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

-Houston-Galveston Area Council
Email: forecast@h-gac.com
Data updated; February 4, 2025

Introduction

H-GAC updates its forecast every two years, covering population, employment, and land use. The forecasting system generates projections in five-year increments from 2020 to 2050, with 2020 as the base year.

The forecast is developed in phases:

1. Estimating the total population and number of households in the region.
2. Forecasting the number of jobs based on the future labor force.
3. Predicting the location, type, and scale of residential and non-residential developments needed to support projected household and job growth.
4. Allocating expected household and job growth across different areas, ensuring every household has a housing unit and every job has a designated 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

Our forecasting system is built on three key features: disaggregation, interrelation, and a data-driven approach.

Disaggregation: Our models operate at the level of individual entities, such as people, households, jobs, land parcels, and buildings. Aggregate statistics, like county population or total jobs in a census tract, are derived from these individual entities. For future years, this data is generated through a process called "simulation." Simulation is a computational technique that mimics real-world dynamics by defining "players" (entities or agents) and "rules" (propensities or parameters), then allowing the system to evolve over time. In this way, our forecasts construct detailed, plausible future scenarios for millions of entities.

Interrelation: Our models are interconnected, meaning changes in one area influence others:

Population growth affects the short-term labor force supply.

Household changes drive housing demand.

The development industry responds to demand for residential and non-residential buildings.

Employers' and households' location choices depend on available options at any given time.

Data-driven approach: Each model contains extensive datasets that define the rules governing the simulation. This makes the forecast a structured "what-if" scenario based on thousands of precise assumptions. As a result, the system can be easily updated when new information becomes available or adjustments are needed.

Demographic Evolution Model

Population change in any area over time results from the addition or removal of residents compared to the previous period. Additions occur through births to local residents and in-migration from other areas, while removals happen through deaths and out-migration. Our model captures all four processes by directly estimating the number of births, deaths, in-migrants, and out-migrants. It also represents household formation and dissolution.

The **Demographic Evolution Model** is a computer simulation that uses a probabilistic approach to replicate biological events (births and deaths) and social events (marriage, divorce, and migration) for synthesized individuals and households.

In the model, the population is categorized by:

- **Race/Ethnicity**: Non-Hispanic White, Hispanic White, Black, and Other
- **Sex**: Male and Female
- **Age**: Single-year cohorts from 0 to 110

The base-year dataset is built using block-level 2020 Census data, consisting of a detailed list of individuals and households. Two key biological parameters—**survival rates** and **birth (fertility) rates**—are provided by the Texas Demographic Center. These rates are used as event probabilities, determining the likelihood of survival or childbirth in a given year, and they evolve over time to reflect trends such as increasing life expectancy and declining fertility.

Migration dynamics are governed by three separate migration rates:

1. **Domestic in-migration** (people moving into the region from other U.S. locations)

2. **Domestic out-migration** (people leaving for other U.S. locations)
3. **Foreign in-migration** (people arriving from outside the U.S.)

Although some foreign out-migration occurs, it is not explicitly modeled due to limited data. Migration rates are derived from the American Community Survey (ACS) Public Use Microdata Sample (PUMS) records. The same data source is used to estimate **marriage and divorce rates**, as well as auxiliary probabilities that help simulate spousal matching.

The core of the simulation involves comparing event probabilities with randomly generated numbers to determine outcomes. For example, if an individual's survival probability is 0.95 (meaning a 5% chance of death), a random number between 0 and 1 is generated:

- If the number is less than 0.05, the individual does not survive.
- If the number is 0.05 or greater, the individual survives.

Repeated over millions of cases, this method ensures that the observed frequency of events aligns closely with expected probabilities.

The **Demographic Evolution Model** provides a comprehensive virtual accounting of future populations and households in the region. In its current specification, the model's outputs are used to generate regional population and household totals.

Employment Model

In the short term, the workforce and job availability are constrained by the resident population. Some individuals do not participate in the labor market due to age, disability, family responsibilities, or other factors. This is governed by the labor force participation rate (LFPR). Others actively seek work but remain unemployed, which is controlled by the unemployment rate (UR). Given population levels and these rates, short-term employment can be estimated.

We derive LFPR and UR values specific to age and ethnicity cohorts using ACS PUMS data. While LFPR remains relatively stable over time, UR fluctuates based on macroeconomic conditions. For forecasting purposes, we assume a long-term average unemployment rate of 5.6%, based on historical regional trends.

The labor force is determined by applying LFPR to the resident population, after which total employment is calculated and distributed across cohorts using the unemployment rate. Jobs are then allocated to 20 two-digit NAICS sectors, using employment projections from Woods & Poole Economics, Inc. and the Texas Workforce Commission.

Next, total jobs by sector are converted into location-specific jobs, which include both wage and salary positions as well as some self-employment roles. "Location-specific" means that jobs are directly linked to individual buildings. In the base year (2020), firm-level job data from Data Axle and other employment sources were matched to specific locations by aligning company addresses with parcel data.

Real Estate Development Model

The **Real Estate Development Model** predicts specific development projects on specific parcels based on land availability, suitability, and economic feasibility.

1. Estimating Demand

The model first calculates the annual demand for housing units and non-residential space based on projected household and job growth. For housing, assumptions are made about the future distribution of **single-family (SF) and multi-family (MF) housing**:

- **2020–2030:** 70% SF / 30% MF
- **2030–2040:** 68% SF / 32% MF
- **2040–2050:** 65% SF / 35% MF

These projections are based on historical building permit data. Demand for non-residential space is determined by allocating forecasted **two-digit NAICS-sector jobs** to different building types (office, retail, warehouse, etc.), applying **space consumption ratios** (square feet per employee).

2. Classifying Development Projects

Aggregate demand is broken down into different project types, such as small single-family subdivisions, mid-sized apartment complexes, and large office buildings.

3. Accounting for Existing Developments

To avoid overestimating demand, the model factors in **"known developments"**—projects that are announced, planned, or under construction—by compiling data from various sources.

4. Generating Development Proposals

Once the number of required projects is established, the model generates a large set of potential **development proposals**, considering all possible projects that could be built on available parcels, regardless of economic feasibility.

For each proposal, the total cost is calculated, including:

- **Land costs**
- **Construction costs** (square footage multiplied by per-square-foot costs)
- **Existing building values and demolition costs** (for redevelopment)

Parcel and building data are primarily sourced from **appraisal records**, covering over **two million parcels** with detailed information on land use, land value, and building types. **Construction costs** are based on industry surveys and published data.

5. Development Constraints

In addition to **zoning regulations**, the model incorporates **real estate development restrictions**, defined based on practical feasibility for specific areas. For example:

- Houston Downtown parcels are restricted from industrial and warehouse projects.
- Houston Ship Channel parcels are limited to industrial and warehouse development.

Certain land types are excluded from development, including:

- Public infrastructure (roads, transit facilities, utilities, water channels, detention ponds, parks, etc.)
- Flood-prone parcels (where **>75% of the area lies within the floodway**)
- Specific land use categories (e.g., single-family homes, mobile homes, government buildings, hospitals, schools, recreational facilities, and unknown-use parcels)

6. Calculating Profitability and Selecting Projects

Using regression models, the model estimates the expected sale price of a project based on per-square-foot prices, intra-urban proximity, and accessibility factors. The return on investment (ROI) is then calculated by comparing expected sale prices to total costs.

From the most profitable proposals, projects are selected for "construction" until demand is met.

7. Allocating Retail, Services, and Public Facilities

An additional procedure ensures retail and service jobs are distributed in primarily residential areas to maintain access to local amenities. Similarly, schools and hospitals are strategically placed in response to population growth.

8. Future Land Use Predictions

The model generates two types of **future land use changes**:

1. **"Model Predictions"** – Forecasted developments based on economic feasibility.
2. **"Known Developments"** – Projects that are already planned or under construction.

Both appear as layers in the **Regional Land Use Information System (RLUIS)**, our web-based mapping application.

Household Location Model

The **Household Location Model** assigns households (demand) to available housing units (supply). Available housing includes newly built units and vacant units resulting from relocation or the death of a single-person household.

Households Requiring Housing

Households in need of a housing unit fall into three categories:

1. **Relocating households** – those moving within the **8-county region**.
2. **In-migrants** – new households moving into the region from domestic or foreign locations.
3. **Households displaced by redevelopment** – those whose buildings were demolished in the **Real Estate Development Model**.

Household Assignment Process

Step 1: County-Level Assignment of In-Migrants

- In-migrating households are first distributed among the eight counties based on **county-level domestic and foreign in-migration rates**, which are derived from:
  - **ACS County-to-County Migration Flows**
  - **U.S. Census Population Change Components data**

Step 2: Grid-Level Assignment Within Counties

- Once assigned to a county, in-migrating households are allocated to individual housing units within a **3-mile grid**.
- The assignment is based on **location probabilities**, which are categorized by:

- **Age**
- **Race**
- **Household size**
- **Income**

- These probabilities are derived from the **base-year household distribution**.

Step 3: Handling Housing Shortages

- If there are **not enough available housing units** within a county, unassigned households are redistributed to other counties.
- This redistribution is based on:
    - **County-level vacancy rates**
    - **Grid-level location probabilities**

Step 4: Assignment of Relocating and Displaced Households

- Similar to in-migrants, relocating households and those displaced by redevelopment are first assigned to counties based on **county-to-county relocation rates** from the **ACS**.
- They are then assigned to specific housing units within each county using the same **grid-level location probabilities**.

At the end of the process, **every household in the region is assigned to a specific housing unit**, ensuring an accurate distribution of population across available housing stock.

<u>Employment Location Model</u>

In the current specification, new jobs are allocated to available space within buildings based on building type and space consumption ratios specific to each building category. This ensures that jobs are placed in appropriate spaces—for example, retail jobs are assigned to retail buildings, with sufficient space to accommodate them.

From a locational perspective, the probabilistic assignment method is designed to preserve the existing sectoral composition of jobs, ensuring a balanced distribution across industries.