A Brief Overview of H-GAC's Regional Growth Forecast Methodology

-Houston-Galveston Area Council
Email: forecast@h-gac.com
Date Updated: May 3rd, 2016

Introduction

H-GAC releases an updated forecast every year. The forecasted items include population, employment, and land use. The forecasting system produces outputs in annual increments from 2015 through 2040. The base year for the forecast is 2014.

The forecast is produced in phases.

1. We forecast the total number of people and households in the region.
2. Based on the future labor force, we forecast the number of jobs.
3. The model makes predictions about the location, type, and size of residential and non-residential development projects which would be needed to accommodate the expected growth in households and jobs.
4. The expected growth in households and jobs is allocated to different areas in such a way that each household has a home (housing unit) and each job has a work site.

These phases correspond to different components of our forecasting system:

- Demographic Evolution Model
- Employment Model
- Real Estate Development Model
- Household Location Model
- Employment Location Model

There are several important features of our forecasting system: disaggregation, interrelation, and property of being data-driven. *Disaggregation* means that our models deal with individual elementary entities: people, households, jobs, land parcels, and buildings. All summary statistics, such as county population or total jobs in a census tract, are derived from data on the individual entities. For a future year, that data is not observed but rather created in a process known as "simulation". A simulation is a computational game-like technique which aims to imitate the dynamics of real life by setting up the "players" (entities or agents) and "rules" (propensities or parameters) and then letting the action unfold over time. In that respect, when we develop a forecast, we construct long lists of plausible future events for millions of entities.

Interrelation means that the different models are connected:

- Population determines the short-term supply of labor force;
- Change in the number of households determine the demand for housing;
- The development industry responds to demand for housing and non-residential buildings; and
- Employers' and households' location choices are limited to what is available at the moment.

Data-driven means that inside each model there are dozens of tables with data elements that control the "rules" which govern the simulation. In that respect, the forecast can be viewed as a particular "what if" scenario predicated upon thousands of very specific "assumptions." That makes it very easy to make updates when new information becomes available or when a correction is needed.

Demographic Evolution Model

Population change over a period of time in any area stems from either addition or removal of residents, as compared with the previous time period. The sources of addition are births to local residents and in-migration of people from other areas. The sources of removal are deaths of local residents and out-migration of local residents to other areas. Our model represents all four processes and estimates directly the number of births, deaths, in-migrants, and out-migrants in the area. Additionally, the model represents the household formation and dissolution processes.

The Demographic Evolution Model is a computer simulation which uses the probabilistic approach to imitate both the biologic events (dying, giving birth) and social events (marriage, divorce, migrating in or out of the area) for the synthesized individuals and households.

Population in the model is stratified into four race/ethnicity categories (non-Hispanic White, Hispanic White, Black, and Other), two sex categories (male, female) and 111 age categories (single year, from 0 through 110). The base-year data is constructed from the block-level 2010 Census data (SF1 tables). The base-year data consists of a list of individuals and a list of households. The two biological parameters, survival rates and birth (fertility) rates, are provided by demographers at the Texas State Data Center. We use these rates as event probabilities, that is the likelihood that a person will live another year, and the likelihood that a female will give birth in a given year. These rates change over time, reflecting the trends for longer life spans and decreasing fertility.

Migration probabilities or rates control the in-flow and out-flow of households. There are three separate migration rates dealing with different origins and destinations of the flows: domestic in-migration, domestic out-migration, and foreign in-migration. Some foreign out-migration does occur; however, we are not modeling it explicitly due to lack of data). We constructed the migration rates from the American Community Survey (ACS) Public Use Microdata Sample (PUMS) records. We used the same source to construct the two parameters that control the household dynamics--marriage and divorce rates--as well as a set of auxiliary probabilities, which are used to "match" brides and grooms.

The essence of the simulation technique is in comparing, for each individual, event probabilities with a randomly generated number. If certain conditions are met, the event (death, birth, migration) “occurs.” For example, assume that an individual’s survival rate is 0.95 (there is 5% probability that a person will not survive into the next year). A randomly drawn number is constrained to be between 1 and 0. The condition that can be established in this case is this:

- If a random number is smaller than 0.05, then the individual will not survive into the next year; or
- If that number is 0.05 or greater, then the individual will survive into the next year.

When this experiment is repeated many times, the observed frequency of the “death” event is very close to 5% (which is the event probability).

The Demographic Evolution Model creates a virtual accounting of all people and households in the region in the future years. In the current release specification, we use the outputs of the Demographic Evolution Model to create regional totals for population and households.

Employment Model

In the short run, the workforce (and jobs) is constrained by the resident population. Due to age, disability, family responsibilities or other factors, some people do not seek employment. The parameter that controls this attrition is the labor force participation rate (LFPR). Other people cannot find work and remain unemployed. The parameter that controls this imbalance is the average (overall) unemployment rate (UR). Therefore, in the short-term framework, employment can be estimated given the population and the rates (LFPR and UR). We use the ACS PUMS data to derive LFPR and UR specific for age and ethnicity cohorts. While LFPR is not expected to fluctuate over time, that is not true with respect to the UR, which reflects macroeconomic conditions.

These macroeconomic conditions are represented with an average (as opposed to cohort-specific) unemployment rate. For forecasting purposes, we assume that the future average unemployment rate will be 5.8% (a historical long-term regional average). Once the labor force (based on the resident population) is established by applying the LFPR to the population, the total employment can be calculated and distributed back to the cohorts using the UR. Total jobs in the region are then allocated to 20 two-digit NAICS sectors using the shares derived from the employment forecast produced by Woods & Poole.

Further, the total jobs by sector are converted to "location-specific" jobs which include wage and salary jobs and some self-employment jobs. “Location-specific” refers to jobs being tied to individual buildings. In the base year (2010), firm-level jobs derived from the Texas Workforce Commission and other employment data sources are linked to specific locations (individual buildings) by matching company and parcel addresses.

Real Estate Development Model

The Real Estate Development Model generates predictions for specific projects on specific parcels, given the physical availability/suitability of land and economic feasibility.

First, we estimate annual demand for housing units and non-residential space based on the forecasted change in the number of households and jobs. With respect to demand for housing, we have to make an assumption about the future shares of the single-family and multi-family housing. In the current specification, we use constant shares, so that 70% of the total demand will be for single-family units and 30% for multi-family units. These shares are based on the historical data for building permits issued in the region. Demand for non-residential space is determined by distributing the forecasted two-digit NAICS-sector jobs into different building types (office, retail, warehouse, etc.) and then applying building type-specific space consumption ratios (square feet per employee).

Second, we break the aggregate demand into different classes of projects (small single-family residential subdivision, medium-size apartment complex, large office building, etc.).

Third, we reduce the demand estimates to account for the supply coming from "known developments" (announced, planned, or under construction projects). We use various sources to compile information on such projects.

Fourth, once the quantity of different classes of projects is established, the model generates a large number of development "proposals." These proposals cover all possible combinations of projects that could be developed on the available parcels, irrespective of the economic feasibility. For each proposal, we calculate a total cost, which includes the cost of land and construction cost (per square foot cost multiplied by the square feet of the project), plus the cost of the existing buildings and demolition in the case of redevelopment. Data related to parcels and buildings is derived primarily from the appraisal records. Our database includes over two million parcels. For each parcel, we know the land use type, value of land, and value and type of buildings. The construction costs are based on the recent industry survey and published data.

Fifth, we calculate the expected sale price of a project using the coefficients from a series of regression models. These models establish the relationships between the per square foot prices for different types of buildings and various intra-urban proximity and accessibility measures. Accessibility is measured in terms of access to a given space (non-residential, residential) or jobs (by type of job) within a certain travel time (15 min and 60 min).

Six, the profitability (return on investment) for all the proposals is calculated (relating the expected sale price to the total costs). Finally, from a pool of most profitable, proposals are selected for "construction" until the demand is met.

We use additional procedure to analyze the distribution of retail and service jobs in primarily residential locations and locate retail and service buildings to ensure adequate levels of access to local retail and services in areas that experience growth in residential buildings. We use a similar procedure to place new schools and hospitals.

Future land use changes generated that way are called "model predictions." The other type of future land changes is called "known developments." Both types appear as layers in Regional Land Use Information System (RLUIS), our web mapping application.

Household Location Model

The household location model assigns households (demand) to housing units (supply). Housing units that are available for household assignment includes newly built housing units and vacant housing units resulting from relocation or death of a 1-person household. Conversely, households that need a housing unit include relocating households (households moving within the 8-county region), in-migrants (domestic and foreign in-migrants moving to the 8-county region) and households whose building was demolished because of redevelopment (in real estate development model). The in-migrating households are first assigned to each of the eight counties using the county-level domestic and foreign in-migration rates derived from ACS County to County Migration Flows and US Census Population Change Components data. In the next step, the count-level in-migrating households are assigned to individual housing unit within a given county using the grid-level (3 mile grid) location probabilities categorized by age-race-household size and income. Location probabilities are calculated based on the base-year household distribution. In case, if there are not enough housing units available to accommodate the new households then those households will be allocated to available housing units in other counties based on county-level vacancy rates and the grid-level location probabilities. Similar to the in-migrating households, relocating households and households whose building was demolished are first assigned to individual counties using the county to county relocation rates derived from the ACS. The households are then assigned to specific housing unit using the grid-level location probabilities specified above. So, at the end each household in the region is tied to a specific housing unit.

Employment Location Model

In the current specification, new jobs are assigned to available space inside buildings controlling for the type of building and applying building type-specific space consumption ratios. This ensures that, for example, retail jobs are assigned to retail buildings and there is enough space to accommodate these jobs. With respect to the locational aspect, the probabilistic assignment is designed to maintain the existing sectoral composition of jobs.