WP. 2
25 September 2007

ENGLISH ONLY

UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES (EUROSTAT)

CONFERENCE OF EUROPEAN STATISTICIANS

Joint Eurostat/UNECE Work Session on Demographic Projections
(Bucharest, 10-12 October 2007)

Agenda item 7: Household projections

ON FUTURE HOUSEHOLD STRUCTURE[1]

Invited Paper[*]

(Draft version. Do not cite)

SUMMARY

We have computed a probabilistic household forecast for Norway to 2030. We have combined a probabilistic population forecast, which predicts numbers of men and women by age, with a probabilistic forecast for the shares of persons who have a particular household position, given age and sex. Point predictions for the shares were computed by means of a deterministic multistate household forecast model. Variances and covariances were estimated from observed forecast errors in an old household forecast. We have restricted ourselves to private households.

We find that prediction uncertainty for future numbers of married couples, cohabiting couples, and one-person households, as indicated by the coefficient of variation, is rather low. Lone parents and other private households show much larger prediction uncertainty.

Key words: household forecast, probabilistic forecast, population forecast, random shares, Norway.

---

[1] This paper is based on the joint work of the international research group "Changing family patterns in Norway and other industrialized countries" at the Centre for Advanced study at the Norwegian Academy of Science and Letters in Oslo during the academic year 2006/2007.

[*] Prepared by Nico Keilman, Department of Economics, University of Oslo and Juha Alho, University of Joensuu.

1. INTRODUCTION

Knowledge about the future number of households, their composition, and changes therein is important for many purposes. Support policies for the elderly depend on numbers of elderly persons who live alone or with others (Grundy 2001; Glaser et al. 2003). Housing planners use estimates of household size and household growth in the future (Holmberg 1987, King 1999, Muller et al. 1999). A recent concern from environmental studies relates to the strong growth in the number of households: Even when the size of a population remains constant, more households imply a larger demand for resources. Household members share space, home furnishings, transportation and energy, leading to significant economies of scale. For instance, members of two-person households in the United States in 1993–94 used 17% less energy per person than one-person households did (O'Neill and Chen 2002).

The number and composition of households in the future is uncertain, but some developments are more probable than others. The standard way to deal with uncertainty is to formulate two or more alternative scenarios for household dynamics or other demographic developments in a household forecast, and trace their consequences for household size and structure (e.g. Jiang and O'Neill 2006). The drawback is that uncertainty is not quantified, and that the results are inconsistent from a statistical point of view (Lee 1999). Therefore, there is a need for probabilistic household forecasts, which give us prediction intervals for future numbers of households and for the household statuses of individual persons.

The purpose of this paper is to show how existing statistical methods can be combined for computing probabilistic household forecasts. To the best of our knowledge, there is only one earlier published example of such a forecast. Alders (1999, 2001) computed a probabilistic household forecast for the Netherlands until the year 2050. His deterministic multistate projection combines fertility, mortality, migration, marriage, and marriage dissolution with shares that distribute the population by age, sex, and marital status over six household positions. Parameters for expected values were obtained from observed time series. Next, uncertainty distributions and uncertainty parameters were assumed on intuitive grounds. Some of the parameters were disregarded, for instance correlations between the sexes, across ages, and across time. As opposed to Alders' approach, we estimate the uncertainty parameters from data on observed forecast errors for an old household forecast.

First, we discuss conceptual issues (Section 2). We argue that there are no generally accepted definitions for the concepts of household and family. We give our working definitions and discuss measurement problems. In Section 3, we briefly describe deterministic and probabilistic forecasts of households and population, and sketch our approach based on random shares. Sections 4 and 5 give two empirical applications of these ideas for the case of Norway. In the concluding section we argue that the method of random shares also can be applied to probabilistic forecasts in other fields, for instance regional forecasts and health forecasts.

2. CONCEPTUAL ISSUES

2.1 Definitions

Studies that analyse numbers of households and families are confronted with problems of definition and of measurement. We start with the notion of "family". Different disciplines approach the definition problem differently. Economists usually ignore it (e.g. Rosenzweig and Stark 1997; Ermisch 2003), while demographers take a pragmatic approach based on available data and the definitions therein (e.g. Keilman 2005). Among sociologists, there is considerable controversy over just what constitutes a family. In earlier times, there was consensus on depicting families: sociologists used the notion of "nuclear family", i.e. a married couple and their minor children, all living apart from other kin (Moen and Forest 1999). This went back to Durkheim's definition of 1921 (Hoffmann-Nowotny 1987; see also International Encyclopaedia of the Social Sciences Vol. 5 of 1968, page 303). However, concomitant with life-course changes, definitions of the family have been broadened to recognize a range of forms. However, there is considerable debate about a number of elements in the classical definition. For instance, does a cohabiting couple qualify for being considered as a family? How about same-sex couples? Is the presence of children a necessity? Does the family definition also cover lone parents? See Settles (1999) for a discussion.

Following Moen and Forrest (1999), this paper takes an inclusive view of the diversity of families and households. Our definition allows considering variations in household and kinship arrangements, as men and women move in and out of various living arrangements. We define a family as a group of two or more persons who live in the same dwelling, and who are related by marriage, cohabitation, blood, or adoption. Hence, we include same-sex couples, childless couples, lone parents, three generation families, and reconstituted families in our definition. A household consists of all persons who live in the same dwelling. Thus, a household may consist of just one person, or it may be a multi-person household. In many cases, the multi-person household consists of a family (possibly with other household members). However, we also allow multi-person households that are not families; for instance two unrelated students who share the same dwelling.[2]

The family and household definitions agree with international statistical practice, see the recommendations formulated by the Conference of European Statisticians (2006). We also follow these recommendations when we define the household positions that individuals may occupy at any given point in time: child, cohabiting partner, married partner, lone parent, and other. Child refers to a blood, step- or adopted son or daughter younger than 25 years of age (but regardless of marital status) who lives in the household of at least one of the parents, and who has no partner or own child(ren) in the same household. Young adults aged 25 or over, as well as persons younger than 25 with a partner or an own child, who live with their parent(s) belong to the category other. The latter category also includes persons who live in a multi-person household but who have no relationship (parent-child, or partner in consensual or marital union) to the other household members. A cohabiting couple (i.e. partners living in a consensual union) is understood as a couple that has a marriage-like relationship while not being married to each other, irrespective of the partners' sexes. Cohabiting persons can have any marital status (including married; in that case they are married to different partners). The category of married couples consists of those who are currently married and live together with the spouse. A lone parent lives together with one or more children as defined above, but without a spouse or cohabiting partner.

The household positions defined here characterize individual persons. Knowing the household positions of the household members, the household type follows immediately: a household may be a married couple household, a cohabiting couple household, a lone parent household, a one-person household, or a household of type "other". Due to data limitations

[2] This is the so-called dwelling definition of a household; see Conference of European Statisticians (2006).

(see below), we will only include private households, and thus ignore institutional and collective households.

2.2 Measurement problems

Given a certain set of definitions for households and families of various kinds, it is not always clear how to operationalize these, and next how to measure their numbers. Together, we refer to these problems as “measurement problems”. These measurement problems imply that we are often unable to assess real household or family trends, but that we obtain measurements that reflect those real trends only to a certain extent. Here are a few examples.

1. The notion of “living in the same dwelling” figures prominently in our definitions of household and family, but it is problematic. It is based on an individual’s place of usual residence (PUR). In our empirical application, we will use Norwegian data from three sources: the Census of 2001, the Survey of Living Conditions, and a household forecast that was computed in the 1990s. The Census is based on the *de jure* definition of PUR, whereas the forecast attempted to reflect the actual (*de facto*) PUR. We expect differences in PUR for some categories of the population, in particular for young adults. Largely, the census followed the rules of the Norwegian Population Register. One of these rules implies that a person who leaves the parental household will not be registered at the new address in Norway, unless he or she marries, gets a child, or receives the major part of his or her income from labour. Moreover, some elderly persons regularly move from a private to an institution and back, for instance because of health problems. It is not always clear whether their PUR is the private or the institutional household.
2. Although most households can clearly be classified as either private or as institutional, the distinction between these two types is not always sharp, because of intermediate forms such as assisted living etc.
3. It is not always clear whether two persons are a cohabiting couple, in particular (but not only) when they are of the same sex. The notion of “marriage-like relationship” leaves room for many subjective interpretations, which may differ from one situation to another, and even between the partners.
4. A lone parent who starts cohabiting with a new partner may still be classified as a lone parent, for instance because the new partner does not take parental responsibility for the children.

3. FORECASTING TOOLS

3.1 The method of random shares

We have computed an illustrative probabilistic household forecast for Norway. The forecast applies to the period 2004-2030. It gives predictive distributions for

- the population broken down by five-year age group (0-4, 5-9, …, 85-89, 90+), two sexes, and the six household positions for individuals in private households defined in Section 2.1: child, living in consensual union, living with spouse, living alone, living as lone parent, and other;
- numbers of private households of five types, defined in Section 2.1: cohabiting couple household, married couple household, one-person household, lone parent household, and other private household.

We have combined two statistical methods:
1. a probabilistic *population forecast*, which gives the predictive distribution of the population broken down by sex and age, irrespective of household status;
2. predicted random shares, which distribute future population numbers by age and sex randomly over *household positions*.

We have used an existing probabilistic population forecast for Norway. This forecast was obtained earlier by means of the scaled model of error (Alho and Spencer 2005), in the framework of the UPE-project (Alho et al. 2006). Methodological details of the UPE forecast can be found at the web site http://www.stat.fi/tup/euupe/, which also contains also forecast results for Norway and 17 other European countries, including details of age and sex for ten-year intervals to 2050. In the empirical part of this paper, we focus on the predictions of the random shares.

Outline of the method

Notation

x: age ($0,1,\ldots\omega$)
s: sex (1=men, 2=women)
t: time
j: household position of an individual person ($1,2,\ldots J$).
$V(j,x,s,t)$: a random variable which gives the population in household position j, age x, sex s at time t.
$W(x,s,t)= \Sigma_j V(j,x,s,t)$: a random variable which gives the population age x, sex s, at time t, irrespective of household position.
$\rho(j,x,s,t) = V(j,x,s,t)/W(x,s,t)$: a random variable which reflects the share of persons age x sex s, who are in household position j at time t. $\rho(j,x,s,t) > 0$, $\Sigma_j \rho(j,x,s,t) = 1$.

Problem

We have a probabilistic forecast for the population $W(x,s,t)$ for future years t. We want to specify a probability distribution for the shares $\rho(j,x,s,t)$ for future years. Given a probability distribution for $\rho(j,x,s,t)$, we compute a probabilistic forecast for $V(j,x,s,t)$ as $V(j,x,s,t)= \rho(j,x,s,t).W(x,s,t)$

Approach

We suppress x and s for notational convenience. In Section 4 we present two empirical specifications for the stochastic processes that predict the shares $\rho(j,t)$: a multivariate random walk, and random walks based on continuing fractions. The main characteristics of the two approaches are as follows.

Multivariate random walk. By definition, the random shares $\rho(j,t)$ are restricted to the interval [0,1]. Their sum over j is 1. A multinomial logit model transforms the ρ-variables into ξ-variables, i.e. $\xi(j,t) = \ln\{\rho(j,t)/\rho(1,t)\}$, with $\xi(1,t) = 0$. Thus, household position j=1 is selected as the reference category. The back transformation is $\rho(j,t) = \exp(\xi(j,t))/[\Sigma_j \exp\{\xi(j,t)\}]$. We specify a multivariate random walk for the $\xi(j,t)$, with expectations (point predictions) $\hat{\xi}(j,t)$. The point predictions are computed by means of a deterministic multistate household forecast; see Section 3.2. The variances for the household positions and the covariances across household positions are estimated from observed errors for the shares $\rho(j,t)$ in an old household forecast (Section 4.2).

Random walks based on continuing fractions. The main features of this method are the following. We order the household positions j according to some criterion that reflects their importance in a way to be discussed later. For the most important household position (j=1) we define the simple logit transform of $\rho(1,t)$ as $\xi(1,t) = \ln\{\rho(1,t)/[1- \rho(1,t)]\}$ and construct a univariate random walk for $\xi(1,t)$. Next, we compute the fraction of the second share relative to all shares except for the first one, i.e. $\lambda(2,t)= \rho(2,t)/\{ 1 - \rho(1,t)\}$, compute the logit transformed value of $\lambda(2,t)$, and construct a univariate random walk in the logit scale. We continue this way with the remaining shares. By construction, the fraction $\lambda(j,t)$ expresses the probability that a person's household position at time t is j, given that it is not 1,2, …, j-1. Thus the univariate random walks in the logit scale are independent across household positions.

For both specifications we also include estimates of correlations across ages and across the sexes. Given the specification of the stochastic process for the random shares, we use simulation to find the multivariate distribution of the shares $\rho(j,t)$ in the future (Section 5).

3.2 Household forecasts: point predictions for Norway

This section presents the point predictions $\hat{\rho}(j,t)$, obtained from a deterministic multistate household forecast. We have updated a household forecast for Norway, which was computed in the 1990s (Keilman and Brunborg 1995). The new forecast applies to the period 2002-2031; we used Van Imhoff's program LIPRO version 4.0 (http://www.nidi.knaw.nl/en/projects/section2/270101/) for multistate household models to compute it (Van Imhoff and Keilman 1991).

Household projection models developed in demography over the past few decades are primarily of the headship rate type, in which the dynamic processes of household formation and dissolution, which underlie changes in household structure essentially, are treated as a black box. LIPRO ("Lifestyle projections") is a dynamic household projection model, which explicitly focuses on the flows underlying household changes. It is based on the methodology of multistate demography, but includes several extensions to solve the particular problems of household modelling. The LIPRO model is of the recursive type $V_{t+1} = P_t.V_t + Q_t.I_t$. Here V_t is a vector of the population at time t, broken down by household position, age and sex, I_t is a vector of immigrants during the time period (t, $t+1$) with the same format as V_t, while P_t and Q_t are square matrices that contain time dependent transition probabilities for the population V_t and for immigrants I_t, respectively. These transition matrices are functions of the forecast parameters, i.e. birth rates, death rates, emigration rates, and rates for household events. All rates are specific for age, sex and household position (in addition to time). The jump-off population for the LIPRO forecast is V_0. Below we report briefly how we arrived at estimates for the jump-off population on 1 January 2002, and for the forecast parameters for the first forecast period, i.e. 2002-2006.

Jump-off population. The jump-off population of the new household forecast is based on observed data taken from the Population and Housing Census of Norway of 3 November 2001. The data consist of a three-way table of the population in private households broken down by sex, five-year age group, and the following six household positions: "child", "cohabiting", "married", "living alone", "single parent", and "other". The Census uses the household definition given in Section 2. Since childless couples (either married or cohabiting) have much higher risks of dissolving their relationship than couples with children, we have additionally distinguished between cohabiting persons with and without children, and likewise for married persons – this way we defined eight household positions. For each combination of sex and five-year age group, we used data from Statistics Norway's Survey of Living Conditions (panel waves of 2000, 2001, and 2002; see below) to compute the shares necessary for splitting up cohabiting persons and married persons into the two groups. We have assumed that the numbers thus obtained apply to the household structure of the population of Norway as of 1 January 2002.

Household events. A household event is defined as an immediate change from one household position to another one. Given the eight household positions, there are 56 theoretically possible events. Many of these imply events that are impossible in practice. For instance, a person cannot change immediately from "married, no children" to "cohabiting, with children". One or more intermediate positions are necessary, for instance "married, with children" and "lone parent". Other events are so rare that we have omitted them for practical reasons.

We have estimated occurrence-exposure rates for household events from the Survey of Living Conditions (SLC). The SLC is a panel survey with annual waves in the spring of each year since 1997; see Normann (2004). We have used data from the six waves conducted between 1997 and 2002.[3] The target population each year was the population aged 16-79 living in private households. The sample size in 1997 was 5000. In later waves, additional persons were drawn from new members of the target population (either immigrants or new persons aged 16). Between 1997 and 2002, 5525 persons were drawn. Among these, 2562 persons took part in all six waves. Remaining persons either refused to respond in one or more rounds, or they left the target population (due to death or emigration; individuals aged 80 years and over were retained in later waves). Response rates were relatively low among the over 80. Otherwise, response was of good quality.

Because the SLC only covers private households, institutional households are not included in this paper.

Households in the SLC are defined slightly differently compared to the Census: a household consists of persons who live in the same dwelling and who have common housekeeping. We assume that the household dynamics in those households is the same as that among households defined according to the household dwelling definition that is used for the Census.

Each respondent reported about the relationship with the other members in his or her household at each interview. We coded the respondent's household position according to the eight categories mentioned above and identified household events by comparing the household positions at subsequent interviews.[4] Due to both left censoring and right censoring, a number of events imply a change from, or to, an unknown household position. Added over all five calendar years, we obtained 3645 events of 27 different types, and 22462 years of exposure. We computed occurrence exposure rates for each of the 27 events, specific for sex and five-year age group. For a number of household events, the age patterns looked irregular, or even unreasonable. In those cases we had to take executive decisions to obtain

3 Starting in 2003, the SLC has a very different format compared to earlier years, and data from before 2003 are not directly comparable to the later data.

4 Individuals may, in principle, experience two or more household events during one calendar year. But this is rare, because a one-year time interval is rather short. We have disregarded these multiple events.

reasonable patterns. We used occurrence exposure rates from the previous household forecast (see below) in some of the adjustments.

External events. In addition to household events, we distinguish the following external events: birth, death, immigration, emigration, entrance into an institutional household, and exit from an institutional household.

Birth rates broken down by mother's age (five-year age groups) and household position were taken from the previous household forecast and adjusted proportionally in such a way that the predicted number of live born children during the years 2002-2006 would correspond to the number that Statistics Norway has registered for that period.

Death rates by five-year age group and household position were estimated based on data from the Norwegian population registers. Øystein Kravdal kindly supplied us with deaths and exposure times broken down by marital status, age and sex for the years 1995-99.[5] We have applied

- death rates for persons with marital status "currently married" to persons with household position "cohabiting" or "married", irrespective of the number of children in the household;
- death rates for never married to persons with household position "child", "living alone", or "other";
- death rates for the divorced to persons with household position "lone parent".

We call these death rates "initial death rates", to be used later.

The target population for this household forecast is the population in private households. Thus, mortality in institutions is disregarded. There are reasons to believe individuals who live in an institution have higher mortality than those who live in private households. Therefore, the (initial) death rates for the whole population cannot be used for persons in private households. The previous household forecast applied to *both* private households and institutions. Its results for the period 1990-1994 indicated that 28.7 per cent of all deaths take place in institutions. We have assumed that this is also the case in 2002-2006. Next, we adjusted the initial death rates such that the number of deaths that the model predicted for the period 2002-2006 was equal to 71.3 per cent of all registered deaths for that period.

Entrance to and exit from institutions was combined with *international migration* into one forecast parameter (for each combination of age, sex, and household position) of *net entries*. The net entries were expressed as absolute numbers (contained in the vector I_t above). We obtained values for the net entries for the period 2002-2004 in several steps.

1. Numbers of net migrants for the period 1990-1994 broken down by age, sex, and household position taken from the previous forecast were proportionally adjusted such that they summed to 83,832, which is the total net immigration registered by Statistics Norway for the years 2002-2006.
2. Initial death rates applied to the jump-off population resulted in an initial number of deaths for the period 2002-2006, specific for age, sex, and household position.
3. The set of death rates obtained after reducing the number of deaths by 28.7 per cent (see above) resulted in a second set of deaths.

5 Mortality data by marital status for Norway are not available for more recent years.

4. The difference between the initial number of deaths (point 2) and the reduced number of deaths (point 3) can be interpreted as the number of exits from the population in private households to the institutionalized population. The implicit assumption in this case is that a person who leaves a private household occupies a place in an institution immediately after the death of an institutionalized person. The total number of such exits from private households in the period 2002-2006 amounted to 60,425 persons, or 12,000 persons per year. The census indicated that 35,000 persons lived in an institution in 2001. Thus, the annual turnover is about one-third.
5. For each combination of age, sex, and household position we subtracted the number of exits from the population in private households from the net immigration numbers (point 1). The result was a set of net entries that showed positive numbers for the young and the middle-aged segments of the population, and negative numbers for the elderly.

Consistency. Parameters for household events and for external events, when applied to the jump-off population, result in certain numbers of events of each type. For some of these events one wants to require that their numbers are constrained to certain values. This is what is generally known as the consistency problem in multistate household forecasts; see Van Imhoff and Keilman (1991) and Van Imhoff (1992). We speak of internal consistency when the numbers of two or more household events are related. An example is that the numbers of new marriages of men and women during a given period should be equal[6]. External consistency applies to the case where the sum of a set of external events is constrained to be a fixed constant. For instance, the total number of births, irrespective of mother's age or household position, may be required to be a certain number.

LIPRO includes a feature that takes account of internal and external consistency relations. LIPRO's consistency algorithm assumes that the relations are linear. We defined 18 consistency relations: 15 for household events related to couple dissolution and to events experienced by children and by parents, and one each for the total number of births, deaths, and net entries. LIPRO adjusted the parameter values for household events and external events in such a way that all relationships were fulfilled.

Entry into a consensual union and marriage are not among the 15 events for which we formulated consistency requirements. There are two reasons.

1. Numbers in the jump-off population are not consistent. The census shows slightly more married men than women who live with their spouse: 834,583 men and 834,225 women. For cohabiting couples the numbers are 204,109 (men) and 204,155 (women).
2. Observed data for new marriages are not consistent. Marriage data from Statistics Norway for the year 2005 show that among the registered population, there were 22,932 men who married, against 20,474 women. When we restrict ourselves to marriages where both partners are registered in the country, there were only 18,976 couples who married in 2005. The surplus of 1918 men implies 10.1 per cent of 18,976 couples. In 1990, the surplus was only 1.1 per cent – the share has increased regularly from that level to the current 10.1 per cent. The regular increase is largely due to increasing numbers of men of Norwegian nationality who marry a woman from Thailand, the Philippines or Russia (Daugstad 2006). At the same time, net immigration numbers are not consistent. During the years 1985-2006, the surplus of married men (compared to married women) fluctuated strongly between a minimum of

[6] Assuming that the net effect of marriages across country boundaries is zero.

-2,600 in 2002 and a maximum of 2,887 in 1987. Business cycles in the Norwegian economy and associated fluctuations in labour migration from EU-countries may explain the variations in the surplus of married men in net migration. Comparable data for cohabiting persons are lacking, but there is no reason to believe that the inconsistencies are less than those for new marriages and net migration for married persons.

Point predictions for household position shares. We have used the exponential version of LIPRO and harmonic mean consistency type to compute a household forecast for the years 2002-2032. We constrained total numbers of births, deaths (net of deaths in institutions) and net entries to corresponding numbers taken from the Medium Variant of Statistics Norway's 2005-based population forecast; see http://www.ssb.no/english/subjects/02/03/folkfram_en/. Parameters for household events were assumed constant over the forecast period (except for possible adjustments due to consistency). This is not to imply that we are convinced that there will be no change in rates for household events in the future. We simply do not have good enough data to extrapolate a possible trend in these events.

Figure 1 summarizes the main results. It gives point predictions of shares for eight household positions for the forecast period, irrespective of age and sex. The year 1990 (Census information) has been added for reference. The figure shows a continuation of changes in household and family structures that have gone on for several decades now: it becomes increasingly more likely for an individual to live alone or in consensual union, and much less likely to live with a spouse and children (Keilman 2005). A new trend is that elderly women are less likely to live alone in the future than they are nowadays, and more likely to live with their husband; compare Figures 2 and 3. The same trend applies to shares of elderly men who live with their spouse (but at a higher level) and elderly men living alone (at a lower level). This new trend is due to improved longevity of both sexes, but with steeper increases in life expectancy among men than among women (Meslé 2004).

3.3 Stochastic forecasts of population: selected UPE-results for Norway

Stochastic population forecast results for Norway were taken from the project "Uncertain Population of Europe" (UPE), which gives predictive distributions of the population by sex and age in 18 European countries (the 15 members of the European Union pre-2004, plus Norway, Iceland, and Switzerland) for the period 2004-2050; see Alho et al. (2006) and Alders et al. (2007). The UPE-project started from analyses of the historical developments in fertility, mortality, and migration in the 18 countries, in which the time trend in these variables were separated from random deviations from the trend. Estimates were given for both the time trends and the random deviations in the future, which made it possible to calculate a cohort component forecast in terms of probability distributions. Below we give a few illustrative results for Norway (see http://www.stat.fi/tup/euupe/no11_results_nor.html).

For Norway, a population of 5.26 million is expected for 2030, very close to Statistics Norway's figure in its most recent official population forecast made in 2005. The latter forecast amounts to 5.37 million (forecast variant with medium population growth; see http://www.ssb.no/emner/02/03/folkfram/tab-2005-12-15-01.html).The 80 per cent forecast interval in 2030 stretches from 4.90 to 5.65 million. There is an 80 per cent probability that the share of the population aged 65 years and over in 2030 will be between 19.5 and 23.3 per cent – at present it is 14.7 per cent.

4. Analysing observed errors in the shares

4.1 Data

We have computed empirical errors in the shares $\rho(j,t)$ emerging from the household forecast that was published in the early 1990s (Keilman and Brunborg 1995). It has jump-off year 1990 and results for 1995, 2000, 2005 and later. All dates are pr. 31 December. As before, we have restricted ourselves to private households. We have analysed predicted shares for six household positions: “child”, “cohabiting”, “married”, “living alone”, “single parent, one or more children”, and “other”.[7] Henceforth, we label these household positions as CHLD, COH, MAR, SIN0, SIN+, and OTHR, respectively. The predicted shares were evaluated against empirical shares observed for the years 1997, 2001, and 2002. This gave us empirical forecast errors, defined as observed shares relative to predicted shares. These forecast errors are approximate, because the empirical data contain sample errors and measurement errors, as discussed in Section 2. We have the following data:

- the 1997-wave and the 2002-wave of the Survey of Living Conditions. The data have sample errors. We interpolated forecasted shares between 1995 and 2000, and between 2000 and 2005.
- data from the Census taken on 3 November 2001 are compared with corresponding forecast results. There are two problems. First, household definitions differ between the forecast (the housekeeping definition of a household, similar to the SLC; see Section 3.2) and the census (dwelling definition). Second, the Census uses the *de jure* definition of household position, whereas the forecast attempted to reflect the *de facto* situation. This creates problems for some categories, in particular for young adults who *de facto* have left the parental household, but who are still registered at their parents’ address (see Section 2.2). We interpolated forecasted shares between 2000 and 2005.

4.2 Empirical errors

The previous household forecast gives us empirical errors for lead times of 6¼ (Spring 1997), 10¾ (3 November 2001), and 11¼ (Spring 2002) years ahead. Figures 4-6 compare, for a selection of the six household positions, predicted shares from the household forecast with observed shares from the SLC (Spring 1997 and Spring 2002) and the Census (November 2001). Two general observations emerge, both of them in accordance with what one could expect. First, forecast errors increase over time. Second, the errors observed from the Census are much smaller than those from the SLC, since the latter include sample and other survey errors. The forecast predicted too few young cohabitants (Figure 4) and young adults living alone (Figure 6), as opposed to too many children living with their parents, and far too many young adults in household position “Other” (not shown).
To a large extent these errors can be explained by the favourable economic development around the turn of the century, which led to many young adults living independently. These errors are less visible in 2001 (Census data) than in 2002 (survey data), because the census is based on the *de jure* household position. Many young adults are registered as living in the parental household, while *de facto* they have left home. The forecast predicted too high mortality and thus too many widows and widowers. This explains the underprediction of the

---

[7] As opposed to the deterministic household forecast in Section 3, we have not distinguished childless married or cohabiting couples from those who live with children, because this would create additional errors in the data for 2001.

number of elderly married men and women (Figure 5), and the overprediction of the number of elderly men and women who live alone (Figure 6).
The group "Other" consists for a large part of adults aged between 25 and 30 who live with their parents (unrelated persons who share a dwelling belong to this group, too). Thus there is a large degree of complementarity between the groups "Child" and "Other". Indeed, when we combined children and persons with household position other into one category, the age patterns for this combined category became were more regular than those for the separate groups.

Observed shares of lone parents aged 60 and older are generally higher than predicted shares for this group (not shown). This is probably due to an artefact in the empirical data, in which elderly persons who live together with an adult child are still categorized as lone parents, even when their youngest child is over 25 years of age. For this reason, we have ignored forecast errors for single parents beyond age 60.

In Section 3.1 we have described the two types of stochastic processes for the errors, namely a multivariate random walk (based on a multinomial logit transformation for the errors) and an approach using continued fractions (based on a simple logit transformation). These two specifications by no means exhaust all possible specifications, but they serve to illustrate that the probabilistic household forecast depends strongly on the stochastic model that one selects.

We have applied the multivariate random walk model to error data from the two SLC rounds and the Census. The continued fractions approach is limited to Census data. Hence, in the latter approach we avoid the idiosyncrasies of the error data from the sample surveys. In particular, the error patterns based on the SLC for the groups "Child" and "Other" are difficult to explain. For the remaining household positions, the errors are generally larger in the SLC than in the Census, as noted above. We believe that the continuing fractions approach gives the best possible representation of the underlying data of the two approaches, given the constraints with which we are working. The multivariate random walk application should be considered as a sensitivity analysis.

4.3 Multivariate random walk

Variances and covariances. At any point in time, we write the share $\rho(j,t)$ as the product of the predicted share and a multiplicative error term: $\rho(j,x,s,,t)$ = $\hat{\rho}(j,x,s,t).e_\rho(j,x,s,t)$, with $e_\rho(jx,s,,t) > 0$. The multiplicative error $e_\rho(j,x,s,t)$ in the share $\rho(jx,s,,t)$ translates into an additive error in the $\xi(jx,s,,t)$-variable, i.e. $\xi(j,x,s,t) = \hat{\xi}(j,x,s,t) + e_\xi(j,x,s,t)$, where $\hat{\xi}(j,x,s,t)$ is the expected value (point prediction) of $\xi(j,x,s,t)$, and $e_\xi(j,x,s,t)$ is the error in ξ. Since household position j=1 is the reference category, we see that $e_\xi(j,x,s,t) = \ln(e_\rho(j,x,s,t)/e_\rho(1,x,s,t))$: the error for share j in the ξ-scale equals the log of the error for that share in the original ρ-scale, relative to the ρ-error for share 1. Similarly, the point prediction $\hat{\xi}(j,x,s,t)$ equals $\ln(\hat{\rho}(j,x,s,t)/\hat{\rho}(1,x,s,t))$. We model the error variances and covariances as a function of household position j and time t, and predict them for lead times other than 6¼, 11¾, and 12¼ years ahead.
Household position "living alone" is the reference category. There are 297 observations: 12 for "child", 81 both for "cohabiting" and for "married", 49 for "lone parent", and 74 for

"other". With only 297 observations we can not entertain a very rich model. We have assumed the following.

We assume that the errors are normally distributed with zero expectation and covariance matrix $\Sigma^2_{\xi}(t)$, the elements of which only depend of household position j and of time t, independent of sex and independent of age. For t = 6¼ and t = 12¼, each observed error can be written as the sum of a sample error and a prediction error. For t = 11¾, there is no sample error. We assume that the two types of error are independent. Prediction errors are assumed to behave like a random walk, the innovation variance of which is written as π_j^2. The corresponding covariance matrix for all five household positions together is written as Π^2. Thus, the prediction error covariance matrix at time t is $t.\ \Pi^2$. The sample error has a constant variance s_j^2, with covariance matrix S^2. Appendix 1 gives details of our estimations of the variances and covariances. We found covariance estimates that were small enough to assume independence in the random walks across household positions. Innovation variance estimates turned out to be equal to

CHLD	COH	MAR	OTHR	SIN+
0.1293	0.0404	0.0330	0.0916	0.1894

Correlations across ages and across the two sexes. For each household position, we estimated AR(1) the autocorrelation coefficient across ages (i.e. for neighbouring five-year age groups) and the correlation across sexes. Median values across household positions were 0.5224 and 0.5463, respectively.

4.4 Random walks based on continued fraction*s*

For this particular application, we adopted the following sequence of positions:
(a) the share of MAR and SIN0 combined
(b) the share of MAR relative to those of MAR and SIN0 combined
(c) the share of COH relative to the sum of shares for COH, SIN+, CHLD, and OTHR
(d) the share of CHLD and OTHR combined relative to the sum of shares for CHLD, OTHR, and SIN+
(e) the share of SIN+.

MAR and SIN0 are the most important groups given their large shares (see Figures 5 and 6); by combining these two we stabilize the error patterns to a great extent. CHLD and OTHR are problematic, given the measurement problems for young adults noted earlier. We did not separate the groups CHLD and OTHR, but in our simulations, we have assumed that predicted shares for the combined group apply to the household position "Child" below age 25, and to the household position "Other" for ages 25 and over. SIN+ displays the most erratic patterns – it is left as a residual. Errors in both the youngest and the oldest age groups are difficult to explain, and hence we have limited the analysis to the age groups 25-59. We have assumed that simulated shares in the logit scale for the youngest and the oldest age groups have probability distributions that are the same as those for the age group 25-59 (except for age-specific locations that are determined by the point predictions).

Based on the sequence of positions listed under (a) – (e) above for the unconditional shares $\rho(j,t)$, we defined the conditional shares $\lambda(i,t)$ $i = 1,\ldots,4$ as follows (suppressing arguments for age x and sex s):

(a) $\lambda(1,t) = \rho(\text{MAR},t) + \rho(\text{SIN0},t)$
(b) $\lambda(2,t) = \rho(\text{MAR},t)/ \lambda(1,t)$
(c) $\lambda(3,t) = \rho(\text{COH},t)/(1 - \lambda(1,t))$
(d) $\lambda(4,t) = (\rho(\text{CHLD},t) + \rho(\text{OTHR},t))/(1- \lambda(1,t) - \lambda(3,t))$

We used observed and predicted values of the unconditional shares ρ based on the Census to compute multiplicative errors in the conditional shares λ for $i = 1, \ldots, 4$, transformed these to the logit scale, and assumed a simple random walk for each of them. Estimates for the innovation variances computed by expression (A2) turned out to be 0.090, 0.098, 0.129, and 0.321. Moreover, we estimated the correlation between the sexes as 0.68, and that between neighbouring five-year age groups (assuming an AR(1) model) as 0.65.

5. Simulation strategy and illustrative results

For both approaches, we assumed normally distributed errors in the ξ-scale with zero mean and variances and covariances as estimated in Section 4. We drew 3000 error values specific for lead time (accounting for correlations across the sexes and across ages) from the assumed normal distributions, added point predictions, and transformed back to the ρ-scale. Since we applied the simulated shares to UPE results, lead time zero corresponds with 1 January 2003. This resulted in a set of 3000 random shares ρ, specific for sex, age, time, and household position. We applied the shares to corresponding UPE results, which we obtained from the UPE website.[8] We focus on the results from the method of continuing fractions, and we will compare a few of these with results from the multivariate random walk method. We inspected the following issues:
- Which household type k has largest prediction uncertainty? How fast does it increase with lead time?
- Which household position j has largest uncertainty? How does this depend of age, of sex, of time?
- How much does the prediction variance in population ($W(x,s,t)$) contribute to the variance in $V(j,x,s,t)$? How large part is due to the variance in the shares $\rho(j,x,s,t)$?
- How fast does the uncertainty in mean household size increase with lead time?

For the continuing fractions method, we computed empirical predictive distributions for three types of variables:

1. The population $V(j,x,s,t)$ in household position j, age x, sex s at time t. We selected men and women in three age groups (20-24, 50-54, and 80-84) and three future years (2010, 2020, and 2030), and computed the results for the following five household positions: "married", "living alone", "cohabiting", "child/other", and "lone parent".
2. The number of households $H(k,t)$ of type k at time t. These were computed for three future years (2010, 2020, and 2030) and five household types. The number of households consisting of a cohabiting couple equals half the total number of persons with that household position, and likewise for married couples. The number of one-person households equals the number of persons with that household position, and

[8] The shares apply to private households only. In contrast, UPE results apply to all persons, both those in private and institutional households. The total number of persons in institutions in Norway has been fairly constant since the 1990s, about 35,000. The Census of 2001 shows that the majority (ca. 23,000) are aged 80 or over. We decided to ignore persons living in institutions, because we do not have good enough data to compute point predictions for their numbers. Thus, our forecast slightly overestimates the number of elderly in various households and the total number of households.

likewise for the number of lone-parent households. The number of other households was estimated as the number of persons with that position[9] divided by 2.5, which is the mean size of households of type other derived from the 2002 wave of the SLC.
3. Mean household size computed as $\sum_{j,x,s} V(j,x,s,t) / \sum_{k} H(k,t)$ for the years 2010, 2020, and 2030.

Households by type. Table 2 summarizes our results for households of various types. The forecast predicts a 28 per cent growth in the number of private households between 2002 and 2030, from 2.026 million to an expected 2.587 million. The UPE forecast expects the population to increase by only 16 per cent, and thus mean household size drops from 2.21 in 2002 to 2.04 in 2030. Forty per cent of the households in 2030 can be expected to consist of only one person, up from 36 per cent in 2002.
Household types that have the smallest innovation variances (i.e. married couples and one-person households; see Section 4.4) have the lowest relative forecast uncertainty as measured by the CV. At the same type, these household types are the most numerous. Relative uncertainty increases regularly between 2010 and 2030.

Table 2. Average value, coefficient of variation (CV), and lower and upper bounds of 80% prediction interval of numbers of households and mean household size in 3000 simulations

	Married couple	One-person household	Cohabiting couple	Lone-parent household	Other private household	All private households	Mean household size
2010							
Average	876,790	828,010	267,348	183,530	42,142	2,197,820	2.15
CV	0.042	0.078	0.108	0.172	0.217	0.019	0.018
80% low	830,542	744,952	231,010	143,924	30,972	2,145,600	2.10
80% high	923,086	911,772	304,906	225,313	54,406	2,251,148	2.20
2020							
Average	918,529	926,450	317,853	199,615	37,138	2,399,585	2.07
CV	0.068	0.117	0.150	0.246	0.348	0.035	0.030
80% low	839,540	791,250	257,524	141,334	21,874	2,292,957	1.99
80% high	999,560	1,068,251	379,653	264,160	54,068	2,509,124	2.15
2030							
Average	955,644	1,043,608	344,575	207,263	35,892	2,586,981	2.04
CV	0.093	0.142	0.181	0.297	0.451	0.052	0.046
80% low	841,856	863,320	266,457	134,089	17,619	2,418,441	1.92
80% high	1,069,151	1,242,075	427,838	288,355	58,376	2,758,973	2.16

Household positions of individuals. We have constructed Table 3 to inspect prediction uncertainty for individuals in different household positions. The table gives the coefficient of variation (CV), which reflects relative uncertainty. Three factors have an impact on the CV. First, there is a strong effect of lead time, as CV's almost without exception increase regularly between 2010 and 2030. Second, the effect of innovation variances of the various household positions is rather strong. However, and this is the third factor, the latter effect is sometimes

[9] That is, the number of persons with household position child or other who are 25 years or older.

mediated by the effect of the level (point prediction) of the particular share. For instance, married men and women were given a low innovation variance, but at ages 20-24 where their shares are low, the CV for married men and women is rather high. In this age group, household position CHLD/OTHR (in practice "child") has the lowest uncertainty among the five household positions. Note also age 80-84. Men are likely to live with their spouse (household position MAR), which gives a low CV, while women frequently live alone (SIN0), which results in a low value for their CV.

Table 3. Coefficient of variation for predictive distributions, men and women in three age groups and five household positions

	20-24	50-54	80-84	20-24	50-54	80-84
		Men			Women	
2010						
MAR	0.312	0.072	0.069	0.300	0.071	0.166
SIN0	0.167	0.217	0.223	0.178	0.230	0.110
COH	0.321	0.288	0.292	0.256	0.280	0.341
CHLD/OTHR	0.106	0.711	0.323	0.150	1.051	0.283
SIN+	1.139	0.331	1.252	0.969	0.258	1.213
2020						
MAR	0.489	0.113	0.128	0.466	0.113	0.224
SIN0	0.246	0.312	0.337	0.261	0.347	0.194
COH	0.453	0.415	0.389	0.358	0.409	0.440
CHLD/OTHR	0.163	1.585	0.626	0.223	1.886	0.565
SIN+	2.093	0.429	2.176	2.000	0.374	1.857
2030						
MAR	0.611	0.160	0.205	0.602	0.152	0.280
SIN0	0.311	0.380	0.395	0.329	0.417	0.245
COH	0.559	0.494	0.489	0.444	0.480	0.507
CHLD/OTHR	0.218	1.969	0.878	0.290	2.465	0.962
SIN+	2.822	0.540	2.368	2.252	0.479	1.748

The results in Table 3 are based on the predictive distributions for *numbers of persons* in various household positions. A similar table can be computed based on *shares* for these household positions. The striking result is that the coefficient of variation for a certain share, almost without exception, is only slightly lower than that for the corresponding number of persons. For instance, Table 3 gives a CV equal to 0.152 for the predictive distribution of married women aged 50-54 in 2030. The corresponding CV based on shares equals 0.146. Hence only 0.006 points of the CV of 0.152 are due to uncertainty in the number of women aged 50-54 in 2030, and the remaining 0.146 are due to uncertainty in the share for household position MAR. How can this be explained?

The number of married women aged 50-54 in 2030 is found as the product of the corresponding share and the total number of women of that age. Write $V = \rho.W$ for short. Since the share ρ is independent of the population W, an approximate value of the variance in V (by the Delta theorem) is $\mathrm{Var}(V) \approx \mathrm{E}^2(W)\mathrm{Var}(\rho) + \mathrm{E}^2(\rho)\mathrm{Var}(W)$, which leads to $\mathrm{CV}^2(V) \approx \mathrm{CV}^2(\rho) + \mathrm{CV}^2(W)$. The table below gives estimated values of the terms in these expressions, based on 3,000 simulations.

$E(W)$	$Var(W)$	$CV^2(W)$	$E(\rho)$	$Var(\rho)$	$CV^2(\rho)$	$CV^2(V)$
151,213	40,741,661	0.001782	0.610966	0.007958	0.021319	0.023112

Of the two CV's, that of ρ determines the CV of V almost entirely in this case. The sum of the two squared CV's amounts to 0.023101, close to the exact value in the last column. Thus, we conclude in general that when forecast uncertainty is expressed in terms of the CV's of the predictive distributions, the factor with the largest CV (share ρ or population W) contributes most to the uncertainty in the population for a specific household position. In many cases, the share ρ has the largest CV, but there are exceptions, in particular for the young and the old. For instance, men aged 80-84 in 2030 have a CV equal to 0.1456, while the CV for the married share of these men is 0.1424. (The square root of their sum equals 0.2037, which is close to the value in Table 3.)

Consistency. When we draw random shares for married or cohabiting men and women from their predictive distributions, there is no guarantee for consistency between resulting absolute numbers for the two sexes. The discussion in Section 3.2 shows that such consistency is not necessary, because in practice there are small differences: the Census of 2001 shows 358 more married men than married women who live with their spouse, and 46 fewer cohabiting men than cohabiting women. We have used a correlation equal to 0.68 across the sexes. What is the impact of this correlation on the differences between future numbers of married or cohabiting men and married or cohabiting women? Table 4 shows simulation results for correlations equal to 0.68 (the reference value), 0.75, and 0.95.

Table 4. Difference between number of men and women by household status, averages and standard errors across 3000 simulations

	Correlation 0.68		Correlation 0.75		Correlation 0.95	
	MAR	COH	MAR	COH	MAR	COH
2010						
Average	-11,023	1,443	-10,992	1,439	-10,883	1,433
Standard error	641	490	581	439	357	235
2020						
Average	-19,000	-3,140	-18,933	-3,241	-18,672	-3,623
Standard error	1,010	778	915	696	564	371
2030						
Average	-31,902	3,834	-31,775	3,654	-31,302	2,967
Standard error	1,388	1,017	1,262	910	802	484

On the whole, an increase in the correlation between the sexes decreases the average difference only marginally, and the standard error of the mean substantially. Note also that differences between men and women are much larger in the forecast than in the base data. They increase with forecast lead time, and the standard errors indicate that the expected differences are significantly different from zero. Yet, compared to the numbers of married couples (about 900,000; see Table 3) and cohabiting couples (about 300,000) the differences are small enough to be ignored.

Comparison with multivariate random walk. The empirical specification of the multivariate random walk is based on observed forecast errors from both the SLC and the

Census. Because of sample and other survey errors in the SLC it is reasonable to expect larger uncertainty in random share forecasts that are driven by the multivariate random walk, then those driven by the continued fraction method. The estimates for the latter are solely based on forecast errors derived from the Census. A comparison of predictive distributions for shares derived from the two methods confirms this. As an example we present Box-and-whisker plots for the distributions of shares for men and women in three age groups and two household positions in 2030; see Figure 7.

Central tendencies in the share distributions are the same (as they should be), but the dispersions in the shares based on the multivariate random walk are clearly much larger than those based on continued fractions. The boxes, the edges of which reflect the first and the third quartile are longer, and the whiskers, the endpoints of which indicate minima and maxima of the distributions, stretch further towards zero and one. In other words, the continued fractions method results in less uncertain forecasts, since the predictive distributions are relatively narrow.

6. Conclusions

We have computed a probabilistic household forecast for Norway. We have combined a probabilistic population forecast, which predicts numbers of men and women by age, with a probabilistic forecast for the shares of persons who have a particular household position, given age and sex. Point predictions for the shares were computed by means of a deterministic multistate household forecast model. Variances and covariances were estimated from observed forecast errors in an old household forecast. We have restricted ourselves to private households.

We find that prediction uncertainty for future numbers of married couples, cohabiting couples, and one-person households, as indicated by the coefficient of variation, is rather low. Lone parents and other private households show much larger prediction uncertainty.

The method that we have used can also be applied to other forecasts in which the population is distinguished by a certain characteristic, in addition to age and sex. We have used household position, but other applications could include health or disability status, region of residence (cf. Wilson and Bell 2007), labour market status, etc. Such applications require that one specify a stochastic process for the random shares, which distribute the population over the various states, given age and sex. A historical deterministic forecast can be used to estimate the parameters of the stochastic process. The random shares are to be combined with the results of a probabilistic population forecast. The latter type of forecast exists for many industrialized countries.

References

Alders, M. (1999) Stochastische huishoudensprognose 1998-2050 (“Stochastic household forecast 1998-2050”). *Maandstatistiek van de bevolking* 1999/11, 25-34.

Alders, M. (2001) Huishoudensprognose 2000-2050: Veronderstellingen over onzekerheidsmarges (“Household forecast 2000-2050: Assumptions on uncertainty intervals”). *Maandstatistiek van de Bevolking* 2001/08, 14-17.

Conference of European Statisticians (2006) Recommendations for the 2010 Censuses of Population and Housing. Geneva, United Nations.

Daugstad, G. (2006) Grenseløs kjærlighet? Familieinnvandring og ekteskapsmønstre i det flerkulturelle Norge. Report nr. 2006/39. Oslo: Statistics Norway.

Ermisch, J. (2003) *An Economic Analysis of the Family*. Princeton, NJ: Princeton University Press.

Glaser, K., E. Grundy, and K. Lynch (2003) Transitions to supported environments in England and Wales among elderly widowed and divorced women: The changing balance between co-residence with family and institutional care. *Journal of Women and Aging* **15** (2/3) 107-126.

Grundy, E. (2001) Living arrangements and the health of older persons in developed countries. *Population Bulletin of the United Nations* 42/43, 311-329.

Hoffmann-Nowotny, H.J. (1987) “The future of the family” Pp. 113-200 in *Plenaries of the European Population Conference*, IUSSP/Central Statistical Office of Finland, Helsinki.

Holmberg, I. (1987) Household change and housing needs: A forecasting model. In J. Bongaarts, T. Burch, and K. Watchter (eds.), *Family Demography – Methods and Their Application*. Oxford: Clarendon Press

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

Keilman, N. (2005) Households and families: Developed Countries. In G. Caselli, J. Vallin, and G. Wunsch (eds.) Demography: *Analysis and Synthesis: A Treatise in Population*. Academic Press.

King, D. (1999) Official household projections in England: Methodology, usage and sensitivity tests. Paper presented at “Joint ECE-EUROSTAT Work Session on Demographic Projections, May 3-7, 1999, Perugia, Italy.

Lee, R. (1999) Probabilistic approaches to population forecasting. In W. Lutz, J. Vaupel, and D. Ahlburg (eds.) *Rethinking Population Projections, A Supplement to Population and Development Review* 24, 156-190.

Meslé, F. (2004) Espérance de vie: Un avantage feminine menace? *Population et Sociétés* 402, 1-4.

Moen, Ph. and K.B. Forest (1999) Strengthening families: Policy issues for the twenty-first century. Pp. 633-663 in M.B. Sussman, S.K. Steinmetz, and G.W. Peterson (eds.) *Handbook of Marriage and the Family* 2nd edition. New York: Plenum Press.

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

Normann, T. M. (2004) Samordnet levekårsundersøkelse 2002 – panelundersøkelsen. Dokumentasjonsrapport. Oslo: Statistisk sentralbyrå (Notater 2004/55).

O'Neill, B. and B. Chen (2002) Demographic determinants of household energy use in the United States. In W. Lutz, A. Prskawitz and W. Sanderson (eds.), *Population and Environment: Methods and Analysis, A Supplements to Population and Development Review* 28, 53–88.

Rosenzweig, M.R. and O. Stark (1997) *Handbook of Population and Family Economics*. Elsevier/ North Holland.

Settles, B. (1999) The future of the families. Pp. 143-175 in M.B. Sussman, S.K. Steinmetz, and G.W. Peterson (eds.) *Handbook of Marriage and the Family* 2nd edition. New York: Plenum Press.

Wilson, T. and M. Bell (2007) Probabilistic regional population forecasts: The example of Queensland, Australia. *Geographical Analysis* 39, 1-25.

Appendix 1

The purpose of this appendix is to outline our estimation strategy for the variances and covariances of the multivariate random walk.

Variances
Observations for 1997 and 2002 consist of sampling errors and prediction errors. The observations for 2001 consist of prediction errors only. Hence we write

$$\text{(A1)}\quad t.\pi_j^2 = \frac{\sum_{x,s}\left(e_\xi(j,x,s,t)\right)^2}{n_{j,t}-1} - \hat{s}_j^2 \quad \text{for } t = 6¼ \text{ and } t = 12¼,$$

$$\text{(A2)}\quad t.\pi_j^2 = \frac{\sum_{x,s}\left(e_\xi(j,x,s,t)\right)^2}{n_{j,t}-1} \quad \text{for } t = 10¾.$$

Here $n_{j,t}$ is the number of observed errors for household position j at time t. Note that we use the Mean Squared Error (MSE) to estimate the prediction variances. In order words we do not subtract the means and thus use conservative estimates of the variances.

Using the Delta theorem, and assuming a multinomial distribution for empirical shares, we compute an approximate value for the estimated sample variance $\hat{s}_j^2$ for household position j as $(\tilde{\rho}_j + \tilde{\rho}_1)/(\tilde{\rho}_j \tilde{\rho}_1 n)$, $j = 2,3,\ldots,6$.[10] Here $\tilde{\rho}_j$ is the empirical share for household position j, $\tilde{\rho}_1$ is the empirical share for the reference group "living alone", and n is the size of the population group to which the shares apply. These sample variances vary strongly across ages, because the empirical shares have very distinct age patterns; cf. Figures 4-6. Population sizes were approximately 200 given age and sex. The median values across age groups and the two sexes turned out to be as follows.

estimated sample variance, median values

	CHLD	COH	MAR	OTHR	SIN+
1997	0.0942	0.1370	0.0646	0.4551	0.2520
2002	0.1115	0.1804	0.0744	0.4542	0.2540

The sample variances are approximately the same in the two years. We subtracted the median values above from the MSE's for 1997 and 2002. Together with the MSE's for 2001 this resulted in the following values

	CHLD	COH	MAR	OTHR	SIN+
1997	0.7786	0.2313	0.1993	0.6009	2.4217
2001	1.3240	0.3790	0.2782	0.7246	1.6197
2002	1.8680	0.6299	0.5371	1.4879	1.7757
column sum	3.9706	1.2402	1.0147	2.8134	5.8171

[10] See Appendix 2 for a proof.

The MSE's for children are relatively large, because we have only four data points for each year. The MSE's for lone parents are high because this household position is difficult to predict. The values vary a great deal across household positions. We divided each number in the table above by its column sum, and ran an OLS-regression without constant with these 15 values as dependent variables, and lead times as independent variables. The resulting OLS-estimate of the coefficient was 0.0326, with a standard error of 0.0026. Residual sums of squares across three observations were lowest for "child" (0.0076), "cohabiting" (0.0181), and "married" (0.0288), and largest for "other" (0.0326), and "lone parent" (0.0649). When multiplied again with the column sums above, this gives us the following estimates of innovation variances:

CHLD	COH	MAR	OTHR	SIN+
0.1293	0.0404	0.0330	0.0916	0.1894

The random walks for the errors in the shares of household positions "married" and "cohabiting" have small innovation variances, compared to those for the other household positions.

Covariances

We proceeded in a similar manner as for the variances above. There are ten covariances to be analysed, one for each pair (*i,j*), $i, j = 2,3,...6; i < j$. First we estimated sample covariances $\hat{s}^2_{i,j}$. A first-order approximation for the latter covariances is $1/\tilde{\rho}_1 n$ for every *i* and *j*; see the appendix. There was very little variation across ages, sexes, and time (years 1997 and 2002 only) in this approximate covariance value, except for extreme ages, which had high values. The median value was 0.0465.

Next we computed mean cross-products of errors similar to the mean squared errors in expressions (A1) and (A2). The numerator is the sum of cross-products of errors. The denominator is $\{\min(n_{i,t},n_{j,t})\text{-}1\}$. In order to obtain an acceptable number of observations, we pooled the data across the three calendar years. Even then, some mean values are based on unacceptably few data points. Mean cross products for the three years combined are given in Table 1. We used these values as an estimate of the corresponding error covariances. The table gives also estimates of the correlations, and the number of data points involved. Correlations are essentially zero, except for covariances (CHLD,MAR) and (CHLD,SIN+). But the latter two estimates are based on too few data points to be trusted. Thus the data do not suggest any particular dependence of the errors across household positions, and we shall assume that they are independent.

When the median sample covariance of 0.0465 is subtracted from the estimated covariances in Table 1, the resulting values are still small enough to assume independence across household positions. Thus the covariance matrix $\Sigma^2_\xi(t)$ is estimated as $\hat{\Sigma}^2_\xi(t) = t\hat{\Pi}^2$, with innovation variances equal to

$$\hat{\Pi}^2 = \begin{array}{ccccc} \text{CHLD} & \text{COH} & \text{MAR} & \text{OTHR} & \text{SIN+} \\ 0.1293 & 0 & 0 & 0 & 0 \\ 0 & 0.0404 & 0 & 0 & 0 \\ 0 & 0 & 0.0330 & 0 & 0 \\ 0 & 0 & 0 & 0.0916 & 0 \\ 0 & 0 & 0 & 0 & 0.1894 \end{array}$$

Appendix 2

The purpose of this appendix is to derive expressions for sample variances and covariances of the ξ_j-variables, when values for the shares ρ_j are estimated from a sample survey.

The share ρ_j for household position j is estimated as $\tilde{\rho}_j = n_j / n$, where n_j represents the number of persons in household position j, and n the total number of persons irrespective of household position. There are J household positions. We assume a multinomial distribution for the counts n_j. Hence the variance of $\tilde{\rho}_j$ is estimated as $\tilde{\rho}_j(1-\tilde{\rho}_j)/n$, and the covariance between any two shares $\tilde{\rho}_i$ and $\tilde{\rho}_j$ as $(-\tilde{\rho}_i\tilde{\rho}_j/n)$, $i \neq j$.

The multinomial logit transformation implies that $\xi_j = \ln(\rho_j/\rho_1)$, $j = 2,3,\ldots,J$. Household position $j = 1$ is taken as the reference. Hence $\frac{\partial \xi_j}{\partial \rho_j} = \frac{1}{\rho_j}$, $j \neq 1$, whereas $\frac{\partial \xi_j}{\partial \rho_1} = -\frac{1}{\rho_1}$. Then the Delta theorem gives the following approximate estimated value for the estimator of the variance of ξ_j:

$$\mathrm{Var}(\tilde{\xi}_j) \approx \left(\left(\frac{1}{\rho_j}\right)^2 \mathrm{Var}(\rho_j) + \left(\frac{-1}{\rho_1}\right)^2 \mathrm{Var}(\rho_1) - 2\frac{\mathrm{Cov}(\rho_1,\rho_j)}{\rho_1\rho_j} \right)_{\rho_i=\tilde{\rho}_i} .$$

Inserting expressions for variances and covariance, we find

$$\mathrm{Var}(\tilde{\xi}_j) \approx \frac{\tilde{\rho}_1 + \tilde{\rho}_j}{\tilde{\rho}_1\tilde{\rho}_j n} .$$

To find the covariance between $\tilde{\xi}_2$ and $\tilde{\xi}_3$, we write first $\xi_2 = \ln(\rho_2/\rho_1) = \xi_2(\rho_1,\rho_2,\rho_3)$, and

$\xi_3 = \ln(\rho_3/\rho_1) = \xi_3(\rho_1,\rho_2,\rho_3)$.

Derivatives are $\frac{\partial \xi_2}{\partial \rho_1} = \frac{-1}{\rho_1}, \frac{\partial \xi_2}{\partial \rho_2} = \frac{1}{\rho_2}, \frac{\partial \xi_2}{\partial \rho_3} = 0$, and similarly $\frac{\partial \xi_3}{\partial \rho_1} = \frac{-1}{\rho_1}, \frac{\partial \xi_3}{\partial \rho_2} = 0, \frac{\partial \xi_3}{\partial \rho_3} = \frac{1}{\rho_3}$.

The Delta theorem gives

$$\mathrm{Cov}(\tilde{\xi}_2,\tilde{\xi}_3) \approx \left(\sum_i \sum_j \frac{\partial \xi_2}{\partial \rho_i} \frac{\partial \xi_3}{\partial \rho_j} \mathrm{Cov}(\rho_i,\rho_j) \right)_{\rho_i=\tilde{\rho}_i}$$, with $\mathrm{Cov}(\rho_i,\rho_i) = \mathrm{Var}(\rho_i)$.

Inserting expressions for derivatives and covariances, we find $\mathrm{Cov}(\tilde{\xi}_2,\tilde{\xi}_3) \approx \frac{1}{\tilde{\rho}_1 n}$. The multinomial logit transformation implies that any pair $(\tilde{\xi}_i,\tilde{\xi}_j)$, $i \neq j;\ i,j \neq 1$ has positive covariance, which is independent of the shares $\tilde{\rho}_2$ and $\tilde{\rho}_3$, at least as a first order approximation.

Figure 1

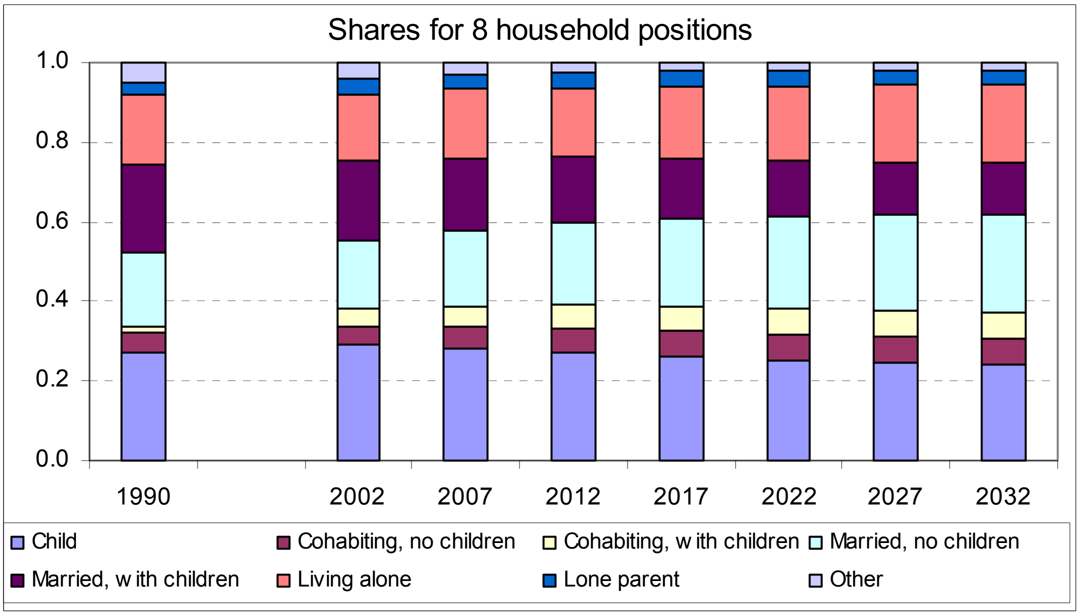


Figure 2

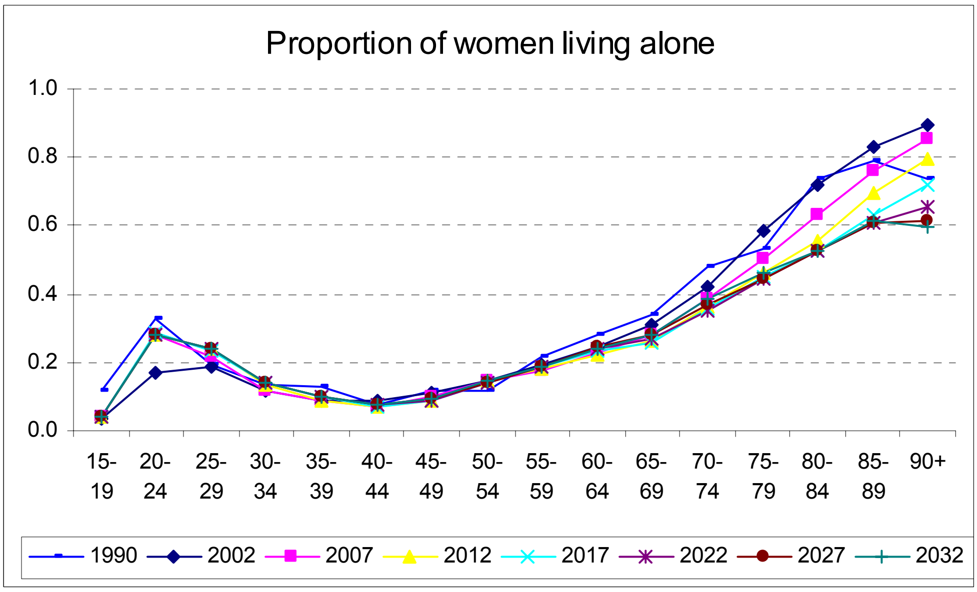


Figure 3

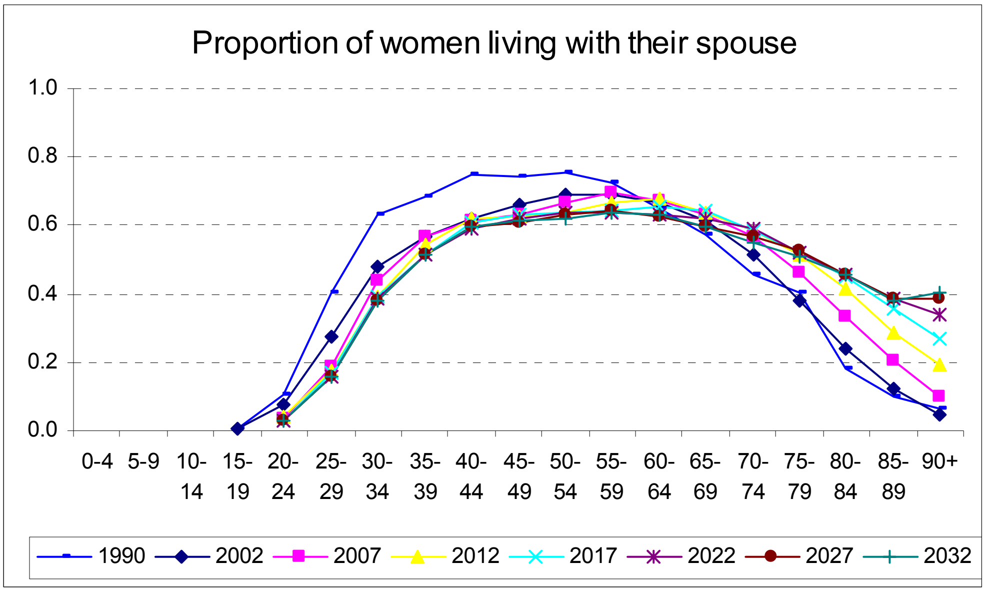
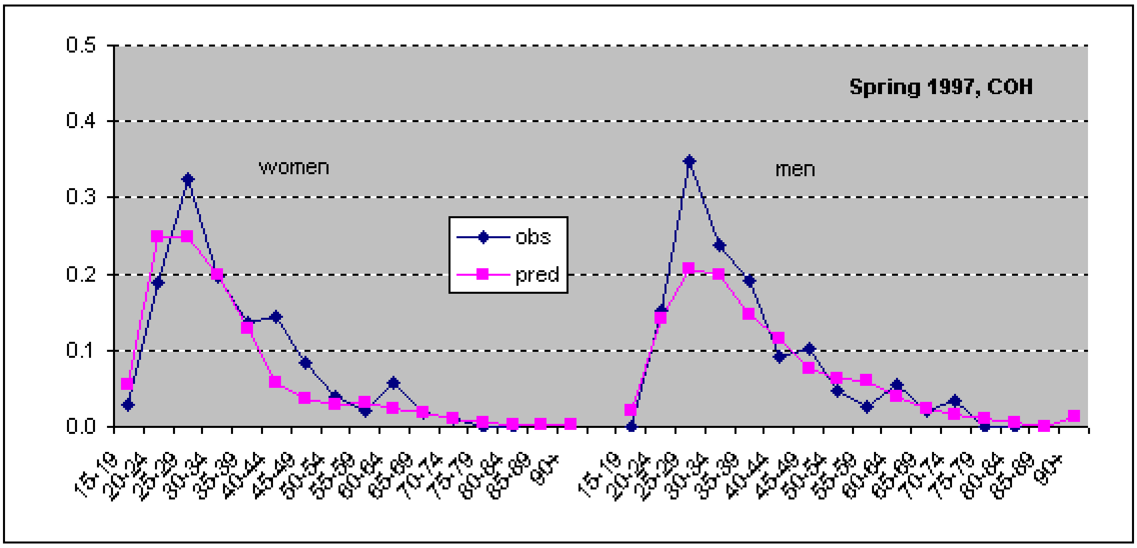
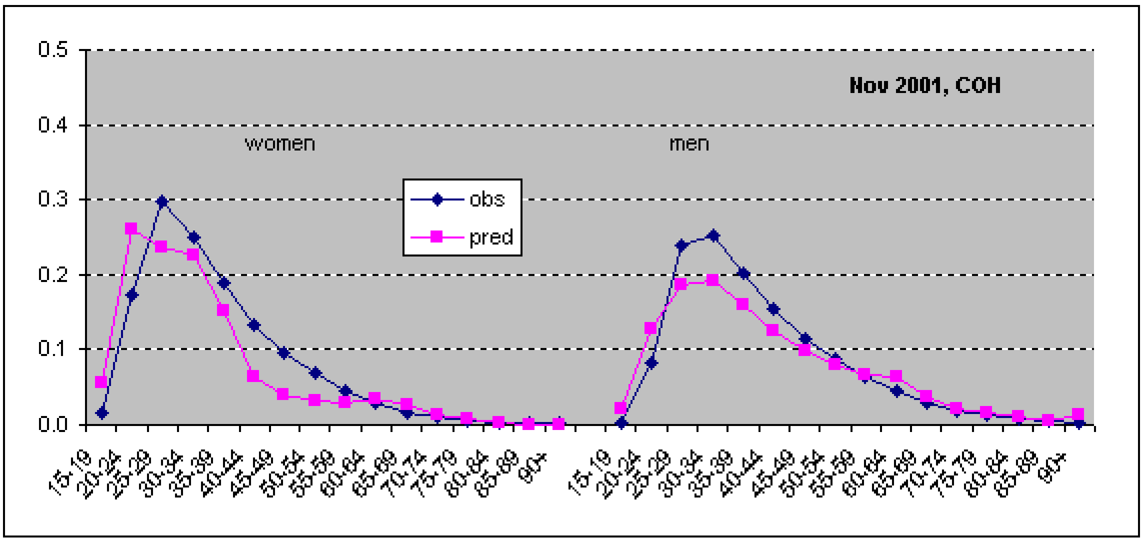


Figure 4





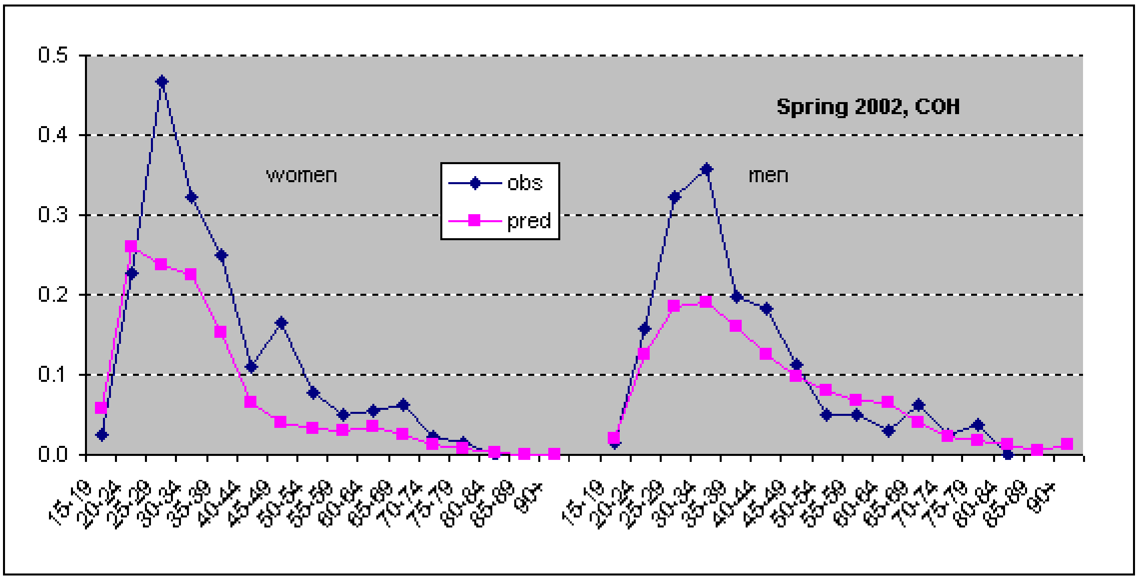

Spring 2002, COH
women
men
obs
pred
0.0
0.1
0.2
0.3
0.4
0.5
15-19
20-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-64
65-69
70-74
75-79
80-84
85-89
90+

Figure 5

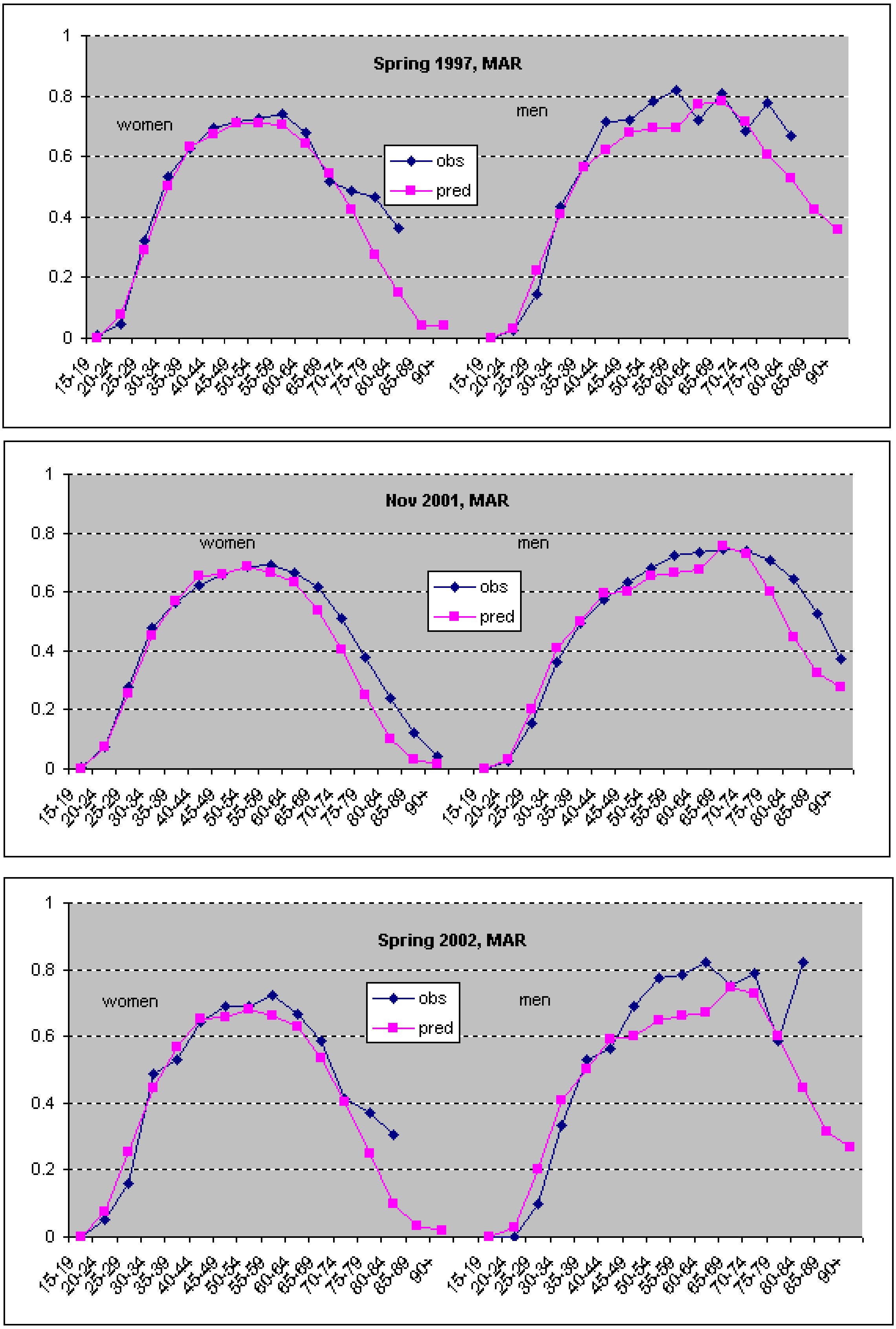

Spring 1997, MAR
women
men
obs
pred
0
0.2
0.4
0.6
0.8
1
15-19
20-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-64
65-69
70-74
75-79
80-84
85-89
90+
Nov 2001, MAR
women
men
obs
pred
Spring 2002, MAR
women
men
obs
pred

Figure 6

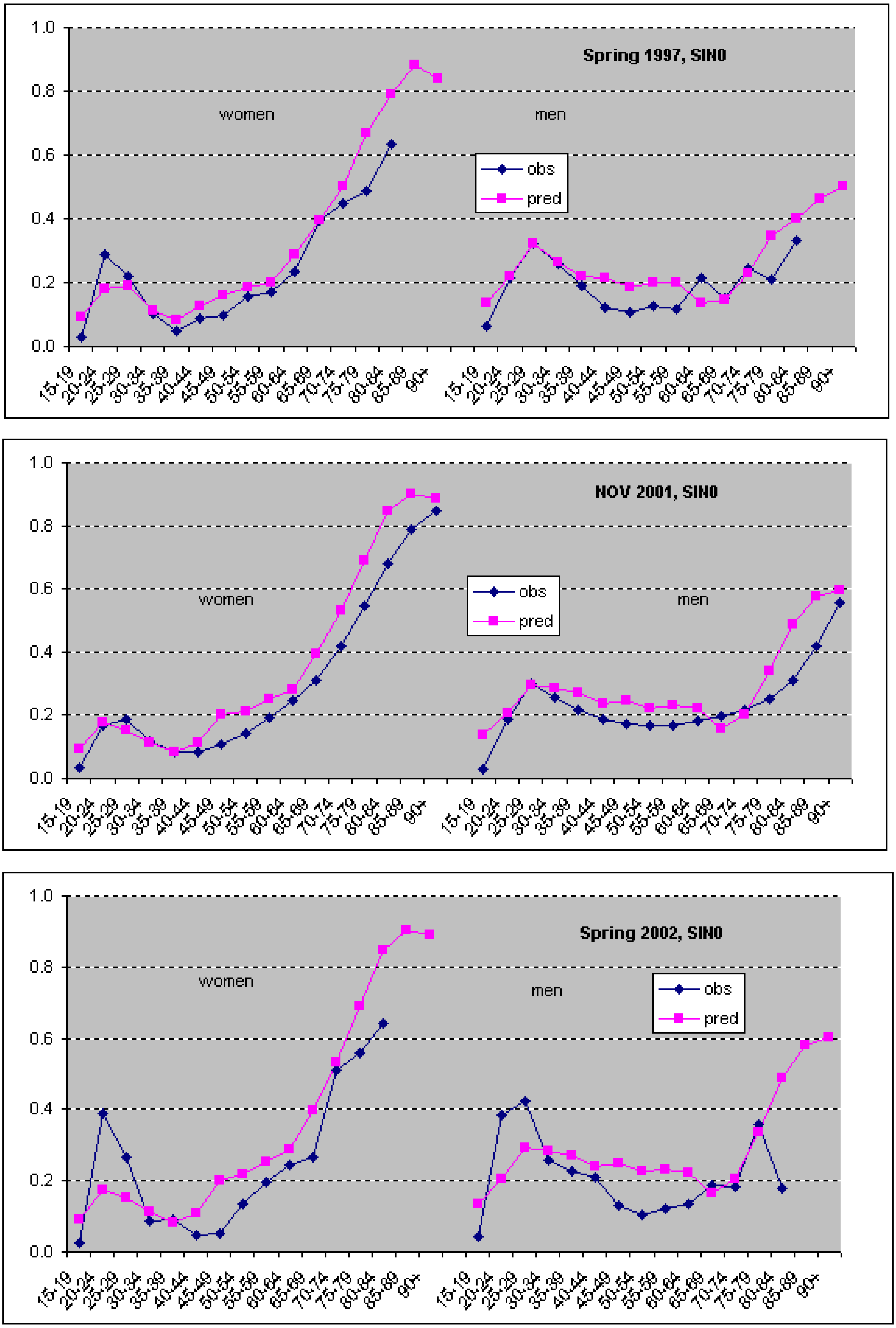
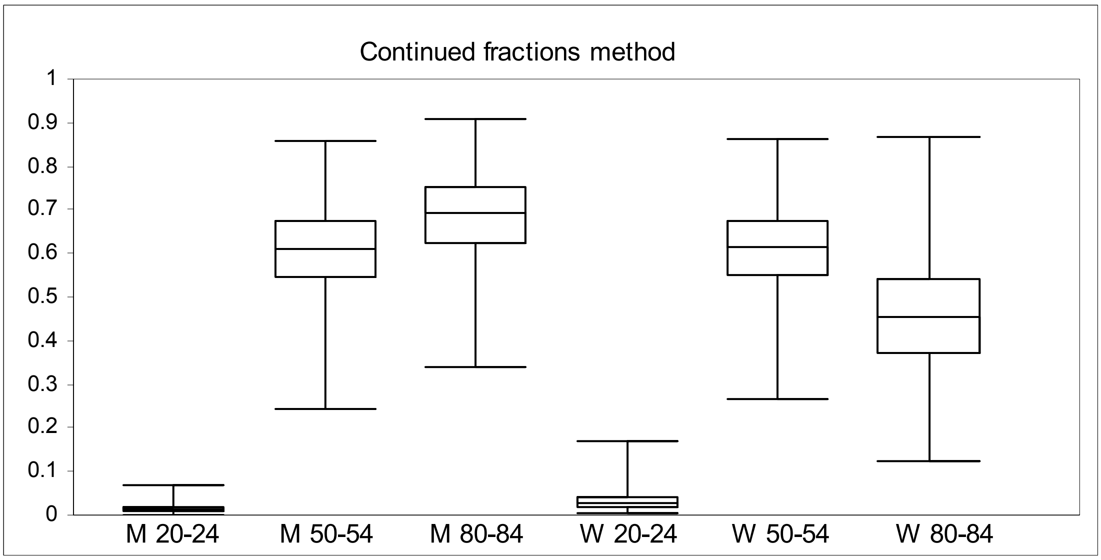

Spring 1997, SIN0
women
men
obs
pred
NOV 2001, SIN0
women
men
obs
pred
Spring 2002, SIN0
women
men
obs
pred
1.0
0.8
0.6
0.4
0.2
0.0
15-19
20-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-64
65-69
70-74
75-79
80-84
85-89
90+

Figure 7a. **Box-and-whisker plots for shares in household position "married" in 2030, men and women, three age groups**



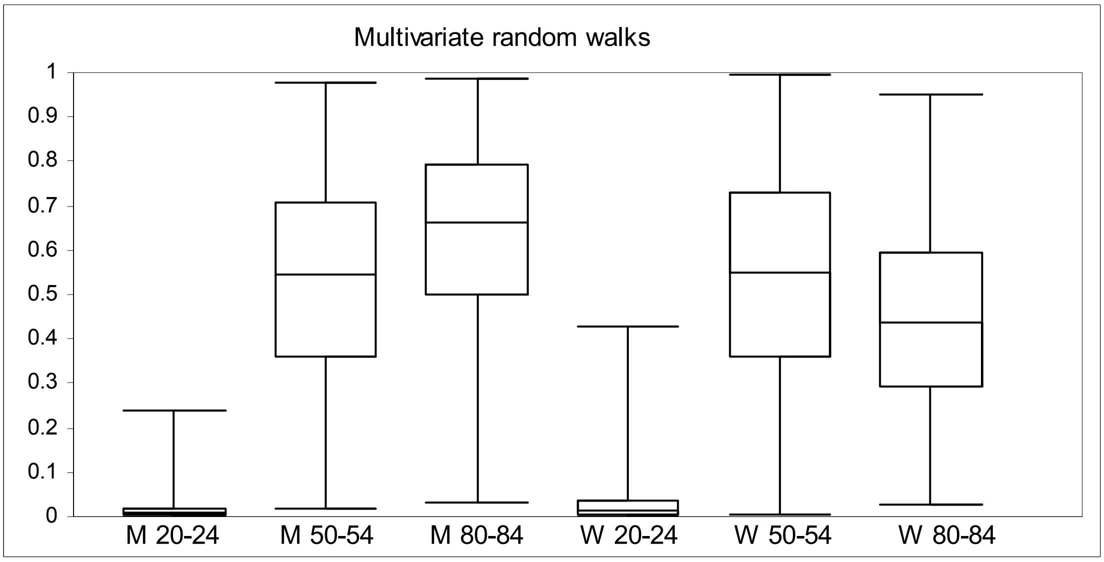
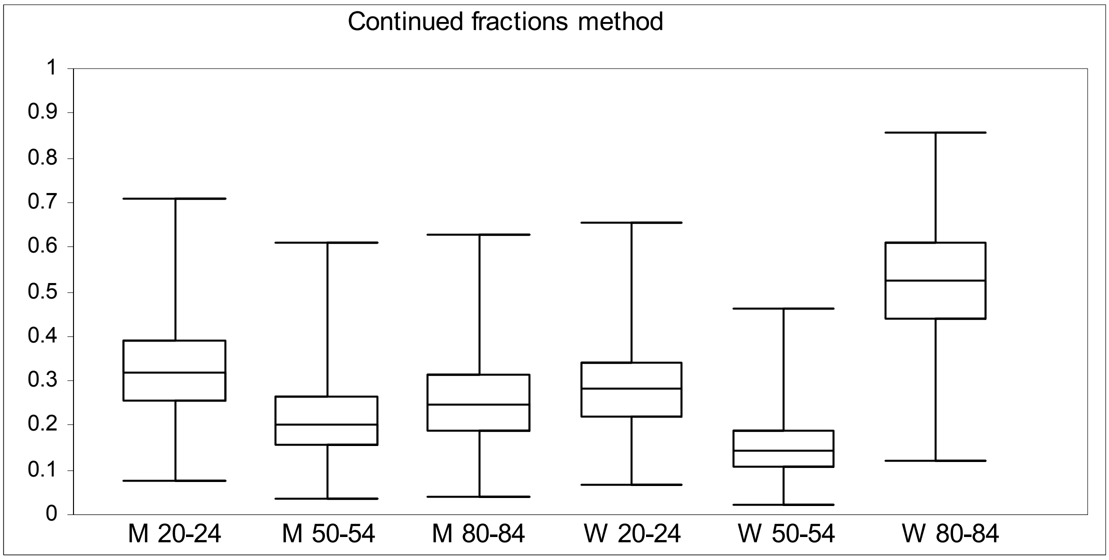


Figure 7b. Box-and-whisker plots for shares in household position "living alone" in 2030, men and women, three age groups



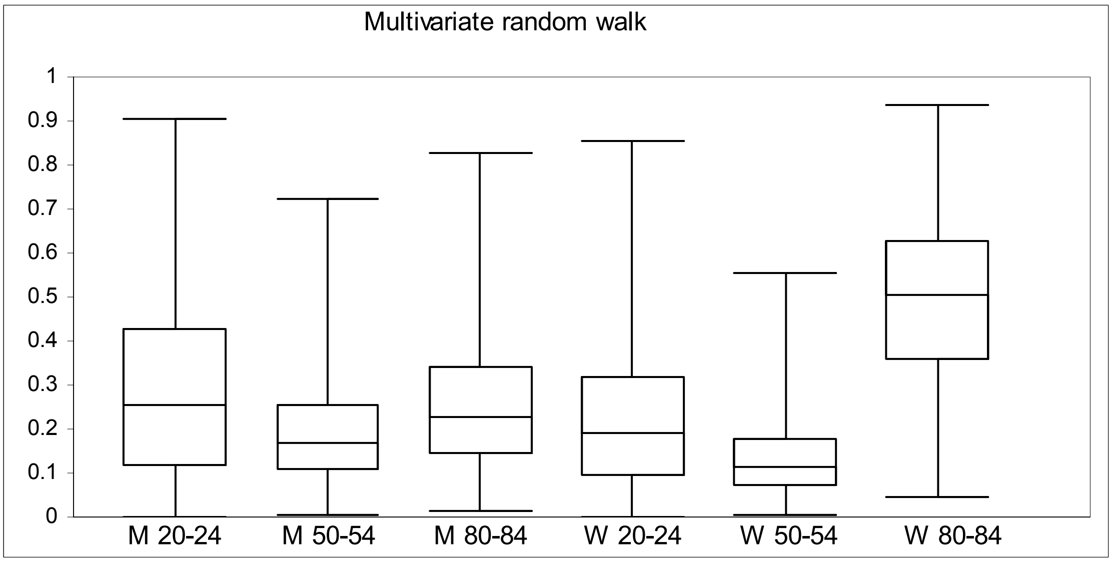


Table 1. Number of observations (n), mean cross products (mcp) and correlations (corr) for errors $e_{\xi}(j,x,s,t)$

	CHLD,COH	CHLD,MAR	CHLD,OTHR	CHLD,SIN+	COH,MAR	COH,OTHR	COH,SIN+	MAR,OTHR	MAR,SIN+	OTHR,SIN+
n	11	5	11	3	75	71	49	68	49	44
mcp	0.0300	0.1286	0.1768	0.6954	0.0030	0.0008	0.0112	0.0002	0.0091	0.0009
corr	0.0484	0.2344	0.0827	0.4729	0.0077	0.0013	0.0101	0.0005	0.0122	0.0008

*****