



Max-Planck-Institut für demografische Forschung
Max Planck Institute for Demographic Research
Konrad-Zuse-Strasse 1 · D-18057 Rostock · GERMANY
Tel +49 (0) 3 81 20 81 - 0; Fax +49 (0) 3 81 20 81 - 202;
<http://www.demogr.mpg.de>

MPIDR WORKING PAPER WP 2010-028
SEPTEMBER 2010

Household and Population Projections at Sub-National Levels: An Extended Cohort-Component Approach

Yi Zeng (zengyi68@gmail.com)
Kenneth C. Land (kland@soc.duke.edu)
Zhenglian Wang (wangzl8886@gmail.com)
Danan Gu (gudanan@duke.edu)

This working paper has been approved for release by: James W. Vaupel (jwv@demogr.mpg.de),
Head of the Laboratory of Survival and Longevity.

© Copyright is held by the authors.

Working papers of the Max Planck Institute for Demographic Research receive only limited review.
Views or opinions expressed in working papers are attributable to the authors and do not necessarily
reflect those of the Institute.

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, National School of Development, 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 the 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 comparing projected with the census-observed counts in 2000 for the 50 states and the District of Columbia and for sets of randomly selected 25 counties and 25 cities which are more or less evenly distributed across the United States. The comparisons show that most absolute percent errors of the main indices of household and population projections and the corresponding census 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 the small town of Chapel Hill, North Carolina 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.

Key words: household projections, cohort-component approach, sub-national household projections, household projections at the state level, sub-state area household projections, aging of households, aging of housing markets.

1. INTRODUCTION

There is a growing demand for projections not only of population size but of the distribution of the 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; Myers 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, sub-state areas such as counties

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.

and cities, and small areas² (Treadway, 1997; Ip and McRae, 1999; Rao 2003; Crowley, 2004).

Sub-national households and population projections are useful for distributing government funds, allocating various types of resources, planning the development of infrastructure and 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; Swanson and Pol, 2009).

Most household projections at sub-national 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 (Burch 1999; Mason & Racelis 1992; Spicer et al. 1992; Murphy 1991; Bell & Cooper 1990). First, the designation of a household head is vague, ill-defined, and an arbitrary choice, making projections difficult (Murphy 1991). Trends in headship rates are thus not easy to model (Mason & Racelis 1992:510). For instance, an increase or decrease in female headship rates may occur because the census or survey was carried out in the daytime or evening, when more or less women were available to complete the questionnaire than were men. Or the increase or decrease may be due to an actual change in women's socioeconomic status. But in either case, it is not due to real changes in demographic conditions. Second, there is no way to link headship rates to demographic rates, given the fact that the numbers of households in future years are projected by cross-sectional extrapolation of the headship rates; graphical illustration of this is given in panel (B) of Figure 1. It thus is impossible to incorporate projected or assumed changes in the propensity/timing of demographic processes into headship rates (Mason & Racelis 1992; Spicer et al. 1992). Third, the headship-rate lumps all household members other than heads

² Sub-state level usually refers to counties or cities/towns. The “small-area” term refers to small towns and places, possibly even tracts or block groups which have a small population size.

into one category “non-head” with no projected information (Burch 1999). This makes it impossible to study family households, marital status and living arrangements of the elderly, adults, and children who are “non-head”, whose forecasted information also is useful for business/academic/policy analysis and planning. Fourth, the information on households produced by headship-rate projections is very limited and inadequate for purposes of more detailed planning and analysis (Bell & Cooper 1990). A typical well-done national household projection using the headship-rate method projected only five household types by age groups of household head or householder (Bureau of the Census 1996), with no projected household sizes available.³ This is, again, disadvantageous since households with various sizes differ substantially in their needs for products and services. For example, Prskawetz et al. (2004) and Wang et al. (2006) found that the headship-rate method yields seriously misleading forecasts of the increase in automobile use in Austria and increase in housing demand in the U.S., through multiplying the forecasted number of households (without size information) by the average number of automobiles or the average housing demand per household derived from the census or a recent survey. This is because future Austrian and American households will comprise many more one- and two-person households (which mostly need only one car, or need smaller housing unit/apartment) than do today's average households, but the headship-rate method cannot forecast households by size.

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

³ Decennial census data could provide a more detailed stratification of headship rates than what was done by the Bureau of the Census, if one assumed that the sex-age specific headship rates are constant over the projection time horizon. This would produce more detailed household projections, but the static approach of assuming constant sex-age-specific headship rates does not take into account changing demographic rates and thus may quickly depart from population trends.

In contrast to the headship-rate method, the extended cohort-component model for households and population projections (abbreviated as “ECC” model hereafter), as discussed and justified in details in Zeng et al. (2006), does not suffer the vague, ill-defined and an arbitrary chosen designation of the household heads; it projects all individuals grouped by cohorts and specified attributes, for instance, a group of persons of the same race, sex, age, marital/union status and co-residence status with parents and children. The calculations proceed iteratively, group-by-group, cohort-by-cohort and time-period-by-time-period, using demographic rates as input (see panel (A) in Figure 1); it projects much more detailed household types, sizes, and living arrangements for all members of the population⁴. Note that detailed forecasts of household sizes by various types are particularly useful in market analysis. For example, after Prskawetz et al. (2004) and Wang et al. (2006) found that the headship-rate method yields seriously misleading forecasts of increases in automobile use and housing demand, they applied the ECC approach and produced much more realistic and detailed forecasts of future households (by size, types, and age of the reference persons) and the resulting automobile use and housing demand market potentials in Austria and the U.S.

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

⁴ The choice between methods with different degrees of comprehensiveness depends on the user’s needs. If a simple static projection of the number of households with limited types without household size information using easily accessible cross-sectional data at a low cost is sufficient for the purpose of the project, the headship-rate approach may be the good choice. For in-depth analyses for socioeconomic planning, market study, academic simulation, or policy scenarios, which need detailed projections of household types, sizes, and elderly living arrangements using various demographic rates as input, the ECC approach is preferable.

Tests of projections from 1990 to 2000 using the ECC model and based on observed U.S. demographic rates before 1991 show that forecast errors measured by discrepancies between the projected values and the U.S. 2000 census observations are reasonably small, validating the new ECC model (Zeng et al., 2006). Using data from national surveys and vital statistics, census micro files, and the extended model, Zeng et al. (2006) presented projections of U.S. households from 2000 to 2050. Medium projections as well as projections based on smaller and larger family scenarios with corresponding combinations of assumptions of time-varying race-sex-age-specific marriage/union formation and dissolution, fertility, mortality, and international migration rates were performed to analyze future trends of U.S. households and their possible higher and lower bounds, as well as enormous racial differentials.

However, Zeng et al. (2006) dealt with households and population projection at the national level only; important questions concerning whether the ECC model also works reasonably well at sub-national level were not addressed. Motivated by the growing scientific and practical needs of household projections at sub-national levels, this paper further develops the ECC model for applications to sub-national household and population projections. We describe in the next section the core ideas of the ECC model for the projection of household composition and population distributions. In the third section, we discuss data and estimation issues. The fourth section presents validation tests of household projections from 1990 to 2000 using the ECC model through comparing the projected and the census observed in 2000 for the 50 states, the District of Columbia (DC), 25 randomly selected counties and 25 randomly selected cities (each state has one randomly selected county or city). 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 MODEL

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 the early version of the ECC model which was a two-sex dynamic macro model for projections of households and living arrangements. Zeng et al. (2006) presented a substantial extension of the ECC model and applied it to U.S. household projections at the national level. The ECC model is built on five core ideas, which now are described.

Core idea 1: A multi-state accounting model. The innermost core of the ECC model 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 $t+1$. In the ECC model, groups of individuals are the basic units of the analysis and all individuals of the population are grouped and projected forward cohort by cohort and by age, sex, marital/union status, parity, and number of co-residing children and parents (Zeng et al., 2006). Why does the ECC model use groups of individuals as the units of analysis, in contrast to the most other macrosimulation models which use groups of households as the units? This is 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 as they are not available in vital statistics, censuses, or ordinary surveys (Ledent 1992; Van Imhoff & Keilman

1992; Keilman 1988). As stated by Van Imhoff and Keilman (1992), such a stringent data demand is an important factor in the slow development and infrequent application of these models. The distribution of households by size is impossible to project unless the size of the household is explicitly incorporated with the household types into the state space (Van Imhoff et al. 1995:348). Incorporation of household size into the household-type state space would greatly increase the size of the transition probabilities matrices that need to be estimated at each age for men and women, which is likely not feasible in practical applications, due to sample size limitations. Furthermore, the household-status-transition-based models cannot directly link changes in household structure with demographic rates. Thus, it is difficult for such a model to identify the impacts of demographic factors on changes in household structure.

The ECC model uses groups of individuals as the units of analysis. Consequently, only conventional and normally available census, surveys 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 ECC model keeps track of individuals' marital/union status, statuses of co-residence with one or two parent(s) and 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 statuses⁵, 3 statuses of co-residence with parents, 6 parity (i.e. parity 0, 1, 2, 3, 4, 5+, while grouping parities 5 and higher into one category of 5+) and 6 co-residence statuses with children are distinguished as in the U.S. household projections of Zeng et al. (2006), one would have to estimate a cross-status transition probabilities matrix with 194,481 elements ($7 \times 3 \times \sum_{p=0}^5 (p+1) = 441 \times 441 = 194,481$)⁶ at each age of each sex for each race group. This is certainly not practical, as it would be impossible to have a huge dataset with appropriate size of the sub-samples to reasonably estimate so many (e.g. 194,481) elements of the cross-status transition probabilities matrix at each age of each sex for each race group. 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 2). This strategy, which was originally proposed by Bongaarts (1987) and further justified

⁵ 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.

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

mathematically and numerically by Zeng (1991: 61-63 & 80-84), circumvents the problems of estimating huge matrices of cross-status transition probabilities.

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

Core idea 3: A judicious use of stochastic independence assumptions. Coupled with core idea 2, the third core idea of the ECC model 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 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 Imhoff and Keilman, 1992; Zeng et al. 1997; 1998).

Core idea 5. Employing national model standard schedules and summary parameters at sub-national levels to specify projected demographic rates of the sub-national area in future years. As Keyfitz (1972) pointed out, a trend extrapolation of each age-specific rate can result in

an excessive concession to flexibility and, therefore, readily produce erratic results.

Consequently, as in many applications of the conventional cohort-component population projection model, the ECC household projection model at national and sub-national levels uses a set of race-sex-age-specific standard schedules of demographic rates (see (2)-a and (2)-b in Table 1) and the summary parameters thereof (see (3) in Table 1) to specify projected demographic rates in future years. 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, as well as changes in the shape of the age-specific schedules towards to be more spread or more concentrated, through adjusting the parameters (mean or median, and interquartile range) of the standard schedules (Zeng et al., 2000). The 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.

Note that 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 ((2)-a in Table 1) for household projection may not be available at the sub-national level. However, once the age-race-sex-specific standard schedules at the national level are prepared (and updated when the new data become available)⁷, they can be employed as *model standard schedules* for projections at the state level. This is similar to the widely practiced application of

⁷ For the U.S., we have estimated model standard schedules based on national datasets. Based on these national model standard schedules and demographic summary parameters (including mean ages at marriage and fertility), estimates of the age-specific demographic rates can be generated for projecting households and population to future years (see Zeng, Stallard and Wang 2004). With the model standard schedules in hand, the analysts can concentrate on projecting future demographic summary parameters. This can be done 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, etc.) also can be used in projecting the demographic summary parameters.

model life tables (e.g., Coale, Demeny, and Vaughn, 1983; U.N., 1982), the Brass logit relational life table model (e.g. Murray, 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 (Brass, 1978; Booth, 1984; 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 which affect the projections or estimations, but the race-age-specific model standard schedules themselves typically are not highly sensitive to the projection and estimation results as long as they reveal the general age patterns of demographic processes of the population.

Two published tests have corroborated the empirical applicability of the core idea 5 of the ECC model which employs national model standard schedules and summary parameters at sub-national level to specify projected demographic rates of the sub-national area in future years. In one of the tests, Zeng et al. (2006) performed two scenarios of U.S. national projections by race from 1990 to 2020 with all race-specific demographic summary parameters being identical each other, but with one scenario using the race-sex-age-specific rates observed in 1990s and the other scenario using the race-sex-age-specific rates observed in 1980s. The 17 main indices of the households and population projections then were compared for the years 2000, 2010, and 2020 for these two scenarios. The results showed that, while the projected input summary parameters 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 relative

discrepancy rates of the projected main indices between these two scenarios are less than 1% and the other one-third are 1.0-3.4% (see Appendix A in Zeng et al. 2006).

In another empirical evaluation, Zeng et al. (2000) applied the extended Brass relational Gompertz model, the observed summary parameters in a population, and a model standard schedule derived from another population with similar age pattern of the demographic process under study to estimate the age-specific demographic rates in the population, and then compared the estimated and observed age-specific rates. The following estimates/comparisons were made:

- Estimates of the age-parity-specific fertility rates in China in 1981-89, using the U.S. average fertility schedule in 1970-80 as a model standard, and then compare the estimated and the observed Chinese fertility rates in 1981-89;
- Estimates of the marriage-duration-specific rates of leaving parental home for females in the 1950-69 and 1970-79 marriage cohorts in each of the twelve provinces of China, using the average of the rates in the other eleven provinces as a model standard, and then compare the estimated and the observed leaving home rates in each of the twelve provinces;
- Estimates of the female age-specific divorce rates in Australia in 1975, using the divorce schedule for the U.S. in 1970 as a model standard, and then compare the estimated and the observed divorce rates in Australia in 1975;
- Estimate of the female age-specific remarriage rates of widows in the Netherlands in 1978, using the remarriage schedule for Australia in 1975 as a model standard, and then compare the estimated and the observed remarriage rates of widows in the Netherlands in 1978.

For these comparisons, it was found that almost all estimated schedules are fairly close to the observed ones and almost all of the Index of Goodness of Fitting (IGF)⁸ values are above 0.95 (see Figures 1f and Tables A-4, A-5, A-6 in the Appendix in Zeng et al. 2000), indicating a remarkable accuracy of the estimates. These results are based on the fact that model standard schedules adopted from another population reflect reasonably well the general age pattern of the schedules in the population being studied. For example, the general shapes of the marriage-duration-specific rates of leaving the parental home in the twelve provinces of China are similar to each other. And the general age patterns of divorce and remarriage in Australia, the Netherlands and the United States in 1970s did not differ substantially. The relevant estimates are, therefore, good. Chinese fertility in the 1980s had already begun to roughly conform to a general age pattern of relatively low fertility comparable to the fertility pattern in the U.S. in the 1970s. Consequently, the average of the age-parity-specific fertility rates in the United States in 1970-1980 could be used as a model standard schedule to estimate Chinese rates in the 1980s, while the observed parity-specific total fertility rates, median ages and interquartile ranges in China in the 1980s were used as input for the estimates, and they adequately measure the differences in fertility quantum, timing, and the shape of the curves between China and the U.S.

Unavailability of detailed race-sex-age-specific demographic schedules of fertility by parity, marriage/union formation and dissolution and leaving parental home is a common

⁸ $IGF = \frac{\text{Total variations} - \text{Residual variations}}{\text{Total variations}}$, where, Total variations = $[\sum_x (f(x) - F)^2] / (N-1)$, Residual variations = $[\sum_x (f(x) - f'(x))^2] / (N-2)$, $f(x)$ are the observed rates; $f'(x)$ are the estimated rates; F is the mean of the observed rates; N is the number of the observed values. Similar to R^2 in the ordinary regression exercises, the closer the value of the IGF to 1.0, the more exact the estimates. If the estimated rates were completely identical to the observed ones, the IGF would be equal to 1.0. If IGF is equal or greater than 0.95, the estimates should be regarded as good.

problem in household projections for sub-national regions. The successful empirical assessments described above show that it is theoretically and empirically justifiable to use the model standard schedules at the national level or even from another region or country with a similar age pattern of the demographic regimes, and use the estimated or projected demographic summary parameters from the sub-national area under study to reasonably estimate the needed sex-age-specific demographic schedules for household projections.

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

3. DATA AND ESTIMATION ISSUES FOR HOUSEHOLD PROJECTIONS AT SUB-NATIONAL LEVELS

3.1. Data requirements

As shown in (1) in Table 1, for population and household projections at the national, state, and sub-state levels employing the ECC 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 of fertility, mortality, and marriage/union formation and dissolution and international migrations (see (2)-a in Table 1) need to be estimated at the national level only, and they can be employed for household projections at the state level (see Core idea 5 in Section 2), while the state-race-age-sex-specific rates of domestic in-migration and out-migration can be estimated based on census or ACS micro data files (see (2)-b in Table 1). Projected (or assumed) demographic summary parameters – 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 ECC model, *household projections at the state level* require a census micro data file and the projected (or assumed) demographic summary parameters 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 household projections for the parental state where the sub-state area is located are done.

3.2. Estimation issues for household projections at the state level

Estimation of the required demographic summary parameters – the TFR, life expectancy at birth, total number of migrants, mean age at first marriage and birth are straightforward. But the estimation of standardized general rates of marriage, divorce, cohabitating, and union dissolution for household projections at the state level merit 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 occurred 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 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 ($i \neq j$, where i and j represent the seven marital/union statuses (see footnote 5 for detailed definition and the codes).

$m_{ij}(x,s,r,T1)$, 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 (newly married couples) including 1st marriages and remarriages occurred in the state in the census year T1.

We assume that the race-sex-race-specific o/e rates of 1st marriage and remarriage in the census year are proportional to the corresponding model standard schedule rates, namely, $m_{i2}(x,s,r,T1) = \gamma(T1) M_{i2}(x,s,r)$ for $i \neq 2$ (see footnote 5 for definition of the codes i and 2). The $\gamma(T1)$ is estimated as:

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

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. We assume that the race-sex-race-specific o/e rates of divorce in the census year are proportional to the corresponding model standard schedule rates of divorce, namely,

$m_{24}(x,s,r,T1) = \delta(T1) M_{24}(x,s,r)$ (the subscript 2 and 4 represent currently married and divorce, respectively; see footnote 5). The $\delta(T1)$ is estimated as:

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

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 by iterative proportional fitting. *First*, we 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 and the model standard schedules of the cohabitation union formation and dissolution. *Second*, we then project the household distributions from the previous census to the most recent census for the state using the previous census data as the base population, the implied GC and GCD and the other estimated demographic rates in the previous census as input. Through such projections, we obtain the projected proportions of households with a cohabiting couple among all households in the most recent census, denoted as PC. *Third*, we then compare PC with the proportion of the households with a cohabiting couple among all households observed in the most recent census, denoted as CC. If the PC is higher (or lower) than CC by a margin of a pre-determined criterion (e.g. 1%), we proportionally reduce (or increase) the implied GC by a factor of CC/PC and increase (or reduce) the implied GCD by a factor of CC/PC, and redo the projection, and compute the new PC and compare it with CC. We repeat this iterative proportional fitting procedure until the projected PC and the observed CC are reasonably close to each other (e.g. with 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 for the 50 states and DC presented in the Section 4.

3.3. 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 (Smith et al. 2001; Rao 2003; Smith and Morrison 2005). Ratio 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 ECC model, we employ the ratio method with either a *constant-share* or a *shift-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 shift-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 formulas of the constant-share and shift-share ratio methods with simple and easily understandable numerical illustrations are presented and well justified in Smith et al. (2001: 177-179), and they can be readily applied for households and population projections of sub-state areas following the ECC approach.

In applying the ratio method for households projections at sub-state level, one additional proportional adjustment is 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 $f(h,z,r,x,t)$ and $f'(h,z,r,x,t)$ denote the preliminarily projected and finally adjusted projected number of households of type h and size z 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. Then

$$f'(h,z,r,x,t) = \theta(t) f(h,z,r,x,t), \text{ where } \theta(t) = PS(t) / \left(\sum_h \sum_z \sum_r \sum_x f(h,z,r,x,t)z \right)$$

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 and 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 (more detail) 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.

Note that the household projection for a sub-state area following the ratio method approach can be done based on its parental state's household projection produced either by the

ECC new approach or the classic headship-rate method. However, as reviewed and discussed earlier, a typical well-done national household projection using the headship-rate method projected only five household types by age groups of household head or householder (Bureau of the Census 1996), with no projected household sizes available (see Table 1 in Zeng et al., 2006: 8). By comparison, the new ECC approach projects much more detailed household types, sizes, and living arrangements for all members of the population (see Table 2 in Zeng et al., 2006: 9). Consequently, using the same ratio method, the household projections for the sub-state areas based on their parental states' household projections produced by the ECC approach would contain much more detailed forecasts information than the case in which the parental states' household projections were produced by the classic headship-rate method. Therefore, the practical usefulness of the ratio method is strengthened if it is applied with the ECC approach as compared with the case in which it is applied with the headship-rate method.

--- 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) conducted such an empirical assessment for 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).

Does the ECC 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 DC⁹, all using the national model standard schedules (see (2)-a in Table 1) estimated based on the pooled national surveys data¹⁰ (Zeng, Morgan et al. 2010), except that the state-race-age-sex-specific domestic migration rates are estimated based on the census 5% micro data files of each of the states. The first test projects from 1990 to 2000 using the 1990 census data as base population and summary parameters based on data before 1991, and compares the projected and the census-observed in 2000. This test assumes that we have no data after 1990 when one projects 1990 to 2000 and tests the accuracy of the projections using the ECC model in the real world (assuming the 2000 census data are accurate). The second test projects from 1990 to 2000 using the 1990 census data as the base population and summary parameters based on data in the 1990s, and compare the projected and the census-observed in 2000. This test assesses the estimation properties of the ECC model based on the assumptions that the input data (observed in the 1990s) and the 2000 census observations (outcome in the exercise) are correct. We also performed household projections from 1990 to 2000 for 25 randomly selected counties and 25 randomly selected cities (each state has one

⁹ The numerical projections reported in this paper were calculated with the ProFamy computer software program; this program incorporates the ECC projection method and contains a demographic database of the U.S. age-specific schedules of demographic rates (see footnote 7) 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/>

¹⁰ 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).

randomly selected county or city), based on the observed/projected household distributions of their parental states in 1990/2000 and the constant-share ratio method.¹¹

We use the percent error (PE), Mean Absolute Percent Error (MAPE), Mean Algebraic Percent Error (MALPE) and Median Absolute Percent Error (MEDAPE), which are the most commonly used measures of forecast errors (Smith, Tayman and Swanson, 2001: 302-304), to assess the validity of the households and population projections at sub-national levels using the ECC approach. More specifically, the PE is defined as the difference between the ECC projected in 2000 and the census observed in 2000 divided by the census observed in 2000 and multiplied by 100. The MAPE and MEDAPE are the average and median of the absolute values of PE, and MALPE is the algebraic mean of PE (in which positive and negative values offset each other), over all 50 states and DC.

The absolute percent errors based on comparisons of the total number of households, average household size, percent of households of 1, 2-3, and 4+ persons households and couple-households, total population size, % of children, % of elderly aged 65+, % of oldest-old aged 80+ and dependency ratios between the projected and the observations in 2000 census for the 50 states and DC are summarized in Tables 3 and 4. Among the first set of tests of the 306 main indices¹² on households of the projected (using data before 1991) and the 2000 census observed in the 50 states and DC, 29.1, 33.9, 17.4 and 12.9 percent of the absolute percent errors are <1.0%, 1.0-2.99%, 3.0-4.99%, and 5.0-9.99%, respectively, and 3.1 and 3.6 percent of the percent errors are 10-14.99% and >15% (see panel (A) in Table 3). As expected, the error rates

¹¹ The general data sources for obtaining the base population and estimating the demographic summary parameters 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.

¹² We compare 6 main indices of households projections and 6 main indices of population projections for each of the 50 states and DC and thus both of the total number of household indices and the total number of population indices under comparisons is 306.

in the second set of the tests which used observed demographic summary measures in 1990s as input are considerably smaller than those in the first set of the tests. More specifically, 45.9, 42.0, 11.5 and 0.6 percent of the absolute percent errors are <1.0%, 1.0-2.99%, 3.0-4.99%, and 5.0-9.99%, respectively, and none is over 10% in the second set of the tests of households projections (see panel (C) in Table 3). The percentage distributions of the absolute percent errors of the main indices of population in the two sets of the tests by comparing the projected with the 2000 census observations in the 50 states and DC are generally similar: about 29.7-34.5, 39.8-43.4, 14.3-16.5, 7.3-9.5, and 0.8-3.9 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 the comparisons) slightly exceeds 15% (see panels (B) and (D) in Table 3).

The Mean Absolute Percent Error (MAPE) and Median Absolute Percent Error (MEDAPE) of the main household indices in comparisons between the projected and census observations in 2000 for the 50 states and DC are all within a reasonable small ranges of 0.8% ~ 4.7% and 0.7% ~ 3.5%. The Mean Algebraic Percent Error (MALPE) of average household size and percent of 2-3 persons households in both Tests I and II, and the MALPE of total number of households in Test II are negative, within a range of -0.33% ~ -1.1%, and all of the other MALPEs are positive, within a range of 0.04% ~ 2.9% (see panel (A) of Table 4). Similar to those error rates of the main indices of household projections, the ranges of the MAPE, MEDAPE and MALPE of the main population indices comparing between projected and census observations in 2000 for the 50 states and DC are all reasonably small (Tests I and II in panel (B) in Table 4).¹³

¹³ We also performed another set of the tests of projections from 2000 onwards using data prior 2001 and comparing the projected and ACS-observed in 2006 for the 50 states and DC. It turns out that 34.2, 35.0, 21.9, and 9.0 percent of the percent errors of the 306 indices of the household projections are <1.0%, 1.0-2.99%, 3.0-4.99%, and 5.0-9.99%, respectively, and none is over 10%. Apart from space limitations, we

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

As shown by the results of validation tests of sub-state level household projections from 1990 to 2000 based on the constant-share ratio method and the parental states' households projections employing the ECC approach and then comparing the projected and the census observed counts in 2000 in Table 5, the MALPE for the counties and cities are all within a small range, and there is no clear different patterns between the counties and cities. However, the differences of the forecasts errors measured by MAPE and MEDAPE between the counties and cities are substantial: the MAPE and MEDAPE are within the ranges of 2.99~7.69 and 2.32~6.29 for the 25 randomly selected counties, and 4.78~14.78 and 2.66~9.81 for the 25 randomly selected cities; it is clear that the percent error rates in the validation tests for the cities are substantially larger than those for the counties. This might be due to the fact that the constant-share assumption is more likely to be violated in the cases of cities' household projections because the demographics in some cities changed substantially during the 1990s due to economic/business events such as opening, closing or reallocating the large enterprises or institutions there.

As shown in Table 6, about one-third of the 25 randomly selected cities have a very small population size of less than 2,000 residents and the population size in about one-third of the 25 randomly selected counties is less than 10,000 persons. We analyzed the forecasts errors of the testing projections from 1990 to 2000 by population size among the 25 randomly selected

did not present the detailed results from these additional tests here (they are available on request) for two reasons. The first reason is that the 2006 ACS data may not be accurate enough to serve as a benchmark standard for the validation tests (Alexander, Davern and Stevenson 2010; Swanson, 2010). Second, the 6-year (2000-2006) projection horizon is relatively short for the purpose of evaluating the forecast errors.

counties and 25 randomly selected cities, and we did not find significant correlation between the population size and the forecasts errors.

---- Tables 5 and 6 about here----

The validation tests results at the state and the sub-state levels summarized in Tables 3, 4 and 5 show that the forecasts errors are within a relatively small range. It is uncertain what portions of the errors are due to the model specification or to inaccuracies of the data. It is clear, however, that the ECC method for simultaneously projecting households and population work reasonably well not only at national level as shown in Zeng et al. (2006) but also at the state level; and in combination with the ratio method, it works at the sub-state level as well.

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

To illustrate some practical applications of the ECC method, we conducted household projections from 2000 to 2050 for the 50 states and DC, as well as household and housing projections for the small town of Chapel Hill (with a total population size of 48,715 persons in 2000), North Carolina¹⁴ 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

¹⁴ The town of Chapel Hill is used as an illustrative application to demonstrate the ability of the ECC approach combined with the ratio method 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 the core idea 5 in Section 2, we apply the model standard schedules of race-sex-age-specific demographic rates (except domestic migration rates) estimated based on national datasets for household projections at the state level. Based on the 2000 census 5% micro dataset, we estimated race-sex-age-specific probabilities of domestic out-migration from each state to the rest of the country and race-sex-age-specific frequencies of in-migration 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 in-migrants and out-migrants 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. 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.¹⁵

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

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 current demographic rates provide a reasonable forecast of future rates.

We summarize the following insights from the nine main indices of the decennial projections on households and elderly living arrangements of the 50 states and DC presented in the Tables 7a, 7b, and 7c.

---- Tables 7a, 7b, and 7c about here----

As shown in panel (A) in Table 7a, 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 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 down considerably after 2020 (see panel (B) in Table 7a). Clearly, American family households are expected to be smaller in all states in the next a couple of 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 panel (C) in Table 7a).

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 states will be higher than that in 2000 by >30%, 20-29.9%, 10-19.9% and <10%, respectively (see panel (D) in Table 7b). 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 panel (E) in Table 7b). 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 the households and elderly living arrangements projections results presented in panels (F) in Table 7b and panels (G), (H) and (I) in Table 7c 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 in-migrants is an exception (see panel (F) in Table 7b). 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 panel (G) in Table 7c). The panel (H) in Table 7c 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 remarkable: 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 a census or ACS micro data file and the projected (or assumed) demographic summary parameters. After projections for the parental state are done, and using the widely recognized ratio method, corresponding projections for a sub-state area requires 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 compared the main indices of the projected and observed in 2000 for the 50 states and DC, 25 randomly selected counties and 25 randomly selected cities. The results show that the ECC model for household and population projection works well at the state level, with most absolute forecast error rates less than three percent of observed values; and the enhanced forecasts outcome at state level produced by the ECC approach can be used to perform more informative household projections for sub-state areas of counties and cities, employing the ratio method.

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; we have not reported interval projections with high and low bounds to reveal uncertainties in the projections for the 50 states and DC, due to space limitations in a single article. 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 parameters 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 and validation tests 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

- Alexander, J. T., M. Davern, and B. Stevenson. "2010 Inaccurate Age and Sex Data in the Census PUMS File." Paper (to be) presented at the 2010 PAA meeting, Dallas, TX (Available online at <http://paa2010.princeton.edu/download.aspx?submissionId=100045>).
- 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.
- Brass, W. 1974. "Perspectives in Population prediction: illustrated by the statistics of England and Wales." *Journal of the Royal Statistical Society*, 137, Series A: 532-583.
- Brass, W. 1978. "The Relational Gompertz Model of Fertility by Age of Women." London School of Hygiene and Tropical Medicine, 1978 (mimeographed).
- Booth, H. 1984. "Transforming Gompertz's function for fertility analysis: the development of a standard for the relational Compertz function." *Population Studies*, 38: 495-506.
- Bureau of the Census, United States (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.
- 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.
- 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. and T. J. Trussell 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 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.
- Keyfitz, N. 1972. "On Future Population." *Journal of American Statistical Association*, 67: 347 - 363.
- 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.
- Murray, C. J. L., B. D. Ferguson, A. D. Lopez, M. Guillot, J. A. Salomon, O. Ahmad 2003, Modified logit life table system: principles, empirical validation, and application, *Population Studies*, Vol. 57: 165 – 182.
- Myers, D., J. Pitkin, and J. Park. 2002. "Estimation of housing needs amid population growth and change." *Housing Policy Debate* 13(3): 567-596.
- Paget, W. John and Ian M. Timaeus, 1994. "A relational Gompertz model of male fertility. development and assessment." *Population Studies*, 48: 333-340.
- 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.
- 1986. "Parameterized Multistate Population Dynamics and Projections." *Journal of American Statistical Association*, 81: 48 - 61.
- 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.

- Swanson, D. A. 2010. 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 2010 PAA meeting, Dallas, TX (Available online at <http://paa2010.princeton.edu/download.aspx?submissionId=100049>).
- Swanson, D. A. and Louis G. Pol. 2009. Applied demography: its business and public sector Components. In: Zeng, Yi (ed.). 2009. “*Demography*” volume of the *Encyclopedia of Life Support Systems* (EOLSS) (www.eolss.net), coordinated by the UNESCO-EOLSS Committee. Oxford: EOLSS Publishers Co.Ltd.
- Treadway, R. 1997. *Population projections for the state and counties of Illinois*. Springfield: State of Illinois.
- United Nations. 1982. *Model Life Tables for Developing Countries*. ST/SEA/SER, A/77, Sales No. E. 81, XIII, 7, New York.
- United Nations. 1983. *Manual X: Indirect Techniques for Demographic Estimation*. 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.
- 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, Paper presented at the 2008 annual meeting of the Population Association of America, April 17-19, New Orleans. Manuscript submitted for consideration of publication.
- Zeng Yi, Coale, Ansley, Minja Kim Choe, Liang Zhiwu and Liu Li. 1994. "Leaving Parental Home: Census Based Estimates for China, Japan, South Korea, The United States, France, and Sweden.." *Population Studies*, Vol. 48, No. 1.

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

<i>Contents of the data</i>	<i>Main data source for U.S. applications</i>
(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)-a Model standard schedules at the national level (may be used for households projections at state level)	
(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 international immigration and emigration.	Census 5% micro data files
(2)-b Model standard schedules at state level	
(f) Race-sex-age-specific rates of domestic in-migration and out-migration for each state	Census 5% micro data, ACS data files
(3) Demographic summary measures for the nation and states	
(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 census micro data, vital statistics and pooled survey datasets

(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: (1) 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. (2) The rural-urban dimension can be included if the rural-urban differentials are substantial and the rural-urban specific data are available.

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

Category	Type (h)	Size (z)	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	3-4	One-couple household with child/other, household size 3-4
9	3	5+	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 percent errors of comparisons between the ECC projected from 1990 to 2000 and the Census observations in 2000 for the 50 states and DC

Absolute percent errors	ECC projected from 1990 to 2000 versus. Census 2000 obs., based on data before 1991				ECC projected from 1990 to 2000 versus Census 2000 obs., including data during 1990s			
	(A) household indices		(B) population indices		(C) household indices		(D) population indices	
	number of Comparison	percent	number of Comparison	percent	number of Comparison	percent	number of Comparison	percent
< 1.00%	89	29.13	91	29.69	141	45.94	105	34.45
1.0-2.99%	104	33.89	133	43.42	129	42.02	122	39.77
3.0-4.99%	53	17.37	51	16.53	35	11.48	44	14.29
5.0-9.99%	39	12.89	29	9.52	2	0.56	22	7.28
10.0-14.99%	9	3.08	3	0.84	0	0	12	3.92
>15.0%	11	3.64	0	0	0	0	1	0.28
Total	306	100	306	100	306	100	306	100

Table 4. The Mean Absolute Percent Error (MAPE), Mean Algebraic Percent Error (MALPE) and Median Absolute Percent Error (MEDAPE) between the ECC projected from 1990 to 2000 and the Census observations in 2000 for the 50 states and DC

	(A) Main indices of household projection					
	Tot. # of household	household aver. size	%1 pers household	%2-3 pers household	%4+ pers household	%Couple household
<u>Test I:</u> ECC projection vs. census obs. in 2000 based on data before 1991						
MAPE	1.63	1.75	4.73	2.63	4.08	2.07
MALPE	0.04	-0.56	2.91	-1.06	0.06	0.56
MEDAPE	1.07	1.16	3.45	2.20	2.76	1.35
<u>Test II:</u> ECC projection vs. census obs. in 2000 including data in 1990s						
MAPE	1.28	0.82	2.01	1.86	2.21	1.23
MALPE	-0.73	-0.33	1.25	-0.79	0.32	0.49
MEDAPE	1.21	0.73	1.67	1.53	2.03	1.08
	(B) Main indices of population projection					
	Pop size	%children <18	% old 65+	% oldest -old 80+	Dep.ratio of children	Dep.ratio of old
<u>Test I:</u> ECC projection vs. census obs. in 2000 based on data before 1991						
MAPE	1.35	1.96	2.52	3.44	2.55	2.80
MALPE	-0.58	1.39	-1.45	-2.02	1.56	-1.05
MEDAPE	1.09	1.82	2.08	2.81	2.21	2.19
<u>Test II:</u> ECC projection vs. census obs. in 2000 including data in 1990s						
MAPE	1.21	1.30	2.77	4.48	1.69	3.76
MALPE	-0.92	0.78	-2.05	-3.90	0.62	-2.34
MEDAPE	1.36	1.03	1.90	3.79	1.30	2.80

Table 5. The MAPE, MALPE and MEDAPE between ECC/ratio method projected from 1990 to 2000 (based on data before 1991) and the Census observations in 2000 for 25 randomly selected counties and 25 randomly selected cities

	population size	Tot. # of household	household aver. size	%1 pers household	%2-3 pers household	%4+ pers household	% married couple hh
MAPE of the 25 counties	7.69	8.62	2.99	5.78	3.45	6.68	4.23
MALPE of the 25 counties	0.40	1.18	-1.55	3.05	-1.87	1.36	2.23
MEDAPE of the 25 counties	4.49	4.57	2.32	4.69	2.79	6.29	3.34
MAPE of the 25 cities	14.78	13.37	5.29	10.73	4.78	9.77	6.60
MALPE of the 25 cities	-4.39	-0.72	-1.24	4.60	-1.89	-0.59	0.16
MEDAPE of the 25 cities	9.81	9.30	4.97	7.58	2.66	9.77	5.53

Table 6. Population size distributions of the 25 randomly selected counties and the 25 randomly selected cities (each of the 50 states has one randomly selected county or city)

Randomly selected 25 cities			Randomly selected 25 counties		
Pop size	number	%	Pop size	number	%
<2,000	8	32.0	<10,000	8	32.0
2,000-4,999	8	32.0	10,000-49,999	9	36.0
5,000-29,999	6	24.0	50,000-99,999	4	16.0
≥ 30000	3	12.0	≥ 100,000	4	16.0
Total	25	100	Total	25	100