HOUSEHOLD AND POPULATION PROJECTIONS AT SUB-NATIONAL LEVELS: AN EXTENDED COHORT-COMPONENT APPROACH*

Yi Zeng, Kenneth C. Land, Zhenglian Wang and Danan Gu

* Zeng Yi, Center for the Study of Aging and Human Development and Geriatric Division of Medical School; Institute of Population Research and Department of Sociology, Duke University; China Center for Economic Research, Peking University; e-mail: zengyi68@gmail.com. Kenneth C. Land, Center for Population Health and Aging of Institute of Population Research and Department of Sociology, Duke University; e -mail: kland@soc.duke.edu. Zhenglian Wang, Senior Research Scientist, Center for Population Health and Aging of Institute of Population Research, Duke University, President, Household and Consumption Forecasting, Inc., NC e-mail: wangzl8886@gmail.com. Gu Danan, Research Scientist, Center for Study of Aging and Human Development, Duke University; e-mail: gudanan@duke.edu. The research reported in this paper was 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), Duke University, Max Planck Institute for Demographic Research, and Sabre System Inc. for supporting related basic and applied scientific research. Correspondence should be sent to: Yi Zeng (zengyi68@gmail.com).

Abstract

This paper describes the core methodological ideas, the required input data, and estimation issues of an extended cohort-component approach to simultaneously project household composition and population distributions at sub-national levels. We assess the projection accuracy of this approach by calculating projections from 1990 to 2000 and from 2000 to 2006 and comparing the projected with the census-observed counts in 2000 and ACS-observed in 2006 for the 50 states and the District of Columbia, the Research Triangle Area, Durham county and a small town of Chapel Hill of North Carolina. The comparisons show that most absolute projection errors between 1,475 pairs of indices of household and population projections and the corresponding census and ACS observations are small – less than three percent – and almost all errors are less than ten percent. We then report illustrative household projections from 2000 to 2050 for the 50 states and the District of Columbia, and household/housing projections for a small town of Chapel Hill up to 2015 in order to demonstrate the practical capabilities of the new approach. Among many interesting numerical outcomes, the aging of American households over the next few decades across all states and the aging of the housing market in Chapel Hill are particularly striking trends in the projections.

1. INTRODUCTION

There is a growing demand for projections not only of population size but of the distribution of the population among household types, sizes, and living arrangements for socioeconomic planning, business, policy and scholarly analysis. For example, there is a need to integrate household size and structure explicitly into projection models of population, environment, and development (Lutz & Prinz 1994; Mackellar et al. 1995). Moffitt (2000) argues that better projections of demographic and household trends could provide valuable intelligence to guide planning and should be a major policy goal. Extant research has well established that demands for energy use (e.g., gas and electricity), automobiles, housing, water, durable goods and other home-related products and services are largely determined by changes in households (e.g., Foncel & Ivaldi 1999; Mayers et al., 2002; Davis 2003; 2004; Keilman, 2003; Liu et al. 2003; Prskawetz et al. 2004; Wang et al. 2005; Dalton et al., 2008). Past research has also established that household and living arrangements are the major determinant of the amount and type of long-term care for the elderly (e.g., Doty, 1986; Chappell 1991; Morris et al. 1998; Soldo et al. 1990; FIFARS 2004). In particular, the use of long-term care varies by family household status (Freedman 1996). With rapid population aging in the U.S. and many other countries around the world, the market for elder-care industries is growing with extraordinary speed, 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 sub-national levels such as states (or provinces)[1], and small areas[2] (Treadway, 1997; Ip and McRae, 1999; Rao 2003; Crowley, 2004). Sub-national population and household projections are useful for distributing government funds, allocating various types of resources, planning the development of infrastructure and

[1] Note that a state in the United States is equivalent to a province or other kind of administrative region immediately underneath the nation in other countries. Accordingly, for applications to other countries, one may replace the word “state” used in this article by “province” or another compatible unit name.
[2] Sub-state level usually refers to counties or cities/towns which have their own Public Use Micro Sample (PUMS) code in the census data file. The “small-area” term refers to small towns and places, possibly even tracts or block groups which have a small population size and normally do not have their own PUMS code.

public facilities, market research and production planning for household-related goods and services, and decisions on the expansion or reduction of local businesses (Smith, Tayman, and Swanson, 2001).

Motivated by these growing scientific and practical needs, we describe in the next section the core ideas of an extended cohort-component approach for the projection of household composition and population distributions at sub-national levels. In the third section, we discuss data and estimation issues. The fourth section presents validation tests of household projections from 1990 to 2000 and from 2000 to 2006 using the extended cohort-component method through comparing the projected and the census-ACS observed in 2000 and 2006 for the 50 states and the District of Columbia (DC), and three sub-state areas (the Research Triangle Area, Durham county and a small town of Chapel Hill of North Carolina). This is followed in the fifth section by a description of illustrative household projections from 2000 to 2050 for the 50 states and the DC, and household/housing projections for a small town (Chapel Hill, North Carolina) up to 2015 in order to demonstrate the practical capabilities of the new approach. A discussion and conclusion section ends the paper.

2. CORE IDEAS OF THE EXTENDED COHORT-COMPONENT METHOD

Building on methodological advances in multidimensional demography (Rogers 1975, 1995; Willekens et al. 1982; Land & Rogers 1982; Schoen 1988), and based on Bongaarts' and Zeng's one-sex family status life table models (Bongaarts 1987; Zeng 1986, 1988, 1991), Zeng, Vaupel, and Wang (1997; 1998) developed a two-sex dynamic macro model for projections of households and living arrangements. Zeng, Land, Wang, and Gu (2006) then presented a substantial extension of this model and applied it to U.S. household projections at the national level.

In the Zeng et al. (2006) model, all individuals of the population are grouped and projected forward cohort by cohort and by age, sex, marital/union status (including cohabitation), parity, and number of co-residing children and parents; in addition, race and rural/urban categories may be optionally included or excluded. Unlike conventional macro-simulation models for household projection, which have stringent data requirements for the estimation of probabilities for transitions among household-type statuses, the Zeng et al. (2006) model *requires as input only conventional and commonly available demographic data*, as listed in Table 1, to compute the individual groups' status changes by cohort and age. These data can be obtained from vital statistics, censuses, and routinely conducted surveys.

The basic mechanism of this model is that projections are made about the changes in demographic components (marriage/union, fertility, leaving parental home, mortality, and migration) for each of the cohorts that produce household distributions in the future years. This is analogous to, and a substantive extension of, the conventional cohort-component population projection model (Preston et al. 2001:119-128). This paper further develops the Zeng et al. (2006) model for applications to sub-national household projections.[3] For brevity, this model hereafter is termed the *extended cohort-component projection model (method)* or, for short, the *extended model (method)*. This model is built on five core ideas, which now are described.

Core idea 1: A multi-state accounting model. The innermost core of the extended cohort-component method is *a multi-state accounting model* for transforming the marital/union statuses and co-residence with children and parents statuses of members of a population in year t into their corresponding statuses in year t+1. Unlike most other macrosimulation models which use the household as the basic unit and require extensive data for the estimation of transition probabilities among household-type statuses (Van Imhoff and Keilman 1992), we use individual as the basic unit of analysis. Consequently, only conventional and normally available census and vital statistics data are required as inputs to project household and population age/sex distributions simultaneously. In addition to identifying the individual members of a population by single years of age, sex, race and rural-urban residence (optional), the extended cohort-component model keeps track of individuals' marital/union status, statuses of co-residence with one or two parent(s) and

[3] The numerical projections reported in this paper were calculated with the ProFamy computer software program; this program incorporates the extended cohort-components projection method and contains a demographic database of the U.S. age-specific schedules of demographic rates to assist users in making projections. A free trial version of the ProFamy software for household forecasting can be downloaded from the Website http://www.profamy.com/.

number of co-residing 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 co-residence statuses with parents/children. For example, a married or cohabiting woman who is not co-residing with parents and whose number of co-residing children is c (c=0,1,2,3,4,5+), is a reference person representing a two-generation and couple household of 2+c family members. If the reference person is not married and not cohabiting, 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, we would face a daunting challenge of estimating very high dimensional matrices of cross-status transition probabilities if we adopted the conventional multistate computation strategy. For example, if 7 marital/union statuses[4], 3 statuses of co-residence with parents, 6 parity and 6 co-residence statuses with children are distinguished as in the U.S. household projections of Zeng et al. (2006), we would have to estimate a cross-status transition probabilities matrix with 194,481 elements (7 x 3 x $\sum_{p=0}^{6}(p+1)$) = 441 x 441 = 194,481) at each age of each sex. This is certainly not practical. Thus, we adopted a computational strategy of calculating individual group marital/union, co-residence (with parents/children), migration and survival status changes by assuming (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 Figure 1). 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.

----- Figure 1 about here ---

[4] The seven marital/union statuses are: 1) Never-married & not-cohabiting; 2) Married; 3) Widowed & not-cohabiting; 4) Divorced & not cohabiting; 5) Never-married & cohabiting; 6) Widowed & cohabiting; 7) Divorced & cohabiting. The multiple dynamic transitions among these seven statuses and death are depicted in Figure 1 in Zeng et al. (2006).

Core idea 3: A judicious use of stochastic independence assumptions. Coupled with core idea 2, the third core idea of the extended cohort-component method greatly simplifies the estimation of the multi-status transition probabilities. This idea, also originally suggested by Bongaarts (1987) and adapted and generalized by Zeng et al. (1997; 1998), is that not all of the elements of the transition probability matrix depend on many of the other elements, and, indeed, some of their real-world dependencies can be reasonably assumed to be stochastically independent. In other cases, the reality of limited data sources available for estimation of the transition probabilities that depend on many other covariates forces the application of an independence assumption. In either case, the consequences of the stochastic 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 model, *marital/union status transitions* depend on age, sex, and race, but are assumed to be independent of parity and co-residence status with parents and children; *fertility* depends on age, race, parity and marital/union status, but are assumed to be independent of co-residence status with parents and children; *mortality* rates are age, sex, race and marital/union status specific, but are assumed to be independent of parity, co-residence 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 of the children.

Core idea 4: Use of the harmonic mean to ensure consistency between the two sexes and between parents and children in the projection model. Consistency of the male and female projections is a basic requirement in any two-sex model dealing with marriage/union of men and women. In any specific year, the number of male marriages is equal to the number of female marriages; the number of male divorces is equal to the number of female divorces; the number of newly widowed women (men) is equal to the number of new deaths of currently married men (women). When cohabiting status is distinguished, the number of cohabiting men is equal to the number of cohabiting women; the number of men (women) who exit from cohabiting status either due to union dissolution or death of partner is equal to the number of their counterparts with opposite sex. For household projections, we also need to ensure the consistency of changes in co-residence status between parents and children. More specifically, the number of parents whose co-residence with children status changed in a year should be consistent with number of children who left parental home or returned to parental home or died. We ensure the consistency between the two sexes and between parents and children by using the harmonic mean approach, which demographers have shown to satisfy most of the theoretical requirements and practical considerations for consistency of two-sex projection models (Pollard, 1977; Schoen, 1981; Keilman, 1985; Van Imholf and Keilman, 1992; Zeng et al. 1997; 1998).

---- Table 1 about here----

Core idea 5: Use of age-sex-status-specific schedules of demographic rates and summary parameters thereof to specify projected demographic rates in future years. In applications of the extended cohort-component household projection model, one needs to use a set of race-sex-age-specific standard schedules of demographic rates (see (2) in Table 1). One then projects or assumes the demographic summary measures (see (3) in Table 1). The standard schedules formulate the age pattern of demographic processes. One may take into account anticipated changes in the age patterns, such as delaying or advancing marriage and fertility, through adjusting the standard schedules to match the projected mean ages of the demographic events in the future years.[5] The demographic reasoning for the core idea 5 is, first, widely replicated research

[5] Based on the standard schedules and demographic summary measures (including mean ages at marriage and fertility), the extended model programmed in the ProFamy software generates estimates of the age-specific demographic rates needed for projecting households and population to future years (see Zeng, Stallard and Wang 2004). With the software and the race-sex-age-specific standard schedules in hand, the analysis can concentrate on projecting future demographic summary measures. This can be done using conventional time series analysis by statistical software (e.g., SAS, SPSS, or STATA) or expert opinion approaches.

indicating that the race-sex-age-specific patterns (the tempo) of the schedules of demographic events are much less variable than the level of the occurrence of the events (the quantum). Second, the less variable age-specific schedules/tempo themselves can be allowed to vary over time in accordance, for example, with steady increases (or decrease) in the mean or median ages of marriage and births and even changes in the shape of the age-specific schedules towards to be more spread or more concentrated (Zeng et al. 2000). Third, summary parameters of the age-specific schedules, e.g, the Total Fertility Rate for a fertility schedule, can be used to "tune" the household and population projections up or down for the development and analysis of various demographic scenarios.[6]

3. DATA AND ESTIMATION ISSUES

3.1. Data Requirements

As shown in (1) in Table 1, for population and household projections at the national, state, and sub-state levels with the extended model, a census micro data file that contains the variables of sex, age, marital/union status, relationship to the householder, and whether living in a private or institutional household is needed (race and rural/urban are optional). Normally, the model standard schedules (see (2) in Table 1) need to be estimated at the national level only, and they can be employed for the household projection at the state level (see Section 3.2.1). The projected (or assumed) demographic summary measures – TFR, life expectancy at birth, general rates of marriage, divorce, cohabitation and union dissolution, total number of migrants, mean age at first marriage and birth in the future years – are needed for projections at both national and state levels (see (3) in Table 1), but not necessarily for sub-state areas projection (see Section 3.3). In sum, using existing national model standard schedules and the extended cohort-component

---

Time series data on other related socioeconomic covariates (e.g., average income, education, urbanization, etc.) also can be used in projecting the demographic summary measures.

[6] More generally, the ProFamy software program allows users to "fine tune" the household and population projections for "local" situations. For example, a state demographer with side information from other data sources about fertility, mortality, or migration patterns for a sub-state area might first conduct a projection for that area using national or state-specific demographic schedules and rates and then refine the projection by incorporating the local information. In this way, projections/forecasts for short- to medium-term (say 5 to 10 years) time horizons can be fine-tuned and the corresponding errors of the projections reduced below those reported in Section 4.

method, *household projections at the state level* require a census micro data file and the projected (or assumed) demographic summary measures for the future years; *household projections for a sub-state area* require only a census micro data file or relevant households distributions obtained from publicly available census cross-tabulations, given that the projections for the parental state where the sub-state area is located are done. Such relatively simple data requirements suggest that household projections at the state and sub-state levels using the extended cohort-component method can be practically produced without much difficulty.

3.2. Household projections at the state level

3.2.1. Application of the model standard schedules

Data for estimating race-sex-age-specific standard schedules of the demographic rates (except migration rates[7]) for household projection ((2) in Table 1) may not be available at the state level. However, once the age-race-sex-specific standard schedules at the national level are prepared (and updated when the new data become available) by a demographer[8], these standard schedules can be used and/or adapted for projections at the state level. This is equivalent to the widely practiced applications of the model life tables (e.g., Coale, Demeny, and Vaughn, 1983) in population projections and estimations. The theoretical foundation of applications of the model life tables and the model standard schedules is that race-specific demographic summary measures are crucial for projections, but race-age-specific model standard schedules typically are not highly sensitive to the projection results as long as they reveal the general age patterns of demographic processes of the population.[9]

---

[7] Data for estimating age-race-sex-specific standard schedules of emigration and immigration at the state level are available from census micro data files.

[8] This has been done for the U.S., with the resulting standard schedules available in the ProFamy software.

[9] This proposition was evaluated in Zeng et al. (2006) by performing two scenarios of U.S. national projections by race from 1990 to 2020 with all race-specific demographic summary measures being identical each other, but one scenario used the race-sex-age-specific rates observed in 1990s and another scenario used the race-sex-age-specific rates observed in 1980s. The 17 main indices of the projections then were compared for the years 2000, 2010, and 2020 for these two scenarios. The main indices of the projections included the total number of households, average household size, percent of 1, 2, 3, 4, 5+ persons households, percent of married couple households, total population size, % of children and elderly, % of people living in group quarters and dependent ratios. The results showed that, while the projected input summary measures are identical, using the standard schedules observed in 1980s and using the standard schedules observed in 1990s produced almost the same projections. About two-thirds of the discrepancy rates of the projected main indices between these two scenarios are less than 1% and one-third of discrepancy rates are 1.0-3.4% (see Appendix A in Zeng et al. (2006)).

3.2.2. Estimation of summary measures at the state level

Estimation of the required demographic summary measures – the TFR, life expectancy at birth, total number of migrants, mean age at first marriage and birth are straightforward. But estimation of standardized general rates of marriage, divorce, cohabitating, and union dissolution for household projections at the state level merits special attention.

The state-specific general rates of marriage, divorce, cohabitating, and union dissolution in the year t are defined by dividing the total number of events of marriage, divorce, cohabitating, and union dissolution occurring in the year in the state by the total number of persons who are at risk of experiencing these events. We use the age-marital/union status distributions of the base population derived from the census dataset of the state as the “standard” to compute the standardized state-specific general rates in future projection years. This is a typical application of the classic demographic procedure of direct standardization (see, e.g. Preston et al. 2001:24). The standardized general rate in future projection years using the census-counted age-marital/union status distributions as the “standard” eliminates possible distortions in measuring the levels of marriage/union formation and dissolution due to changes in population structure in the future. For example, the un-standardized general marriage (or divorce) rate would decrease/increase solely due to the structural growth/decline of the numbers of elderly persons even if the age specific marriage (or divorce) rates did not change, because the risks of marriage (or divorce) of the elderly are substantially lower than those of younger people. Detailed discussion of rationales for applying the standardized general rates of marriage, divorce, cohabitating, and union dissolution, such as why they are male-female combined to ensure the two-sex constraints and how the gender differentials in the age-specific rates are retained, were given in Zeng et al. (2006: 6). We present below the procedures for estimating the race-specific general rates of marriages, divorce, cohabitation and union dissolution at the state level.

Procedure for estimating the race-specific general rates of marriages and divorce. Given that we have the all-race combined but do not have the race-specific total numbers of marriages and divorces for each of the 50 states and DC, we employ the following procedure to estimate the state-race-specific general rate of marriage (GM) and divorce (GD). Let:

N_i(x,s,r,T1) denote the known number of persons of age x, race r, marital/union status i and sex s counted in the census year T1 in the state (we omit subscript index of state to simplify the presentation);

M_{ij}(x,s,r), the national model standard schedules of the race-sex-age-specific o/e rates of transition from marital/cohabiting status i to j ($i \neq j$; i and j represent the seven marital/union statues, see footnote 3);

m_{ij}(x,s,r,T1), the estimated race-sex-age-specific o/e rates of transition from marital/union status i to j in the census year T1 ($i \neq j$) in the state; and

TM(T1), the published all-races-combined total number of marriages (couples) including 1st marriages and remarriages occurred in the state in the census year T1.

Then,

$$m_{i2}(x,s,r,T1) = M_{i2}(x,s,r)\frac{2TM(T1)}{\sum_{x=\alpha}^{\beta}\sum_{s=1,2}\sum_{r}\sum_{i}N_i(x,s,r,T1)M_{i2}(x,s,r)}, \; i \neq 2;$$

where α (usually taken as 15) and β are the low and the upper boundary of the age range in which the events of marriage/union formation and dissolution occur.

We then use the estimated m_{i2}(x,s,r,T1) and N_i(x,s,r,T1) to compute the estimated race-specific GM in year T1 for the state.

Let TD(T1) denote the published all-races-combined total number of couples who divorced in the state in the census year T1.

Then,

$$m_{24}(x,s,r,T1) = M_{24}(x,s,r)\frac{2TD(T1)}{\sum_{x=\alpha}^{\beta}\sum_{s=1,2}\sum_{r} N_2(x,s,r,T1)M_{24}(x,s,r)},$$

We then use the estimated $m_{24}(x,s,r,T1)$ and $N_2(x,s,r,T1)$ to compute the estimated race-specific GD in year T1 for the state.

Procedure for estimating the race-specific general rates of cohabitation and union dissolution. As we do not have data on the total numbers of events of formation and dissolution of cohabitation unions, we cannot estimate the race-specific general rate of cohabitating (GC) and union dissolution (GCD) at the state level directly, and we need to employ an indirect estimation approach. We first estimate the so-called standard-schedules and previous-base-population implied GC and GCD using the base populations of the state derived from the previous census (e.g.1990) and the model standard schedules of the cohabitation union formation and dissolution. We then project the household distributions from 1990 to 2000 for the state using the 1990 census data as the base population, the implied GC and GCD and the other estimated demographic rates in 1990s as input. Through such projections, we obtain the projected proportions of households with a cohabiting couple among all households in 2000, denoted as PC. We then compare PC with the proportion of the households with a cohabiting couple among all households observed in the 2000 census, denoted as CC. If PC is higher (or lower) than CC, we proportionally reduce (or increase) the implied GC and increase (or reduce) the implied GCD and redo the projection, until the projected PC and the observed CC are reasonably close to each other (e.g. a relative difference between -1% and +1%). Given the cohabitation data constraints at the state level, this procedure produces reasonably good estimates of the GC and GCD, as shown by the results of the validation testing projections from 1990 to 2000 and from 2000 to 2006 for the 50 states and DC presented in the Section 4.

3.2. Household projections at the sub-state level

It is usually very difficult to have the adequate data at the sub-state level to estimate the demographic parameters which are necessary to apply the classic and extended cohort-component methods for population and household projections, except if the sub-state area has a large population size with its own strong data collection programs. Indeed, even census micro datasets, although valuable, cannot provide information at the sub-state level (e.g., a small county) for all the characteristics that are of interest and needed for cohort-component projections. Therefore, most researchers use “indirect” methods that “borrow strength” based on a projection for the parental state to increase the stability and accuracy of population and household projections for sub-state areas (Rao 2003; Smith & Morrison 2005). Ratio trend extrapolation methods (e.g., extrapolating sub-state’s shares of state population) are frequently used for the sub-state areas projections because their data requirements are minimal, they are easy to apply, and their projections have often to be reasonably accurate (Smith 2003). In household projection using the extended cohort-component method, we employ the ratio trend extrapolation method with either a constant-share or a shifting-share specification (Smith et al. 2001) to conduct household projections for the sub-state areas, in combination with the household projections of the parental state. The household projections of the parental state must be done first as a basis. We then compute the race-sex-age-specific proportions of households with various types/sizes of the sub-state area among the corresponding households of the parental state. We assume that the proportions are constant or changing following past trends or projected new trends, and then multiply the existing parental state’s household projections by the proportions to derive the household projections for the sub-state area. The assumption imposed and the rationale of such constant-share or shifting-share approaches in household projections for sub-state areas are the same as those generally used for sub-state area population projections, which has been shown to be accurate (Smith et al. 2001). The following procedure is designed for household projections at the sub-state area level, based on the parental state projections using the extended cohort-component method.

Let *T1* denote the starting year of the projection; *T1* may be the most recent census year; *T1* may also be a year after the most recent census if one decides to use more recent large survey data such as American

Community Survey (ACS) as baseline for the projection[10]. Denote by *t* the future years for the time period of the projection.

Input:

L(h,s,r,x,T1) and *L(h,s,r,x,t)*: the number of households of type *h* and size *s* with a reference person of race *r* and age *x* in the parental state observed in the baseline year *T1* and projected in year *t*, respectively; Age *x* can be 5-year, 10-year, 15-year, or 20-year age groups, depending on the user's choice based on the population sample size.

p(h,s,r,x,T1): the proportions of the households of type *h* and size *s* with a reference person of race *r* and age *x* of the sub-state area among the corresponding households of the parental state in the baseline year *T1*;

PL(T1) and *PL(t)*, the total population size in the parental state observed in the baseline year *T1* and projected in year *t*, respectively;

PS(T1), the total population size in the sub-state area observed in the baseline year *T1*;

L(h,s,r,x,T1), p(h,s,r,x,T1), PL(T1) and PL(T1) are derived from census or ACS data sets; *L(h,s,r,x,t)* and *PL(t)* are available from the existing household projections for the parental state.

Output:

f(h,s,r,x,t) – projected number of households of type *h* and size *s* with a reference person of race *r* and age *x* in a sub-state area in the year *t*;

PS(t) – total population size projected in the future year *t* in the sub-state area.

Computation:

Based on the constant-share method (Smith et al. 2001):

$$f(h,s,r,x,t) = L(h,s,r,x,t)\, p(h,s,r,x,T1)$$

---

[10] For the major demographic indicators, the annual ACS is representative at the state level, and 2-, 3-, 4-, or 5-year moving averages of the ACS data are representative for the sub-state areas, depending on their population size.

$PS(t) = PL(t)\ (PS(T1) / PL(T1))$

Based on the shift-share method (Smith et al., 2001):

$$f(h,s,r,x,t) = L(h,s,r,x,t)\ (p(h,s,r,x,T1) + ((t\text{-}T1)/(T1\text{-}T0))\ (p(h,s,r,x,T1)\text{-}\ p(h,s,r,x,T0)),$$

$$PS(t) = PL(t)\ (Ps(T1) / PL(T1) + ((t\text{-}T1)/(T1\text{-}T0))\ (Ps(T1) / PL(T1) - Ps(T0) / PL(T0)),$$

where $T0$ refers to the previous census year preceding to the starting year of the projection ($T1$).

Additional adjustments are necessary to maintain consistency between the total population size implied by the projected numbers of households by type/size/race/sex/age and the projected total population size in the sub-state area. Let $f'(h,s,r,x,t)$ denote the finally adjusted projected number of households of type h and size s with a reference person of race r and age x in a sub-state area in the year t.

$$f'(h,s,r,x,t) = f(h,s,r,x,t)\ (PS(t) / (\sum_h \sum_s \sum_r \sum_x f(h,s,r,x,t) \times s)$$

Given sample size limitations for sub-state areas, we classify the household projection output into 11 categories (see Table 2) by household types and sizes including persons living in private households or group quarters by age/sex of the reference person and by race (if sample size allows) in the illustrative application to be presented later. The classification of these 11 categories as shown in Table 2 is for illustration only, as one may group the available detailed outcomes of household projections by types and sizes (see Table 2 in Zeng et al. 2006) into larger (detailed) or smaller (simplified) number of the categories, depending on the population size of the sub-state area under study and the purpose of the analysis.

--- Table 2 about here ---

4. AN EMPIRICAL ASSESSMENT

A useful validation exercise for a population projection model is to project between two past dates for which the observations are known, and then compare the observed data with the projected data. Zeng et al. (2006) provided this form of empirical assessment of the extended cohort-component U.S. national household and population projections from 1990 to 2000. The assessment found that most discrepancies between the projections and census observations in 2000 were small, which provides evidence towards the validation of the new method at the national level (see Section 2.3 in Zeng et al. 2006).[11]

Does the extended cohort-component method and software also work reasonably well at the state and sub-state levels? To address this issue, we conducted two sets of the empirical assessments of validation tests of household projections for the 50 states and DC, all using the national model standard schedules (see (2) in Table 1) estimated based on the pooled national survey data[12] (Zeng et al. 2009). The first test compares projections from 1990 to 2000 using the 1990 census data as the base population and summary measures based on data in the 1990s with census-observed data for 2000. This test assesses the simulation properties of the extended model based on the assumptions that the input data (observed in the 1990s) and the 2000 census observations (outcome in the exercise) are correct. The second test projects from 2000 to 2006 using the 2000 census data as base population and summary measures based on data before 2001, and compares the projected and the ACS-observed in 2006. This test assumes that we have no data after 2000 when one projects 2000 to 2006 and tests the accuracy of the projections using the extended method/software in the real world (assuming the 2006 ACS data are accurate). We also performed household projections from 1990 to 2000 and from 2000 to 2006 for the Research Triangle Area[13] (a middle-size sub-state area), Durham

---

[11] Zeng et al. (2006) also presented a detailed comparative analysis showing that the extended cohort-component method has substantial merits as compared to the classic headship-rate method, which is still widely used for household projections at national, state and sub-state levels. One of the problems inherent in the headship rate approach is that it does not link to demographics rates and it is very hard to conduct reasonable projections based on understanding of the underlying demographic changes. For example, increases in divorce and cohabitation rates and decreases in marriage and remarriage rates imply that the prevalence of future households with a married couple will decrease (Goldscheider 1990). Changes in the availability of children, due to changing fertility, mortality and mobility, will affect household structure and long-term care for the elderly (Doty 1986; Himes 1992). The extended cohort-component method can incorporate such demographic changes into household projections, but the headship-rate method cannot. Also, the headship-rate method projects a few household types without size, and does not deal with household members other than heads, but the extended method projects much more detail on households types and sizes as well as family households structure and living arrangement of all members of the population (Zeng et al., 2006).

[12] The marriage/union history data from the following four national surveys are pooled to estimate the race-sex-age-specific model standard schedules: (a) National Survey of Family Households (NSFH) conducted in 1987-1988, 1992-1994, and 2002; (b) National Survey of Family Growth (NSFG) conducted in 1983, 1988, 1995, and 2002; (c) Current Population Surveys (CPS) conducted in 1980, 1985, 1990, 1995; (d) Survey of Income and Program Participation (SIPP) conducted in 1996 (see Zeng, Morgan et al. 2010 for discussions on justifications of pooling data from the four surveys).

[13] The Research Triangle Area includes six counties that share a super PUMS code in the census micro data files. The Research Triangle Area had a total population size of 541,922 persons in 2000.

County (a smaller sub-state area) and Chapel Hill (a small town) in North Carolina (NC), based on the projected household distributions of NC in 1990-2000 and 2000-2006 and the constant-share approach.[14]

The absolute relative discrepancy rates based on comparisons of the total number of households, average household size, percent of households of 1, 2-3, and 4+ persons and couple-households, total population size, % of children, elderly, oldest-old and dependency ratios and group quarters residents between the projected and the observations in 2000 census and the 2006 ACS for the 50 states and DC are summarized in Tables 3 and 4. Among the first set of tests of the 357 pairs of the main indices on family households of the projected and the 2000 census observed in the 50 states and DC, 45.9, 42.0, 11.5 and 0.6 percent of the absolute discrepancy rates are <1.0%, 1.0-2.99%, 3.0-4.99%, and 5.0-9.99%, respectively, and none is over 10% (see panel (A) in Table 3). The absolute relative discrepancy rates of the household main indices in the second set of the tests using data prior 2001 and compare the projected and ACS-observed in 2006 are relatively larger than those in the first set of the tests: 34.2, 35.0, 21.9, and 9.0 percent of the 357 pairs are <1.0%, 1.0-2.99%, 3.0-4.99%, and 5.0-9.99%, respectively, and none is over 10% (see panel (C) in Table 3). The percentage distributions of the absolute relative discrepancy rates of the main indices of population in the two sets of the tests (i.e. comparing the projected with the 2000 census observations and with the 2006 ACS observations) in the 50 states and DC are generally similar: about 34, 40-41, 14-15, 7-8, and 2-4 percent are <1%, 1.0-2.99%, 3.0-4.99%, 5.0%-9.99, and 10-14.99%, respectively, and only one (0.3 percent of the total number of projections) slightly exceeds 15% (see panels (B) and (D) in Table 3). The fourteen average rates of the absolute discrepancies of the main household indices between projected and Census 2000/ ACS 2006 observations for the 50 states and DC are all within a range of 0.8%-2.9% (see panel (A) of Table 4). The range for the fourteen averages rates of the absolute discrepancies of the main population indices between projected and Census 2000/ ACS 2006 observations for the 50 states and DC is

[14] The general data sources for obtaining the base population and estimating the demographic summary measures at state and sub-state levels used in the validation tests presented in this section and in the illustrative applications to be presented in the next section are indicated in Table 1.

somewhat larger: 0.6%- 5.4%, but the high bound (percent of oldest-old) is still reasonably small (see panel (B) in Table 4).

---- Tables 3 and 4 about here----

Among the 47 pairs of absolute relative discrepancy rates of the main indices for family households and population of the projected and the 2000 census/ 2006 ACS observed in the Research Triangle Areas, Durham county and small town of Chapel Hill, 21.2, 25.5. 19.2 and 34.1 percent of the discrepancy rates are <l.0%, 1.0-2.99%, 3.0-4.99%, and 5.0-9.99%, respectively, and none is over 10% (see Table 5).

---- Table 5 about here----

These results show that the differences are within a relatively small range. It is uncertain what portions of the discrepancies summarized in Tables 3, 4 and 5 are due to the model specification, programming, or to inaccuracies of the data. It is clear, however, that the extended cohort-component method/software for simultaneously projecting households and population work reasonably well not only at national level as shown in Zeng et al. (2006) but also at state and sub-state levels.

5. SOME ILLUSTRATIVE PROJECTIONS AT THE STATE AND SUB-STATE LEVELS

To illustrate some practical applications of the extended method, we conducted household projections from 2000 to 2050 for the 50 states and DC, as well as household and housing projections for Chapel Hill, North Carolina[15] from 2000 to 2015. Due to space limitations and given the nature of illustrative applications, we present here only summaries of these projections.

5.1. Data and parameter assumptions

[15] The town of Chapel Hill with a total population size of 48,715 persons in 2000, is used as an illustrative application to demonstrate the ability of the extended model to make household projections for very small sub-state areas.

The data sources for the projections are listed in the last column of Table 1. As discussed in Section 3.2.1, we apply the model standard schedules of race-sex-age-specific demographic rates (except migration rates) estimated based on national surveys for household projections at the state level. Based on the 2000 census 5% micro data set, we estimated race-sex-age-specific probabilities of domestic emigration from each state to the rest of the country and race-sex-age-specific frequencies of immigration from the rest of the country to each of the states.

The race-sex-specific life expectancies at birth and the race-parity-specific TFRs from 2000 to 2050 for each of the state and DC are estimated based on the medium assumptions of the Census Bureau population projection (Hollmann, Mulder, and Kallan 2000). The numbers of domestic immigrants and emigrants as well as the international net migrants for each of the states and DC are estimated based on the combined data of ACS from 2000 to 2004, which are assumed to be constant after 2004. The procedures for estimating the general rates of marriage/union formation and dissolution in 2000 for each of the 50 states and DC are discussed in Section 3.2.2. In our illustrative application of the medium projections, the race-specific general rates of marriage/union formation and dissolutions from 2000 to 2050 are simply assumed to be constant at the 2000 level.[16]

Note that we cannot derive the baseline households and population distributions for the small town of Chapel Hill from the census micro dataset, as it has no its own PUMS code. However, the following eight categories of households by race, income and the age groups <35, 35-64, 65-79 and 80+ are available from Census Bureau publicly available on-line census tabulations for states, counties, county subdivisions, towns (including Chapel Hill), subbarios, census tracts, and block group data: I, single-man only; II, single-woman only; III, a not-married-man and children/other, size 2+; IV, a not-married-woman and children/other, size 2+; V, a married couple only or a married couple with children/other, size 2+; VI, a cohabiting couple only or a cohabiting couple with children/other, size 2+; VII, men in group quarters; VIII, women in group

[16] 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 current rates provide a reasonable forecast of future rates.

quarters. Note that the on-line available categories I, II, VII, VIII are the same as the needed categories (1),(2),(10),(11) as listed in Table 2; but the available categories III, IV, V, VI are not the same as the needed (3)-(9) listed in Table 2. We use the proportional distributions of household categories (3)-(9) derived from the 5% micro data of the surrounding census PUMS area (Orange and Chatham counties combined), which includes the smaller area Chapel Hill, to decompose Chapel Hill's on-line available categories III, IV, V, VI to reasonably approximate its needed categories (3)-(9).

5.2. A summary of household projections from 2000 to 2050 for the 50 states and DC

To depict the patterns of demographic changes in family households and elderly living arrangements in the next 50 years revealed by the output of the projections for the 50 states and DC, we utilize the demographic tool of contour maps in Figures 2 (a) to (h) which present more than 20 thousand numbers in 8 maps.[17] The vertical and horizontal axis of the contour maps in Figure 2 represents the states and single calendar years from 2000 to 2050, respectively, while the different grays represent the values of the main indices of the projections. To show future patterns of changes in households and elderly living arrangements, we rank the 50 states and DC from the lowest to the highest value of the summary index being depicted in the starting year of the projection (2000). By such ranking in the starting year, it can be clearly seen how the states with lower (or higher) value of the summary index will change in the future years and the overall evolving patterns across the states and years can be learned. We summarize the following insights from the huge numerical details of the projections presented in Figures 2 (a) to (h).

---- Figure 2 about here----

As shown in Figure 2(a), the average household size would decrease considerably and pervasively in almost all states in the first half of this century, and the speed of decrease before 2020 is faster than that in the period 2021-2050. The percent of one-person households would increase substantially in all states. As

[17] The first application of contour maps tool in demographic data analysis was presented in Vaupel, Gambill, and Yashin (1985).

compared to 2000, the proportion of one-person households in 2020 in about 39.2, 33.3, and 27.5 percent of the states would increase by more than 15%, 10-14.99% and <10%. But the trend of increase in one-person households would slow considerably after 2020 (see Figure 2(b)). Clearly, American family households are expected to be smaller in all states in the next a few decades, based on the results of our household projections at the state level. Note that, adopting the Census Bureau's medium fertility variant, we assume that the Total Fertility Rates in the United States would increase slightly in the next a few decades and thus continuously retain at a relatively high level as compared to Europe and Eastern Asian (including China, Japan and Korea), etc. If the U.S. fertility level dropped to the low level in the forthcoming decades as did European and Eastern Asian countries, American family households would become even smaller.

Husband-wife households would decrease moderately. As compared to 2000, the proportion of married-couple households among all households in 41.1, 41.2, and 17.7 percent of the states in 2050 would decrease by <10%, 10-14.99% and >15% (see Figure 2(c)). However, cohabiting-couple households would increase considerably up to 2020 or so: its proportion in about 23. 5, 11.8, 25.5 and 39.2 percent of the state will be higher than that in 2000 by >30%, 20-29.9%, 10-19.9% and <10%, respectively (see Figure 2(d)). The proportion of cohabiting households would remain relatively stable after 2020 in all states.

Directions of changes in the percent of single-parent households among the two-generation households in the first half of this century are diversified, increase moderately in some states but decrease moderately or remain more or less unchanged in the other states (see Figure 2 (e)). Such pattern may be explained by the opposite effects of moderate decline in marriages and considerable increase in cohabitation, plus the stable divorce and union dissolution rates.

The aging of American households trends shown in contour maps in Figures 2(f), 3(g) and 3(h) are striking. As compared to 2000, the proportion of elderly households (with at least one member aged 65+) in about 21.6, 39.2, 27.5 and 11.8 percent of the states in 2020 will increase by >30%, 20-29.9%, 10-19.9% and <10%; and in 2050, the elderly households in 37.3, 37.2, 23.5 percent of the states will increase by >60%, 40-59.9%, 23-40%, and only DC which attracts a lot of young immigrants is an exception (see Figure 2(f)).

By the middle of this century, elderly households will have doubled in Colorado, Hawaii and New Hampshire and nearly tripled in Alaska. Very similarly to the pattern of increase in elderly households, the proportion of elderly aged 65+ living alone will increase dramatically and pervasively across all states (see Figure 2(g)). Figure 2(h) demonstrates that the oldest-old aged 80+ living alone will even more dramatically increase in the next a few decades across all states: as compared to 2000, the proportion of oldest-old living alone will be more than triples in 5 states (Louisiana, South Carolina, Hawaii, New Hampshire and Alaska), more than doubled in slightly more than half of the states, increase by 45-99.9% in 37.3 percent of the states, and increase by 17.3 percent in DC.

5.3. A summary of household and housing projections for Chapel Hill

Based on the trend extrapolation method with constant-share assumption and the household projections of the parental state NC, we projected the eleven categories of the households by type and size (as shown in Table 2) for the Chapel Hill. We use the census data to estimate age-sex-household type/size-income specific homeownership rates and home-renter rates (the sum of these two rates of one household category is equal to one). We then multiply the homeownership rates and home-renter rates by the corresponding type/size/age/race/income specific projected numbers of households to yield projected future housing demands up to 2015. Due to space limitations, we briefly summarize the main results here, and the detailed numerical results and discussions are referred to Wang, Gu and Zeng (2006).

The results show that the relative increase in households who own a housing unit over the period 2005-2015 in Chapel Hill would be 17.2%, and about 1,400 households who own a housing unit will be added in Chapel Hill. The number of households with a rental unit in Chapel Hill will increase from 12,176 in 2005 to 14,176 in 2015, representing a 16.4 percent increase. The aging of households and owned-housing market trend in Chapel Hill is remarkably striking: the households with housing units owned by elderly aged 65+ will occupy the largest share (49%) of the increase in households who own a housing unit after 2012. The projection results also show that the majority of the cumulative increase in households with a rental housing unit will be mainly one-person or 2-person households with younger householders aged less than 35 in Chapel Hill where The

University of North Carolina is located.

6. DISCUSSION AND CONCLUSION

Applying the 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 household and population projections at the state level requires only a census or ACS micro data file and the projected (or assumed) demographic summary measures. After projections for the parental state are done, and using the widely recognized ratio trend extrapolation method, corresponding projections for a sub-state area requires only a census micro data file (for those with a PUMS code) or the publicly available on-line census tabulations (for small areas without PUMS codes). We conducted validation tests of household projections from 1990 to 2000, and from 2000 to 2006 and compared the main indices of the projected and observed in 2000 and in 2006 for the 50 states and DC, as well as the sub-state areas of the Research Triangle Area, Durham County and the small town of Chapel Hill. The results show that the extended cohort-component method for household and population projection works well, with most absolute discrepancy error rates less than three percent of observed values, at state and sub-state area levels.

For the purpose of illustrative applications, we calculated household and population projections from 2000 to 2050 for each of the 50 states and DC. To our knowledge, these are the first comprehensive household projections by race, age of the householders, and various types/sizes using conventional demographic rates as input for all of the states and DC in the United States. As an illustrative application to the small sub-state areas, we also calculated household and housing projections for the small town of Chapel Hill.

Some important limitations in our illustrative projections should be noted. First, the results are our "medium projections" only; interval projections with high and low bounds to reveal the uncertainties have not yet been conducted. These will be investigated in further study, as was done for U.S. national household projections by Zeng et al. (2006). Second, our projections are based on trend extrapolations or expert opinions regarding the demographic summary measures using available demographic data. Thus far, we have not included other socioeconomic factors relevant to changes in demographic components. In sum, this article has focused on a presentation of the methodological core ideas of the extended method, data and estimation issues, and the presentation of illustrative main results without much detail, due to space limitations. Further studies could take advantages of using the detailed outcomes of household projections produced by the new approach to conduct more sophisticated investigations of socioeconomic planning and household consumption market analysis, as well as incorporating social and economic factors into the model.

References

Bongaarts, J. 1987. "The projection of family composition over the life course with family status life tables." Pp189-212 *Family demography: Methods and applications,* edited by J. Bongaarts (eds.). Oxford; Clarendon Press.

Brass, W. (1983), "The formal demography of the family: An overview of the proximate determinates." In *The Family*, Occasional Paper 31. London: Office of Population censusese Surveys.

Chappell, N. 1991. "Living arrangement and sources of caregiving." *Journal of Gerontology: Social Sciences 46:* S1-S8.

Coale, A.J., P., Demeny, and B. Vaughn, B. 1983. *Regional model life tables and stable populations.* New York: Academic Press.

Coale A.J. 1984. "Life table construction on the basis of two enumerations of a closed population." *Population Index* 50: 193-213.

----- 1985. "An extension and simplification of a new synthesis of age structure and growth." *Asian and Pacific Forum* 12 : 5-8.

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 September 2, 2004, http://www.ppacg.org/Trans/2030/Volume%20III/Appendix%20G%20-%20Small%20Area%20Forecasts.pdf.

Dalton, M., B. O'Neill, A. Prskawetz, L. Jiang, and J. Pitkin. 2008. "Population aging and future carbon emissions in the United States." *Energy Economics* 30, 642-675.

Davis, L. W. 2003. *A dynamic model of demand for durable goods that consume energy.* University of Wisconsin – Madison. Retrieved August 29, 2004, http://www.ssc.wisc.edu/~ldavis/es.pdf

Davis, L.W. 2004. *The role of durable goods in household water and energy consumption: The case of front loading clothes washers.* University of Wisconsin – Madison. Retrieved September 3, 2004, from http://rider.wharton.upenn.edu/~wred/Seminar%20Papers%2004-05/Lucas%20Davis%20cw.pdf

Doty, P. 1986. "Family care of the elderly: The role of public policy." *Milbank Memorial Fund Quarterly* 64:34-75.

Federal Interagency Forum on Aging Related Statistics (FIFARS). 2004. *Older Americans 2004: Key indicators of well-being.* Federal Interagency Forum on Aging-Related Statistics. Washington, DC: U.S. Government Printing Office. November 2004.

Foncel, J., and M. Ivaldi. 1999. *Econometric modeling of differentiated durable goods markets: An application to telephone.* The Institut d'Economie Industrielle (IDEI) working paper. Retrieved August 12, 2004, from http://idei.fr/activity.php?a=700.

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 Ww know and what do we need to know?" *Research on Aging* 12(4) 4: 531-545.

Himes, C. L. 1992. "Future caregivers: Projected family structures of older persons." *Journal of Gerontology: Social Sciences* 47(1) : S17-26

Hollmann, F. W., T.J. Mulder and J.E. Kallan. 2000. *Methodology and assumptions for the population projections of the United States: 1999 to 2100.* Population Division Working Paper No. 38. Washington, D.C.: U.S. Bureau of the Census.

Ip, F., and D. McRae. 1999. *Small area household projections - A parameterized approach.* Population Section, Ministry of Finance and Corporate Relations, Province of British Columbia, Canada. Retrieved September 1, 2004, from http://www.bcstats.gov.bc.ca/data/pop/pop/methhhld.pdf

Keilman, N. 1985. "Nuptiality Models and the Two-Sex Problem in National Population Forecasts." *European Journal of Population* Vol. 1(2/3), pp. 207-235.

Keilman, N. 2003. “The threat of small households.” *Nature* 421: 489-490.

Land, K. C., and A. Rogers. 1982. *Multidimensional mathematical demography*. New York: Academic Press.

Liu, J., C.C. Dally, P.R. Ehrlich, and G.W. Luck. 2003. “Effects of household dynamics on resource consumption and biodiversity.” *Nature*, 421,530-533.

Lutz, W., and C Prinz. 1994. “The population module.” Pp 221-231 in *Population-Development - environment: Understanding their interactions in Mauritius*, edited by W. Lutz. Berlin; Springer-Verlag Press.

Mackellar, F. L., W.Lutz, C. Prinz, and A . Goujon. 1995. “Population, households, and CO2 emissions.” *Population and Development Review* 21: 849-866.

Mofitt, R. 2000. “Demographic change and public assistance expenditures.” Pp 391-425 in *Demographic change and public assistance expenditures, edited by* A.J. Auerbach and R. D. Lee. Cambridge, UK: Cambridge University Press.

Morgan, Philip. 2004. Between-state differentials in demographic rates are mostly caused by differences in racial compositions rather than different age patterns of race-specific rates.” Personal email communication.

Morris, R., F.G. Caro, and J.E. Hansan 1998. *Personal assistance: The future of home care.* Baltimore: the Johns Hopkins University Press.

Myers, D., J. Pitkin, and J. Park. 2002. “Estimation of housing needs amid population growth and change.” *Housing Policy Debate* 13(3): 567-596.

Pollard, J.H. 1977. “The continuing attempt to incorporate both sexes into marriage analysis.” In Volume 1 of the papers of the General Conference of the International Union for the Scientific Study of Population, Mexico City, 1977.

Preston, Samuel H., Patrick Heuveline and Michel Guillot. 2001. *Demography: Measuring and Modeling Population Processes* London: Blackwell Publishers.

Prskawetz, A., L. Jiang, and B. O'Neill. 2004. "Demographic composition and projections of car use in Austria." Pp.274-326 in *Vienna yearbook of population research 2004,* edited by T. Fent and A. Prskawetz. 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.

---- 1995. *Multiregional Mathematical Demography: Principles, Methods, and Extensions*. New York: Wiley.

Schoen, R. 1988. *Modeling multi-group populations.* New York: Plenum Press.

Schoen, R.1981. "The harmonic mean as the basis of a realistic two-sex marriage model." *Demography* 1981, 18, 201-216.

Schoen, Robert, and Nicola Standish. 2001. "The Retrenchment of Marriage: Results from Marital Status Life Tables for the United States, 1995." *Population and Development Review* 27(3):553-563.

Smith, S. K. 2003."Small area analysis," Pp.898-901 in *Encyclopedia of population, edited by* P. Demeny and G. McNicoll . Farmington Hills, MI: Macmillan Reference.

Smith, S.K., and P.A. Morrison. 2005. "Small area and business demography." Pp.761-786 in *Handbook of Population, edited by* D.L. Poston and M. Micklin. New York: Kluwer Academic/Plenum Publishers.

Smith, S. K., J. Tayman, and D.A. Swanson. 2001. *State and local population projections: Methodology and analysis.* New York: Kluwer Academic/Plenum Publishers.

Soldo, J., D.A. Wolf, and E.M. Agree. 1990. "Family, households, and care arrangements of frail older women: A structural analysis." *Journal of Gerontology: Social Sciences* 45: S238-249.

Stupp P.W. 1988. "A general procedure for estimating intercensal age schedules." *Population Index* 54: 209-234.

Treadway, R. 1997. *Population projections for the state and counties of Illinois*. Springfield: State of Illinois.

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.

Wang, Z.L. Y. Zeng, and D. Gu. 2005. *The vehicle ownership rates by household type, size, income, race, and region.* GM project phase II report No.2. The Household and Consumption Forecasting, Inc. Chapel Hill, NC.

Wang, Z., D. Gu, and Y. Zeng. 2006. Households and Housing Forecasts at State and Small Area Levels --A New Approach and Applications to the State, Two Counties and One Town of North Carolina. Paper selected for presentation at the Session on "Business Demography" of the Annual Meeting of the Population Association of America, March 30 – April 1, 2006, Los Angeles. Available on line: http://*paa2006.princeton.edu/download.aspx?submissionId=61200*

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(1): 129-144.

Zeng, Y. 1986. "Changes in family structure in China: A simulation study." *Population and Development Review* 12(4): 675-703.

---- . 1988. "Changing demographic characteristics and the family status of Chinese women." *Population Studies* 42: 183-203.

---- . 1991. *Family dynamics in China: A life table analysis.* Madison, WI: The University of Wisconsin Press.

Zeng, Y., A. Coale, M.K. Choe, Z. Liang, and L. Liu. 1994. "Leaving parental home: census based estimates for China, Japan, South Korea, the United States, France, and Sweden." *Population Studies* 48(1):65-80.

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(6): 525-549.

Zeng, Y., E. Stallard, and Z. Wang. 2004. "Computing Time-Varying Sex-Age-Specific Rates of Marriage/Union Formation and Dissolution in Family Household Projection or Simulation." *Demographic Research* 11: 264-299.

Zeng, Y., K.C. Land, Z. Wang, and D. Gu. 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., J.W. Vaupel, and Z. Wang. 1997. „A multidimensional model for projecting family households -- with an illustrative numerical application.” *Mathematical Population Studies* 6(3): 187-216.

—— 1998. “Household projection using conventional demographic data.” *Population and Development Review, Supplementary Issue: Frontiers of Population Forecasting* 24: 59-87.

Zeng, Yi, S. Philip Morgan, Zhenglian Wang, Danan Gu, Chingli Yang (2010). A Multistate Life Table Analysis on Union Regimes in the United State -- Trends and Racial Differentials, 1970-2002, manuscript submitted for consideration of publication.

Table 1. Data and data sources for projecting U.S. households using the extended cohort-component method at the national, state, and sub-state levels

Contents of the data	*Main data source for U.S. applications*
(1) Base Population for the nation, states and sub-state areas. A census or a exceptionally large survey micro data file 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. If a sample data set is used, published 100% census tabulations of age-sex-marital status distributions of the entire population including the elderly and 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.	Census 5% micro data or more recent and cumulative American Community Survey (ACS) data files and the published online 100% census or ACS cross- tabulations.
(2) Model standard schedules at the national level (not necessary for the states and sub-state areas)	
(a) Race-sex-age-specific death rates (marital-status specific, if possible).	Census Bureau's estimates, Schoen and Standish (2001)
(b) Race-sex-age -specific o/e rates of marriage/union formation and dissolution (c) Race-age-parity-specific o/e rates of marital and non-marital fertility	Pooled NSFH, NSFG, CPS, SIPP data sets, see Zeng and Land et al. (2006) and Zeng and Morgan et al. (2010).
(d) Race-sex-age -specific net rates of leaving the parental home, estimated based on two adjacent census micro data files and the intra-cohort iterative method (Coale1984; 1985; Stupp 1988; Zeng, Coale et al., 1994).	The 1980, 1990, and 2000 censuses micro data files
(e) Race-sex-age-specific rates of emigration and immigration.	Census 5% micro data or ACS data files
(3) Demographic summary measures for the nation and states (not necessary for sub-state areas)	
(a) Race-specific general rates of marriage and general rates of divorce (b) Race-specific general rates of cohabiting and general rates of union dissolution	Based on vital statistics and pooled survey data sets
(c) Race-specific Total Fertility Rates (TFR) by parity (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	Based on estimates released by the Census Bureau and the National Center for Health Statistics

Note: the race dimension is distinguished in the U.S. households forecasting, but it can be omitted in the other countries' applications if the race differentials are not crucial, or the race-specific data are not available, or the sub-population sizes of the minority race groups are too small.

Table 2. Illustrative example of eleven categories of households by type and size for forecasting at the sub-state level, using the ratio trend extrapolation methods

Category	Type (*h*)	Size(*s*)	Characteristics of the household category
1	1	1	One-single-man only, household size 1
2	1	2-3	One-single-man with child/other, household size 2-3
3	1	4+	One-single-man with child/other, household size 4+
4	2	1	One-single-woman only, household size 1
5	2	2-3	One-single-woman with child/other, household size 2-3
6	2	4+	One-single-woman with child/other, household size 4+
7	3	2	One-couple only, household size 2
8	3	3	One-couple household with child/other, household size 3-4
9	3	4	One-couple household with child/other, household size 5+
10	4	1	Men living in group quarters
11	5	1	Women living in group quarters

Note: "Single-" refers to not-married (including never-married, divorced, or widowed) and not-cohabiting persons. "One-couple" refers to a married or cohabiting couple.

Table 3. Distributions of the absolute relative discrepancies rates of comparisons between extended cohort-component projections and Census 2000/ ACS 2006 observations for the 50 states and DC

	ProFamy projected and Census 2000 obs.				ProFamy projected and ACS 2006 obs.			
Absolute Relative Discrepancy rates	(A) households		(B) population		(C) households		(D) population	
	# pairs Compa.	percent	# pairs Compa.	percent	# pairs Compa.	percent	# pairs Compa.	percent
< 1.00%	164	45.94	123	34.45	122	34.17	120	33.61
1.0-2.99%	150	42.02	142	39.77	125	35.02	145	40.62
3.0-4.99%	41	11.48	51	14.29	78	21.85	54	15.13
5.0-9.99%	2	0.56	26	7.28	32	8.96	30	8.40
10.0-14.99%	0	0	14	3.92	0	0	7	1.96
>15.0%	0	0	1	0.28	0	0	1	0.28
Total	357	100.00	357	100.00	357	100.00	357	100.00

Table 4. Averages of absolute discrepancy rates (%) between projected and 2000 Census 2000 /ACS 2006 observations for the 50 states and DC

	(A) Av. abs. discrepancy rate (%) of main indices of household projection						
	Tot. # hh	Av. hh size	%1 pers hh	%2-3 pers hh	%4+ pers hh	%Couple hh	Group q.
ProFamy proj. vs. census obs. 2000	1.30	0.81	2.01	1.86	2.19	1.23	0.81
ProFamy proj. vs. ACS obs. 2006	2.84	2.92	2.68	1.75	2.66	1.27	0.94
	(B) Av. abs. discrepancy rate (%) of main indices of population projection						
	Pop size	%child <18	%old 65+	%oldest-old	DR of children	DR of old	Total DR
ProFamy proj. vs. census obs. 2000	1.23	1.29	2.74	4.43	1.68	3.69	1.72
ProFamy proj. vs. ACS obs. 2006	0.60	2.02	2.06	5.35	2.53	2.38	1.38

Note: DR – dependency ratio.

Table 5. The relative discrepancies rates (%) of comparisons between projected and Census 2000/ ACS 2006 observations for the Triangle Area and Durham

		Relative discrepancy rate (%) of main indices of household projection						
	Pop size	Tot. # hh	Av. hh size	%1 pers hh	%2-3 pers hh	%4+ pers hh	%Couple hh	Group q.
Proj. vs. census obs. 2000 Triangle Area	-0.4	-0.2	0.0	-2.1	-2.0	7.9	-0.9	-4.8
Proj. vs. ACS 2006 Est. Triangle Area	0.8	5.6	-4.8	4.8	-3.7	2.1	-1.0	6.2
Proj. vs. census obs. 2000 Durham	-2.7	0.2	-2.7	-0.8	0.5	-0.1	7.6	-7.0
Proj. vs. ACS 2006 Est. Durham	-3.0	2.0	-5.4	4.0	-6.0	8.4	-2.2	8.6
Proj. vs. census obs. 2000 Chapel Hill	-0.5	-4.0	-2.1	4.7	-9.9	4.0	1.7	-5.0
Proj. vs. ACS 2006 Est. Chapel Hill	-2.1	-1.6	-7.8	8.9	-5.0	-5.0	-5.3	**

** Reliable 2006 ACS data on group quarters residences in Chapel Hill are not available.

Figure 1. Computational strategy to calculate changes in marital/union, co-residence with parents/children, migration and survival statuses

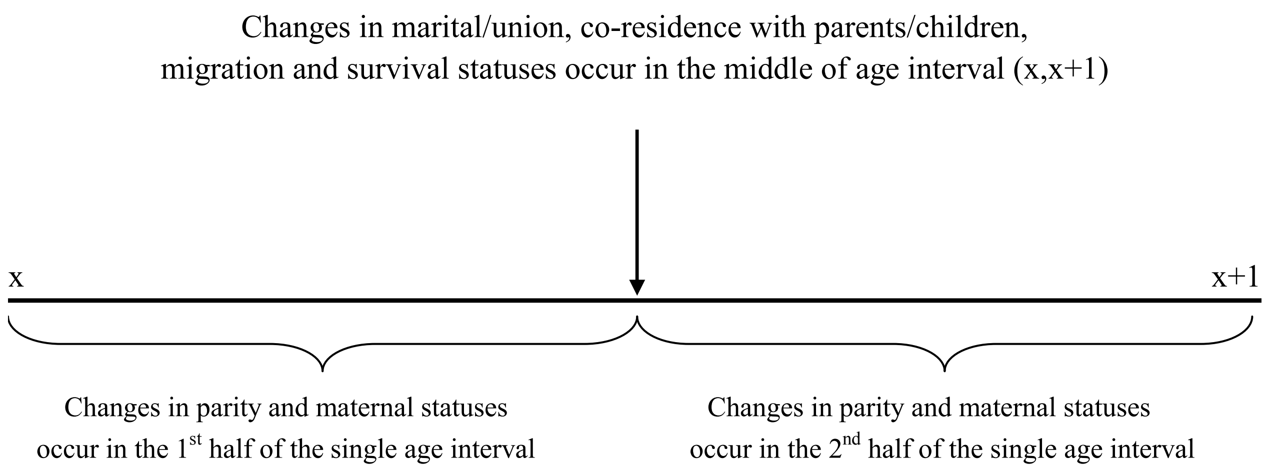
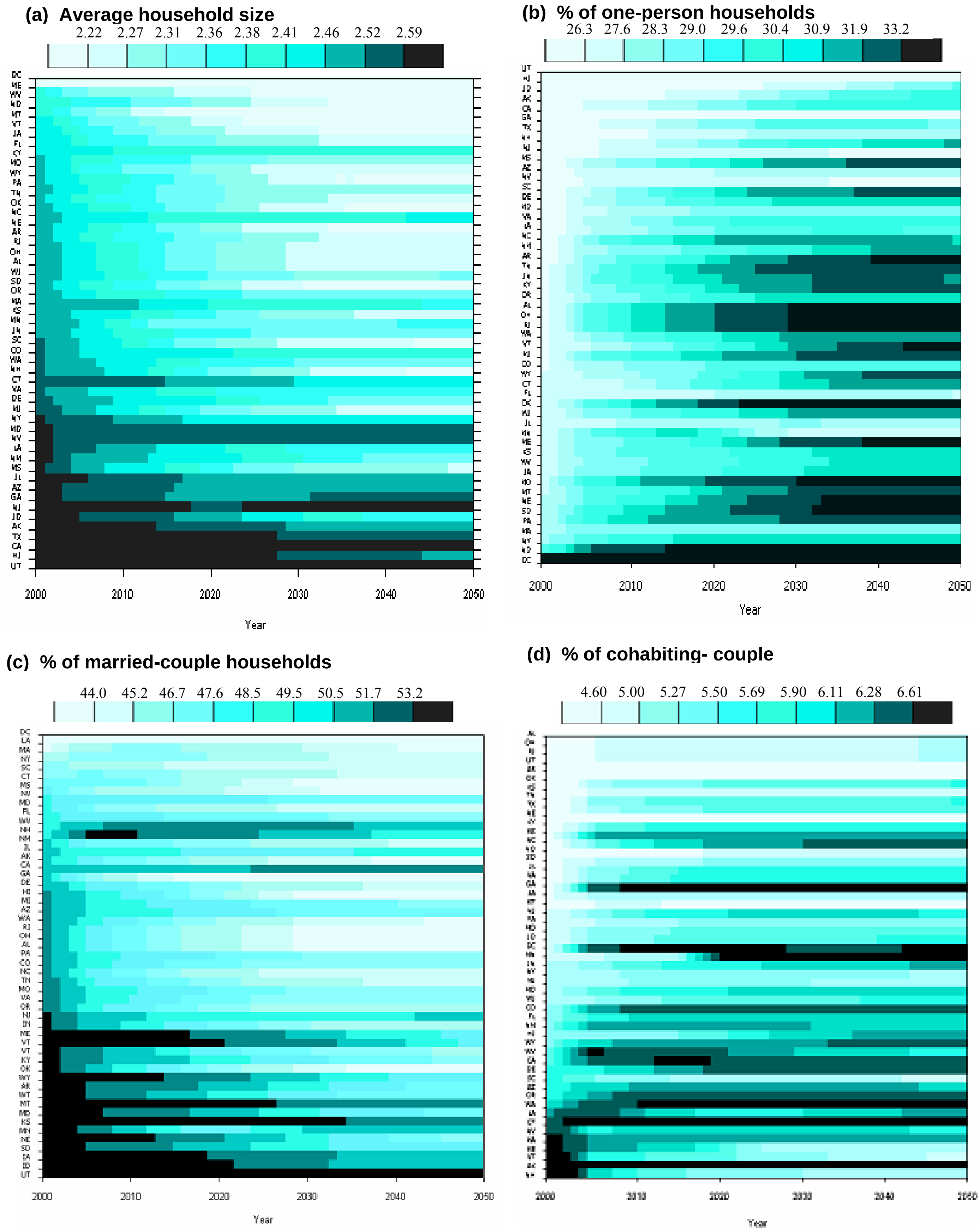
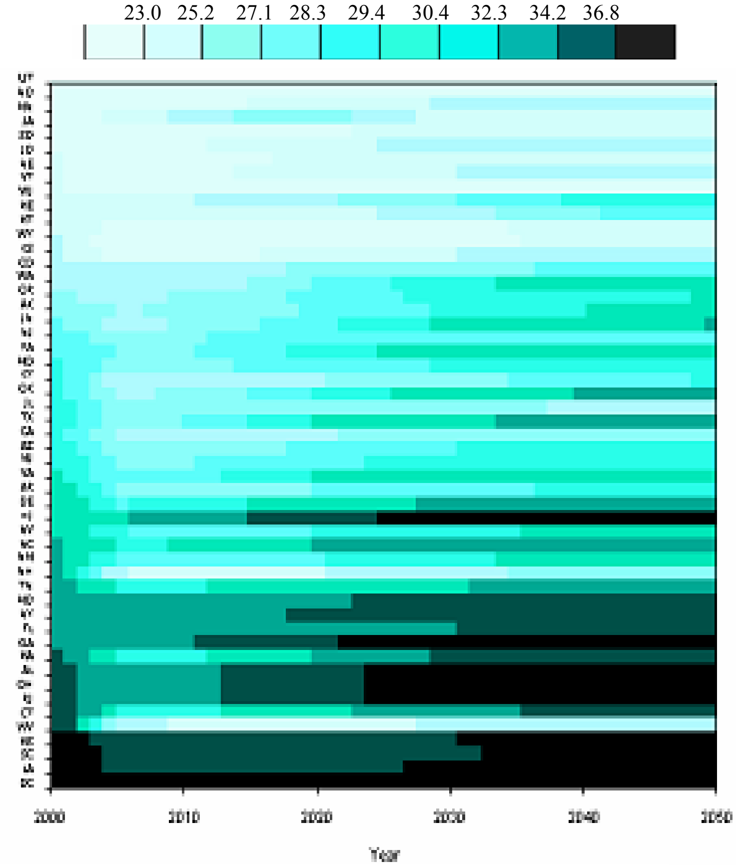


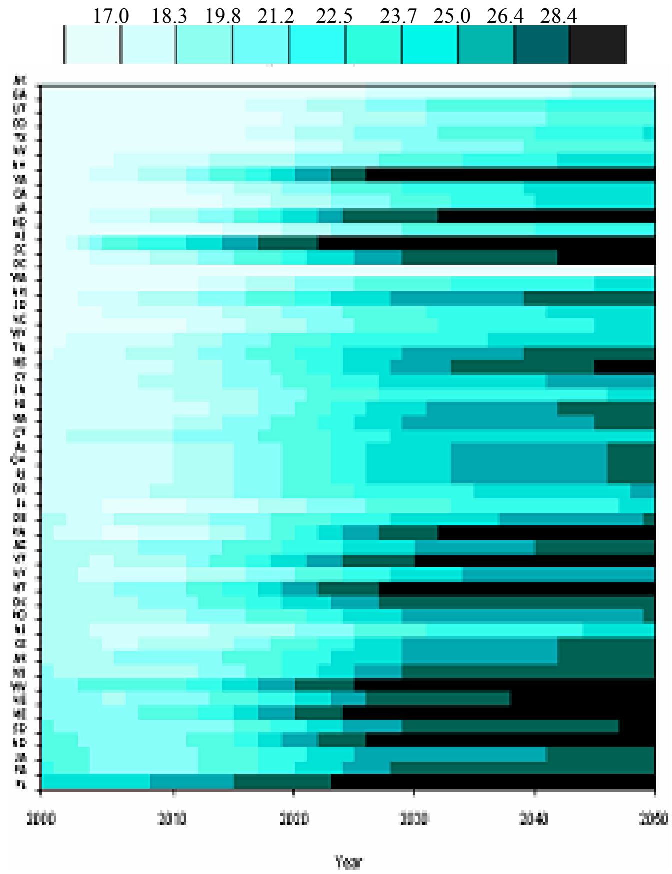
Figure 2. Projections of households and elderly living arrangement for 50 states and DC, 2000-2050



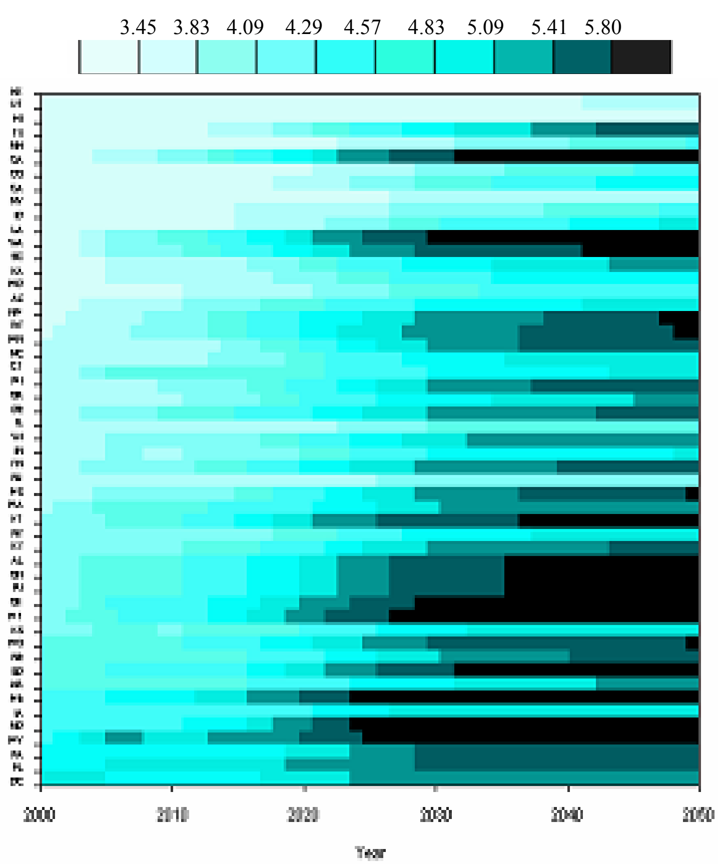
(e) % of single-parent households among 2-generation households



(f) % of elderly households



(g) % of the elderly aged 65+ living alone



(h) % of the elderly aged 80+ living alone