

The EASI Research Methodology - Our Philosophy

Our intent is to establish a proper benchmark or starting point for a data series, which ensures a reliable and reasonable source for updating. We then find and develop a logical and consistent set of information, from reliable sources, which we then use to develop procedures, models, and algorithms to update and forecast the data elements in a manner that allows for accountability and accuracy.

The following is a general description of the methodology used by EASI to update the demographic and economic characteristics for the United States, States, Counties, ZIP Codes, Census Tracts, Block Groups, and ZIP Plus 4's.

The purpose of this explanation is not to divulge any proprietary methods but to illustrate the efforts made on your behalf to create accurate updates. EASI statistician's and programmers have over 30 years of experience updating these types of data. By industry standard EASI estimates would be considered of the highest quality.

Input Files

1. With the current release EASI will benchmark at the Block Group and higher levels all of the details supplied with the 2000 Census (all related releases at the Block Group level). All data are now derived from BG data from SF3 and for certain variables SF4 (Ancestry). If SF4 data are the benchmark EASI develops BG data use. Starting with the 2007 release EASI is adjusting various Census files to conform to the American Community Survey data. Call for details.
2. In all current estimates EASI racial data including Black Population, Asian Population, White Population, and Other Population. These are based on different questions with the 2000 Census and these data are not compatible with the 1990 Census (multiple race categories are now possible and are part of the Other group). EASI uses the 2000 Census Block Group data as our benchmark and makes adjustments for consistency with age and sex and with household counts.
3. EASI has collected from the Census Bureau all current local (counties etc.) and national updates and estimates for all the key demographic information. All these official estimates have been analyzed and then incorporated into our estimates and projections using a variety of EASI models.
4. EASI has summarized from the United States Postal Service (USPS) mailable Households at a County, ZIP Code, Census Tract, and Block Group level. These data have been used as the primary input to estimate local current change within a small area such as a Block Group. Mailable households are not the same as Census Households but are used to indicate recent change in household formations. These changes are combined with an EASI proprietary model for updating and forecasting at the Block Groups.
5. The Mailable Household data match starts by identifying, for every ZIP Plus 4, (ZIP+4), which Block Group it belongs to. EASI develops a split file and a plurality file of these matches using the latest Tiger file, to determine which

Block Group (primary) they should be assigned to. One of the key goals is to identify all correct ZIP Codes and ZIP+4's and then assign them to the correct Block Group that these Mailable Households should be assigned to. EASI has also reconfigured the 1990 Block Groups into the 2000 Block Group configuration to estimate the 1990 population for comparative purposes. An analysis of this decade change is also included in our model.

6. EASI has also analyzed the 2000 Census Block files in order to create a **population** Centroid for each Block Group. The results of that analysis are used for all ring study analysis.
7. Specific other sources include:
 - a. Bureau of the Census - 2000 Census PL 94 - 171; 2000 Census SF1 and SF3. (These Census data are the EASI benchmark or starting point for the demographic updates and the forecasts.) Other related sources are: Annual Demographic Survey, Current Population Reports (P20; P25; P60; and numerous special Census reports.
 - b. ZIP and County Business Patterns (US Department of Commerce - Economics and Statistics Administration- Bureau of the Census.).
 - c. US Department of Justice - Federal Bureau of Investigation. (2006).
 - d. National Center for Education Statistics - Common Core of Data (CCD)
 - e. National Oceanic and Atmospheric Administration - National Environmental Satellite, Data and Information Service - National Climatic Data Center.
 - f. United States Department of the Interior - Geological Survey - Office of Earthquakes, Volcanoes, and Engineering.
 - g. Bureau of Labor Statistics - Department of Labor.
8. **New Orleans - Katrina Analysis.** In past years, EASI has incorporated the latest Census estimates for the Katrina affected areas. EASI has analyzed a variety of maps and input data primarily using ZIP codes as the key geography where more information is know. Specifically, Block Groups in the area of destruction are then controlled to ZIP codes. Call 800 469 3274 (HOW EASI) for more details.

Data Preparation

The steps in creating the ZIP Plus4 (ZIP+4) and Block Group mailable Households include:

1. Start with a USPS ZIP+4 file for December (year prior to the estimating year) which includes all valid residential ZIP+4's in the country whether are residential mail or not
2. For each ZIP+4's, we add Census Blocks Groups based upon the Tiger file distance formula. Approximately 20 million records are processed by this direct match (about 75%).

3. For each remaining ZIP+4, we match against our internal geocode file (latitude and longitude). This file is based on running through address matching/geocoding software. Approximately 18% of total are matched to their Block Group this way.
4. For each remaining ZIP+4 that cannot be geocoded by b) or c), we use a calculated carrier route or Block Group centroid. We weight the geographies to a larger area and calculate a latitude and longitude. We then determine which is the closest (distanced) Block Group. This is done for approximately 5% of total.
5. If a ZIP4 is still unassigned then, we use nearest neighbor ZIP+4. There are approximately 2% or total are done through this approach (recent, 6 months old ZIP+4s are often in this category).
6. Block Groups assignments are from the most recent Census Tiger file. Tiger errors, where identified (such as wrong FIPS Codes) have been corrected.
7. ZIP Plus 4's are assigned data based upon the data of the Block Group that it has been assigned to. (Note: There are no official Census Bureau data for ZIP+4.)
8. These mailable household data analysis are for residential ZIP+4's (no business-exclusive ZIP+4's are included).

Analysis

EASI has developed a series of models which use the relationship between the count of the current mailable households at the Block Group level to develop estimates of the change in the household size relationships at the BG compared to the county and to the ZIP Code. In addition, EASI analyzes the change in relationships between these mailable households over time and compares them to the county and to the ZIP Code households using a proprietary formula. Care is taken in this approach since there can be ZIP4 definitional changes. The analysis relates the current estimate of mailable households to the number of mailable households at the time of the 2000 Census (4/1/00) and as time progresses.

One key component of the analysis is a proximity site review of all ZIP+4's based upon their Block Group assignment (208,790 Block Groups). This analysis *prepares* our input data before use in EASI demographic models.

Newly released Census county estimate information are analyzed compared to prior releases.(P-25 and P-26) to develop current and forecasted county control totals through an analysis of population component changes (births, deaths, migration, etc.).

Annually, EASI also incorporates relevant national and state data as control totals. This is done for a variety of demographic factors. EASI derives this from analysis of national data, over time, from the Annual Demographic Survey, the Current Population Survey, American Community Survey, and the Annual Housing Survey. There are also from a variety of sources at the Census Bureau web site (www.census.gov).

ZIP Code results are independently compared to the USPS current ZIP Code file of residential deliveries. Additional updating sources include: USPS AMS files and Postal bulletins (the ZIP Alert); these record any annual changes that take place to ZIP codes including name changes, delivery or branch changes as they become official. Other sources include: U.S. Postal Service City-State File (monthly) and Delivery Statistics File. These CD ROM's incorporate main inventory of ZIP Codes and the post office and other names associated with them. Each year EASI conducts a complete review of these files to maintain a current ZIP Code roster. EASI inventories the old ZIP Codes as well.

Updates to the current year and a 5 year projections are first done at the United States level and for key variables at the county level as well. Block Group (BG) level estimates are all controlled to the county control totals. That is, the Block Group data will add to the separately generated county data for all data elements. In a similar manner, other geographies are summarized from the Block Group level. However, parts of BGs are added to get ZIP Codes and to get cities.

Consistency - year to year changes

Each year EASI uses **all** available sources to maintain the highest quality of our estimates. Sometimes the new information will makes year to year changes less meaningful. e.g. a current ZIP Code may have a different definition of BG's because of postal changes in the last year. However, the changes from our 5 year forecast, within an EASI calendar year, are consistent from the current estimate but changes from last's years estimates are not necessarily so. EASI geography estimates are all based on the same geography, which is all ZIP Code estimates for April 1, 1990; April 1, 2000, 1/1/current year; and a 5 year forecast are all based on the same geographic definition.

Starting in 2007 the Census Bureau has begun releasing The American Community Survey (ACS - www.census.gov) to supplement its Census 2000 SF3 data. EASI has incorporated all of these estimates into our 2007 updates and these will be carried forward and will use the latest released data. Note: These new ACS estimates offer an improvement for many of the previous data series especially for housing and related data. If you have questions or concerns about the impact of ACS, please call EASI 800 HOW EASI (469 3274) for a thorough and complete discussion.

Users must be use caution when comparing data from prior censuses or even releases of new Census data. For example, the 2000 Census has a new Race question (e.g. White Alone, Black Alone, Asian Alone, etc.) which allows for multiple races and is not compatible with previous estimates.

Another factor in consistency is that with some data sources information becomes available annually but with others data elements may not be released but once every two or even three years.

Occasionally a post Census estimates can be subject to revision for several reasons. In one instance a data series may be deemed more important by Congress and as result a

sample size can be expanded to allow for more detailed results. Another change could be that the sample is framed against any new data such as the 2000 Census. EASI with decades of experience analyzes all information and then EASI incorporates the results into our estimates.

ZIP Code Details - As mentioned above, ZIP Codes even if they seem to be the same (same 5 digits) are especially difficult for consistency from year to year (they are always consistent within the EASI data and software.) Since each ZIP Code area may change from year to year EASI spends considerable time and effort to develop new ZIP Code data for each and every year. That is, EASI assigns a portion of each Block Group to a ZIP Code based on the latest information for each year (1990, 2000, current and five year forecast). Note: Annually EASI's creates a proprietary ZIP to Block Group (partial) analysis and we also allocate all land area to create each ZIP Code.

Income

There are many different definitions of income that are available for analysis. With the release of the 2000 Census EASI has been using the actual 2000 Census Income estimates (for the year 1999) as our starting point. These estimates are then modeled using the P60 Money Income in the United States (Current Population Reports - Consumer Income) as well as other data. EASI income models are based on race and by family characteristics to obtain a current estimate. All use the 2000 Census definition of income as a benchmark.

EASI income estimates are controlled to analysis from the Money income data after analyzing the differences in that sample compared to the actual 2000 Census. EASI estimates inflation (current dollars) in all of our estimates and forecasts. EASI also maintains Income distributions based on gross income (includes all taxes).

Consumer Expenditure Survey (CEX)

The results of the CEX are analyzed annually by EASI and then combined with EASI estimates at the Block group level. The Bureau of Labor Statistics and the Bureau of the Census conduct the CEX. There are two parts to the survey. The first part is a diary, which is completed by respondents for two consecutive 1-week periods. The second part is an interview survey, which are conducted quarterly (3 months) for five quarters. The interview survey includes about 95% of all expenditures and includes large expenditures such as property, automobiles, major appliances, rent, utility payments insurance premiums, and many others.

EASI annually models these results of about 550+ categories of expenditures against our updated demographic estimates. EASI's models use our own BG demographic estimates to update these potential sales

An example:

EASI models the age of respondent, income of respondent, and tenure (own home versus rent). Then for each demographic characteristic we have an average expenditure for the previous calendar year (e.g. a respondent earning \$50,000 to \$75,000 spent \$210 (for example only) and we might then see that a respondent with income of \$35,000 to \$50,000 spent \$150 (for example only).

We take all the values for the demographics and then develop a model for this CEX characteristic that combines the factors to get one BG level estimate.

Retail Sales and Store Groups and Minor Stores and Major Merchandise Lines

EASI's Retail Sales Estimates include Food Service - Total Retail Sales includes the standard 12 major stores plus Food Service; 55+ Minor Stores, and 45 Major Merchandise Lines. All data are based on an extensive review of County and ZIP Code Retail Trade data for 2002. EASI created a file of benchmark data from the released Census data which is used for our annual update.

Each year, EASI creates a new consistent file of benchmark and updated for 2002, current, and a 5 year forecast. EASI re-benchmarks estimates for each update to a new set of Block Group estimates for all retail categories based on new information so our data over time is consistent. These estimates are based on our current analysis of the latest NAICS employment data for each retail store and food service. Note: EASI resolves any inconsistencies between sources as part of this annual process.

The 13 store groups that comprise Total Retail Sales are:

- Motor Vehicle and Parts Dealers
- Furniture and Home Furnishings Stores
- Building Material and Garden Equipment and Supplies Dealers
- Electronics and Appliance Stores
- Food and Beverage Stores
- Health and Personal Care Stores
- Gasoline Stations
- Clothing and Clothing Accessories Stores
- Sporting Goods, Hobby, Book, and Music Stores
- General Merchandise Stores
- Miscellaneous Store Retailers
- Nonstore Retailers
- Food Services

Call EASI for Minor stores and Major Merchandise Line information.

Benchmark Methodology and Assumptions

These retail data are benchmarked at the county level from the 2002 Census. Then EASI develops a ZIP code version of this file. EASI models these actual store locations at the Block Group level using a business employment relationship developed from the latest ZIP Business Patterns. This is done in order to allow the retail sales estimates to be used as part of standard database summaries. Note: EASI does not know the actual locations of stores at the Block Group. Other geographies are estimated by adding up the Block Group estimates.

The updates are modeled against estimated changes based upon the ZIP Business Patterns. Therefore, the sum of the BG's retail sales estimates within a ZIP Code is consistent to the ZIP Code Business employment data. Any inconsistencies between sources are reviewed and made consistent to the most current data from ZIP Business Patterns.

EASI models the retail trade data to a Block Group based on a proximity model. The model assigns exclusive Business or Retail ZIP Codes to the closest Block Group. For example, from ZIP Business Patterns EASI can identify point business locations and the retail configuration within each.

Accuracy

With all estimates and with ours as well, the higher the level of data (national is the highest) the more accurate the estimate. Our data follows standard demographic techniques, all developed with over 35 years of experience in this industry. It is considered a highly accurate technique.

EASI data has also been "field tested". That is, portions of our updated data are available at our web site and have been used by hundreds of thousands of users. These users raise questions about our updates, which we investigate. This input does help us to review and check results and makes our estimates better.

Here are some common questions:

Why are the Post Office mailable households different than EASI's?

One reason is that the differences between the counts of ZIP households in the Census and the mailable households from the post office is that there are differences in definitions between mailable households and Census households. There can be two mailable households in a residence but only one household. The Census will call it a single household if there is a relationship and the post office does not keep track of relationships

How close is EASI updated data to other sources?

EASI has made an extensive effort to obtain all relevant information and to incorporate it in a logical statistical manner. Other companies who use similar sources and statistical approach should give similar results. One method of comparison is a circle or ring study. An analysis of comparable ring studies has shown a current population difference of less than 2 percent. In denser population areas the results of the ring analysis are within .005 percent analysis. With the release of the 2000 Census an analysis showed that EASI ZIP Code estimates were in over 98% of the cases within .005 percent.

Validity Checking

EASI has made numerous checks for internal and external consistency in all our estimates. There are 3 types of checks that are rigorously reviewed. These include; Census internal consistency, controlling updates to definitions of estimates, and correcting for, or preventing, rounding errors, especially in small geographies.

2000 Census validity check is an analysis and comparison of the results of SF1 vs. SF3 estimates at the Block Group level. Due to sample sizes and Census procedures for disseminating the Census results there frequently are Census results which are inconsistent. These results are analyzed and EASI has developed a series of algorithms to adjust these estimates to make them consistent. (Examples are mostly in small BGs where there might be a single household, by total or by race, found in SF3 but no population in SF1. Or a value for a single cell in a detailed by race age distribution won't re-add to the SF1 distribution for the same results.) EASI strives to correct all of these problems with the Census data and remove these as issues that could affect EASI updates.

EASI updating validity checks involve controlling all Census 2000 distributions that require a controlling definition. EASI then makes the same checks on the EASI updates in order to prevent inconsistencies from coming into the updates.

EASI updating validity checks involve controlling all Census 2000 distributions that require a controlling definition. EASI then makes the same checks on the EASI updates in order to prevent inconsistencies from coming into the updates.

EASI updating validity checks involve controlling all Census 2000 distributions that require a controlling definition. EASI then makes the same checks on the EASI updates in order to prevent inconsistencies from coming into the updates.

The next issue is the controlling of distribution to the correct sum. A basic example of that is population by age and sex must add to population. This same issue is where the sum of the male age 0 to 5 for White, Black, Asian, and Other must add to total 0 to 5. Another example is that education attainment is defined as for the population 25. Note: Each distribution has a requirement like this. Many must add to population 16+ or population 3+, or households, or population, etc. Other key ones are that Hispanic must be less than or equal to Total Population less White Non-Hispanic Population. Also key

is that White Non-Hispanic Population must be less than or equal to White Population. These conditions for updates apply across all estimates including individual age groups (0 to 5, 6 to 11, etc.) and individual income groups (\$0 to \$15k, 15k to 25k, etc.).

The last part of the validity check is to find and fix rounding errors. Rounding errors are introduced in all estimates since results for the sum of a distribution will frequently not exactly add to the require estimate. To accommodate the rounding error EASI has developed various ways of adjusting the error into the most likely cell (in EASI rounding errors are calculated simultaneously as the distribution is being estimated, so when a group or cell sum is off by 1 (high or low) EASI immediately makes the adjustment in that actual group or cell.

These checks are performed at the BG, City, and ZIP Code levels. This is required since EASI splits BGs to create cities and ZIP Codes. Since splitting of BGs can introduce these validity issues the EASI methodology require the BG checks described above to be repeated at both cities and ZIP Codes as well.

Life Stage Clusters - The Basics

1. Begin with a collection of neighborhood (Census Block Groups) demographic data series to learn about what comprises a "neighborhood".
2. Through thousands of multivariate analyses, EASI synthesized and identified the independent variables, and their relationship to each other, that form the foundation of the clusters. This statistical foundation of neighborhoods form the basis of "Life Stages".
3. Based on the unique variables characterized by the Life Stages concept of independent clusters, EASI was able to replicate and verify the accuracy and utility of their neighborhood prediction model.
4. Create EASI Life Stages, an understandable, explainable, and statistically relevant group of clusters which comprise a highly predictive neighborhood model of location.