

Household and Living Arrangement Projections at the Subnational Level: An Extended Cohort-Component Approach

Yi Zeng · Kenneth C. Land · Zhenglian Wang ·
Danan Gu

Published online: 4 December 2012
© Population Association of America 2012

Abstract This article presents the core methodological ideas and empirical assessments of an extended cohort-component approach (known as the “ProF-amy model”), and applications to simultaneously project household composition, living arrangements, and population sizes–gender structures at the

Electronic supplementary material The online version of this article (doi:10.1007/s13524-012-0171-3) contains supplementary material, which is available to authorized users.

Y. Zeng (✉)

Center for the Study of Aging and Human Development and Geriatric Division of Medical School,
Population Research Institute, Duke University, Box 3003, Durham, NC 27710, USA
e-mail: zengyi68@gmail.com

Y. Zeng

Center for Healthy Aging and Development Studies, National School of Development, Peking
University, Beijing, 100871, China

K. C. Land

Population Research Institute and Department of Sociology, Center for Population Health and Aging,
Duke University, Durham, NC, USA
e-mail: kland@soc.duke.edu

Z. Wang

Center for Population Health and Aging of Population Research Institute, Duke University, Household
and Consumption Forecasting, Inc., Durham, NC, USA
e-mail: wangzl8886@gmail.com

D. Gu

Center for Study of Aging and Human Development, Duke University, Durham, NC, USA
e-mail: gudan@duke.edu

subnational level in the United States. Comparisons of projections from 1990 to 2000 using this approach with census counts in 2000 for each of the 50 states and Washington, DC show that 68.0 %, 17.0 %, 11.2 %, and 3.8 % of the absolute percentage errors are <3.0 %, 3.0 % to 4.99 %, 5.0 % to 9.99 %, and ≥ 10.0 %, respectively. Another analysis compares average forecast errors between the extended cohort-component approach and the still widely used classic headship-rate method, by projecting number-of-bedrooms-specific housing demands from 1990 to 2000 and then comparing those projections with census counts in 2000 for each of the 50 states and Washington, DC. The results demonstrate that, compared with the extended cohort-component approach, the headship-rate method produces substantially more serious forecast errors because it cannot project households by size while the extended cohort-component approach projects detailed household sizes. We also present illustrative household and living arrangement projections for the five decades from 2000 to 2050, with medium-, small-, and large-family scenarios for each of the 50 states; Washington, DC; six counties of southern California; and the Minneapolis–St. Paul metropolitan area. Among many interesting numerical outcomes of household and living arrangement projections with medium, low, and high bounds, the aging of American households over the next few decades across all states/areas is particularly striking. Finally, the limitations of the present study and potential future lines of research are discussed.

Keywords Household and living arrangement projections · Cohort-component approach · Subnational projections · Aging of households · Forecast errors

Introduction

There is a growing demand for projections of the distribution of household types, sizes, and living arrangements for socioeconomic planning, environment, development, business and market research, and policy and scholarly analysis (Keilman 2003; Liu et al. 2003; Lutz and Prinz 1994; Mackellar et al. 1995; Moffitt 2000; Myers et al. 2002). For example, past research has established that household status and living arrangements are two of the major determinants of the amount and type of long-term care for the elderly (e.g., Federal Interagency Forum on Aging Related Statistics (FIFARS) 2004; Freedman 1996; Soldo et al. 1990). With rapid population aging in many countries around the world, the importance of elder care research and planning is growing quickly, which creates a strong demand for projections of household and elderly living arrangements (Goldscheider 1990; Himes 1992).

In recent years, more researchers and policymakers are demanding demographic projections at subnational levels, which are useful for distributing government funds and resources, planning the development of infrastructure and public facilities, market research and planning for household-related goods and services, and decisions on the expansion or reduction of local businesses (Crowley 2004; Ip and McRae 1999; Rao 2003; Swanson and Pol 2009; Treadway 1997).

Most household projections at subnational levels by statistical offices and market analysis agencies employ the classic headship-rate approach. However, the headship-rate method suffers several serious shortcomings and has been criticized widely by demographers for about two decades (see Bell and Cooper 1990; Mason and Racelis 1992; Murphy 1991; Spicer et al. 1992). First, the designation of a household head is a vague, ill-defined, arbitrary choice¹ that is not easy to model, making projections difficult (Mason and Racelis 1992:510; Murphy 1991). Second, given its nature of cross-sectional extrapolations (see panel B of Fig. 1), the headship-rate method cannot be linked to demographic rates, and thus it is impossible to incorporate projected or assumed changes in the propensity and timing of demographic processes into headship rates (Mason and Racelis 1992; Spicer et al. 1992). Third, the information on households produced by headship-rate projections is very limited and inadequate for purposes of more detailed planning and analysis (Bell and Cooper 1990). For example, a typical well-done national household projection using the headship-rate method projected only five household types by age groups of household head (U.S. Census Bureau 1996), with no projected household sizes available. This, again, is limiting given that households with various sizes differ substantially in their needs for products and services. Fourth, the headship rate lumps all household members other than heads into one category—“non-head”—with no projected information (Burch, personal communication, Sept. 10, 1999). This makes it impossible to study the household status and living arrangements of the elderly, adults, and children who are “nonhead” but who may also themselves be significantly relevant to business, academic, and policy analysis and planning.

In contrast with the headship-rate method, the extended cohort-component model for projecting family households and living arrangements (abbreviated as “ProFamy” model hereafter), as initially developed in Zeng et al. (1997, 1998), further extended and justified in Zeng et al. (2006)² and employed in (among others) Prskawetz et al. (2004), Smith et al. (2008, 2012), Dalton et al. (2008), and Feng et al. (2011), does not suffer the vague, ill-defined, and arbitrary chosen designation of the household heads; instead, it projects all individuals grouped by ages/cohorts and specified attributes (e.g., a group of persons of the same race, sex, marital/union status, and coresidence status with parents and children). The calculations of the ProFamy model proceed iteratively, group by group, cohort by cohort, and time period by time period, using demographic rates as inputs (see panel A of in Fig. 1), and it projects much more detailed household types, sizes, and living arrangements for all members of the population. Note that detailed projections of household sizes by various types are particularly useful in market analysis.

As we describe in the next section, the basic mechanism of the ProFamy model is that projections of changes in demographic components (marriage/union formation and dissolution, fertility, leaving parental home, mortality, and migration) are made for each cohort that produce household distributions in future years. This is analogous to, and a substantive extension of, the conventional cohort-component population projection model

¹ For example, changes in headship rates may depend on whether the census or survey was carried out in the daytime or evening and whether more women or men were available to complete the questionnaire.

² The ProFamy model was built on methodological advances in multidimensional demography (Land and Rogers 1982; Rogers 1975, 1995; Schoen 1988; Willekens et al. 1982) and based on Bongaarts’s and Zeng’s one-sex family status life table models (Bongaarts 1987; Zeng 1986, 1988, 1991).

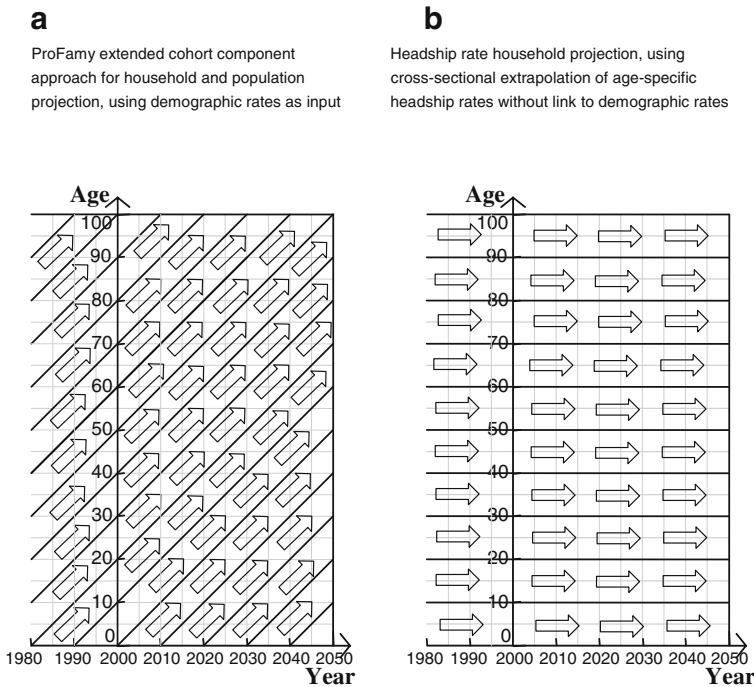


Fig. 1 A comparison of the basic framework of the ProFamy extended cohort-component approach and the headship-rate method

in the sense that the ProFamy model simultaneously projects households, living arrangements, and population age/sex distributions. Prior assessments of the accuracy of projections at the national level from 1990 to 2000 using the ProFamy model show that forecast errors measured by discrepancies between the projected values and the U.S. 2000 census observations are reasonably small, validating the ProFamy model (Zeng et al. 2006). Similar validation tests of projections of Chinese households, living arrangements, and population from 1990 to 2000 have shown that the discrepancies between the projected and the 2000 census observations are again within a reasonable range (Zeng et al. 2008). It thus has been established that the ProFamy method for producing household, elderly living arrangement, and population projections works reasonably well at the national level.

The question remains, however, as to whether the ProFamy extended cohort-component model also works well for subnational areas, which may not have their own observed age-/sex-/race-specific demographic rates. Motivated by the growing scientific and practical needs, this article further develops, applies, and empirically evaluates the extended cohort-component model for household and living arrangement projections at the subnational level.³ The next section describes the core ideas of the ProFamy model. The third section discusses data and estimation issues. The fourth section presents empirical validation tests of projections from 1990 to 2000 by comparing projected with census counts in 2000, using the ProFamy model for each of the 50 states

³ The “subnational level” referred in this article does not include small counties/cities/towns and other kinds of small areas (possibly even tracts or block groups) that do not have reasonably reliable data from which to estimate the demographic summary parameters.

and Washington, DC. We also compare average forecast errors of several major indices of the headship-rate method and the ProFamy approach, both applied to project household housing demands by number of bedrooms from 1990 to 2000 with results compared with the census counts in 2000 for each of the 50 states and Washington, DC. The fifth section describes a summary of illustrative households and living arrangement projections in the United States, with medium-, small-, and large-family scenarios from 2000 to 2050 for each state; Washington, DC; each of the six counties of southern California (“SC” hereafter); and the Minneapolis–St. Paul metropolitan area (abbreviated as “M-S area” hereafter). A discussion and conclusion section ends the article.

Core Ideas of the Extended Cohort-Component Model

The ProFamy model is built on four core ideas: (1) a multistate accounting model; (2) distinguishing continuously occurring from periodic demographic accounting processes; (3) judicious use of independence assumptions; and (4) using national model standard schedules and summary parameters at the subnational level to specify projected demographic rates of a subnational area in future years.

Core Idea 1: A Multistate Accounting Model

The innermost core of the ProFamy model is a multistate accounting model for transforming the marital/union statuses and coresidence with children and parents statuses of members of a population in year t into their corresponding statuses in year $t + 1$. In the ProFamy model, groups of individuals are the basic units of the analysis, and all individuals of the population are grouped and projected forward by age, sex, marital/union status, parity, and number of coresiding children and parents; consequently, only conventional and normally available census, surveys, and vital statistics data are required (Zeng et al. 2006). The ProFamy model uses *groups of individuals* as the units of analysis, in contrast with the *groups of households* used in most other macrosimulation models for household projections, because using households as units of analysis requires data on transition probabilities among household-type statuses—data that have to be collected in special surveys because they are not available in vital statistics, censuses, or ordinary surveys (Keilman 1988; Van Imhoff and Keilman 1992). This strong data requirement is an important factor in the slow development and infrequent application of these models (Van Imhoff et al. 1995). Furthermore, household status transition-based models cannot directly link changes in household structure to demographic rates. Thus, identifying the effects of demographic factors on changes in household structure is difficult for such a model.

In addition to identifying the individual members of a population by single years of age, sex, race, and rural–urban residence (optional), the ProFamy model keeps track of changes in individuals’ living arrangements, including marital/union status, statuses of coresidence with one or two parent(s), and number of coresiding children in each year of the projection. To derive the distributions of household types and sizes, we follow Brass’s (1983) basic concept of using a marker or reference person to identify and classify households based on the individuals’ marital/union and coresidence statuses with parents/children. For example, a married or cohabiting woman who is not coresiding with parents

($k = 0$) and whose number of coresiding children is c ($c = 0, 1, 2, 3, 4, 5+$; $c > 0$ in this example) is a reference person representing a two-generation and couple household of $2 + c$ family members. If this married or cohabiting woman's status of coresidence with parents is 1 ($k = 1$, living with one parent) or 2 ($k = 2$, living with two parents), she represents a three-generation household with a size of $2 + c + k$. If the reference person is not married and not cohabiting (can be a man⁴ or a woman), he or she is the reference person for a single-parent household of $1 + c$ family members.

Core Idea 2: Distinguishing Continuously Occurring From Periodic Demographic Accounting Processes

With the model design and individual statuses identified, a conventional multistate computation strategy would require estimation of very high dimensional matrices of cross-status transition probabilities. For example, if seven marital/union statuses,⁵ three statuses of coresidence with parents (i.e., $k = 0, 1, 2$), six parities (i.e., parity 0, 1, 2, 3, 4, 5+), and six coresidence statuses with children are distinguished as in the U.S. household and living arrangement projections of Zeng et al. (2006), one would have to estimate a cross-status transition probabilities matrix with $194,481 (= 441 \times 441)$; where $441 = 7 \times 3 \times \sum_{p=0}^5 (p+1)$ elements⁶ at each age of each sex for each race group. This is certainly not practical because it would be impossible to have a sufficiently large data set with appropriate sizes of the subsamples to reasonably estimate so many elements of the cross-status transition probabilities matrix at each age of each sex for each race group, although there are considerable numbers of structural zero elements, such as transitions to lower parity. Thus, we adopt a computational strategy of calculating changes in individual group marital/union, coresidence (with parents/children), migration, and survival status by assuming the following: (a) births occur throughout the first and second half of the single-year age interval, and (b) marital/union status changes, leaving parental home, migration, and death occur in the middle of the age interval (see Fig. 2). This strategy, which was originally proposed by Bongaarts (1987) and further justified mathematically and numerically by Zeng (1991:61–63, 80–84), circumvents the problems of estimating huge matrices of cross-status transition probabilities.

Core Idea 3: Judicious Use of Independence Assumptions

Coupled with Core Idea 2, the third core idea of the ProFamy model greatly simplifies the estimation of the multistatus transition probabilities. This idea, also originally suggested by Bongaarts (1987) and adapted and generalized by Zeng et al. (1997, 1998), is that not all elements of the transition probability matrix depend on many of the other elements; and,

⁴ A married or cohabiting man cannot be a reference person because we already chose the married or cohabiting woman as the reference person, and one household cannot have two reference persons.

⁵ The seven marital/union statuses are (1) never-married and not cohabiting, (2) married, (3) widowed and not cohabiting, (4) divorced and not cohabiting, (5) never-married and cohabiting, (6) widowed and cohabiting, and (7) divorced and cohabiting.

⁶ Because number of coresiding children is equal to or less than parity, the number of composite statuses of parity and coresiding children is $\sum_{p=0}^5 (p+1)$ rather than (6×6) .

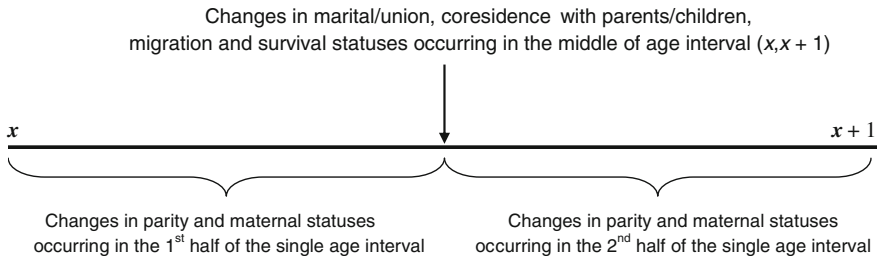


Fig. 2 Computational strategy to calculate changes in marital/union, coresidence with parents/children, migration, and survival statuses

indeed, some of their real-world dependencies are sufficiently small enough that they can be reasonably assumed to be independent. In other cases, the reality of limited data sources available for estimation of transition probabilities that depend on many other covariates forces the application of an independence assumption. In either case, the consequences of the independence assumption are that either (a) some statuses do not affect or condition the risks of transition between other statuses or (b) marginally or partially conditioned estimates of risk for each of two or more statuses can be multiplied to estimate the corresponding transition probabilities. More specifically, in the extended cohort-component model, marital/union status transitions depend on age, sex, and race but are assumed to be independent of parity and coresidence status with parents and children⁷; fertility rates depend on age, race, parity, and marital/union status but are assumed to be independent of coresidence status with parents and children; mortality rates are age-, sex-, race-, and marital/union-status-specific but are assumed to be independent of parity and coresidence status with parents and children; the probability of two parents dying in the same year is estimated by multiplying the corresponding probabilities of death of the mother and father; and the probability of more than one child leaving home in the same year is estimated by multiplying the corresponding probabilities of leaving home of each child.

Core Idea 4: Using National Model Standard Schedules and Summary Parameters at the Subnational Level to Specify Projected Demographic Rates of a Subnational Area in Future Years

Data for estimating race-/sex-/age-specific standard schedules of the demographic rates of fertility, mortality, marriage/union formation and dissolution, and leaving the parental home (see panel 2-I in Table 1) may not be available at the subnational level. However, after the age-/race-/sex-specific standard schedules at the national level are prepared (and updated when new data become available),⁸ they can be employed as model

⁷ Ideally, one may wish to differentiate the marital-/union-status transition probabilities by parity and coresidence status with children. Such differentiation is, however, not practically feasible because it would require a data set with a very large sample size (not available to us currently but not theoretically impossible at some future time point for some specific populations) for estimating the parity-/coresidence-/marital-status-/union-status-specific transition probabilities at each single age for men and women of each race group, with a reasonable accuracy.

⁸ With the model standard schedules in hand, analysts can concentrate on projecting future demographic summary parameters. This can be done by using conventional time series analysis by statistical software (e.g., SAS, SPSS, or STATA) or expert opinion approach. Time series data on other related socioeconomic covariates (e.g., average income, education, urbanization) also can be used in projecting the demographic summary parameters.

Table 1 Data needs for projecting households and living arrangements using the ProFamy extended cohort-component method and the data sources for the U.S. applications

Contents of the Data	Data Source for U.S. Applications
1. Base population at national and subnational levels. A census or an exceptionally large survey microdata file ^a with a few needed variables, including sex, age, marital/union status, relationship to the householder, and whether living in a private or institutional household, and optional categories (race and rural/urban). ^b	Census 5 % microdata or more recent and cumulative American Community Survey (ACS) data files and the published online 100 % census or ACS cross-tabulations.
2-I. Model standard schedules at national level (may be used for the projections at subnational level)	
a. Race-/sex-/age-specific death rates (marital-status specific, if possible).	a: Census Bureau's estimates, and Schoen and Standish (2001).
b. Race-/sex-/age-specific o/e rates of marriage/union formation and dissolution.	b and c: Pooled NSFH, NSFG, CPS, SIPP data sets. See Zeng et al. (2006) and Zeng et al. (2012).
c. Race-/age-/parity-specific o/e rates of marital and nonmarital fertility.	
d. Race-/sex-/age-specific net rates of leaving the parental home, estimated based on two adjacent census microdata files and the intracohort iterative method (Coale 1984, 1985; Stupp 1988; Zeng et al. 1994).	d: The 1990 and 2000 censuses microdata files.
e. Race-/sex-/age-specific rates of international immigration and emigration.	e: Census 5 % microdata files.
2-II. Model standard schedules at subnational level	
f. Race-/sex-/age-specific rates of domestic in-migration and out-migration for the subnational regions (estimated based on the census data).	f: Census 5 % microdata, ACS data files.
g. Race-/age-/sex-specific proportion of persons who live in group quarters.	g: Census data.
3. Demographic summary parameters for the nation and subnational regions	
a. Race-specific general rates of marriage and divorce.	a and b: Based on census microdata, vital statistics, and pooled survey data sets.
b. Race-specific general rates of cohabitation union formation and dissolution.	
c. Race-specific total fertility rates (TFR) by parity.	c–f: Based on estimates released by the Census Bureau and the National Center for Health Statistics.
d. Race-/sex-specific life expectancies at birth.	
e. Race-/sex-specific total numbers of male and female migrants	
f. Race-/sex-specific mean age at first marriage and births.	
g. Race-/sex-specific proportion of those aged 45–49 who do not live with parents (measuring level of race-/sex-/age-specific net rates of leaving the parental home).	g: Census 5 % microdata files.
h. Race-/household size-specific average number of other relatives (than spouse/partner, parents, and children) and nonrelatives living in the same household.	h: Census 5 % microdata files.

^a If a sample data set is used, published 100 % census tabulations of age/sex/marital-status distributions of the entire population including those living in group quarters, as well as the aggregated numbers of households, will be used. This is to ensure accurate total population size and total numbers of households in the starting year of the projection, while the sample data set provides more detailed distributions.

^b The race dimension is distinguished in the U.S. household and living arrangement projections, but it can be omitted in the other countries' applications if the race differences are not crucial, the race-specific data are not available, or the subpopulation sizes of the minority race groups are too small. The rural-urban dimension can be included if the rural-urban differences are substantial and the rural-/urban-specific data are available, such as in the applications for China (Zeng et al. 2008).

standard schedules for projections at the subnational level. This is similar to the widely practiced application of model life tables (e.g., Coale et al. 1983; United Nations 1982), the Brass logit relational life table model (e.g., Murray et al. 2003), the Brass relational Gompertz fertility model (Brass 1974), and other parameterized models (e.g., Coale and Trussell 1974; Rogers 1986) in population projections and estimations. Numerous studies have demonstrated that the relational parameterized models consisting of a model standard schedule and a few summary parameters offer an efficient and realistic way to project or estimate demographic age-specific rates (Booth 1984; Brass 1978; Paget and Timaeus 1994; Zeng et al. 1994). The theoretical foundation of applications of the model life tables and the other model standard schedules is that the demographic summary parameters are crucial for determining changes in level and age pattern of the age-specific rates that affect the projections or estimations. At the same time, the projection and estimation results themselves are typically not highly sensitive to the race-/sex-/age-specific model standard schedules as long as the possible changes in the general shape of the standard schedules and timing of the demographic events are properly modeled by the relevant summary parameters (e.g., mean or median age, interquartile range). Two tests in our previous publications (Zeng et al. 2000, 2006) have corroborated the empirical applicability of Core Idea 4 of the ProFamy model, which employs national model standard schedules and summary parameters at the subnational level to specify projected demographic rates of the subnational regions in future years. A relatively detailed summary of these two tests is presented as Online Resource 1 (Appendix A).

Data and Estimation Issues

For household and living arrangement projections at the national and subnational levels employing the ProFamy model, a census microdata file that contains the variables of sex, race (optional), age, marital/union status, relationship to the householder, and whether living in a private or institutional household is required (see panel 1 in Table 1). Normally, the model standard schedules of fertility, mortality, and marriage/union formation and dissolution and international migration (see panel 2-I in Table 1) need to be estimated at the national level only, and they can be employed for projections at subnational level. The age-/sex-specific rates of domestic in-migration and out-migration at the subnational level can be estimated based on census or large survey microdata files (see panel 2-II in Table 1). Projected (or assumed) demographic summary parameters (e.g., the total fertility rate (TFR); life expectancy at birth (e_0); general rates of marriage, divorce, cohabitation, and union dissolution; total number of migrants; and mean age at first marriage and birth in the future years) are needed for projections at both national and subnational levels (see panel 3 in Table 1). In sum, using existing national model standard schedules and the ProFamy model, household and living arrangements projection at the subnational level require a census microdata file and the projected (or assumed) demographic summary parameters for the future years.

Estimation of the required demographic summary parameters—the TFR, life expectancy at birth, total number of migrants, and mean age at first marriage and

birth—is straightforward. Definitions and methods of estimation and standardization of general rates of marriage, divorce, cohabitating, and union dissolution are given in Zeng et al. (2006:6); the main points are summarized in Appendix B in Online Resource 1. As an illustration of the application, we also present in Appendix B the procedures for estimating the U.S. race-specific general rates of marriages, divorce, cohabitation, and union dissolution at the state level.

Empirical Assessments

Comparisons of Projections of Household and Living Arrangements and Census Enumerations in 2000 for Each of the 50 States and Washington, DC

A useful validation exercise for a demographic projection model is to project between two past dates for which the observations are known, and then compare the observed with the projected. To test whether the ProFamy extended cohort-component method and software work reasonably well at the subnational level, we conducted a set of empirical assessments of validation tests of household and living arrangement projections for each of the 50 states and Washington, DC,⁹ all using the national race-/sex-/age-specific model standard schedules estimated based on the pooled national surveys data,¹⁰ except that the race-/age-/sex-specific domestic migration rates are estimated based on the census 5 % microdata files for each of the 50 states and Washington, DC. The tests project from 1990 to 2000 using the 1990 census data as base population and summary parameters based on data before 1991 and then compare the projections with the census observations in 2000. These tests assume that we have no data after 1990 when projecting 1990 to 2000 and assess the accuracy of the projections using the ProFamy model in the real world (assuming that the 2000 census data are accurate) at the subnational level.

We use the percentage error (PE), mean absolute percentage error (MAPE), mean algebraic percentage error (MALPE), and median absolute percentage error (MEDAPE), which are the most commonly used measures of forecast errors (Smith et al. 2001:302–304), to assess the validity of the household and living arrangement projections at subnational levels using the ProFamy approach. More specifically, the PE is defined as the difference between the ProFamy projections in 2000 and the census observations in 2000, divided by the census observations in 2000 and multiplied by 100 for each of the 50 states and Washington, DC. The MAPE and MEDAPE are, respectively, the average and median of the absolute values of PEs across all states and Washington DC, and

⁹ The numerical projections reported in this article were calculated with the ProFamy computer software program, which contains a demographic database of the U.S. age-specific schedules of demographic rates to assist users in making projections. The ProFamy software for household and living arrangements projections can be downloaded (<http://www.profamy.com/>).

¹⁰ National Survey of Family Households (NSFH) conducted in 1987–1988, 1992–1994, and 2002; National Survey of Family Growth (NSFG) conducted in 1983, 1988, 1995, and 2002; Current Population Surveys (CPS) conducted in 1980, 1985, 1990, and 1995; Survey of Income and Program Participation (SIPP) conducted in 1996. (See Zeng et al. (2012) for discussions on justifications of pooling data from the four surveys.)

MALPE is the algebraic mean of PEs (in which positive and negative values offset each other) across all states and Washington, DC.

Figure 3 and Table 2 summarize the forecast errors based on comparisons of the total number of households; average household size; percentage of households with one, two to three, and four or more persons and couple-households; total population size; percentages of children, elderly (aged 65 and older), and oldest-old (aged 80 and older); and dependency ratios between the projections and the observations in the 2000 census for all states and Washington, DC. Among the set of tests of the 306 main indices¹¹ on household and living arrangements of the projections and the 2000 census observations in all 50 states and Washington, DC, 29.1 %, 33.9 %, 17.4 %, 12.9 %, and 6.7 % of the forecast errors are <1.0 %, 1.0 % to 2.99 %, 3.0 % to 4.99 %, 5.0 % to 9.99 %, and ≥10 %, respectively (Fig. 3, panel a). The percentage distributions of the forecast errors of the main indices of population in the set of the tests comparing the projected with the 2000 census observations in all states and Washington, DC, are 29.7 %, 43.4 %, 16.5 %, 9.5 %, and 0.8 % are <1 %, 1.0 % to 2.99 %, 3.0 % to 4.99 %, 5.0 % to 9.99 %, and ≥10.0 %, respectively (see Fig. 3b).

The MAPE and MEDAPE of the main household indices in comparisons between the projections and census observations in 2000 for all states and Washington, DC are all within reasonably small ranges of 1.6 % ~ 4.7 % and 1.1 % ~ 3.5 % (see the second and third columns of panel A in Table 2). The MALPE of average household size and percentage of 2 to 3 persons household are negative—at -0.56 % and -1.06, respectively—and all other MALPEs for household projections are positive, within a range of 0.04 % ~ 2.91 % (see the fourth column of panel A in Table 2). Similar to those error rates of the main indices of household projections, the ranges of all forecast errors of the main population indices comparing projections and census observations in 2000 for all states and Washington, DC, are all reasonably small (see columns 2–4 of Table 2, panel B). No significant associations between the forecast errors and population sizes of the states were found. This is similar to what was found in some other projections (ESRI 2007).

Note that there are no fixed guidelines for the evaluation of population forecast accuracy, but we may compare ours with the others. Our household and population forecast errors from 1990 to 2000 at the subnational level are close to or even smaller than those population forecast errors by the U.S. Census Bureau (Campbell 2002) and some other institutions (e.g., ESRI 2007). According to previous studies, it is fairly common for some countries in the United Nations population projections to have 2 % to 5 % forecast errors for total population and 5 % to 10 % forecast errors for age-specific subpopulations in a 10-year projection period (e.g., Khan and Lutz 2008). These prior forecasts and forecast evaluations provide the framework within which it can be stated that the validation tests results summarized in Fig. 3 and Table 2 show that the forecast errors of household and population projections at the subnational level using the ProFamy

¹¹ We compare six main indices of household projections and six main indices of population projections for each of the 50 states and Washington, DC, and thus both of the number of household indices and the total number of population indices under comparisons are 306.

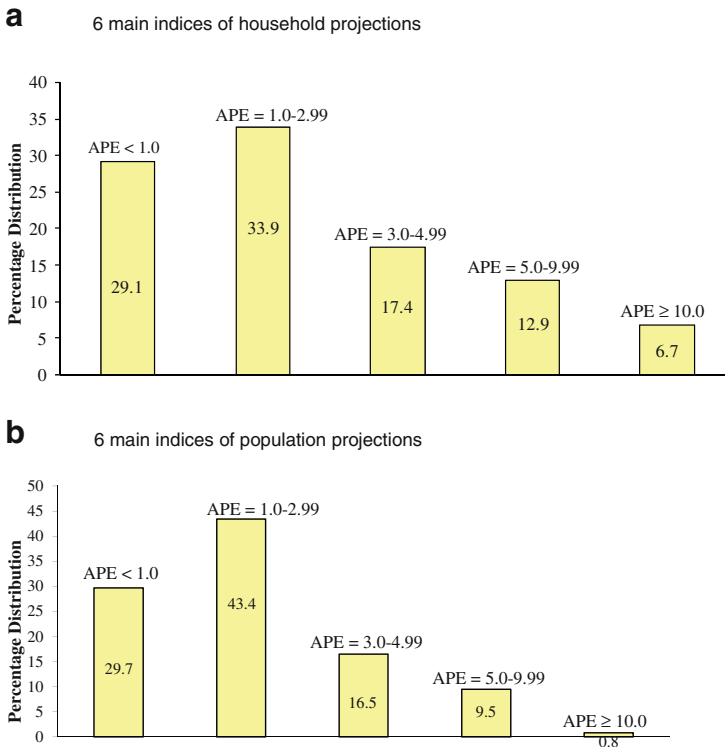


Fig. 3 Percentage distributions of the mean absolute percent errors (APE) of comparisons between the ProFamy projections from 1990 to 2000 (using the data before 1991) and the census observations in 2000, six main indices of households (Fig. 3a) and six main indices of population (Fig. 3b), for each of the 50 states and Washington, DC, based on 306 ($= 6 \times 51$) pairs of comparisons between forecasts and the observations in Figs. 3a and b, respectively

extended cohort-component method are within a reasonably and relatively small range.¹² It is uncertain what portions of the errors are due to the model specification versus to inaccuracies of the data. It is clear, however, that the ProFamy extended cohort-component approach for simultaneously projecting households, living arrangements, and population age/sex distributions work reasonably well not only at the national level, as shown in previous publications, but also at the subnational level.

¹² We performed another set of the tests of projections from 2000 onward using ProFamy approach and data prior to 2001 and comparing the projections and the American Community Survey (ACS) observations in 2006 for each of the 50 states and Washington, DC. It turns out that 34.2 %, 35.0 %, 21.9 %, and 9.0 % of the percentage errors of the 306 indices of the household projections are <1.0 %, 1.0 % to 2.99 %, 3.0 % to 4.99 %, and 5.0 % to 9.99 %, respectively, and none is more than 10 %. A similar scale and pattern of forecast errors were also found in tests of projections from 2000 onward using ProFamy approach and data prior to 2001 and comparing the projected and ACS observations in 2006 and 2009 for the six SC counties and the M-S area (Wang 2009a,b, 2011a,b). We did not present detailed results from these additional tests here (they are available upon request), mainly because the 2006 and 2009 ACS data may not be accurate enough to serve as a benchmark standard for the validation tests (Alexander et al. 2010; Swanson 2010).

Table 2 Forecasting errors based on comparisons between the ProFamy projections from 1990 to 2000 (based on data before 1991) and the census observations in 2000 across the 50 states and Washington, DC

	MAPE	MEDAPE	MALPE
A. Main Indices of Household Projection			
Total number of households	1.63	1.07	0.04
Average household size	1.75	1.16	-0.56
% of 1-person households	4.73	3.45	2.91
% of 2- to 3-person households	2.63	2.2	-1.06
% of 4+-persons households	4.08	2.76	0.06
% of couple households	2.07	1.35	0.56
B. Main Indices of Population Projection			
Total population size	1.35	1.09	-0.58
% of children (aged ≤18)	1.96	1.82	1.39
% of elderly (aged 65+)	2.52	2.08	-1.45
% of oldest-old (aged 80+)	3.44	2.81	-2.02
Dependency ratio of children	2.55	2.21	1.56
Dependency ratio of elderly	2.8	2.19	-1.05

Comparisons of Housing-Demand Forecast Errors Between the Headship-Rate Method and the ProFamy Extended Cohort-Component Approach

As discussed earlier, compared with the still widely used classic headship-rate method, the ProFamy approach is theoretically advantaged and projects much more detailed household types/sizes and living arrangements. However, the ProFamy approach needs substantially more data than does the classic headship-rate method. This raises the question, Is it worthwhile to employ the ProFamy approach rather than the classic headship-rate method if users simply need projections of the home-based consumption demands, such as numbers of housing units by number of bedrooms, but do not care about the details of the household characteristics, including household types and sizes, marital/union status, coresidence status with parents, and children of the reference persons? The following assessments are designed to answer this question.

We projected from 1990 to 2000 the numbers of housing units by number of bedrooms for each of the 50 states and Washington, DC, using the constant headship-rate approach (see Online Resource 1, Section C-1) and the ProFamy extended cohort-component approach with data before 1991 (see Online Resource 1, Section C-2). By comparing the projected and census-observed numbers of housing units occupied by private households in 2000, we estimated the error rates of forecasts of housing units by number of bedrooms. The error rates were estimated to evaluate the projections produced by employing the headship-rate method and the ProFamy approach, respectively. As shown in Table 3, the MALPE of forecasts based on the constant headship rates for the zero-/one-bedroom,¹³ two-bedroom, three-bedroom,

¹³ The zero-bedroom housing unit term means that the bedroom is mixed with the living room.

and four-bedroom housing units are -18.67% , 5.01% , 4.30% , and -3.23% , respectively; this contrasts with -6.25% , 2.51% , 1.38% , and 1.15% , respectively, based on the ProFamy approach. The MAPE and MEDAPE of forecasting based on the constant headship rates for the zero-/one-bedroom, two-bedroom, three-bedroom, and four-bedroom housing units are, respectively, about 114.4% to 128.4% , 20.4% to 37.0% , 13.2% to 27.9% , and 24.4% to 53.4% higher than those based on the ProFamy approach.

Even if one uses the changing headship rates based on regression or another trend extrapolation method to correctly project numbers of households, it is still possible that the headship rates may result in biased projections of household consumption demands, which largely depend on household size (Myers et al. 2002) because the headship-rate method excludes household size. To test this hypothesis, we conducted another assessment in which the changing headship rates are assumed to produce the same numbers of households as those observed in the 2000 census in each of the 50 states and Washington, DC (for details, see Online Resource 1, Sections C3–C4). We estimated average forecast errors by comparisons between the 2000 census observations and the adjusted projections of the housing units in 2000 based on the changing headship-rate method and the ProFamy approach, respectively, across all states and Washington, DC; see Table 4.

The empirical assessments show that after making the adjustments described earlier herein, the average negative-forecast error of zero-/one-bedroom housing units by the headship-rate method is reduced by 3.3 percentage points, but the forecast error is still substantially larger than that of the ProFamy approach. More specifically, the forecast errors of zero-/one-bedroom housing units measured by MALPE, MAPE, and MEDAPE are, respectively, -15.35% , 15.45% , and 16.50% for the headship rate, in contrast with -6.37% , 8.24% , and 7.73% for the ProFamy approach (see Table 4). The forecast error rates of MALPE for two-bedroom and three-bedroom housing units by the headship-rate method are 5.96% and 3.71% , respectively, in contrast with 2.46% and 1.41% by the ProFamy approach. On the algebraic average criterion, the headship-rate method downwardly projected four-bedroom housing units by -4.23% , while the error rate for four-bedroom housing units by the ProFamy approach is 1.22% . The forecast errors listed in Table 4 show that, compared with the ProFamy approach, the headship-rate method produced substantially more serious negative forecast errors for the zero-/one-bedroom and four-bedroom units, and positive forecast errors for the two-bedroom and three-bedroom housing units.

Decennial census data facilitate understanding and interpretation of these results. Compared with 1990, the one-person, two-person, three-person, four-/five-person, and six-or-more-person households in 2000 increased by 20.6% , 16.9% , 9.2% , 9.3% , and 15.1% , respectively. Clearly, American households with one person (which likely need one bedroom), two persons (which more likely need two bedrooms), and six or more persons (which more likely need four bedrooms)¹⁴ increased substantially faster during this decade than

¹⁴ Our research indicates that the increase in proportion of American households with six or more persons in 2000 compared with 1990 is due to the changing racial composition of the population, given that Hispanic, Asian, and other nonwhite and nonblack minority groups have higher proportions of large households with six or more persons and are growing substantially faster.

Table 3 Forecasting errors of housing demands projections from 1990 to 2000 (compared with the 2000 census observations) produced by the ProFamy cohort-component approach and the constant headship rates

	Zero-/One-Bedroom Units			Two-Bedroom Units			Three-Bedroom Units			Four-Bedroom Units		
	ProFamy	Headship Constant	% Diff.	ProFamy	Headship Constant	% Diff.	ProFamy	Headship Constant	% Diff.	ProFamy	Headship Constant	% Diff.
MALPE	-6.25	-18.67	198.7	2.51	5.01	99.6	1.38	4.3	211.6	1.15	-3.23	-380.9
MAPE	8.71	18.67	114.4	5.87	7.07	20.4	3.94	5.04	27.9	6.52	8.11	24.4
MEDAPE	8.11	18.52	128.4	4.86	6.66	37.0	3.10	3.51	13.2	5.19	7.96	53.4

Table 4 Forecasting errors of housing-demand projections from 1990 to 2000 (compared with the 2000 census observations) produced by the ProFamy cohort-component approach and the adjusted changing headship rates, both resulting in the same projected total number of households as that observed in the 2000 census

	Zero-/One-Bedroom Units			Two-Bedroom Units			Three-Bedroom Units			Four-Bedroom Units		
	ProFamy	Headship Changing	% Diff.	ProFamy	Headship Changing	% Diff.	ProFamy	Headship Changing	% Diff.	ProFamy	Headship Changing	% Diff.
MALPE	-6.37	-15.35	141.0	2.46	5.96	142.3	1.41	3.71	163.1	1.22	-4.23	-446.7
MAPE	8.24	15.45	87.5	5.48	7.53	37.4	4.23	4.37	3.3	6.97	8.27	18.7
MEDAPE	7.73	16.50	113.5	5.06	7.88	55.7	2.80	3.03	8.2	4.57	7.64	67.2

the three-person and four-/five-person households (which more likely need two or three bedrooms). Consequently, the headship-rate method, which does not project household size, resulted in substantially more serious negative forecast errors for the zero-/one-bedroom and four-bedroom units and upward forecast errors for the two-bedroom and three-bedroom housing units, as compared with the ProFamy approach, which projects detailed household size. This is consistent with what Prskawetz et al. (2004) found for the Austria vehicle projections using the ProFamy approach compared with the headship-rate method.

Illustrative Applications at the Subnational Level

To further illustrate the potential of the ProFamy extended cohort-component method, we conducted household and living arrangement projections from 2000 to 2050 for each of the 50 states and Washington, DC, each of the six counties of SC, and the M-S area. Because of space limitations and given the nature of illustrative applications, we present only the main results of the projections.

Data and Parameter Assumptions

The data sources for the projections are listed in the last column of Table 1. As discussed in Core Idea 4 earlier, we apply model standard schedules of race-/sex-/age-specific demographic rates (except domestic migration rates) estimated based on national data sets for household and living arrangement projections for each state; Washington, DC; SC counties; and the M-S area. Based on the 2000 census 5 % microdata, we estimated race-/sex-/age-specific probabilities of domestic out-migration from each state; Washington, DC; SC counties; and the M-S area to the rest of the country. We also estimated and race-/sex-/age-specific frequencies of in-migration from the rest of the country to each state; Washington, DC; SC counties; and the M-S area (see Appendix D in Online Resource 1 for details).

The race-/sex-specific life expectancies at birth and the race-/parity-specific TFRs from 2000 to 2050 for each state; Washington, DC; SC counties; and the M-S area in the baseline and future years were estimated/projected based on the regional data in reference to the medium assumptions of the Census Bureau population projections (Hollmann et al. 2000; U.S. Census Bureau 2008). The numbers of domestic in-migrants and out-migrants as well as the international net migrants for each state; Washington, DC; SC counties; and the M-S area are estimated based on the combined data from the ACS from 2000 to 2006; the migration parameters are assumed to be constant after 2006.

The procedures for estimating the general rates of marriage/union formation and dissolution in 2000 at subnational level are discussed in the third section of this article and Online Resource 1, Appendix B. Instead of constant assumptions, time-dependent changes were specified for some of the parameters for the period of 2000–2010 so that the projected values for 2010 were consistent with the 2010 census' corresponding main results. These parameters include the race-specific general rates of marriage/union formation and dissolution, race-/age-/sex-specific proportion of persons who live in group quarters (PGQ), race-/sex-specific proportion of

those aged 45–49 who do not live with parents (PNP), race-/household size-specific average number of other relatives (other than spouse/partner, parents, and children) and nonrelatives living in the same household (ARNR). This is similar to the practice adopted by other demographic projections in which an earlier census year is the starting point of the projection, the most recent census year is within the projection period, and the main results of the most recent census are published, but its detailed microdata to derive the base population of the projections are not yet available (U.S. Census Bureau 2008). The race-specific general rates of marriage/union formation and dissolutions, PGQ, PNP, and ARNR from 2010 to 2050 are simply assumed to be constant at the 2010 level in our medium projection.¹⁵

Four race groups (non-Hispanic white, non-Hispanic black, Hispanic, and non-Hispanic Asian and others) defined by the Census Bureau are distinguished in 14 states in which each of the four racial groups has a sufficiently large population size for the projection. Because of the small population sizes of the minority groups, three racial groups (non-Hispanic white, non-Hispanic black, and Hispanic plus Asian and others) are distinguished in 13 states, and two racial groups (non-Hispanic white, and all other races combined) are distinguished in another 13 states. Only one race group (white and others) is distinguished in 11 states because the population size of all nonwhite groups combined in these states is not sufficiently large for distinct projections.

Low and High Bounds of Household and Living Arrangement Projections

To explore the possible low and high bounds of household and living arrangement projections, we examined small- and large-family scenarios. The small-family scenario assumes that, compared with the medium projections, the general rate of divorce and general rate of cohabitation union dissolution are 15 % higher in 2025 and 25 % higher in 2050, and the general rates of marriage and of cohabitation are 15 % lower in 2025 and 25 % lower in 2050. To obtain the TFR, e_0 , and number of international migrants in 2025 and 2050 for the small-family scenario, we multiplied our medium fertility (TFR), medium mortality (e_0), and medium international net migration in 2025 and 2050 by the ratios of the low TFR, low mortality (high e_0), and low international net migration to the corresponding medium variants adopted by the Census Bureau's population projection in 2000 (Hollmann et al. 2000). This small-family scenario assumes increasing marriage/union dissolution, decreasing marriage/union formation, decreasing fertility and mortality,¹⁶ and receipt of fewer international immigrants. We expect that such a combination of demographic rates results in low bounds of household size and

¹⁵ One common approach in population projection is to hold some of the current demographic rates constant throughout the projection horizon (e.g., Day 1996; Treadway 1997). Smith et al. (2001:83–84) argued that neither the direction nor the magnitude of future changes can be predicted accurately, and thus if upward or downward movements are more or less equally likely, the constant demographic rates provide a reasonable forecast of future rates.

¹⁶ Low mortality may (1) reduce the U.S. average household size through increasing number of elderly households that are mostly small (one or two persons) and (2) increase the size of some households by increasing the survivorship of adults and children in these larger households. The effects of the latter may be smaller than those of the former because a further decrease in adult and child mortality in the United States is limited, but the prolongation of elderly life span may have larger effects.

percentages of married- or cohabiting-couple households, and high bounds of percentages of one-person households, single-parent households, and so on.

The large-family scenario assumes that, compared with the medium projections, the general rates of divorce and of cohabiting-union dissolution are 15 % lower in 2025 and 25 % lower in 2050, and the general rates of marriage and of cohabitation are 15 % higher in 2025 and 25 % higher in 2050. To obtain the TFR, e_0 , and number of international migrants in 2025 and 2050 for the large-family scenario, we multiplied our medium fertility (TFR), medium mortality (e_0), and medium international net migration in 2025 and 2050 by the ratios of the high TFR, high mortality (low e_0), and high international net migration to the corresponding medium variants adopted by the Census Bureau's population projection in 2000 (Hollmann et al. 2000). This large-family scenario assumes that the family will regain its traditional values with decreasing marriage/union dissolution, increasing marriage/union formation, and increasing fertility, accompanied by a larger number of international immigrants and relatively higher mortality. The combination of demographic rates in the large-family scenario results in high bounds of household size and percentages of married- or cohabiting-couple households, and low bounds of percentages of one-person households, single-parent households, and so on.

The general rates of marriages, divorces, cohabitations and union dissolutions, TFR, e_0 , and number of international migrants for all individual years between 2010, 2025, and 2050 in all scenarios are linearly interpolated. The assumptions that there will be 15 % and 25 % increases (or decreases) in general rates of marriage, divorce, cohabitation, and union dissolution in 2025 and 2050, respectively, constitute our educated guesses about the largest possible changes in marriage/union formation and dissolution in the next few decades. Although we made these guesses with reference to the available time series data of the general rates, they are largely arbitrary because of uncertainties about future trends. Nevertheless, similar to conventional deterministic population projections of low and high variants that formulate possible bounds of population growth, our small- and large-family scenarios formulate possible low and high bounds of future household and living arrangements distributions.¹⁷

A Summary of the Projection Outcomes

The relatively detailed numerical outcome of the medium projections of the main indices in 2010, 2020, 2030, 2040, and 2050 and the low and high bounds after 2010 for each of the 50 states; Washington, DC; the six SC counties; and the M-S area are presented in the tables in Online Resource 2. These tables also include the available census observations for comparing the projected and observed in 2010. We summarize here the insights from the main indices of the decennial projections but cannot present the details because of space limitations.

¹⁷ The race-/sex-specific demographic parameters (TFR is parity-specific) (see parameters (a–h in Table 1, panel 3) in the medium-, small-, and large-family scenarios in selected years from 2000 to 2050 for each of the 50 states; Washington, DC; each SC county; and the M-S area can be listed in one large table. Including them in this article would require an unfeasibly large number of pages.

As shown in Table S1 of Online Resource 2, the average household size would decrease moderately and pervasively in almost all states and Washington, DC, in the 2000–2020 period from an overall mean of 2.58 for the whole country in 2000 to 2.52 (SI: 2.48~2.55) in 2020; here and throughout, SI refers to the scenarios interval of the low and high bounds, and the number preceding the SI is the medium variant. The average household size may continue to decline after 2020 in 46 states but at a slightly slower rate compared with the period of 2000–2020, from an average of 2.44 (SI: 2.40~2.48) in 2020 to an average of 2.36 (SI: 2.14~2.61) in 2050 in these states. In the other four states (California, Colorado, Maryland, and New Jersey) and Washington, DC, the trend after 2020 is different: the average household size may slightly increase from 2.54 (SI: 2.50~2.57) in 2020 to 2.60 (SI: 2.36~2.86) in 2050.

The proportion of one-person households would increase substantially in almost all states in the first two decades. Compared with 2000, the proportion of one-person households in 2020 would increase by <5.0 % in 8 states, 5.0 % to 9.99 % in 17 states, 10.0 % to 14.99 % in 22 states, and ≥ 15.0 % in four states. However, during the period 2020–2050, the trend of increase in the proportion of one-person households would slow down considerably in 44 states, from 0.275 (SI: 0.265~0.284) in 2020 to 0.301 (SI: 0.238~0.361). The proportion of one-person households in 2020–2050 in six states (California, Maryland, New Jersey, Arizona, Hawaii, and Nevada) would decline slightly from 0.275 (SI: 0.265~0.284) in 2020 to 0.266 (SI: 0.210~0.317) in 2050 (see Table S2, Online Resource 2). The pattern of changes in average proportions of one-person households in the six SC counties in the first half of this century is totally different than the overall trends in most of the states: it declines from 0.227 in 2000 to 0.211 (SI: 0.204~0.217) in 2020 and 0.187 (SI: 0.157~0.221) in 2050 (see the eighth line from the bottom of Table S2).

The projected declines in the proportion of the one-person households in the six SC counties during the entire period of the first half of this century after 2020 are likely due to the large racial differences and changes in the racial compositions. Although the proportions of one-person households will increase substantially among all racial groups in the next few decades, the Hispanic population—which has the lowest proportion of one-person households and largest average household size—would compose a substantially higher percentage of the total population in the future in the six SC counties. For example, the proportion of one-person households of the Hispanic group was 0.099 in 2000 in the six SC counties, but the corresponding figures for the non-Hispanic white, non-Hispanic black, and non-Hispanic Asian and others groups were 0.30, 0.288, and 0.19, respectively. The Hispanic group, which made up 40.6 % of the total population in 2000, will be the majority in 2020 (53.5 %) and 2050 (65.8 %) in the six SC counties. Consequently, such racial composition changes would result in the decline of proportion of one-person households for all races combined.

Husband-wife households will decrease moderately in almost all states (with a few exceptions of slight increases). Under the medium scenario, compared with 2000, the proportion of married-couple households among all households in 2020 will decrease by <5.0 % in 6 states, 5.0 % to 9.99 % in 15 states, 10.0 % to 14.99 % in 14 states, and ≥ 15.0 % in 13 states. The decrease in the proportion of married-couple households will slow down considerably in the period 2020–2050: the proportion in 2050 would decrease by <5.0 % in 14 states, 5.0 % to 9.99 % in 19 states, 10.0 % to

14.99 % in 9 states, and ≥ 15.0 % in 6 states (see Table S3, Online Resource 2). The proportion of cohabiting-couple households among all households would increase dramatically in the first two decades of this century: the proportion in 2020 would be higher than that in 2000 by < 10 % in 3 states, 10.0 % to 29.0 % in 7 states, 30.0 % to 49.9 % in 11 states, 50.0 % to 69.9 % in 18 states, and 70.0 % to 89.9 % in 11 states under the medium scenario (see Table S4). The proportion of cohabiting households would remain relatively stable after 2020 in almost all states.

Directions of changes in the percentage of single-parent households in the first half of this century are diversified, increasing moderately in some states but decreasing moderately or remain more or less unchanged in the other states. The average percentage of single-parent households across all states and DC was 30.9 in 2000; it will be 30.8 (SI: 29.3~32.2) in 2020 and 33.1 (SI: 24.7~44.4) in 2050 (see Table S5, Online Resource 2). Such patterns may be explained by the opposite effects of moderate declines in marriages and substantial increases in cohabitation, plus the stable divorce and union dissolution rates.

The aging trends shown in the results of household and living arrangement projections presented in Tables S5–S8 are striking. Under the medium scenario, compared with 2000, the proportion of elderly households (with householder ages 65 and older) in 2020 would increase by < 10 % in 11 states, 10 % to 19.9 % in 17 states, 20 % to 29.9 % in 15 states, 30 % to 39.9 % in 4 states, and ≥ 40 % in 3 states. During the period 2020–2050, households aging will further accelerate. More specifically, compared with 2020, the proportion of elderly households in 2050 will increase by < 20.0 % in 2 states, 20 % to 29.9 % in 24 states, 30 % to 39.9 % in 20 states, and ≥ 40 % in 4 states; Washington, DC, which attracts a lot of young in-migrants, is the exception (see Table S6). Compared with 2000, elderly households will have slightly more than doubled in Hawaii and New Hampshire and nearly tripled in Alaska by the middle of this century. Similar to the general pattern of increase in elderly households, the proportion of elderly aged 65+ and older living alone will increase dramatically and pervasively across all states (see Table S7). Table S8 demonstrates that the oldest-old (aged 80 and older) living alone will increase even more dramatically in the next a few decades across all states. The average percentage of the oldest-old living alone across all states and Washington, DC, will be 1.48 (SI: 1.44~1.51) in 2020 and 2.41 (SI: 1.85~2.96) in 2050, representing a 23.8 % increase in 2020 and slightly more than a doubling in 2050 compared with 2000, under the medium assumption. Under the medium scenario, between 2000 and 2050, the percentage of oldest-old living alone will increase by 44.5 % to 79.9 % in 7 states and by 80 % to 99.9 % in 14 states; will more than double (less than triple) in 24 states; and will more than triple in 5 states (Louisiana, South Carolina, Hawaii, New Hampshire, and Alaska).

Discussion and Concluding Remarks

Applying the ProFamy extended cohort-component method and national model standard schedules of age-/sex-/race-specific demographic rates based on the commonly available survey and census data, we have demonstrated that comprehensive and simultaneous projections of household, living arrangements, and

population at the subnational level requires using a census microdata file and the projected (or assumed) demographic summary parameters. We conducted validation tests of household, living arrangements, and population projections from 1990 to 2000, and compared the main indices of the projections and the observed data in 2000 for each of the 50 U.S. states and Washington, DC. The results show that the extended cohort-component approach works well at the subnational level, with most absolute forecast error rates less than 3 % of observed values.

Our empirical assessments of the average forecast errors of demands of housing units by number of bedrooms through projections from 1990 to 2000 and comparisons of the projections with the census observations in 2000 for each of the 50 states and Washington, DC, have demonstrate that compared with the headship-rate method, the extended cohort-component approach results in substantially smaller forecast errors. This is because there will be more smaller households in the future in the United States and many other countries, but the headship-rate method cannot project households by size, whereas the ProFamy approach projects detailed household size information. This advantage of using the extended cohort-component approach over the classic headship-rate method is also applicable to other kinds of projections of household consumption demands—such as vehicles (Feng et al. 2011; Prskawetz et al. 2004), energy use (Dalton et al. 2008), and other home-based products—which largely depend on household size.

For the purpose of illustrative applications, we calculated household and living arrangement projections from 2000 to 2050 with medium-, small-, and large-family scenarios for each of the 50 states; Washington, DC; six SC counties; and the M-S area. Among many interesting numerical outcomes of household and living arrangement projections with medium, low, and high bounds, the aging of American households over the next few decades across all state/areas is particularly striking. To our knowledge, these are the first comprehensive household and living arrangement projections by race, age of the householders, and various household types and sizes using conventional demographic rates as input for each of the state; Washington, DC; and some counties/areas in the United States.

Note that the ProFamy approach requires much more work for preparing the input data than does the headship-rate method. The choice between methods with different degrees of comprehensiveness and work amount depends on the user's needs. For analyses of socioeconomic planning, home-based consumption/service market study, academic simulations, or policy scenarios—which need detailed projections of household types, sizes, and living arrangements using demographic rates as input—using the ProFamy approach is preferable. If a simple and quick projection of the number of households with no need for detailed household types and size information is sufficient for the purpose of a projection and no requests for home-based consumption/service forecasts, the headship-rate approach—which requests fewer data, with a lesser amount of work at a low cost—may be a rational choice.

Limitations of the present study and potential for further investigations should also be noted. First, our projections of the demographic summary parameters are based on trend extrapolations and expert opinions. Thus far, we have not included other socioeconomic factors relevant to changes in demographic parameters, and this can be done in future research. Second, we discussed the main results of all races

combined because of space limitations, and analysis of many other detailed state-/race-specific projection outcomes may be an interesting topic for the further research. Third, the ProFamy model cannot be directly applied to project households and living arrangements for small areas, which do not have adequate data to estimate the needed demographic parameters. However, it is possible to project household and living arrangements for small areas, employing the well-established ratio method and the projection of the small area's parental region (a state or province or large county/city) produced by the ProFamy approach. The ratio method is frequently used for population projections of small areas because its data requirement are minimal, it is easy to apply, and the accuracy of its projections often is reasonably acceptable (Rao 2003; Smith 2003; Smith and Morrison 2005; Smith et al. 2001). In applications of the ratio method combined with the ProFamy approach, household and living arrangement projections for the parental region must be done first. An analyst then can calculate the proportions of the indices of households and living arrangements of the small area in the parental region in the baseline census year. Assuming that the proportions are constant (constant-share) or changing (shift-share) for the projection period, the projected indices of the parental region produced by the ProFamy approach then can be multiplied by the proportions to derive the relatively detailed household and living arrangement projections for a small area.¹⁸

Acknowledgments The research reported in this article was mainly supported by NIA/NIH SBIR Phase I and Phase II project grants. We also thank the Population Division of U.S. Census Bureau, NICHD (grant 5 R01 HD41042-03), NIA (grant 1R03AG18647-1A1) and NSFC international collaboration project (grant 71110107025), Duke University, Peking University, and the Max Planck Institute for Demographic Research for supporting related basic and applied research. We thank Huashuai Chen for preparing the graphics.

References

- Alexander, J. T., Davern, M., & Stevenson, B. (2010, April). *Inaccurate age and sex data in the census PUMS file*. Paper presented at the annual meeting of the Population Association of America, Dallas, TX.
- Bell, M., & Cooper, J. (1990, November). *Household forecasting: Replacing the headship rate model*. Paper presented at the Fifth National Conference, Australian Population Association, Melbourne.
- Bongaarts, J. (1987). The projection of family composition over the life course with family status life tables. In J. Bongaarts (Ed.), *Family demography: Methods and applications* (pp. 189–212). Oxford, UK: Clarendon Press.
- Booth, H. (1984). Transforming Gompertz's function for fertility analysis: The development of a standard for the relational Gompertz function. *Population Studies*, *38*, 495–506.
- Brass, W. (1974). Perspectives in population prediction: Illustrated by the statistics of England and Wales. *Journal of the Royal Statistical Society*, *137*, 532–583. Series A.
- Brass, W. (1978). *The relational Gompertz model of fertility by age of women*. Unpublished manuscript, London School of Hygiene and Tropical Medicine.

¹⁸ Zeng et al. (2010) preliminarily assessed the projection accuracy of the combined approach using the ratio method and the ProFamy approach by calculating projections from 1990 to 2000 and comparing projections with census-observed counts in 2000 for sets of randomly selected 25 counties and 25 cities that are more or less evenly distributed across the United States. The comparisons show that most forecast errors are reasonably small, at less than 5 %.

- Brass, W. (1983). *The formal demography of the family: An overview of the proximate determinates* (Occasional Paper 31). London, UK: Office of Population census and Surveys.
- Campbell, P. R. (2002). *Evaluating forecast error in state population projections using Census 2000 counts* (U.S. Census Bureau Population Division Working Paper Series No. 57). Washington, DC: U.S. Census Bureau. Retrieved from <http://www.census.gov/population/www/documentation/twps0057/twps0057.html>
- Coale, A. J. (1984). Life table construction on the basis of two enumerations of a closed population. *Population Index*, 50, 193–213.
- Coale, A. J. (1985). An extension and simplification of a new synthesis of age structure and growth. *Asian and Pacific Forum*, 12, 5–8.
- Coale, A. J., Demeny, P., & Vaughn, B. (1983). *Regional model life tables and stable populations*. New York: Academic Press.
- Coale, A., & Trussell, T. J. (1974). Model fertility schedules: Variations in the age structure of childbearing in human population. *Population Index*, 40, 185–258.
- Crowley, F. D. (2004). *Pikes Peak Area Council of Governments: Small area estimates and projections 2000 through 2030*. Southern Colorado Economic Forum, University of Colorado at Colorado Springs. Retrieved from <http://www.ppacg.org/Trans/2030/Volume%20III/Appendix%20G%20-%20Small%20Area%20Forecasts.pdf>
- Dalton, M., O'Neill, B., Prskawetz, A., Jiang, L., & Pitkin, J. (2008). Population aging and future carbon emissions in the United States. *Energy Economics*, 30, 642–675.
- Day, J. (1996). *Population projections of the United States by age, sex, race, and Hispanic origin: 1995 to 2050* (Current Population Reports P25-1130). Washington, DC: U.S. Census Bureau.
- ESRI. (2007). Evaluating population projections—The importance of accurate forecasting (ESRI White Paper). Retrieved from <http://www.esri.com/library/whitepapers/pdfs/evaluating-population.pdf>
- Federal Interagency Forum on Aging Related Statistics (FIFARS). (2004). *Older Americans 2004: Key indicators of well-being*. Washington, DC: U.S. Government Printing Office.
- Feng, Q., Wang, Z., Gu, D., & Zeng, Y. (2011). Household vehicle consumption forecasts in the United States, 2000 to 2025. *International Journal of Market Research*, 53(5).
- Freedman, V. A. (1996). Family structure and the risk of nursing home admission. *Journal of Gerontology*, 51B(2), S61–S69.
- Goldscheider, F. K. (1990). The aging of the gender revolution: What do we know and what do we need to know? *Research on Aging*, 12, 531–545.
- Himes, C. L. (1992). Future caregivers: Projected family structures of older persons. *Journal of Gerontology: Social Sciences*, 47(1), S17–S26.
- Hollmann, F. W., Mulder, T. J., & Kallan, J. E. (2000). *Methodology and assumptions for the population projections of the United States: 1999 to 2100* (Population Division Working Paper No. 38). Washington, DC: U.S. Bureau of the Census.
- Ip, F., & McRae, D. (1999). *Small area household projections—A parameterized approach*. Population Section, Ministry of Finance and Corporate Relations, Province of British Columbia, Canada. Retrieved from <http://www.bcstats.gov.bc.ca/data/pop/pop/methhhld.pdf>
- Keilman, N. (1988). Dynamic household models. In N. Keilman, A. Kuijsten, & A. Vossen (Eds.), *Modelling household formation and dissolution* (pp. 123–138). Oxford, UK: Clarendon Press.
- Keilman, N. (2003). The threat of small households. *Nature*, 421, 489–490.
- Khan, H. T. A., & Lutz, W. (2008). How well did past UN population projections anticipate demographic trends in six South-east Asian countries? *Asian Population Studies*, 4, 77–95.
- Land, K. C., & Rogers, A. (1982). *Multidimensional mathematical demography*. New York: Academic Press.
- Liu, J., Dally, C. C., Ehrlich, P. R., & Luck, G. W. (2003). Effects of household dynamics on resource consumption and biodiversity. *Nature*, 421, 530–533.
- Lutz, W., & Prinz, C. (1994). The population module. In W. Lutz (Ed.), *Population-development-environment: Understanding their interactions in Mauritius* (pp. 221–231). Berlin, Germany: Springer-Verlag Press.
- Mackellar, F. L., Lutz, W., Prinz, C., & Goujon, A. (1995). Population, households, and CO2 emissions. *Population and Development Review*, 21, 849–866.
- Mason, A., & Racelis, R. (1992). A comparison of four methods for projecting households. *International Journal of Forecasting*, 8, 509–527.
- Moffitt, R. (2000). Demographic change and public assistance expenditures. In A. J. Auerbach & R. D. Lee (Eds.), *Demographic change and public assistance expenditures* (pp. 391–425). Cambridge, UK: Cambridge University Press.

- Murphy, M. (1991). Modelling households: A synthesis. In M. J. Murphy & J. Hobcraft (Eds.), *Population research in Britain, A supplement to population studies* (Vol. 45, pp. 151–176). London, UK: Population Investigation Committee, London School of Economics.
- Murray, C. J. L., Ferguson, B. D., Lopez, A. D., Guillot, M., Salomon, J. A., & Ahmad, O. (2003). Modified logit life table system: Principles, empirical validation, and application. *Population Studies*, 57, 165–182.
- Myers, D., Pitkin, J., & Park, J. (2002). Estimation of housing needs amid population growth and change. *Housing Policy Debate*, 13, 567–596.
- Paget, W. J., & Timaeus, I. M. (1994). A relational Gompertz model of male fertility: Development and assessment. *Population Studies*, 48, 333–340.
- Prskawetz, A., Jiang, L., & O'Neill, B. (2004). Demographic composition and projections of car use in Austria. In *Vienna yearbook of population research* (pp. 274–326). Vienna, Austria: Austrian Academy of Sciences Press.
- Rao, J. N. K. (2003). *Small area estimation*. New York: John Wiley & Sons Inc.
- Rogers, A. (1975). *Introduction to multi-regional mathematical demography*. New York: John Wiley & Sons.
- Rogers, A. (1986). Parameterized multistate population dynamics and projections. *Journal of the American Statistical Association*, 81, 48–61.
- Rogers, A. (1995). *Multiregional mathematical demography: Principles, methods, and extensions*. New York: Wiley.
- Schoen, R. (1988). *Modeling multi-group populations*. New York: Plenum Press.
- Schoen, R., & Standish, N. (2001). The retrenchment of marriage: Results from marital status life tables for the United States, 1995. *Population and Development Review*, 27, 553–563.
- Smith, S. K. (2003). Small area analysis. In P. Demeny & G. McNicoll (Eds.), *Encyclopedia of population* (pp. 898–901). Farmington Hills, MI: Macmillan Reference.
- Smith, S. K., & Morrison, P. A. (2005). Small area and business demography. In D. L. Poston & M. Micklin (Eds.), *Handbook of population* (pp. 761–786). New York: Kluwer Academic/Plenum Publishers.
- Smith, S. K., Rayer, S., & Smith, E. A. (2008). Aging and disability: Implications for the housing industry and housing policy in the United States. *Journal of the American Planning Association*, 74, 289–306.
- Smith, S. K., Rayer, S., Smith, E., Wang, Z., & Zeng, Y. (2012). Population aging, disability and housing accessibility: Implications for sub-national areas in the United States. *Housing Studies*. Advance online publication. doi:10.1080/02673037.2012.649468
- Smith, S. K., Tayman, J., & Swanson, D. A. (2001). *State and local population projections: Methodology and analysis*. New York: Kluwer Academic/Plenum Publishers.
- Soldo, J., Wolf, D. A., & Agree, E. M. (1990). Family, households, and care arrangements of frail older women: A structural analysis. *Journal of Gerontology: Social Sciences*, 45, S238–S249.
- Spicer, K., Diamond, I., & Ni Bhrolchain, M. (1992). Into the twenty-first century with British households. *International Journal of Forecasting*, 8, 529–539.
- Stupp, P. W. (1988). A general procedure for estimating intercensal age schedules. *Population Index*, 54, 209–234.
- Swanson, D. A. (2010, April). *The methods and materials used to generate two key elements of the housing unit method of population estimation: Vacancy rates and persons per household*. Paper presented at the annual meeting of the Population Association of America, Dallas, TX.
- Swanson, D. A., & Pol, L. G. (2009). In Y. Zeng (Ed.), *Demography, Encyclopedia of Life Support Systems (EOLSS)* (www.eolss.net), coordinated by the UNESCO-EOLSS Committee. Oxford, UK: EOLSS Publishers Co. Ltd.
- Treadway, R. (1997). *Population projections for the state and counties of Illinois*. Springfield: State of Illinois.
- U.S. Census Bureau. (2008). *Interim projections of the U.S. population by age, sex, race, and Hispanic origin: Summary methodology and assumptions*. Washington, DC: U.S. Census Bureau. Retrieved from <http://www.census.gov/population/www/projections/usinterimproj/idbsummeth.pdf>
- U.S. Census Bureau. (1996). *Projections of the number of households and families in the United States: 1995 to 2010*. U.S. Department of Commerce Economics and Statistics Administration (Current Population Reports P25-1129). Washington, DC: U.S. Government Printing Office.
- United Nations. (1982). *Model life tables for developing countries* (ST/SEA/SER, A/77, Sales No. E. 81, XIII, 7). New York: United Nations.
- Van Imhoff, E., & Keilman, N. (1992). *LIPRO 2.0: An application of a dynamic demographic projection model to household structure in the Netherlands*. Amsterdam, Netherlands: Swets & Zeithinger.

- Van Imhoff, E., Kuijsten, A., Hooimeiger, P., & van Wissen, L. (1995). Epilogue. In E. Imhoff, A. Kuijsten, P. Hooimeiger, & L. van Wissen (Eds.), *Household demography and household modeling* (pp. 345–351). New York: Plenum.
- Wang, Z. (2009a). *Households and living arrangements forecasting and the associated database development for southern California six counties and the whole region, 2000–2040*. (in response to RFP No. 09-043-IN of SCAG: Household Projection Model Development: Simulation Approach) (Technical Report No. 09–02). Chapel Hill, NC: Households and Consumption Forecasting Inc.
- Wang, Z. (2009b). *Household projection and living arrangements projections for Minneapolis-St. Paul Region, 2000–2050* (Technical report No. 09–03). Chapel Hill, NC: Households and Consumption Forecasting Inc.
- Wang, Z. (2011a). *The 2011 updating of households and living arrangements forecasting for Southern California six counties and the whole region, 2000–2040* (Technical report No. 11–01). Chapel Hill, NC: Households and Consumption Forecasting Inc.
- Wang, Z. (2011b). *The 2011 updating of household projection and living arrangements projections for Minneapolis-St. Paul region, 2000–2050* (Technical report No. 11–02). Chapel Hill, NC: Households and Consumption Forecasting Inc.
- Willekens, F.J., I. Shah, J.M. Shah, and P. Ramachandran. (1982). Multistate analysis of marital status life tables: theory and application. *Population Studies*, 36, 129–144.
- Zeng, Y. (1986). Changes in family structure in China: A simulation study. *Population and Development Review*, 12, 675–703.
- Zeng, Y. (1988). Changing demographic characteristics and the family status of Chinese women. *Population Studies*, 42, 183–203.
- Zeng, Y. (1991). *Family dynamics in China: A life table analysis*. Madison: The University of Wisconsin Press.
- Zeng, Y., Coale, A., Choe, M. K., Liang, Z., & Liu, L. (1994). Leaving parental home: Census based estimates for China, Japan, South Korea, the United States, France, and Sweden. *Population Studies*, 48, 65–80.
- Zeng, Y., Land, K. C., Wang, Z., & Gu, D. (2006). U.S. family household momentum and dynamics—Extension and application of the ProFamy method. *Population Research and Policy Review*, 25, 1–41.
- Zeng, Y., Land, K. C., Wang, Z., & Gu, D. (2010, April). *Extended cohort-component approach for households projection at sub-national levels—With empirical validations and applications of households forecasts for 50 states and DC*. Paper presented at the annual meeting of the Population Association of America, Dallas, TX.
- Zeng, Y., Morgan, S. P., Wang, Z., Gu, D., & Yang, C. (2012). A multistate life table analysis on union regimes in the United States—Trends and racial differentials, 1970–2002. *Population Research and Policy Review*, 31, 207–234.
- Zeng, Y., Vaupel, J. W., & Wang, Z. (1997). A multidimensional model for projecting family households—With an illustrative numerical application. *Mathematical Population Studies*, 6, 187–216.
- Zeng, Y., Vaupel, J. W., & Wang, Z. (1998). Household projection using conventional demographic data. *Population and Development Review, Supplementary Issue: Frontiers of Population Forecasting*, 24, 59–87.
- Zeng, Y., Wang, Z., Jiang, L., & Gu, D. (2008). Future trend of family households and elderly living arrangement in China. *GENUS—An International Journal of Demography*, LXIV(1–2), 9–36.
- Zeng, Y., Wang, Z., Ma, Z., & Chen, C. (2000). A simple method for estimating α and β : An extension of Brass relational Gompertz fertility model. *Population Research and Policy Review*, 19, 525–549.