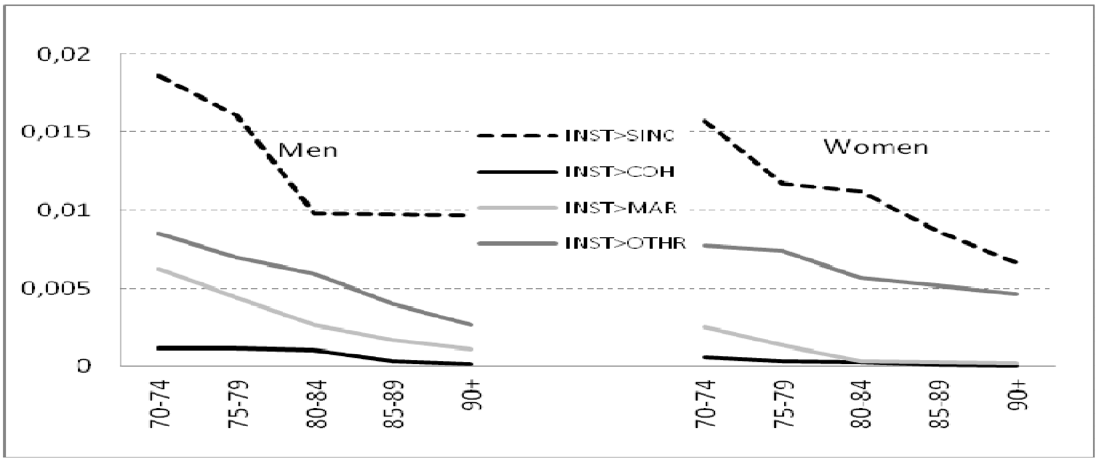


Figure 7 One-year transition probabilities. People leaving an institution, 2004-2008, Finland[6]



[6] Note different scale

Table 1 Percentage of life time spent in various household positions, and number of children by mother's household position, Denmark 2007-2011

	CHLD	SIN0	COH	MAR	SIN+	OTHR	INST	All (=100%)
				%				years
Men	34	17	11	30	1	7	0.1	76.2
Women	31	20	10	29	4	5	0.2	80.8
				children				
	0.01	0.08	0.61	0.79	0.07	0.14	0.00	1.71

Table 2 Percentage of life time spent in various household positions, and number of children by mother's household position, Finland 2009-2013

	CHLD	SIN0	COH	MAR	SIN+	OTHR	INST	All (=100%)
				%				years
Men	26	20	11	35	1	6	1	74.8
Women	22	22	11	34	4	5	1	82.3
				children				
	0.00	0.09	0.42	1.27	0.10	0.05	0.00	1.93

Table 3 Average value, coefficients of variation and lower and upper bounds of 80% prediction intervals, for the number of private household, by household type. Denmark

	Married couple	One-person household	Cohabiting Couple	Lone parent household	Other private household	All private households
2007 Observed	990299	944405	283197	168944	91148	2477992
2017 Average	968171	1036930	302350	181323	86936	2575710
CV(%)	3.3	4.9	8.9	21.3	12.5	1.8
80% low	926441	972475	268641	135234	72368	2517057
80% high	1009953	1103045	336773	234197	100781	2637122
2027 Average	962468	1167539	321254	177936	90419	2719616
CV(%)	7.2	9.7	17.7	29.5	19.9	3.5
80% low	873445	1025302	251565	115541	68362	2602674
80% high	1051393	1314627	397698	249010	114750	2839892
2037 Average	957762	1244238	324567	179700	90555	2796823
CV(%)	7.8	17.8	17.8	29.1	20.2	4.8
80% low	862518	1084466	254229	117327	68352	2626514
80% high	1052869	1413500	402241	250377	114791	2968712

Table 4 Average value, coefficients of variation and lower and upper bounds of 80% prediction intervals, for the number of private household, by household type. Finland

	Married couple	One-person household	Cohabiting Couple	Lone parent household	Other private household	All private households
2009 Observed	924692	1014974	292381	127534	90830	2450410
2019 Average	1012967	1166789	321919	142903	70801	2715379
CV(%)	1.9	3.0	4.9	8.2	6.5	1.1
80% low	988438	1123281	301699	128596	64752	2677248
80% high	1037475	1211384	342188	158311	76866	2753627
2029 Average	1037753	1279715	325467	141195	71540	2855671
CV(%)	3.7	5.0	9.4	13.7	12.2	2.6
80% low	988412	1197593	286651	117146	60886	2762293
80% high	1087005	1359412	364671	166342	82429	2947925
2039 Average	1043100	1530345	330394	139418	74851	3118108
CV(%)	4.8	5.8	10.2	14.3	12.9	4.1
80% low	980039	1415572	288291	115461	63024	2953258
80% high	1108612	1641940	373195	165715	87043	3278831

Table 5 CVs for the number of people in different household positions for selected age groups, by sex. Denmark

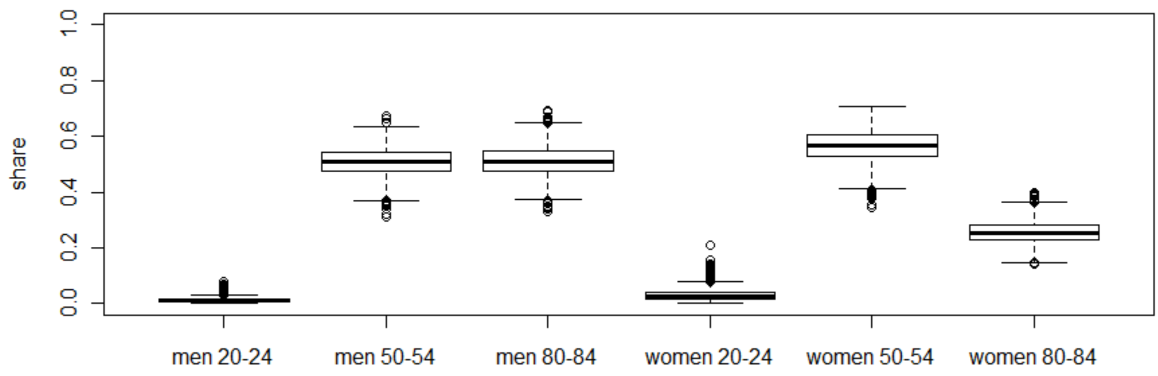
	20-24 years	50-54 years	80-84 years
Men 2017			
MAR	0.390	0.045	0.065
SIN0	0.104	0.084	0.080
COH	0.113	0.179	0.320
SIN+	1.890	0.290	-
OTHR	0.386	0.183	0.530
INST	-	-	0.345
Men 2027			
MAR	0.755	0.098	0.142
SIN0	0.261	0.136	0.171
COH	0.198	0.415	0.618
SIN+	1.963	0.524	-
OTHR	0.490	0.407	0.865
INST	-	-	0.750
Men 2037			
MAR	0.762	0.110	0.177
SIN0	0.273	0.142	0.192
COH	0.212	0.417	0.596
SIN+	1.969	0.525	-
OTHR	0.499	0.405	0.840
INST	-	-	0.822
Women2017			
MAR	0.349	0.042	0.079
SIN0	0.115	0.091	0.050
COH	0.093	0.190	0.346
SIN+	1.170	0.202	-
OTHR	0.626	0.237	0.541
INST	-	-	0.322
Women 2027			
MAR	0.655	0.098	0.175
SIN0	0.275	0.145	0.110
COH	0.166	0.432	0.710
SIN+	1.230	0.414	-
OTHR	0.686	0.447	0.876
INST	-	-	0.680
Women 2037			
MAR	0.660	0.105	0.196
SIN0	0.287	0.150	0.138
COH	0.181	0.431	0.682
SIN+	1.124	0.416	-
OTHR	0.697	0.448	0.839
INST	-		0.744

Table 6 CVs for the number of people in different household positions for selected age groups, by sex. Finland

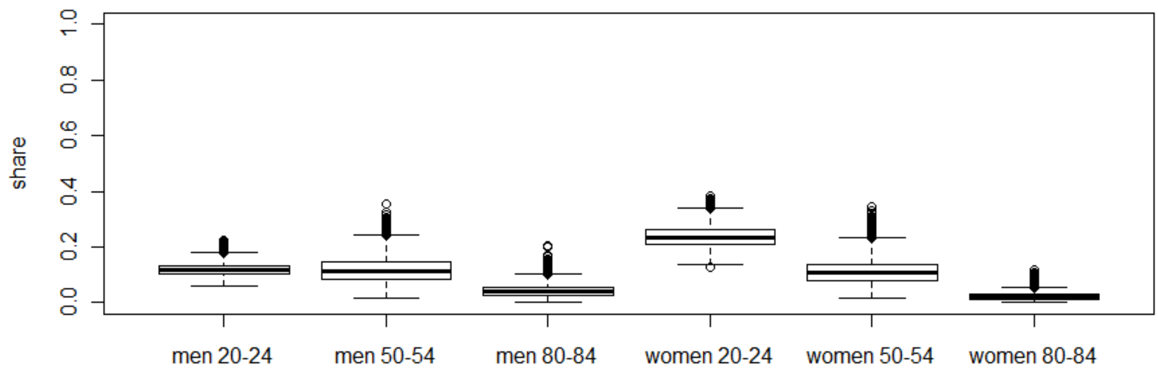
	20-24 years	50-54 years	80-84 years
Men 2019			
MAR	0.262	0.033	0.051
SIN0	0.142	0.050	0.072
COH	0.149	0.104	0.582
SIN+	-	0.160	-
OTHR	0.214	0.123	0.323
INST		-	0.365
Men 2029			
MAR	0.432	0.066	0.123
SIN0	0.246	0.086	0.157
COH	0.281	0.235	0.829
SIN+	-	0.398	-
OTHR	0.364	0.288	0.618
INST		-	0.783
Men 2039			
MAR	0.439	0.075	0.155
SIN0	0.254	0.093	0.169
COH	0.290	0.240	0.780
SIN+	-	0.400	-
OTHR	0.367	0.289	0.634
INST	-	-	0.866
Women 2019			
MAR	0.244	0.034	0.068
SIN0	0.177	0.055	0.044
COH	0.118	0.112	0.715
SIN+	0.710	0.121	-
OTHR	0.377	0.135	0.297
INST	-	-	0.306
Women 2029			
MAR	0.357	0.071	0.158
SIN0	0.306	0.095	0.094
COH	0.219	0.249	1.060
SIN+	0.817	0.285	-
OTHR	0.493	0.332	0.544
INST	-	-	0.664
Women 2039			
MAR	0.360	0.077	0.166
SIN0	0.332	0.080	0.170
COH	0.228	0.253	0.989
SIN+	0.821	0.287	-
OTHR	0.499	0.335	0.550
INST	-	-	0.726

Figure 8 Box- and whisker plots of the shares living in the household types married, cohabiting and single person household, for selected age groups in 2037. Denmark

Married couple households



Cohabiting couple households



One-person households

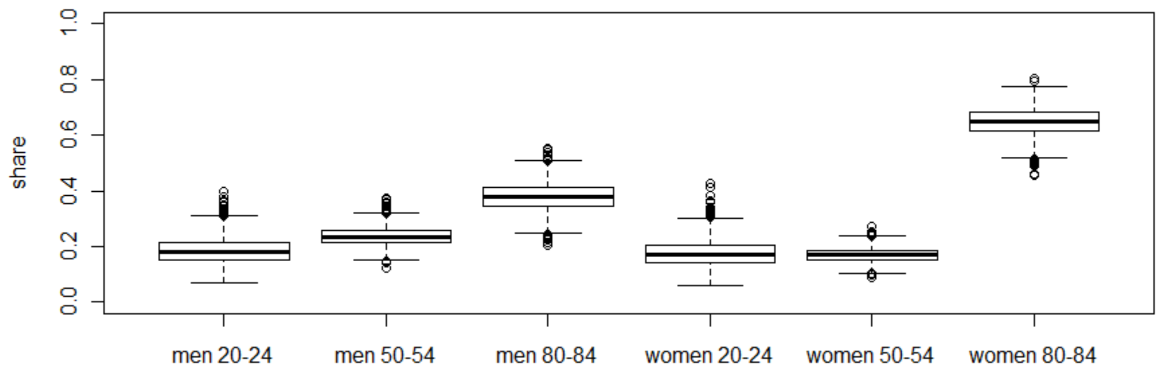
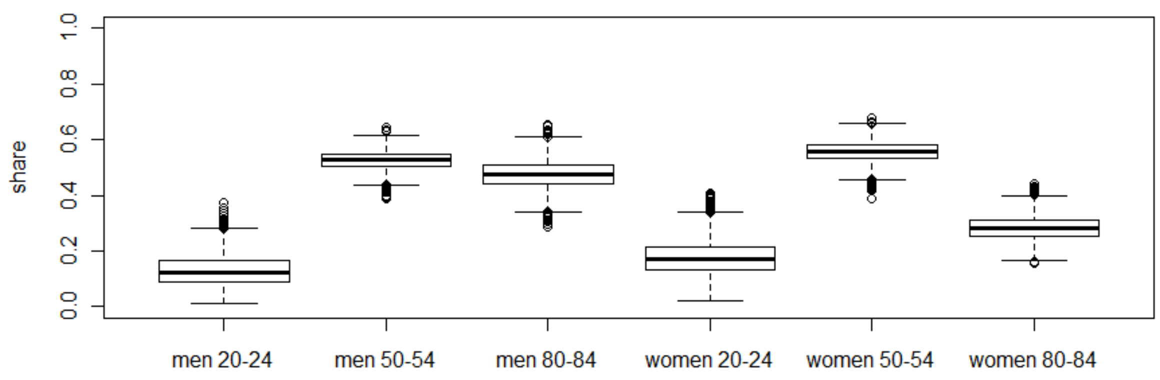
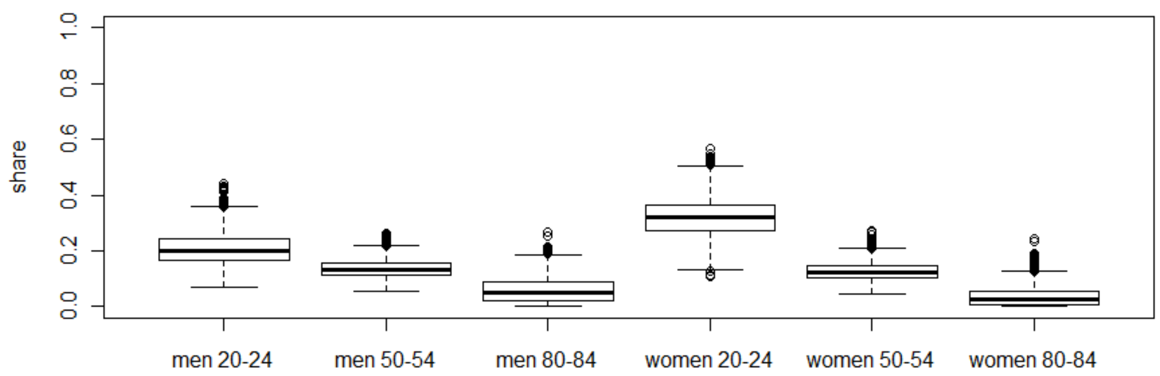


Figure 9 Box- and whisker plots of the shares in the household types married, cohabiting and single person household, for selected age groups in 2039. Finland

Married couple households



Cohabiting couple households



One-person households