Publications, Reports and Presentations   Population Research Center

2015

Methods and Data for Developing Coordinated Population Forecasts

Portland State University. Population Research Center

Let us know how access to this document benefits you.

Follow this and additional works at: http://pdxscholar.library.pdx.edu/prc_pub

Part of the Urban Studies and Planning Commons

Citation Details

Methods and Data for
Developing Coordinated Population Forecasts

Prepared by

Population Research Center

College of Urban and Public Affairs

Portland State University

Population Research Center

PORTLAND STATE UNIVERSITY

Table of Contents

Introduction

This document is to accompany the coordinated population forecasts and their corresponding reports prepared by the Population Research Center (PRC).

Developing long-term coordinated population forecasts for a county and its sub-areas (UGBs and areas outside UGBs are referred to sub-areas in this document), requires these main stages: 1) compiling and evaluating historical and recent data to ascertain demographic characteristics and trends in the study area and to obtain a population base from which the forecasts may be launched; 2) making assumptions about the future and adjusting the data or rates in the forecasting models (calibrating the models) to incorporate predicted rates or trends; and 3) reconciling, or adjusting the forecasts so the sum of the sub-area forecasts is consistent with the forecast for the county total.

We first develop population projections and then we make adjustments to these projections in order to produce population forecasts. Population projections are developed by extending historical and current demographic and housing trends into the future. Forecasting population requires that assumptions be made about future population change, and adjusting the projection models to account for circumstances that perhaps skewed past trends or may affect future change. Such circumstances in the past could be a building moratorium or the opening of a new group quarters facility. Events affecting future change would be, for example, planned future housing development that is higher than usual, a foreseen change in an area's physical ability to accommodate growth (buildable land available is approaching capacity or improvements to infrastructure are underway), anticipated changes in the economy (the location of a new employer or the upswing or downturn of the economy in general), or an expected change in the local population and household composition (age, ethnicity, and average household size).

Models for Population Forecasts

Two different types of primary demographic models were utilized to develop the population forecasts for a county and its sub-areas. For a county and its larger sub-areas—those with populations greater than 8,000—a cohort-component model was used. For each smaller sub-area, a housing unit model was relied upon. The cohort-component model best predicts population over the long-term for areas with larger populations. The housing unit model is better suited for smaller populations and incorporates recent annual data that account for more variability in population growth over the forecasting period. The forecasting models are described in more detail below.

Cohort-Component Method

Separate cohort-component models were developed for each county and for each larger sub-area. These forecasts are 2010-based projections. However, adjustments were made to the model to incorporate into the forecasts the 2011-2014 PRC's certified population estimates and capture trends from the most recent data available.

The cohort-component model predicts future populations as outcomes of the life events that occur over time. These events are comprised of births, deaths, and migrations. Thus, an area's population grows when births outnumber deaths or when more people move into the area than leave it; or when some

combination of these circumstances occurs. These events occur more often in certain age groups, or cohorts, than in others. For example, people tend to move around the most when they are in their 20s, or the elderly have lower chances than people in their 40s to survive over the next five years. Applying appropriate age- and gender-specific rates of birth, death and migration to the existing population cohorts of the County produces its future population projection. Our cohort-component model produces forecasts in 5-year time intervals[1].

$$Pop_{t2} = Pop_{t1} + Births_{(t1\ through\ t2)} - Deaths_{(t1\ through\ t2)} + NetMig_{(t1\ through\ t2)}$$

Pop: Population
NetMig: Net migration
t: time

The cohort-component method relies on the availability of accurate data on the age and gender composition of an area's population. The most precise information about population age structure in our study areas is provided by the U.S. Census Bureau. Age-specific rates of life events are derived from data provided by the U.S. Census Bureau and the Oregon Center for Health Statistics. These rates are then applied to known population cohorts from one Census, which moves them forward through time (aging the population) to meet the next cohort in the following Census. These baseline demographic rates are modified to account for the most recent trends as well as for ones assumed to occur in the future. Examples of such trends that may affect the future population of an area include the recent tendency among women of childbearing ages to delay having their first child, or a predisposition of young men (ages 20 to 29) to be more mobile than women in the same age cohort. A set of assumptions must be developed to address likely changes in the baseline rates of life events and require judgment about how the trends might evolve in the study area. The existing population structure mostly determines the future population composition of the area, but it may change some depending on age-specific migration rates predicted for the future. Trends detected from historical and recent data, such as housing; land use; employment; income tax returns and exemptions; driver license issuances; and school enrollment data, help determine these future migration rates.

The population and housing data came from the 2000 and 2010 Censuses of Population and Housing and PRC's Population Estimates Program. Additional housing information and land use data were obtained from local governments, both at the county and UGB levels of geography; the Oregon Center for Health Statistics provided information on births and deaths; and some administrative data sets were collected from Oregon state agencies.

The 2000 and 2010 population and housing data from the Census were compiled for census blocks by age group and gender. The census blocks were allocated and aggregated to UGB boundary areas (the UGBs were as existing in the forecast launch year) using Geographic Information Systems (GIS). The 2000 population for each UGB, in five-year age cohorts (0-4 years, etc.), was then "survived", or aged

---

[1] To obtain population projections for single-year time intervals, we prepared an interpolation worksheet. The worksheet and instructions about how to use the worksheet to calculate a single year population forecast based on the forecasts for 5-year time intervals is located on our website at this link: http://www.pdx.edu/prc/region-1-documents.

into the corresponding older cohort in the year 2010. "Surviving" the cohorts is accomplished by applying age- and sex-specific survival rates. These rates represent the proportion of population in each younger cohort that would survive during a given time period (such as the five years between 2000 and 2005, or 2005 and 2010) to become the next older age cohort. This process is repeated for each five-year age group and five-year time interval over the forecast period. Forecasting a known population (the Census 2010 population) and its age distribution enables appropriate adjustments to be made to the model so that the forecasted population becomes aligned with the actual population and ensures the accuracy of the model's projections.

For each five-year interval, we compiled the number of live births that occurred to women of childbearing ages. To calculate the number of newly born residents of a county and its larger sub-areas, age-specific fertility rates (calculated from past Censuses and data on actual births that occurred) were applied to the numbers of women in childbearing cohorts—five-year age groups starting at 10-14 and ending at 45-49. Age-specific fertility rates indicate how many children women in a given age group are likely to give birth to during each five-year period. Once born, children become subject to survival rates and are "moved," or "aged," through the system like all the other cohorts.

County fertility rates were projected into the future by relying upon four factors: historical total fertility rates (TFRs), fertility trends used in the U.S. Census Bureau's national population projections, the ratios between national and statewide TFRs, and the ratio between statewide and the county TFRs. The historical total fertility rates anchor the starting point for each larger sub area. The future TFR trend following the starting point is jointly influenced by the national trend the ratio between national and statewide TFRs, and the ratio between statewide and county TFRs. This method assumes that Oregon's TFRs will be influenced by national TFR trends in the future, and each county's TFRs will be influenced by the statewide TFR trends.

Survival rates were projected using two factors: historical survival rates, and the survival rates used in the Census Bureau's national population projections. The historical three-year average survival rates (observed for the 2009 to 2011 period) for each study area were used as the starting point. The Census Bureau's projected survival rates for the White non-Hispanic population in 2060 are used for both the state and counties. Survival rates for individual forecast years are either an interpolation between the historical survival rates and those projected for 2060, or an extrapolation (for years beyond 2060) from rates projected for 2060. This method assumes each county's life expectancy will merge to meet the state's projected life expectancy, and the state's life expectancy will merge to meet the nation's projected life expectancy.

The most difficult part in forecasting population is estimating the in- and out-migration of an area. Due to a scarcity of reliable data available to study in- and out-migration, it's best to use net migration rates. Net migration is the balance between in- and out-migration. Net migration can be calculated as a residual if the population is known at the beginning and the end of a time period, along with the number of births and deaths that occurred during that same time period. Net migration is positive when more people move into the area than leave it; it is negative if the opposite is true. Net migration rates used in the cohort-component model can be interpreted as the number of people per 1,000 who are added to

(or subtracted from) a given cohort due to migration over a given period of time (in this case, five years). The initial net migration rates for the cohort-component model were derived from the 2000 and 2010 population cohorts that are located within the county and larger sub-areas, as well as from births and deaths that occurred in the same areas from 2000 to 2010. The rates were adjusted so that the "forecast" population for the year 2010—using 2000 as the launch year—fit the actual known population obtained from the 2010 Census. The net migration rates—used to forecast the population in the county and in its larger sub-areas beyond 2010 were further modified to reflect the likely future migration patterns. Demographic trends identified in post-2010 data from PRC's annual population estimates and the U.S. Census Bureau's ACS data had some bearing on the adjustments made to the model in the initial forecast period. In addition, migration patterns are greatly influenced by the local economy and by housing growth in the area, both current and assumed. When making the final adjustments to the net migration rates, consideration was also given to planned future development, as well as to local planners' and citizens' shared knowledge about accommodations or limitations to population growth in the area.

A unique set of demographic data were used for each of the cities, and trends specific to each of them were considered when making adjustments to their cohort component models.

Housing Unit Method

A housing unit model was used to prepare forecasts for the smaller sub-areas. This method requires that a current housing inventory for each area be compiled and that historical and recent rates of change in each inventory be known. Other housing and population data are also needed for the components of the housing unit model, they are: occupancy rates, the average number of persons per household (PPH), and group quarter's population. In this method, the number of housing units in an area is first projected. Next, assumptions about future housing occupancy and average household size are made to forecast household population. Persons residing in group quarters (such as in college dormitories, prisons, and nursing homes) are also projected and then added to the household population to arrive at the forecast for the total population. This process is carried out for five-year intervals throughout the forecast period. An area's total population is calculated in the housing unit method by multiplying the projected number of housing units by the occupancy rate and PPH and then adding to that product, the group quarters population. This process is carried out for five-year intervals throughout the forecast period[2].

$$Pop = HU * OCC * PPH + GQ$$

HU: Housing unit stock
OCC: Occupancy rate
PPH: Persons per household
GQ: Group quarter population

---

[2] To obtain population projections for single-year time intervals, we prepared an interpolation worksheet. The worksheet and instructions about how to use the worksheet to calculate a single year population forecast based on the forecasts for 5-year time intervals is located on our website at this link: http://www.pdx.edu/prc/region-1-documents.

Data used in the housing unit models are from the 2000 and 2010 Censuses of Population and Housing, and from recent and historical tax lot data that were obtained from county tax assessors in the forecast region. Other housing data and group quarters population data were collected from the local jurisdictions by PRC's Population Estimates Program (we send a housing and population questionnaire to Oregon's cities and counties and request that they complete and return the form to us each year).

Population and housing data from 2000 and 2010 Censuses were compiled for each geographic part in the study area. An allocation of data was made to the 2014 jurisdictional boundaries using the same GIS methods as described previously in the cohort-component model section. Housing inventories were created from the 2000 and 2010 Census data and from counts derived from county tax assessor data. Housing growth trends were detected from the Census data, the residential tax lot data, and PRC's Population Estimates Program housing data for cities. The numbers of units added to the inventory each year were extracted from the residential tax lot data using the 'year built' information, and from housing data collected in PRC's Population Estimates Program.

The numbers of housing units are projected into the future based on past housing growth trends. Housing growth rates were calculated using the housing inventories and the magnitude of annual or periodic change they experienced. The housing trends were extrapolated into the future and applied to the 2014 housing inventory to predict the numbers of housing units in the future. Adjustments were made to the models to accelerate or curb growth based on current conditions compared to the past, or plans for future change. In sub-areas where future growth is expected to be very different than in the past, adjustments were made to the housing unit model by calculating a weighted average from annual or periodic growth rates, giving more bearing to the years believed to have more influence on what likely will occur in the future.

Adjustments were made to the model's growth rates to account for known planned future housing. The numbers of housing units scheduled to be constructed and completed in an area during the forecast period were accounted for in the model by adding in the numbers of planned housing units in the five-year time period that construction is planned to be completed.

Census data for 2000 and 2010 were also used to calculate average household sizes (PPH) and housing occupancy rates. Recent data for PPH and housing occupancy from the U.S. Census Bureau's American Community Survey are also available and were also considered. However, these data are sample data and are not always reliable, especially for areas with small populations.

Methods used to forecast PPH and occupancy rates were based on recent historical and current trends. These trends were projected into the future using extrapolation methods, or were based the most recent Census rates and were held constant throughout the forecast period. Occupancy rates vary more than PPH and are more dependent on local economic conditions, which can fluctuate. Demand for housing, and thus occupancy rates also depend on population growth. We were conservative when forecasting occupancy rates to avoid forecasting extreme rates in a wrong direction.

To forecast PPH, along with historical trends, additional data on factors that influence PPH, such as births by race and ethnicity, changes in school enrollment, fertility rates, and age structure of the population were considered.

Occupancy rates for the sub-areas within each county were predicted for the entire 50-year forecast period based on the most recent Census data (2010), and adjusted according to past occupancy trends detected from the 2000 and 2010 data. In addition, population and housing composition, and the residential and non-residential classification of tax lots were considered to forecast changes the occupancy rates will undergo in the future. Sub-area occupancy rates for 2014 were also estimated using a ratio method that assumes changes in housing occupancy in each county are related to change in its sub-areas. Rather than assume that rates widely fluctuate the occupancy rates for the 50-year forecast period are a weighted average of 2000 and 2010 occupancy rates which are held constant beginning in 2015.

The number of persons residing in group quarters is a component of population that is added to the number of persons residing in households to arrive at the total population.

After the population residing in housing units was forecasted for each sub-area, the group quarters population was projected for the corresponding areas. The projection of future group quarters populations was based on historical and recent trends, and known planned future group quarters facilities. The projected group quarters populations were then added to the forecasted household populations to obtain total population forecasts.

Reconciliation of the Forecasts

We developed separate population forecasts for each county in the forecast region and for each county's sub-areas. For consistency, the sum of the parts must equal the whole, which means here that the sum of the individual forecasts of each county's sub-areas should be equal to the corresponding county-level forecast. In some cases, the county-wide forecast served as the control total to which the sum of the individual forecasts for the sub-areas was reconciled, or adjusted to match (called a "top down" approach). In other cases, the county total was adjusted to equal the sum of the forecasts for the sub-area (a "bottom up" technique). This means that some minor adjustments were made to the sub-area forecasts so that the sum of the parts is the same as the whole.

Adjustments were made to reconcile the forecasts using control factors. The control factors were calculated based on the relationship between the control total and the sum of the parts. In the top-down approach, the actual difference between the county forecast and the sum of the forecasts for the parts was proportionately distributed to each of the individual sub-area forecasts. This distribution was made by calculating a ratio of the county total to the sum of the parts (the control factor) and by multiplying each individual sub-area forecast by the control factor. In the bottom-up approach, the county total is adjusted to meet the sum of the parts and the control factor is the ratio of the sum of the parts to the county total. This control factor is then multiplied by the county total to arrive at the final reconciled result.

Please note that in some instances, fluctuations in the forecast growth rates are at least partially attributed to the reconciliation of the sub-areas to the county-level forecasts.

Supporting Data and Projections Produced from Other Demographic Models

In addition to evaluating demographic trends detected from the data we used in our forecasting models, we reviewed other data and information to obtain a better understanding of the dynamics of population change specific to our study areas. This supporting information helps us to make better, or more realistic, assumptions about future population growth and helps us to use better judgment when making adjustments to our demographic models. Most of the supporting data and information were available either at the county level of geography, or for other large geographic areas. The sources include labor force data and economic profiles from the Oregon Employment Department, school enrollment data for school districts from the Oregon Department of Education, and demographic and socioeconomic data from the U.S. Census Bureau's American Community Survey (ACS). Also, population forecasts from the Oregon Office of Economic Analysis (OEA), and employment projections from the Oregon Employment Department were used to gauge our county-wide results and for comparison.

Also, to help make our forecasts more accurate, we developed additional sets of population projections from demographic models other than the primary models employed in this study. Secondary sets of projections were produced to serve as an evaluation tool to verify that the numbers forecast from the primary models are reasonable. The additional projections were used to detect and evaluate, and adjust if necessary, any inconsistencies that those primary forecasts may have had.

Additional housing unit models were developed for other geographic sub-areas in this study, not only for the smaller sub-areas in this study. For areas where a cohort-component model was created to produce its population forecast, the forecast results generated from the two models were checked and compared.

We also related and compared our county population forecasts to employment projections produced by Oregon's Employment Department. We developed a simple economic model to forecast county-wide net migration based on the projected demand for additional workers in the employment projections. The projected net migration was compared to the net migration forecasted in our model.

Data Sources and Description

Equivalent types of datasets were compiled for most of the geographic parts in the forecast region. Some data, such as those from the American Community Survey (ACS), are only available for one or three-year periods for geographic areas whose population is a minimum of 20,000. This means for many geographic parts within the forecast region, ACS data for one and three-year periods were not available.

The population forecasts are based on data obtained from several sources. Most of the data were aggregated to the county and its sub-areas by PRC staff. The data sources include:

- The U.S. Census Bureau's Decennial Census. The decennial Census is the only source of data collected for small areas across the nation. We used 2000 and 2010 Census data to obtain the population, by age and sex, residing in each county and its sub-areas. We compared the changes from 2000 to 2010 to develop an initial estimate of the age-sex profile for net migrants in the cohort-component models. Female population ages 10-49 were used with birth data to calculate fertility rates. In addition, data for population by race/ethnicity, and housing were obtained from the two Censuses.

- American Community Survey. These U.S. Census Bureau survey data are available for one and three-year periods for areas with a population of 20,000 or more. The American Community Survey asks the same or similar questions as the 2000 census. We used the 2000 census and ACS tables to develop estimates of housing and population change, including estimates of net migration.

- Birth and Death Data. Information on births and deaths were obtained from the Oregon Center for Health Statistics. The data were used for three purposes: 1) Calculate overall fertility and mortality rates for each county and its larger sub-areas. 2) Note the number of births and deaths in order to examine their trends and the correspondence between births, deaths and population change. 3) Derive age-specific net migration rates together with census population data.

- Annual Population Estimates. Annual population estimates for cities and counties of Oregon are prepared by the Population Research Center at Portland State University as part of its Population Estimates Program. Data on State income tax returns, births, deaths, Medicare and school enrollment, and information about changes in housing stock and group quarters population are utilized in developing the population estimates. We used historical and recent population estimates of the counties and cities within the forecast region to help approximate growth trends throughout each county.

- Group Quarters and Annexation Data. Data for the population residing in group quarters facilities and for the numbers of persons living on properties recently annexed into cities from the county were available from PRC's Population Estimates Program.

- UGB Boundary Files. Local governments and agencies, and the Oregon Geospatial Enterprise provided the boundary files for UGBs within the forecast region, and we requested that they be verified by local government staff. The boundaries are those that were current in the forecast launch year. These files are used for mapping and for aggregating demographic and other data unique to each geographic part in the forecast region.

- Land Use Data. Tax lot data were collected from county tax assessors and were used to create current housing unit inventories for the geographic parts in our study area. Tax lot data were used to identify housing units.

- Local Employment Dynamics Data. These data provide background information about commuting patterns of workers. An area's availability of employment or draw of workers, influences population and housing changes. These data were evaluated to detect changes in commuting patterns.

- Oregon Labor Force Data and Employment Projections. Labor force data from the Oregon Employment Department were evaluated to determine trends and their relation to population change. The employment projections, also from the Employment Department, were available for

the economic regions which approximate multiple counties in the forecast region. The employment projections were compared to our county level forecasts.

- Regional Economic Profiles and Reports. Background and current economic information for the economic regions and counties within the forecast region were obtained from the Oregon Employment Department. The information was used to provide us with an understanding of historical and recent economic trends and the general economic climate in our study area. Ultimately, the information enabled us to make more rational assumptions when forecasting future population for each county and its sub-areas.
- Other Background Information. County and city comprehensive plans, amendments to those comprehensive plans, other population forecasts prepared for counties or cities, and other planning reports and documents were examined to obtain background information. Additional information that county and city officials and staff thought might have bearing on the population forecasts were collected from most jurisdictions in the forecast region.

General Comments about Population Forecasts

The longer the time-span of the forecast, the more likely it is that conditions change, and thus will increase the uncertainty in rates and assumptions. It is crucial to have recent data that allows for testing or calibrating of the assumptions used in the forecasting models. The study area's historical population helps to calibrate and adjust original migration rates and growth rates in the forecast models so that a better fit between actual and predicted number of persons can be achieved. In the long-run, however, the local economy and conditions affecting populations are likely to change in ways not currently anticipated.

All population forecasts are based on a combination of a beginning population; various known, estimated, and predicted rates; and the forecaster's judgment about future trends. The forecasts may err through imprecise data or unexpected shifts in demographic trends. Generally, forecasts for larger geographical areas, such as the entire county are more reliable than those for small areas, such as for a small city with fewer than 1,000 persons. These forecasts may be used as a guide to population growth over the next few years. However, changes in local areas will surely affect populations in some cities and actual populations will deviate from those forecasted. The differences between the forecast and actual populations will vary in magnitude and perhaps direction.

A Note of Caution about the Forecasts Themselves

Given that these forecasts are developed for long-term trends, they are conservative. This means that they do not assume drastic changes to the population trends, such as seen during a recession.

Policy makers should view population forecasts as one of several available sources of information about likely future conditions. PRC's forecasts are based on assumptions developed from analysis of historical trends and expectations for the future. While the past gives some indication of what is likely to happen in the future, there is always the possibility of the occurrence of unforeseen events that could have a significant impact on population change. Thus, users of these forecasts should be aware that unexpected changes could happen and that it is wise to evaluate forecasts periodically in future years.

Given the uncertainty of the timing, occurrence, and magnitude of future events, several points should be kept in mind when interpreting the population forecasts in this report.

First, the population forecasts represent a most-likely scenario derived from assumptions representing our best judgment as to the possibilities for future conditions. It is not possible to judge at this time which of the assumptions, or combinations of assumptions, may best forecast future populations. The initial few years of the forecast period will better reveal whether the modeled demographic trends are likely to occur. If different conditions arise, then it would be appropriate to revise the population projections, taking into account new assumptions.

Second, variations in forecasts become larger in the long run. As years go by, the population forecasts depend increasingly on assumptions about who and how many persons will move into and out of the study area and the number of births that will occur annually to parents who reside in the study area. The population forecasts become less certain over longer periods of time.

Third, the smaller the population the harder it is to develop an accurate forecast. Slight unpredicted variations in demographic trends can cause larger fluctuations in these population forecasts than those for larger populations. Forecasts for large cities and counties tend to be more precise than forecasts for small cities or towns.

Finally, there is a temptation in interpreting forecasts to ask: "Which is the correct forecast?" Asking such a question implies that there is need to pick one forecast at present and then base future plans on it. The more appropriate use of the forecasts is to consider that there is likely to be some variation around the numbers and that we will want to update them as conditions evolve. Instead of deciding what exact outcome will occur over the forecast horizon, we urge government officials and the public to "monitor and manage" the changing conditions that will affect future populations. The most-likely forecast can best serve as a guideline in this process of monitoring and managing.