STI: PopStats - Methodology

Every market researcher knows that many of today's markets are changing rapidly - especially high-growth markets. In particular, new businesses rapidly move in and begin securing their share of the local consumers' business, loyalty, and dollars. As a result, the first retailers to drive in new markets have the singular opportunity of winning the lion's share of business in their categories from the local consumers - no matter how many competitors arrive second or third, or later. That's why the goal of many of today's smartest marketers is to be the first business to move into today's scarce and often difficult-to-find new high-growth areas.

Aggressive competition is not the only challenge facing today's marketers. Consumer demographics are also volatile - experiencing ongoing changes in population counts, ages, and income levels.

Adding to the challenges is the fact that the majority of today's population estimates are updated only once every 12 months. As a result, the data being used to target rapidly changing new markets and new consumer segments is essentially out-of-date. This has put marketers at a serious disadvantage for decades. That's why Synergos Technologies created the industry's first "quarterly" population statistics - and, as a result, has completely revolutionized population estimating. Instead of being a generalized process that adds marginal value to market selection, population estimates with STI: PopStats make a powerful contribution to identifying highly valuable new locations and consumer groups.

A Historical View of Population Estimates

In the past, all demographic information, including population estimates, were released once a year (typically in May or June). The method used to generate the annual estimates was a "top-down" approach, which started by calculating a national population estimate, then a state estimate, then a county, etc., until the calculations reach the local level. This national-to-local estimating process used by the U.S. Census Bureau, was adopted and subsequently enhanced by demographers to make population estimates at the block group level. However, this method has serious drawbacks for businesses that want to pinpoint the best new markets for opening new businesses.

The top-down approach is used by the U.S. Census Bureau for research at the macro level and is, therefore, unsuitable for use at a micro level, like block groups, which are greatly influenced by singular events, like the opening of an apartment complex or a building demolition. To compensate for this significant shortcoming, demographers developed spreading techniques that broad-stroked areas of growth and decline at the sub-county level. While the broad-stroke market analyses of growth or decline are often accurate, this methodology also has a significant defect: It tends to mask specific, block-group-level areas of growth or decline, thereby, concealing hidden opportunities of market development.

A New Approach to Population Estimates

In 1997, Synergos Technologies approached the task of creating population estimates with the dual goal of overcoming the significant problems with the traditional methodology and greatly elevating the accuracy of the estimate itself. As a result, we have taken population estimates in a whole new direction - literally.

To create STI: PopStats we use a "bottom-up" methodology, which gives us a unique advantage and gives market researchers several significant benefits. We start at the lowest level of geography possible, which is the ZIP+4™ level, and then we move up the ladder of standard U.S. Census Bureau geography (block group, tract, county, and state). The ZIP+4 level targets markets with greater precision, because it:

- Is extremely detailed, containing over 28 million records,
- Covers all major population centers,
- Can be manipulated statistically, and
- Is easily consolidated into any geography necessary.

Plus, ZIP+4 targets areas as small as a specific group of houses - typically four to 12 - or a building. As a result, we can literally see structures come online as they are finished being built and occupied.

To further ensure accuracy, we use a series of checks-and-balances to validate our results, including consulting with multiple state and federal agencies whose data is independently gathered and calculated. Because our method works from the bottom-up, and our controls come from entirely different and mutually exclusive sources, we are able to provide the most unbiased estimates possible with STI: PopStats.

STEP 1 - Estimating Households

Our research has shown a unique and quantifiable relationship exists between USPS (United States Postal Service) data and U.S. Census Bureau household counts. Due to this relationship, we model population shifts quickly and accurately using a proprietary technique that leverages the correlation between the two. (Note: To limit bias in the data due to extraneous figures, such as errors in the raw data, we use a variety of data filtering techniques to limit data irregularities.)

The process is initiated by base-lining the ZIP+4 data and its associated statistics as they existed in April 2000. Then, as new ZIP+4 data is provided (new data and statistics are delivered monthly) we can model and derive a USPS growth factor for every ZIP+4 in the country. Given the USPS growth factor over April 2000, we are then able to apply the growth factor to U.S. Census household counts as they existed in April 2000. This application occurs via our proprietary model that uses this information as well as other pertinent factors to generate a current estimate. In generating the five-year household forecast, best-worst-most-likely probability scenarios are created using several independently derived trend analysis curves. These curves then act as controls or boundaries on a sophisticated simulation technique called Monte Carlo Simulation.

STEP 2 - Estimating Household Populations

A variety of U.S. Census Bureau and private studies have shown that the relationship of persons-to-households remains relatively stable over time. Therefore, we take the Census 2000 persons-per-household-per-block group figures, and adjust the ratio to reflect any changes in the county estimated per-sons-per-household generated by the U.S. Census Bureau. We then apply these new figures to our estimated households to derive an estimated household population.

STEP 3 - Apply Controls

To ensure that the base population and household estimates are reasonable, we compare the information to the U.S. Census Bureau's annual population estimates released every Spring. If any major discrepancies occur between the two numbers, our model applies a set of heuristics to determine the most probable population figure. In addition, selected cities throughout the U.S. are field-surveyed to further validate our model's results.

Age/Sex/Race Breakouts

Once the base population has been estimated, our model then determines or "breaks out" the demographic components of the population. Age and sex are determined through a traditional cohort survival model. This sub-model to the main model looks at each age distribution within a race category and applies the appropriate birth and survival rates as determined by the NCHS (National Center of Health Statistics). These results are then balanced back to the base population using an iterative approach. In addition, in-formation from the NCES (National Center for Education Statistics) is applied to validate the age distri-bution of school-age children. U.S. Census Bureau estimates are used to validate all other age ranges.

Race is calculated by ratio analysis of April 2000 observed and annual U.S. Census estimated. In areas of high growth we use race information gathered by the FFIEC (Federal Financial Institutions Examination Council). This agency collects information from financial institutions concerning loans and race issues. We have found it to be a reasonable source for understanding race percentages in high-growth areas. As a final check for race, our model also consults with the NCES race data for elementary school children, and checks that data against our own figures.

Group Quarters

Another component of population estimates is Group Quarter data. In layman's terms, group quarter is defined as a collection of unrelated people where no one individual can claim "head of household." Generally speaking, group quarter data can be divided into three categories: institutions (state homes, hospi-tals, and prisons), colleges, and military bases.

We determine a group quarter estimate by estimating each category individually, then combine the results for a total estimate. Military group quarters are determined based on a direct data feed received from the DMDC (Defense Manpower Data Center). College student dormitory information is derived from the NCES (National Center for Education

Statistics) and its annual college survey. Institutionalized persons are estimated by using historical trends as provided by the U.S. Census Bureau.

Income Estimates

Income estimates are based on a two-step process. First, household incomes at the county level are estimated. Our estimates are based on a blend of information from the Survey of Income from the IRS, income estimates from the U.S. Census Bureau's March CPS, and personal income estimates from the BEA (Bureau of Economic Analysis).

Once the county estimate is derived, we are then ready to estimate data at the block-group-level. This low-level estimate is done in two parts. First, we separate existing households from new-growth households. The reason for the separation is because our research has found that in high-growth areas existing households do not lend themselves to be a good base for determining the income of the new households entering the area. Therefore, we use a typical income growth approach that resembles the growth of county income. Then we add to that a separate income growth for new households. New household income is modeled on mortgage data transactions received from the FFIEC.

Housing Values

Housing Values are determined in a fashion similar to incomes. Housing, and their associated values, that existed as of April 2000 are updated using data from the OFHEO (Office of Federal Housing Enterprise Oversight). This federal organization performs a very detailed analysis of same home selling prices that occur over time. We use the resulting growth factors and apply them to existing April 2000 Owner Occupied homes. New home values (homes built after April 2000) are determined by ratio analysis of mortgage values from the FFIEC and actual selling price.

Sources of Information:

United States Postal Service (USPS)
United States Department of Defense (DMDC)
United States Census Bureau
National Center for Education Statistics (NCES)
Federal Financial Institutions Examination Council (FFIEC)
Internal Revenue Service (IRS)
Bureau of Economic Analysis (BEA)
Bureau of Labor Statistics (BLS)
Office of Federal Housing Enterprise Oversight (OFHEO)