New York Metropolitan Transportation Council

NYMTC 2050
Socioeconomic and Demographic (SED) Forecast

White Paper on Methodology
County Level Forecasts

Technical Memorandum | Task 2.1

Final | June 23, 2014

Disclaimer

The preparation of this report has been financed through the U.S. Department of Transportation's Federal Transit Administration and Federal Highway Administration.

This document is disseminated under the sponsorship of the New York Metropolitan Transportation Council in the interest of information exchange. The contents of this report reflect the views of the authors who are responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Federal Transit Administration, Federal Highway Administration or the State of New York. This report does not constitute a standard, specification or regulation.

This report, White Paper on Methodology County Level Forecasts, was funded through the New York Metropolitan Transportation Council project 2050 Socio-Economic/Demographic (SED) Forecasts PTCS13D00.D01 in the 2013-2014 Unified Planning Work Program. This project was funded through matching grants from the Federal Transit Administration and from the Federal Highway Administration.

Submitted by:

THE Louis Berger Group, INC.

NYC PLANNING
DEPARTMENT OF CITY PLANNING CITY OF NEW YORK

Table of Contents

List of Tables

List of Figures

EXECUTIVE SUMMARY

1.0 PROJECT SUMMARY

The New York Metropolitan Transportation Council (NYMTC) is a regional council of governments (COG) that is the metropolitan planning organization (MPO) for New York City, Long Island and the lower Hudson Valley. NYMTC provides a collaborative planning forum to address transportation-related issues, develop regional plans and make decisions on the use of federal transportation funds. Socioeconomic and demographic (SED) forecasts are crucial components in the development of regional transportation plans, air quality conformity analysis and decision-making on the use of these funds. In order to produce the 2050 SED forecasts, NYMTC retained the services of The Louis Berger Group (LBG), in collaboration with the New York City Department of City Planning (DCP), to create population, employment, labor force, and household forecasts for the region at the county level, and to develop a tool that allocates the county-level SED projections down to the Transportation Analysis Zone (TAZ) level.

The following White Paper describes elements of the proposed methodology developed by the Project Team (LBG and DCP) to be used in developing the forecasts for the 31 counties in the New York metropolitan area, as shown in Figure 1, which are included in NYMTC's modeling efforts through the New York Best Practices Model (NYBPM).

FIGURE 1   31-COUNTY NEW YORK METROPOLITAN REGION

For the purposes of this project, the LBG Team is responsible for producing a region-wide employment forecast for all 31 counties. For variables other than employment, DCP will be responsible for producing forecasts for the five boroughs comprising New York City, while the LBG Team will prepare forecasts for the 26 counties outside of New York City. The responsibilities of each of the two different Teams follow:

Table 1. LBG Team and DCP Roles

Geography	LBG Team	DCP
5 NYC Counties: New York, Kings, Queens, Bronx, Richmond	• Employment Forecasts (reconciled between DCP and LBG through the FWG process)	• Population Forecasts • Household Forecasts • Labor Force Forecasts • Employment Forecasts (reconciled between DCP and LBG through the FWG process)
Other New York Counties: Nassau, Suffolk, Westchester, Rockland, Putnam, Orange, Dutchess, Sullivan, Ulster	• Population Forecasts • Household Forecasts • Labor Force Forecasts • Employment Forecasts	N/A
NJ & CT Counties: Bergen, Passaic, Hudson, Essex, Union, Morris, Somerset, Middlesex, Monmouth, Ocean, Hunterdon, Warren, Sussex, Mercer, Fairfield, New Haven, and Litchfield	• Population Forecasts • Household Forecasts • Labor Force Forecasts • Employment Forecasts	N/A

Data sharing and coordination between DCP and LBG throughout the demographic and employment forecast development will provide the overall Project Team with an additional opportunity to re-evaluate the broad methodological assumptions for the region and to identify potential changes in assumptions to ensure greater consistency across the whole region.

The development of NYMTC's SED forecast requires consensus among a diverse group, including relevant stakeholders, the public, the Technical Advisory Committee (TAC) and the Forecast Working Group (FWG). The FWG consists of approximately 90 representatives from all 31 counties involved in the forecast, transportation agencies at the local, regional, state and federal level, as well as labor and economic development organizations. The representatives bring input from the stakeholders and serve as a direct line to more localized data within their region.

This White Paper addresses how numerous factors influence the proposed forecasting methodology, including the current economic environment's impact on future conditions and structural changes in the economy that may be underway; how land use and development potentially constrain the forecasts; how LBG will work with DCP to incorporate their forecasts while maintaining a regional approach to forecast development; and incorporating input from regional stakeholders. The remainder of this document discusses the issues and methodological approaches with distinctions made between the two regions:

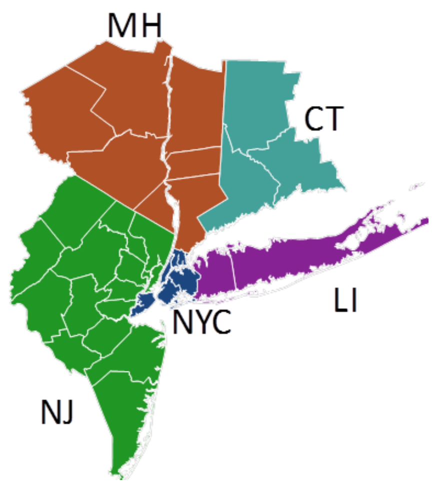
1. *NYC Forecast – comprised of the 5 boroughs of New York City*
2. *Outside NYC Forecast – comprised of 26 counties outside New York City*

1.1 Summary of Key Points

NYMTC's current forecasting methodology is a combination of top-down and bottom-up approaches. Population forecasts begin with a cohort-component (CC) model that operates individually for each NYMTC sub-region (see Table 2). These sub-region totals are then allocated to counties based on historical trends and TAZs following the zonal allocation program (ZAP) process. In contrast, employment forecasts begin with national macroeconomic projections from IHS Global Insight (a global information company). National employment and income are estimated for NYMTC sub-regions with an econometric model that "shares" national totals based on sub-region characteristics and their relation to national-level benchmarks. The Project Team undertook a detailed examination of the prior methodology to develop the current approach that will be employed in producing the 2050 SED forecasts.

Table 2. NYMTC Sub-Regions

New York City	Long Island	Mid-Hudson	New Jersey	Connecticut
Bronx	Nassau	Dutchess	Bergen	Fairfield
Kings	Suffolk	Orange	Essex	Litchfield
New York		Putnam	Hudson	New Haven
Queens		Rockland	Hunterdon	
Richmond		Sullivan	Mercer	
		Ulster	Middlesex	
		Westchester	Monmouth	
			Morris	
			Ocean	
			Passaic	
			Somerset	
			Sussex	
			Union	
			Warren	



NYC Forecast Highlights:

- *Household and Population Projections*
    - The population projections will be created using a cohort-component model. The CC model is based on the premise that the population can be broken down into three main components: births are added to the population; deaths subtracted; and migration can either be added or subtracted depending on whether there is overall positive or negative migration. Due to the variability of migration rates, these rates will be adjusted to ensure that the city's land use and zoning are taken into account when projecting population. This is done at three time points: 2020, 2040 and 2050.
    - Households will be projected forward by applying occupancy rates and average household size from the decennial census to the forecasted population.
    - A planning component, which examines land use and zoning, will be included to check the reasonability of the demographic model, ensuring that the city's housing can accommodate the demographic projections.
    - Due to numerous limitations, including problems of census coverage, separate CC projections for mutually exclusive race/Hispanic groups will not be conducted. Rather, a race-specific "overlay" will be provided on the output of the projections by using a proportional allocation method. Historical data, local knowledge, immigration patterns and national projections will be used to construct assumptions about the relative distribution of these groups going forward and then applied to the total population for each projection point.
- *School Enrollment (K-12)*
    - Data will be collected from the New York City Department of Education for current public and private school enrollment by school location for K-12 students.

- *University Enrollment*
  - University enrollment by school location will be determined from administrative data taken from two sources: the Department of City Planning Facilities Database, which is a main source of data on institutions of higher education; and the New York State Education Department data system known as ORIS that collects and distributes information on the status of higher education in New York State.
- *Employment* – Total employment will be determined through both a population-based method and an independent aggregate employment method.
  - The *population-based method* will use the historical record of resident labor force participation ratios; resident employment and unemployment; and data on net commuting to project total employed. Adjustments to resident labor force participation ratios and commuting patterns will be made based on existing research and past patterns; however, total growth in employment will be subject to population growth, as determined by the cohort-component projections model.
  - An *independent aggregate employment method* will also be used, which will take into account survey data; administrative data; and projected national trends. Total employment and labor force are determined as a function of selected economic variables, where population is not used as the primary input. In this model, historical rates of employment, along with existing indicators of future trends from the Bureau of Labor Statistics (BLS) and IHS Global Insight, among others, will be used to project growth in labor force and employment. Selected industry trend information will be incorporated as required and appropriate.
  - The population and independent aggregate employment models will generate different outputs that need to be reconciled. Historical data, for example, will be used to create an adjustment factor to account for the discrepancy between employed persons and jobs. This discrepancy exists as the economic model draws employed persons from the American Community Survey (with those holding multiple jobs represented just once), while the population-based method included some number of multiple job holders.
- *Household Income*
  - A model of median household income will be developed and distributed by borough and Public Use Microdata Areas (PUMA), and the population will be multiplied by this distribution to obtain household income for projected years.

Outside NYC Forecast Highlights:

- *Summary of Proposed Methodology* – The population, household formation, employment and labor force models draw from NYMTC's current SED models, while incorporating effective practices discussed later in this White Paper. The four models can be briefly summarized as follows:
  - *Population* – Consistent with the NYC region methodology, a disaggregate cohort-component model will be used to forecast population, employing detailed demographic data from the U.S. Census Bureau and state vital statistics data. The model produces estimates of population change attributable to natural increase (births minus deaths) and net migration (net population flows into or out of the county).

- *Household Formation* - The Household Module relies on base-year headship rates to forecast household formation. In addition, the Household Module estimates average household income for each age cohort categorized by sex.
- *Employment* – The employment model will apply a top-down allocation of national economic growth, provided by IHS Global Insight to the county level through a four-step process:
    1. Allocate national employment growth to 31-county region
    2. Reconcile regional employment with population via labor supply and demand constraints
    3. Allocate regional employment to counties
    4. Reconcile county employment with population via land use and commuting constraints
- *Labor Force* – In the 31-county region and other areas with a strong and varied employment base, a primary determinant of net migration is labor force demand. Net migration of the labor force for each county reflects the residual required to balance labor demand (available jobs) and labor supply (labor force) after accounting for unemployment. Labor force net migrants are then converted into total persons by analyzing dependency ratios and trends in very young and elderly migration (i.e., persons 0-15 years and adults over 65 years of age). The demographic composition (i.e., the age, sex, race/ethnicity characteristics) of net migration is allocated in proportion to each cohort's share of observed county migration during the 2000 to 2010 period.

- *Additional considerations* – Key topics were identified in the evaluation of the proposed approach that serve to augment the overall methodology:
    - *Economic Considerations* – The forecast methodology is influenced by a variety of economic factors including sources of structural change, risk analysis and specific input variables. Forecasting tools such as multiple feedback loops and shift/share analysis will be used to understand the duration of trends as well as determine the sensitivity of structural changes to future employment. Simulations provide a risk analysis tool designed to minimize uncertainty by defining the range of potential outcomes, and by providing estimates of the probability of attaining each outcome. The likelihood of the baseline, optimistic and pessimistic scenarios will be estimated using @Risk software to conduct an independent evaluation of probabilistic future outcomes.
    - *Forecast Stability* – Relying on a purely top-down or bottom-up approach leaves ample room for error; top-down approaches tend to have diminished accuracy at lower levels of geography, while bottom-up approaches can yield implausible regional totals, and frequently do not incorporate seasonal and cyclical effects or predictable changes in economic structure. To ensure the proper balance between both approaches, the methodology documents accompanying IHS forecasts will be reviewed to determine the steps and data involved with down-allocation to counties as well as calibrating and reconciling the population and employment models at multiple spatial scales: the entire 31-county region, sub-regions, and individual counties.
    - *Land Use Considerations* – Potential constraints on the SED projections exist, as conditions on the ground may not be conducive to population or employment numbers anticipated in a region. The Project Team will conduct an outreach effort to the 26

counties outside New York City seeking development and land use data that can inform the accuracy of the SED forecasts.

- *Incorporating Expert/Stakeholder Judgment and Input* – A review of MPO practices indicated that the Southern California Association of Governments (SCAG) conducted a series of panel workshops intended to reduce forecasting errors by leveraging the collective opinions of several experts on critical factors and economic/demographic assumptions. Similarly, the Delphi Method is a structured technique that aggregates expert opinions of likely future outcomes. In addition to regular Forecast Working Group (FWG) meetings, these methods will be incorporated as consensus is sought across numerous topics.

Key Differences Between NYC and Outside NYC Forecast Methodology:

The following table highlights the distinctions between the methodologies used for the NYC region and outside NYC region:

Model	NYC	Outside NYC	Key Differences
Population	Cohort-Component with land use component (*Part II, Section 4*)	Cohort-Component (*Part III, Section 2.2*)	Both the NYC and Outside NYC population forecasts will use a CC methodology. The NYC forecast includes an additional land use analysis which projects the broad number of housing units and will be used to adjust the population projection generated form the CC model.
Employment	Reconciliation between population driven method and economic based projections (*Part II, Section 7*)	Top-down allocation of national employment projections using shift-share methods (*Part III, Section 2.4*)	The NYC method is a bottom-up approach, with total employment being derived at the county level. The Outside NYC forecast is a top-down approach, with national economic growth being down-allocated to the counties via a shift-share.
Household	Average household size applied to forecasted population (*Part II, Section 4*)	Household formation based on age-specific headship rates (*Part III, Section 2.3*)	The NYC method will hold average household size constant, and apply them to the forecasted population to derive households. Outside NYC, age-specific headship rates (the ratio of household heads to the total household population) will be applied to the population to forecast household formation.
Labor Force	Labor force participation ratios applied to the projected populations to produce the resident labor force (*Part II, Section 7*)	Labor force participation ratios applied to the projected populations to produce the resident labor force (*Part III, Section 2.5*)	Both the NYC and Outside NYC forecast will apply labor force participation ratios to projected populations.

2.0 NEXT STEPS

This White Paper will be followed by a trends analysis paper and a more detailed methodology technical memorandum. The trends analysis paper will review historical data of key input variables intended for use in the SED county level forecasts, with the goal of informing the methodological assumptions (e.g. future rates of fertility, immigration, etc.) applied in each of the modules supporting this forecast effort. The methodology technical memorandum will describe in greater detail the modeling approach and the theoretical basis underlying the modeling method for each of the variables. Issues in development of the ZAP will be examined in an upcoming white paper on allocation.

I. Literature Review of Forecasting Methods

The purpose of this section is to provide a context for NYMTC's current forecasting effort. The section outlines forecasting methods employed by large MPOs and Councils of Government (COGs), with a focus on the elements of the proposed methodology most likely to differ from NYMTC's previous forecast method. After a brief review of the various approaches employed by MPOs across the country, a more detailed discussion of recent forecasts done by the Metropolitan Washington Council of Governments (MWCOG) and Southern California Association of Governments (SCAG) will follow.

1.0 MPO Practice

Across MPOs, employment forecasts are most frequently generated using a top-down approach, where forecasts are estimated *first* at the region (or sometimes county) level before being allocated to smaller areas (see Table 3). This approach typically begins with the development of a regional macroeconomic forecast, which can come from three general sources: (1) a macroeconomic forecasting model such as REMI or the IHS Global Insight model, (2) another quantitative method such as shift/share or in-house statistical models, or (3) a consultant or expert panel (with or without quantitative support).

Table 3. Regional socioeconomic forecast methods for select MPOs

MPO	Forecasting Method	Macroeconomic model or quantitative technique used (if any)
ABAG (San Francisco/Oakland)	Top-down	Consultant using shift/share
ARC (Atlanta)	Top-down	REMI (with review/modification)
BaltoMetro (Baltimore)	Bottom-up	
CMAP (Chicago)	Bottom-up	
DVRPC (Philadelphia)	Bottom-up	
H-GAC (Houston/Galveston)	Bottom-up	Woods & Poole forecasts used for industry splits
MAG (Phoenix)	Top-down	From Arizona State Demographer
MAPC (Boston)	Bottom-up	
MARC (Kansas City)	Top-down	REMI
MetroCouncil (Minneapolis/St. Paul)	Top-down	REMI (with review/modification)
MRCOG (Albuquerque)	Top-down	REMI (in concert with UNM model)
MWCOG (Washington, DC)	Top-down/bottom-up	IHS Global Insight
NCTCOG (Dallas/Ft. Worth)	Top-down	Consultant
PSRC (Seattle)	Top-down	Fair macroeconomic model
SANDAG (San Diego)	Top-down	Econometric system
SCAG (Los Angeles)	Top-down	Shift/share
SEMCOG (Detroit)	Top-down (from counties)	REMI (with review/modification)

It should be noted that in the cases of the first two sources, the regional forecasts are rarely used "out of the box." Typically, a review committee works iteratively with the modeling agency to modify assumptions or results until the forecasts satisfy the review process. This review/modification process can govern the forecasts' overall levels as well as ratios to households and population (particularly when these other variables are forecast using alternative methods). Therefore, even for organizations with a primarily top-down approach, a review process can ensure some preliminary conformity with bottom-up considerations – prior to formal quantitative down-allocation modeling.

Alternatively, a few organizations – such as CMAP and MAPC – apply a bottom-up approach, where employment forecasts are based on land use/population projections or independent local forecasts without regard to broader national or regional economic trends. In the case of CMAP, employment forecasts are model-driven, based on land use projections. In contrast, MAPC does not use formal modeling, but instead aggregates employment projections from local planners within each town.

Finally, some organizations employ a combination of bottom-up and top-down methods with an explicit process to reconcile differences

1.1 Relevant Case Studies

Case studies offer important examples of how other MPOs and regional planning agencies have combined analytical techniques and collaborative processes to produce forecasts. This section summarizes a review of two recent relevant examples.

1.1.1 Metropolitan Washington Council of Governments (MWCOG)

MWCOG is one of the MPOs surveyed that combines a top-down and bottom-up approach with an explicit reconciliation process. The top-down approach uses IHS Global Insight forecasts "out of the box" for regional population, household, and employment by North American Industry Classification System (NAICS) sector.

Independent of the regional forecasts, small-area TAZ-level employment, household, and population forecasts are prepared by local planning staff in each COG member jurisdiction for the 30-year forecasting period. Short-term local forecasts are based on current construction, building permits, approved development plans, and rezoning applications. Longer-term local forecasts are based on adopted and approved area master plans, jurisdictional comprehensive or general plans, current zoning capacity, and past and current trends in market absorption rates.

Once local forecasts are complete, the process moves into "reconciliation." Here, a MWCOG subcommittee reviews the regional and local forecasts, and is given the opportunity to solicit further information from local planners. Local forecasts are modified until they sum to within about 3 percent of the regional forecasts for each of households, population, and total employment.

1.1.2 Southern California Association of Governments (SCAG)

SCAG's forecasting approach is relevant to NYMTC's present effort in several regards. First, it uses a shift/share process to generate employment forecasts at both the regional and county level.[1] Second, bottom-up population and top-down employment forecasts undergo a calibration process to ensure consistency of the final results at both the regional and county level.[2] Finally, the process incorporates scenario-based risk assessment to develop high, medium and low scenario projections.

The process begins with a bottom-up demographic cohort-survival-migration model to project population for the entire SCAG region (with sub-models estimating household formation and labor force participation). The same framework is then used to project the baseline population, households, and labor force at the county level. The sum of the county trend projections is then compared to the regional projections. If the results are significantly divergent, input data at the county level is adjusted (primarily via the migration model) to bring the sum of county projections and the independent, regional projections more closely in line. Complete agreement between the two projections is not mandatory. After analysis, the sum of counties constitutes the regional baseline projections.

Initial SCAG employment forecasts are developed completely independently from population forecasts. The process is top-down, where a shift/share model is used to project the region's share of forecasted national growth by 2-digit NAICS sector. For the national projection, SCAG uses the U.S. Bureau of Labor Statistics (BLS) short-term national employment projection (a 10-year forecast) in combination with in-house modeling techniques for extending the forecast to the horizon year.

Once the national forecast is established, a set of shift/share models are calibrated to predict the SCAG share of national employment for each 2-digit NAICS sector. This step involves analyzing historical data trends to determine the most robust share method for each sector. This analysis yielded six share methods across 20 NAICS industries (summarized in Table 4).

Table 4. Summary of SCAG employment shift/share methods

Share Method	Applies to NAICS sectors	Description
Constant Share of Levels	Mining	Assumes regional employment levels will remain constant historically-observed share of national levels
Trend Share of Growth	Agriculture; Utilities; Wholesale; Finance & Insurance; Arts & Entertainment; Accommodation & Food Services	Extrapolates historical trend of region's share of national growth
Constant Share of Growth	Manufacturing; Transportation; Information; Professional, Scientific, Technical Services	Assumes regional employment growth will maintain a constant historically-observed share of national growth
Econometric (employment)	Real Estate; Management of Companies; Business Services	Industry employment is a function of levels and growth of employment in other sectors
Econometric (population)	Construction; Education Services; Health & Social Services; Other Services; Government	Industry employment is a function of population levels and growth
Econometric (other)	Retail Trade	Industry employment is a function of levels and growth of multiple variables

Source: SCAG (http://rtpscs.scag.ca.gov/Documents/2008/fGrowthForecast.pdf)

The same shift-share methods developed to down-allocate national growth to the SCAG region are used to down-allocate SCAG employment to counties.

Finally, population and employment projections are brought before an "Expert Review Panel" to review the assumptions in the national forecast and the internal consistency of the shift/share and cohort model results. Consistency is determined by comparing regional labor supply and demand. The first step in this process is projecting regional trends in labor force participation rate and multiple job-holding rate. Next,

> "the implied unemployment rate is derived by matching labor supply estimated from population projections with workers estimated from job projections. (In the case of the 2008 forecast) the panel of experts suggested that the acceptable implied unemployment rate ranges from 5 percent to 8 percent."[3]

If the implied unemployment rate is outside the acceptable range, "it is corrected by adjusting the domestic migration assumptions of the demographic projection model."[4] A final step in the review process is for the expert panel to ensure that the forecasts imply a "reasonable share of national jobs."

1.1.3 Key Takeaways

Both MWCOG and SCAG offer insights into how NYMTC can generate logical and accurate SED forecasts. Like MWCOG and SCAG, NYMTC incorporates outside experts and opinions in their forecasting process. In addition to regular meetings with the FWG, three additional experts, chosen for their knowledge of regional economic and demographic trends, help ensure both the legitimacy of the forecasting methodology and inclusion of localized knowledge as additional input (the names of the experts have not yet been finalized, as their input will be critical in defining the models' final assumptions). The methodology employed by MWCOG and SCAG illustrates the importance of a reconciliation process between various methods and stakeholders. Broad methodological assumptions for the development of NYMTC's population, household, employment and labor force models will be reviewed and evaluated.

II. NYC FORECAST

1.0 OBJECTIVE

To produce projections for 16 Key socioeconomic and demographic indicators at TAZ (Transportation Analysis Zone) level, listed below, for each of the five boroughs of New York City.

Goals:

- Create projections that are informed by past trends that identify patterns and relationships in the data;
- Evaluate these patterns in the context of current information on population, employment, school enrollment, land use, and zoning to form reasonable assumptions about the future;
- Consult with members of the FWG to evaluate future scenarios from demographic, economic and planning perspectives;
- Apply assumptions using commonly accepted scientific techniques that allow for a cogent and transparent evaluation of results by the FWG;
- Apply assumptions regarding rates and ratios that can be regarded as "low," "medium" and "high" to create a range of projections, where required and appropriate.

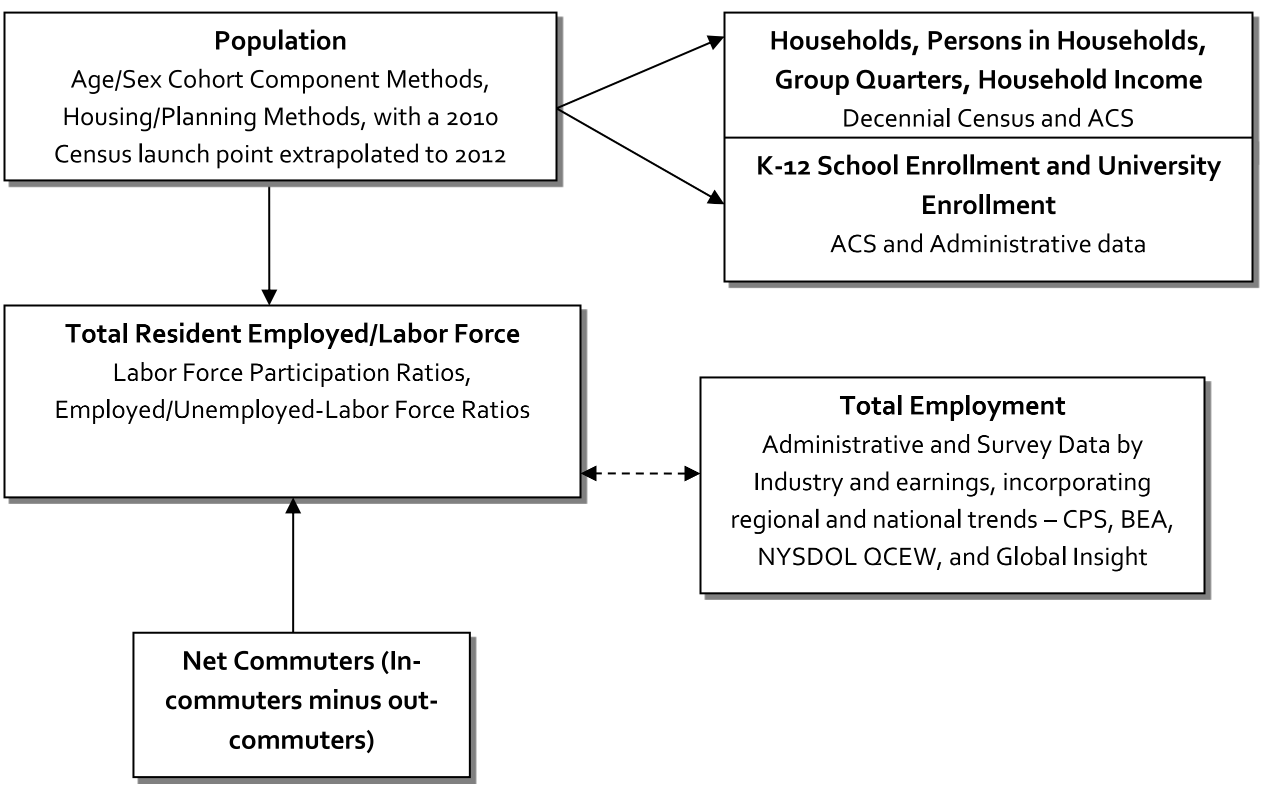
2.0 PROJECTED POPULATIONS AND PRINCIPAL METHODS (16 VARIABLES)

1. Total and Household Population (2 variables)
    - A cohort-component model, tied to a housing unit/land use planning method for corroboration
2. Households and Household Size (2 variables)
    - Application of Occupancy Rates and Average Household Size from decennial census will be held constant. Using these rates, households will be projected forward by deriving them from the forecasted population
3. Group Quarters Population (4 variables)
    - Group Quarters population consists of persons who do not live in housing units, but in facilities, such as prisons, nursing homes, dormitories, group homes and other similar arrangements. In 2010, about 186,000 New Yorkers were classified as living in group quarters. The model uses group quarters population from the most recent census enumeration – persons in service-based locations from the 2010 Census. Group quarters population will be held constant at 2010 levels, given the absence of any firm basis for group quarters forecasts, both from the standpoint of total numbers and the locations of group quarters facilities.
4. K-12 School Enrollment (1 variable)
    - Department of Education administrative data and school enrollment projections by grade level, linked to population projections by age/sex
5. University Enrollment (1 variable)
    - ACS, New York State Department of Education administrative data and the Department of City Planning's Facilities Database

6. Employment/labor force forecasts by selected industry for residents and total employed (5 variables)
    - Two parallel methods:
        i. A *population-based method* using historical record of resident labor force participation ratios, resident employment and unemployment, and data on net commuting to project total employed.
        ii. An *independent aggregate employment method* using survey data, administrative data, and projected national and local trends incorporated with the outside NYC forecasts. Worker earnings are also included here.
7. Household Income (1 variable)
    - Current 2012 household income distribution applied to projected households. For each projected time point, the aggregate household income associated with each quintile will be generated.

3.0 POPULATION AND EMPLOYMENT PROJECTIONS OVERVIEW

FIGURE 2 POPULATION AND EMPLOYMENT PROJECTIONS OVERVIEW



4.0 TOTAL AND HOUSEHOLD POPULATION (ITEMS 1, 2, 3)

The objective of this analysis is to produce population projections by age and sex at five-year intervals for 2010 through 2050, for each of New York City's boroughs.[5] The projections will be created using a cohort-component (CC) model. The CC model is based on the premise that the population can be broken down into three main components: births are added to the population, deaths subtracted, and migration can either be added or subtracted depending on whether there is overall positive or negative migration. Therefore:

Projected Population (P1) = Base Population (P0) + Births – Deaths + Net Migration

The power of the CC model is based on its use of age and sex cohorts as a starting point. Births, deaths, and migrants are all used to move these cohorts forward through time, creating a new age/sex distribution at each five year time point. A particular cohort's ability to grow or decline is dependent on how the components of change impact each age and sex group. It is essentially an "accounting approach" to modeling population changes, one that applies birth, death, and migration rates to these age and sex cohorts.[6]

The CC Model is a widely used projection method because the components of change for each cohort interact, providing a more realistic outcome. For example, if there were large numbers of 25-29 year old female in-migrants, the 25-29 year old female cohort would grow from net positive migration; but there would also be increased growth in the 0-4 cohort since these women are in their prime child-bearing ages. These types of relationships within the model make it both realistic and complex, because each of the components of change *interacts* with the others to affect the age structure of the population.

Any method for projecting population is subject to error because the future is uncertain. So many variables can act to alter the population of New York City over the next few decades, including federal immigration policy; economic cycles; changes in transportation infrastructure and networks; shifts in the size of households; changes in patterns of domestic migration; shifts in patterns of fertility and mortality; changes in housing demolition and construction; local government policies and priorities involving land use and sustainability; and conditions in other parts of the nation and world. There is no definitive method for looking forward. The best that anyone can do is adopt a series of assumptions and then look forward employing some tests of reasonability, based on what is known at the current time about the factors mentioned above.

Thus, the fundamental difficulty in applying the cohort-component model rests in identifying appropriate fertility, mortality, and migration rates for persons of different ages going forward. In general, the starting point for most projections consists of shaping the future to be in line with historical patterns. Such a baseline *projection* differs from a *forecast.* Forecasts alter the historical or baseline projections using scenarios that demographers believe may be likely for the future. For example, it is likely that over time increased longevity will become reality, based on the opinions of medical experts at the National Center for Health Statistics and the Social Security Administration. As a result, as part of a forecast, demographers may decide to increase rates of survival for persons in the older ages. Similarly, knowledge of new housing development may spur planners to lower net migration losses, keeping more people in a jurisdiction over time. Of course, since no one can predict the future with certainty, the ultimate success of a projection depends upon the degree to which the assumptions in the model hold. Therefore, three sets of rates based on different assumptions will be employed throughout the analysis, yielding low, medium and high projections.

With the cohort-component method, the starting point will be an evaluation of change by age/sex for 2000 to 2010. Each age group is "survived" from 2000 to 2010 at regular intervals, and the residual ("survived" population versus the population reported in the census) is derived for each age group. Normally the residual for each age group is an estimate of the effects of net migration; however, any errors associated with the population counts also will be subsumed into the net migration residual. The method assumes that coverage of the population (i.e., undercounts/overcounts in census-taking) remains constant over time. For example, with higher coverage rates in 2000 compared with 1990, differences in net census "error" between the two time points will likely result in distorted estimates of net migration by age for that decade. Therefore, the net migration rates have been adjusted for coverage change using data from the 1990 Census Post-Enumeration Survey, the basis for what would have been the 1990 Census adjustment.[7] More recently, a review of population changes and net migration rates for 2000-2010 in Queens and Brooklyn revealed distortions in selected areas, associated with the erroneous enumeration of housing units as vacant and what were likely coverage problems for selected age groups in 2010. As a consequence, adjustments to the 2010 Census will be required to compensate for these coverage issues and to make net migration rates more consistent with the general patterns displayed by age and geographic area.[8]

Finally, the number of persons in group quarters (GQ) will be held constant at 2010 levels, since an evaluation of changes to that population did not yield many large changes since 2000. Moreover, the questionable utility of the American Community Survey (ACS) data on group quarters for sub-state areas,[9] the precarious nature of predicting group quarters living arrangements going forward, and the difficulties of determining the locations of future GQ facilities make for few alternatives. Further along in Task 6, the group quarters population will be divided and allocated into TAZ geographies by institutional and non-institutional categories, based on the distribution from the 2010 Census. Despite the fact that the initial considerations called for estimates of the homeless GQ population, estimates from the decennial census, the ACS and the Department of Housing and Urban Development (HUD) are all very tenuous, with substantial issues that limit coverage of populations regarded as homeless. Thus, given their marginal ties to the labor force for most, a decision was made in concert with NYMTC's Forecasting Working Group not to expend resources on creating separate estimates of the homeless population. It is important to remember, however, that at least some part of the homeless population (e.g. persons in shelters) will be included in the total group quarters population estimates.

4.1 Population Projections and Housing Constraints: Introducing the Planning Component

The cohort-component method is a purely demographic method, meaning that it responds solely to changes in assumptions about fertility, mortality, and migration. By themselves such methods are insufficient for the creation of useful projections, which need to be examined in the context of the city's planning environment. Of the three demographic components used in the cohort component model, births and deaths can be modeled by demographers with a high degree of confidence, based on the age structure of the population; but migration is far more variable and unpredictable. When projecting a population increase over several decades, for example, demographers need to select a rate of net migration from a range of possibilities. For these projections, migration rates will be closely analyzed to ensure that the resulting projection of population and housing could be reasonably accommodated, given the city's current land use and zoning. In addition to land use and zoning constraints, high population densities, fixed political boundaries, and the high cost of residential construction also act to limit population growth. Thus, migration rates in the demographic model will be adjusted based on this planning component, which ensures that the city's land use and zoning are taken into account when projecting population. This will be done at three time points: 2020, 2040 and 2050.

4.2 Creating 2015 and 2020 Populations from Projected Housing Permits and Certificates of Occupancy

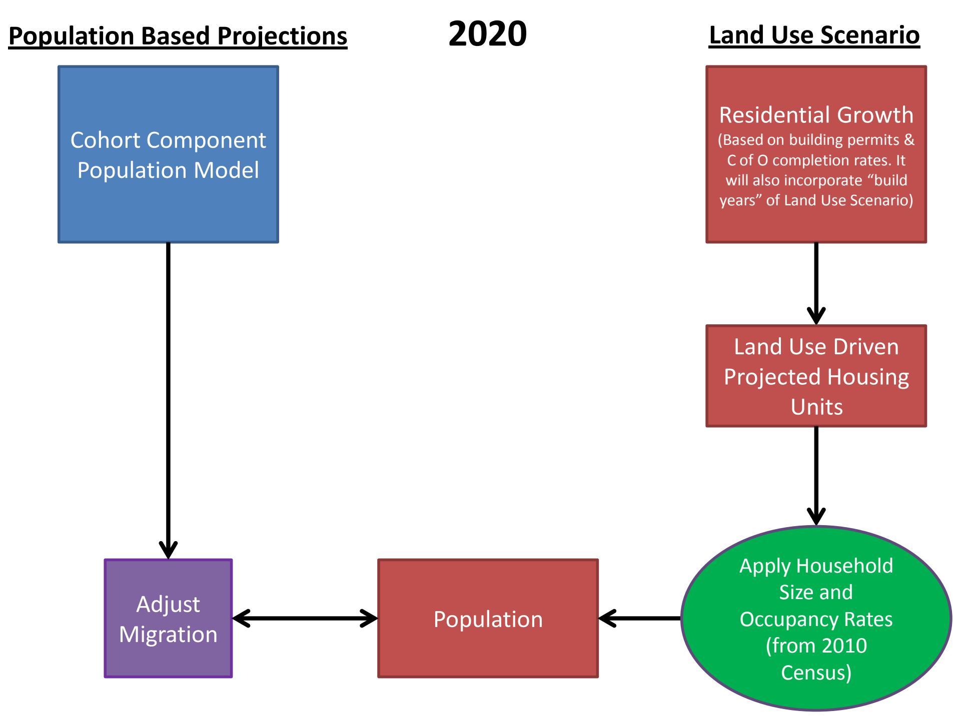
Total Population can be derived from housing units based on the following:

*Population = (Occupied Housing Units * Average Household Size) + Persons in Group Quarters*

Figure 3 below shows the methodology that will be used to create a projected 2020 population. The approach takes advantage of the fact that actual population and housing data exist for part of the 2010-2020 period. A 2020 housing target will be created using three inputs: 1) Certificates of Occupancy (C of Os) for 2010 to 2013;[10] 2) demolitions for the years 2005 to 2012; and 3) the average number of building permits issued between 2005 and 2012, which will be used as the estimated number of building permits for each of the years between 2014 to 2020. Permits issued between 2005 and 2012 will be examined and a net number will be created by subtracting the demolitions from the permits for each year. C of Os that were tied to the permits issued in each of the above years will be tracked, resulting in a determination of units that were completed and occupied.[11] An annual rate of completion/occupancy will be applied to the estimated number of building permits for 2014-2020, to obtain the estimated number of C of Os for each of these years. The actual number of C of Os for 2010 to 2013 will then be summed, along with the projected number of completions for 2014 to 2020, to obtain the number of housing units for the 2010 to 2020 period.

The housing number will be then converted into population by using average household size and occupancy rates for each borough from the 2010 Census. The baseline crude migration rates (CMR) will be adjusted for 2010 to 2020 in order to bring the cohort-component population projection for each of the boroughs in line with the population based on the 2020 housing target.

FIGURE 3 2020 POPULATION PROJECTION METHODOLOGY



4.3 Using a Land Use Analysis to Inform the 2020-2050 Population Projections

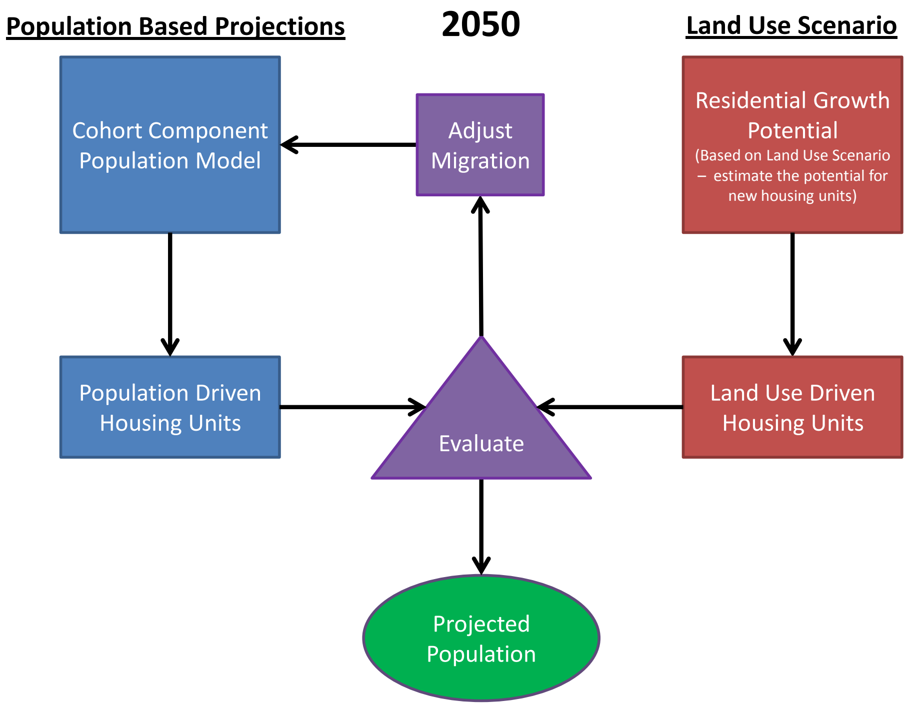
Confidence in the method for 2020, just described, is based on the fact that data on average household size are derived from a relatively recent census enumeration and that actual data on housing are available for at least part of the decade. Such confidence wanes in succeeding decades with the passage of time. Since recent data on building permits and certificates of occupancy (C of Os) are not available, the housing evaluation for the out years—2040 and 2050—will assess whether each borough can accommodate the population generated by the cohort-component model. Therefore, for these time points, the population will not be created as a direct function of housing, since any exact number of housing units is impossible to determine going forward. Instead, the population projections from the cohort-component model will be evaluated to determine whether they are reasonable, given broad land use and zoning conditions related to the capacity for housing growth. This method is best characterized as population projections informed by a land use analysis: a projected population using the cohort model that can be adjusted based on the broad number of housing units produced from a land use analysis.

Figure 4 below shows the methodology that will be used to create the 2020 to 2050 population projections. Projections of the housing needed to accommodate the cohort-component demographic projections will be developed using the current average household size, vacancy rates, and group quarters population, all held constant for the 2020 to 2050 period and applied to the 2050 cohort-component population projection:

Occupied Housing Units = (Population - Persons in Group Quarters)/Average Household Size

This is referred to as the "Population Driven Housing Units" and is shown on the left side of Figure 4. The number of housing units from the population driven method will then be compared with the number of housing units generated from an analysis of land use, which will seek to provide an estimate of capacity—high, medium and low. The land use scenario will evaluate housing potential—the housing opportunities and constraints presented by each borough's land use and zoning (right side of Figure 4). This is referred to as the *planning component,* or a check of the reasonability of the demographic model, to ensure that the demographic projections can be accommodated by the city's housing. In general, it is this planning component that dampens the tendency for demographic projections to show linear increases as the lack of capacity for new residential development acts to constrain growth moving forward. Assumptions regarding high, medium and low growth of the housing stock will be the key to setting the stage for multiple population projections.

Figure 4 2050 Population Projection Methodology



As part of the planning component, there are five general categories of opportunities for future growth:

1. DCP initiatives/re-zonings (enacted between 1/1/2002 and 1/1/14): rezoned areas providing increased housing opportunity;
2. Large private applications: large private developments approved as of 1/1/14, that would generate new housing;
3. Large public site projects: New York City Housing Authority, Department of Housing Preservation and Development, and city-owned sites, approved as of 1/1/14;
4. Opportunity areas: potential locations for future rezonings that would permit increased housing opportunity;
5. Background growth: potential sites for residential development, based on zoning in place as of 1/1/2002 and existing vacant or underutilized status.

Similarly, the analysis will also look at constraints in the boroughs that could prevent growth from occurring:

1. Carrying capacity of the infrastructure: existing water supply and sewer capacity;
2. Land development needs of public facilities: including schools, parks, libraries, police, and fire departments;
3. Non-residential uses: prohibits residential development.

Each development or opportunity identified in the planning component is accompanied by the projected number of housing units that could be built when the development is completed.[12] The number of housing units can then be summed up for each borough and, using current occupancy rates and average household size, populations associated with the projected number of housing units in 2040 and 2050 will be derived. The CMRs will be adjusted for the period between 2020 and 2050 to ensure that the projected population could be accommodated, given the prospective housing potential in each borough.

Average household size will remain constant at 2010 Census levels because of the lack of stability that will be introduced when attempting to alter the average household size for future years. Increases in immigration may portend increases in household size, while an increase in the number of non-family households with heads under the age of 65 may portend a drop in household size going forward. It is a valid approach to use current person per household counts obtained from the most recent census, in light of the fact that data to update average household size on a regular basis does not have enough precision to be useful at the local level.[13] It is important to note that even small changes to average household size can have a major effect on a population as large as that of New York City.[14] Therefore, household size will be held constant throughout the forecast years.

Vacancy rates will also be held constant at 2010 levels for many of the same reasons that household size will be held constant, and taking into account the housing unit anomalies cited above, e.g. erroneous enumeration of housing units as vacant and what were likely coverage problems for selected age groups in 2010.

4.4 Population Projections by Race and Hispanic Origin

After the population projections are produced using the methods just discussed, projections at the borough level will be prepared for four major mutually exclusive race/Hispanic groups:

1. White Non-Hispanics
2. Black Non-Hispanics
3. Asian Non-Hispanics
4. Hispanics

While it is technically possible to produce separate cohort-component projections for each of these groups, the process is fraught with many important limitations. This includes the compilation of vital statistics and census data in categories that are race-specific with known problems of census coverage and unknown race/Hispanic vital events; the tenuous nature of projecting race-specific rates; the absence of any way of implementing residential land use scenarios with a race/Hispanic overlay (something that may be conceptually unfounded to begin with); and the problem of making assumptions about groups whose internal ethnic composition has changed dramatically and is likely to continue to change in the years to come. Since 1970, for example, the ethnic composition of the White, Black, Asian and Hispanic populations has undergone substantial changes in New York City. Creating rates and assumptions based on the past, therefore, cannot be justified.

Moreover, there is evidence that the marginal gains of using race-specific inputs for a cohort-component model are limited or non-existent in practical terms. Research by Grip (2008) for school districts in New Jersey showed that the use of race-specific rates, instead of an overall rate, to project school-age populations frequently does not yield more accurate results.[15] And, in fact, earlier often-cited work by Keyfitz (1977) showed that there is a tendency for projections created as a sum of separate race/ethnic-specific groups to overstate the size of populations.[16] Grip points out that there are a number of reasons for creating race-specific projections, for example in fair treatment litigation or for program compliance purposes. However, there remains only limited statistical justification for the contention that separate race/Hispanic rates will result in lower levels of error than the use of a single overall rate when constructing projections.

Therefore, instead of using race-specific inputs, a race-specific "overlay" on the output of the projections will be applied by using a proportional allocation method. The percent of the population associated with each major race/Hispanic group back to 1970 will also be examined, along with the components of change by race/Hispanic origin as a way of evaluating the impetus for changes over time. Using this local knowledge, along with current patterns of immigration and national projections by race/Hispanic origin, assumptions will be developed about the relative distribution of these groups going forward. These distributions will then be applied to the total population for each projection point, which will yield projections by race/Hispanic origin. Groups will be further sub-divided by using the 2010 age/sex distribution of race/Hispanic groups.

5.0 School Enrollment K-12 (Item 4)

The New York City Department of Education (DOE) periodically issues school enrollment information, along with projections of future enrollment levels. This is done for New York City as a whole, each of the five boroughs, and for community school districts. The most valuable data will be current public and private school enrollment by school location for K-12 students from the administrative data collected and compiled by the New York State Department of Education (NYSED ORIS). (Further, school enrollment by school location can be geo-coded and then allocated to PUMAs and TAZs in Task 6).

Previous SED forecasts included information from the *Public Use Microdata Sample (PUMS)* data of the 2000 Census and 2005 American Community Survey. *PUMS* cross-tabulations of population by single year of age and enrollment status were created at the sub-regional level for years 2000 and 2005. Enrollees were estimated as a percent of all races for both sexes by five-year age groups, using the following age-grade equivalency:

- Under 5 years: Pre-K
- Ages 5-9: Kindergarten & Lower Elementary (Grades 1-4)
- Ages 10-14: Middle School or Upper Elementary (Grades 5-8)
- Ages 15-19: High School

Although the population model age cohorts do not strictly match school-going ages by grade level, future forecasts will need to be related to the age structure of the population and the coarse retention rates derived as a percent of population enrolled by age cohort in 2010. The average or trended retention rates derived from 2010 data could then be applied to the population forecasts to project future enrollment from 2020 to 2050.

6.0 University Enrollment (Item 5)

University enrollment by school location will be determined from administrative data. The *Department of City Planning Facilities Database* is a main source of data on institutions of higher education. In addition, the New York State Education Department manages a comprehensive data system (ORIS) that collects and distributes information on the status of higher education in New York State.

The *Facilities Database* is accessible to City Planning staff members through the local area network (DCP-LAN) as ArcGIS mapping files and in Microsoft Access. Each university is geocoded for tax block, tax lot, 2010 census tract, city council, district, community district, school district, police precinct, health area, zip code, borough, and X and Y coordinates. The New York State ORIS institutional enrollment data are accessible through their website:

http://eservices.nysed.gov/orisre/r_creditenrollment.jsp.

Users can access enrollment reports from 1980-2010 on statewide and county levels. These geo-coded facilities' enrollment records will be tallied down to the Traffic Analysis Zone (TAZ) level as part of Task 6.

7.0 EMPLOYMENT (ITEM 6)

Total employment will be determined two ways:

1. A ***population-based method*** using the historical record of resident labor force participation ratios, resident employment and unemployment, and data on net commuting to project total employed. Adjustments to resident labor force participation ratios and commuting patterns will be made based on existing research and past patterns; however, total growth in employment will be subject to population growth, as determined by the cohort-component projections model.
2. An ***independent aggregate employment method*** using survey data, administrative data, and projected national trends. *Total* employment and labor force are determined as a function of selected economic variables, where population is not used as the primary input. In this model, historical rates of employment, along with existing indicators of future trends from the Bureau of Labor Statistics (BLS) and IHS Global Insight among others, will be used to project growth in labor force and employment. Selected industry trend information will be incorporated as required and appropriate.

7.1 Population-Based Method for Determining Total Employment

7.1.1 Resident Labor Force

Figure 5 below shows the two key stages of the population-based model. One involves the examination of historical patterns of labor force participation, and the ratios of employed and unemployed to persons in the labor force. These data are important in formulating assumptions that drive the second stage of the model, which consists of the projections. In this part of the model, key assumptions built upon a foundation of historical research and assumptions about the future are combined with the age/sex cohort-component projections from the population model, to produce forecasts of resident labor force, employed and unemployed persons. Using the population projections imposes a reality test on the size and scale of the work force projections.

Studying changes in these measures over time is very important in the formation of assumptions about the future. Figure 6 shows participation rates over time, starting with 1970, by age/sex. For men, labor force participation rates have been declining, with a fairly consistent pattern by age. The exception is for 2000, which may be attributed to high levels of imputation resulting in poor data and erroneously low levels of labor force participation. Among women, participation rates have increased with each decade, and the trough that was present in the childbearing ages in 1970 has all but disappeared from the age-specific rates. All told, however, the patterns by age for both, men and women, reveal a picture that is more similar than different from 1980 through the present. Participation ratios by age will be modeled and applied to the projected populations to produce the resident labor force. A similar process will be executed for persons who are employed using ratios to resident labor force of employed and unemployed persons by age/sex.

FIGURE 5 POPULATION MODEL: RESIDENT LABOR FORCE, EMPLOYED AND UNEMPLOYED

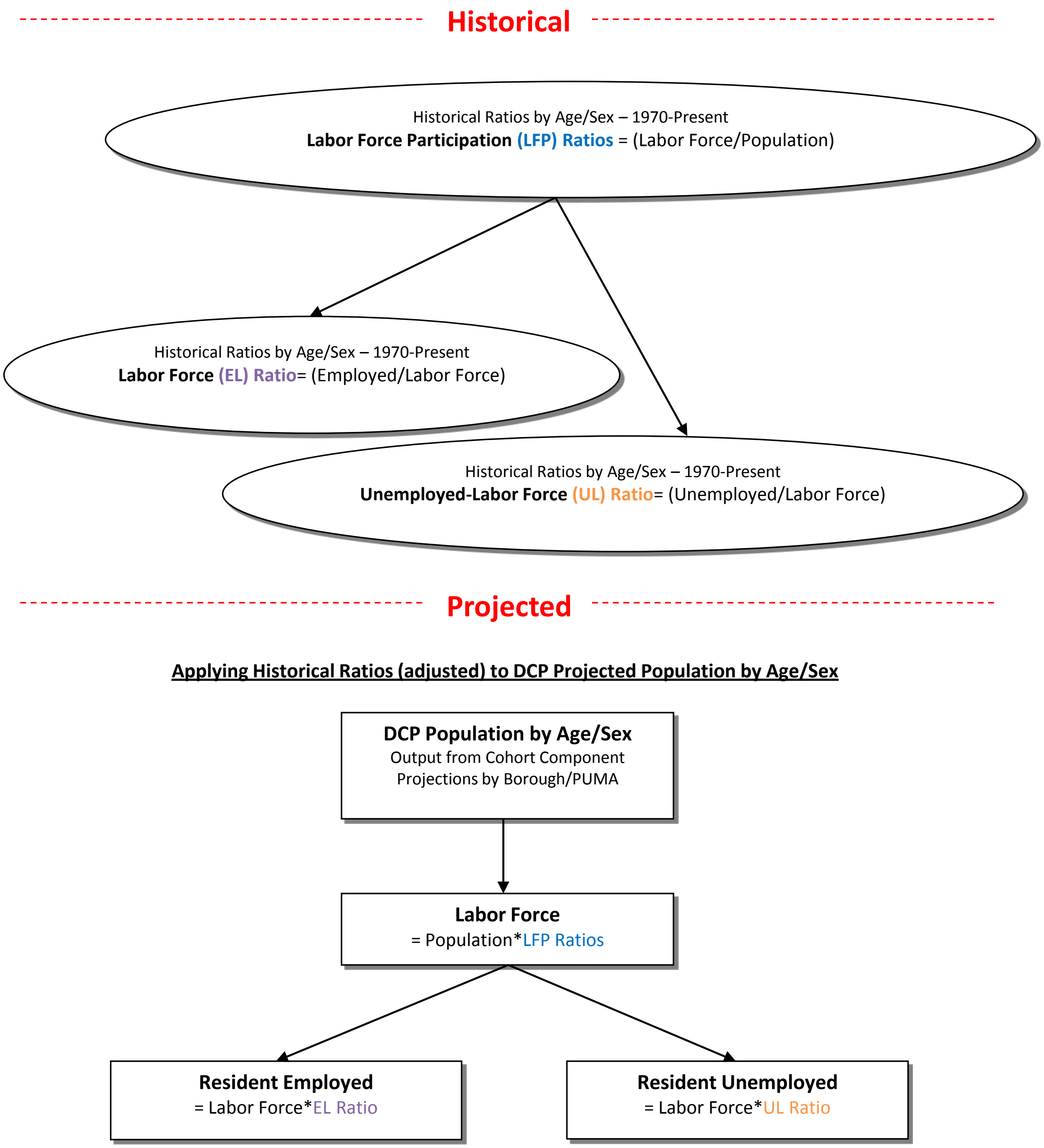
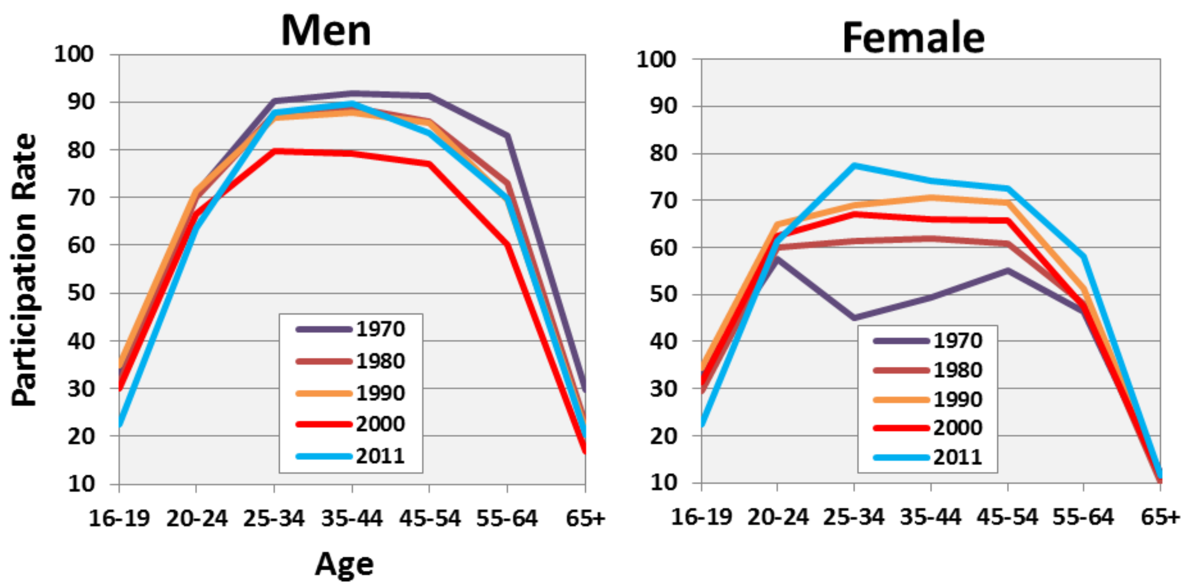


Figure 6 Labor Force Participation Rates by Sex and Age, New York City, 1970 to 2011



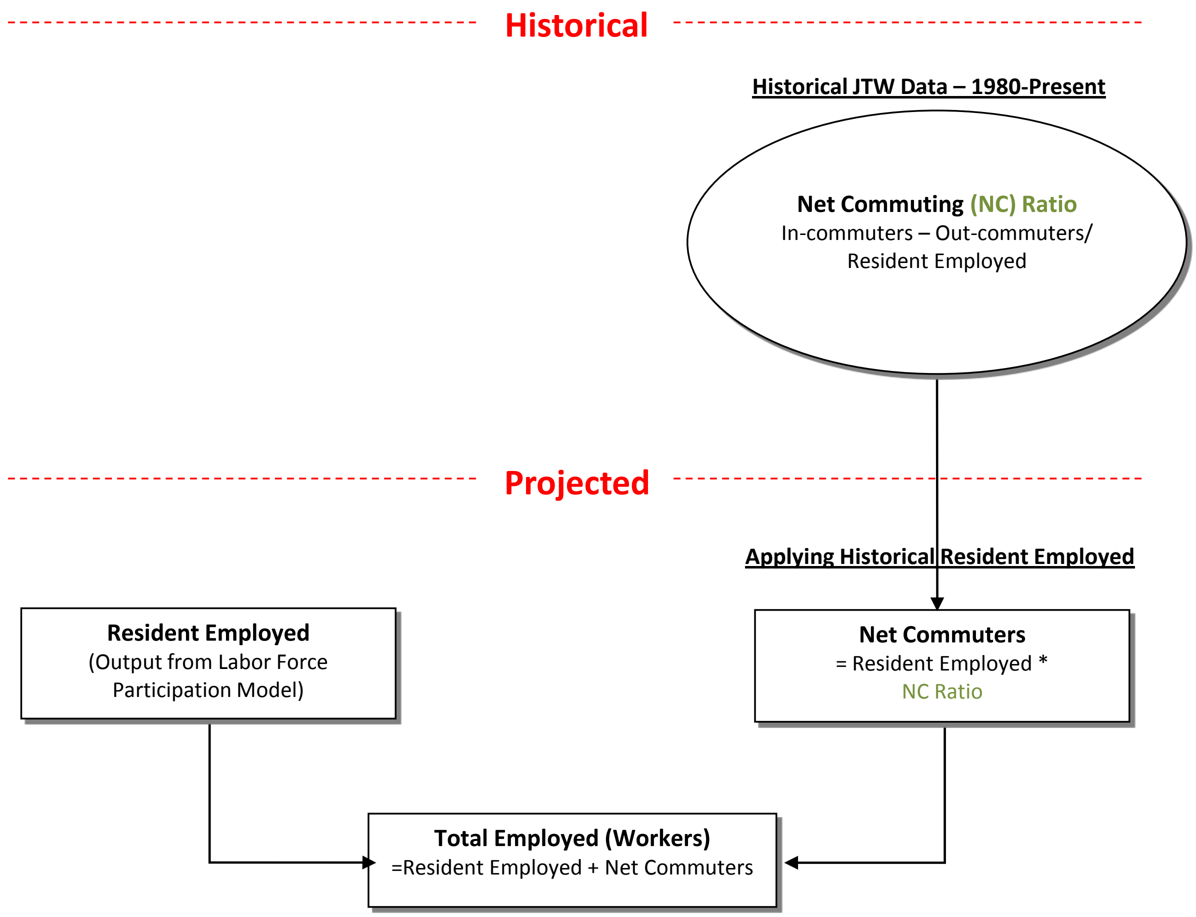
Source: 1970-2000 Censuses and 2011 American Community Survey

In addition to using history to guide calculations, data will be used from a number of sources to formulate assumptions about future participation rates. The Bureau of Labor Statistics of the U.S. Department of Labor does national analyses of participation rates by age and has a regular program of projections that currently goes out to the year 2018.[17] In addition, there is extensive literature on expected changes in participation rates, based on anticipated shortages of workers in selected occupations and industries, increased longevity, and an increasing proclivity among older workers to delay retirement because of reduced or nonexistent pension benefits, insufficient personal savings, and a lack of health care coverage.[18] This research, along with data for the city and region, is expected to be used to look forward and make informed judgments about the utility of using national rates to calibrate city-specific changes in participation ratios. Much of this will be based on an analysis of national ratios by age/sex compared to those in New York City's boroughs, with its relatively larger share of immigrants. This exercise should provide upper and lower limits on the number of persons in the resident labor force and the number of employed and unemployed at the city and borough levels.

7.1.2 Incorporating Commuters to Create Total Employment

Total employment is the sum of both resident employed (see above) and net-commuters available from Journey to Work (JTW) data (Figure 7). These data are collected by the Census Bureau in the decennial census long form and, more recently, in the American Community Survey. Historic data are available through the 1980 Urban Transportation Planning Package (UTPP) and its successors, the 1990 and 2000 Census Transportation Planning Packages (CTPP). For more current worker data, ACS 2006-2010 special tabulations on the Journey to Work will be used. These data files will be used to identify patterns that can serve as a basis for assumptions moving forward.

Figure 7 Population Model: Total Employed Derived From Resident Employed and Net Commuters

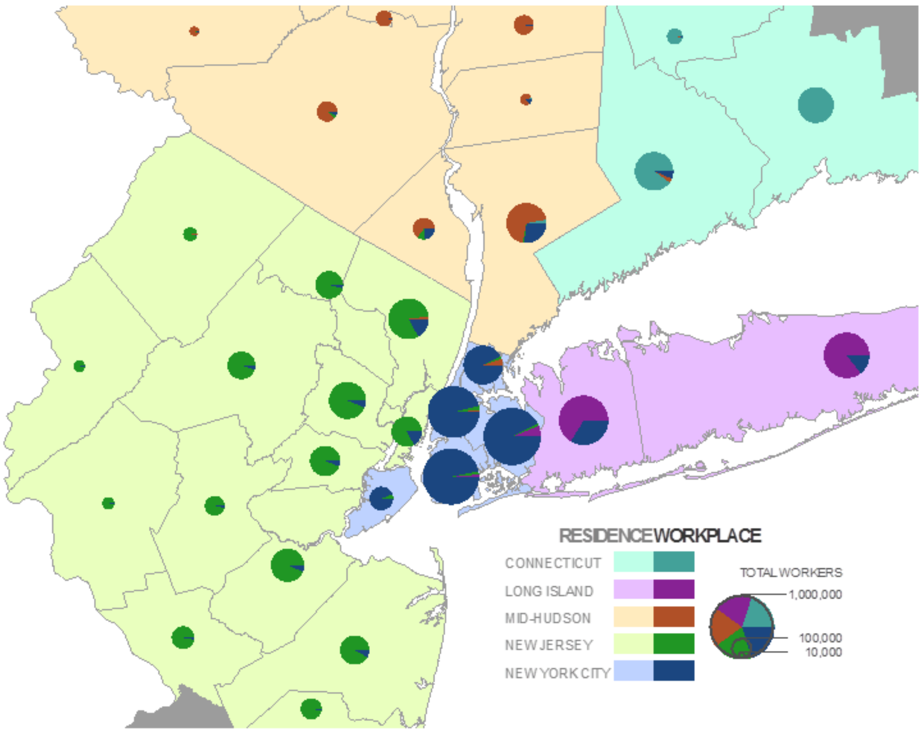


The basis for this analysis will be the Net Commuter Ratio, defined as in-commuters minus out-commuters, divided by total workers. The number of net-commuters will be calculated using these ratios and will be added to the resident employed, derived from the labor force participation ratios described earlier. Having established some baseline net commuter ratios, these can be multiplied by the resident employed to create numbers for net commuters going forward. Total employed is then derived by adding net commuters to resident employed.

A look at historical data on commuting between the region and New York City shows a pattern of remarkable stability. Since 1980, the percentage of all workers in New York City who are residents has remained in the range of 70 to 80 percent. Figure 8 through Figure 11 shows employment by county of residence from 1980 through 2006-2010. Pie charts show changes in absolute employment, as well as change in resident and out-commuting shares of employment by sub-region of workplace. The light coloring shows the residence sub-region and the dark color is the workplace sub-region; the workplace sub-region pies show the share of the residence sub-region that commutes into a county. Overall, resident and out-commuting shares remain fairly steady, with most noticeable changes occurring between 1980 and 1990. From 1980 to 2006-2008, the Mid-Hudson sub-region has seen a slight shift of workers from that sub-region to the adjacent sub-regions of Connecticut and New Jersey. This shift is most prevalent in counties directly bordering those sub-regions (e.g., more residents of Rockland and Orange counties now working in New

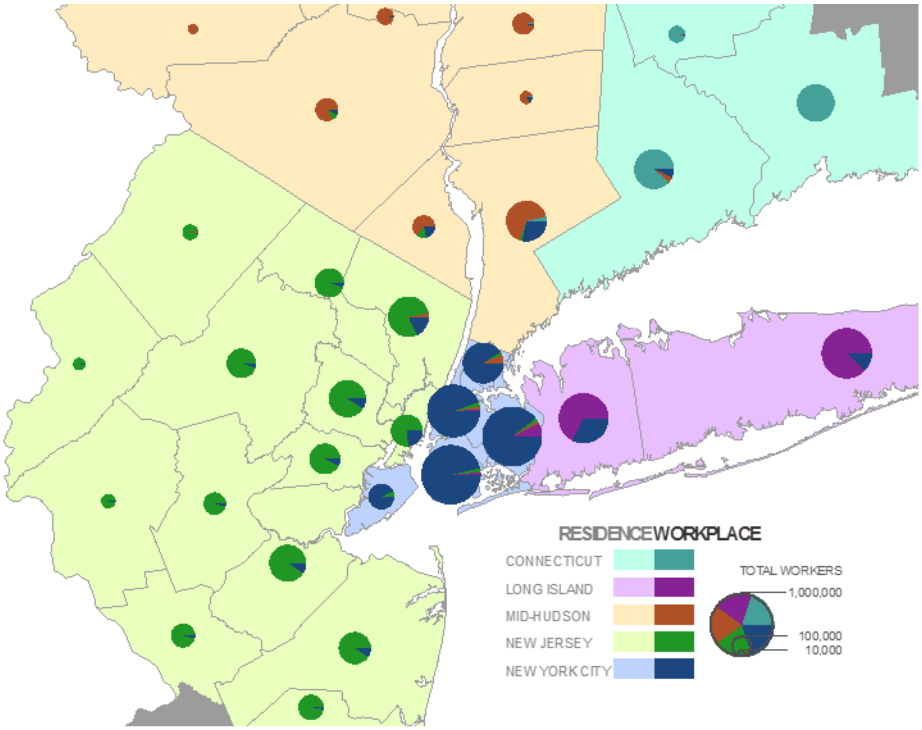
Jersey). Also, notable is a small shift in Long Island toward a greater share of resident employment over time. Throughout the region, the relative share of New York City's workers who have come from other points in the region has not changed to any appreciable degree, in spite of increases in overall employment. Figure 12 summarizes the trends in workplace shares by sub-regional place of residence over the period 1970 to 2006-2010.

FIGURE 8 1980 REGIONAL WORKERS: SUB-REGION-TO-COUNTY COMMUTER FLOWS



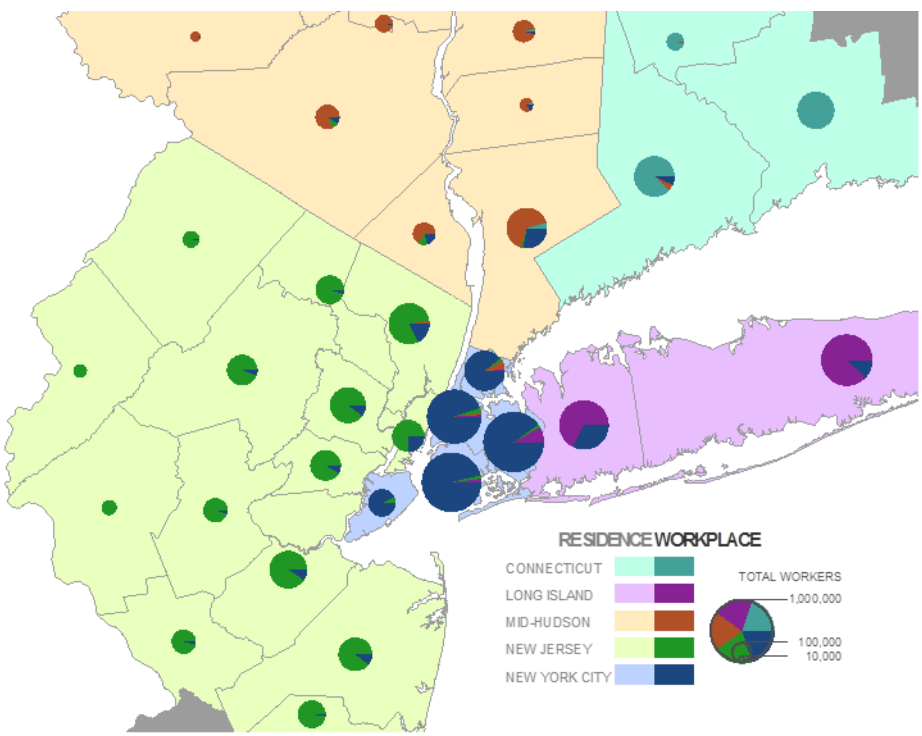
Source: 1980 UTPP

Figure 9 1990 Regional Workers: Sub-Region-to-County Commuter Flows



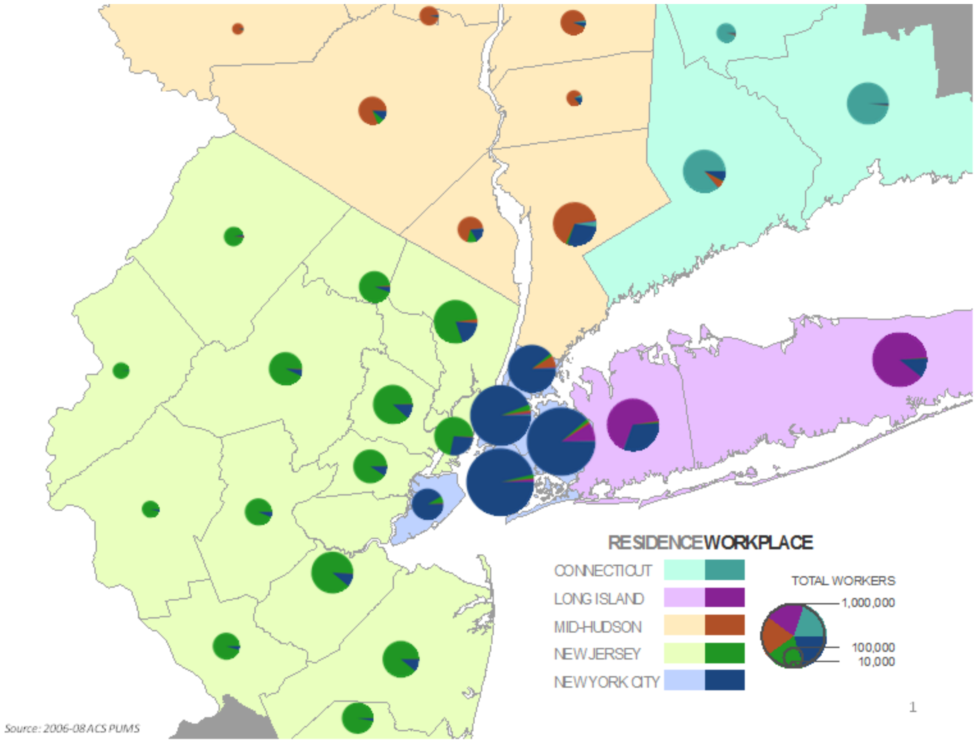
Source: 1990 CTP

Figure 10 2000 Regional Workers: Sub-Region-to-County Commuter Flows



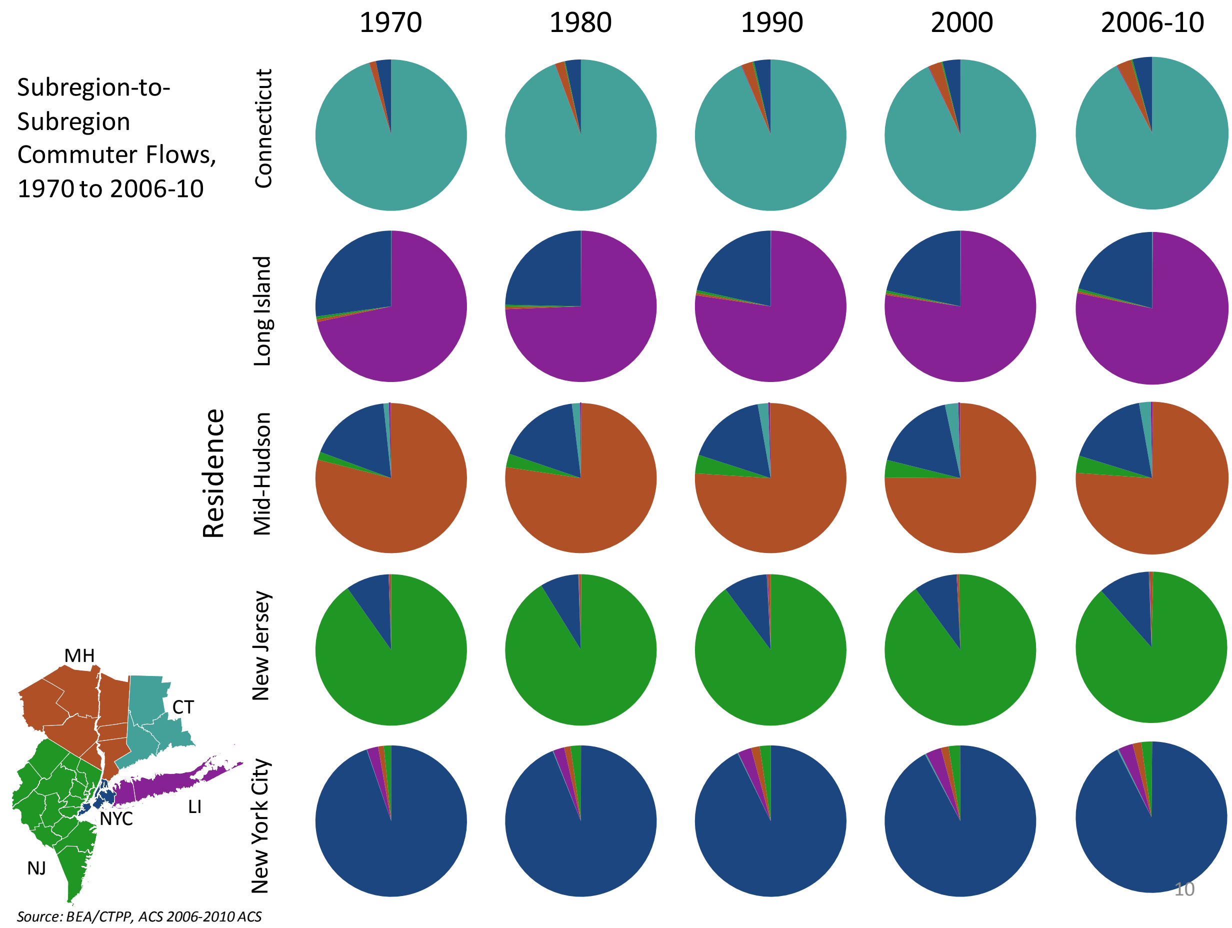
Source: 2000 CTPP

Figure 11 2006-2010 Regional Workers: Sub-Region-to-County Commuter Flows



Source: 2006-2010 ACS

Figure 12 Sub-region-to-Sub-region Commuter Flows, 1970 to 2006-2010



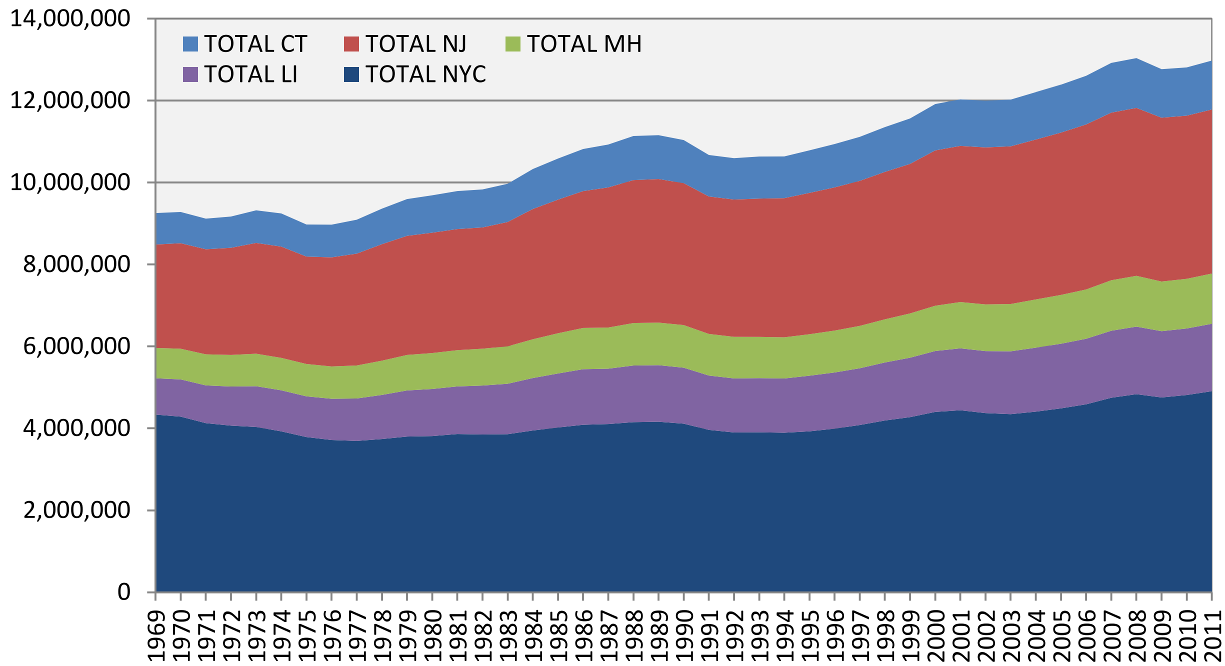
Source: BEA/CTPP, ACS 2006-10 ACS

Figures 13 and 14 show annual workplace employment by sub-region for the years 1969 to 2006-2010. Figure 13 shows employment in absolute numbers, with total regional employment change made up of component change in each of the five sub-regions.

Figure 14 shows sub-regional employment change converted into percentages of total regional employment. According to this figure, New York City's share of total regional employment declined through the 1970s and the 1980s, leveling off around the early 1990s. Sub-regional shares of regional employment have remained at relatively stable levels since that time.

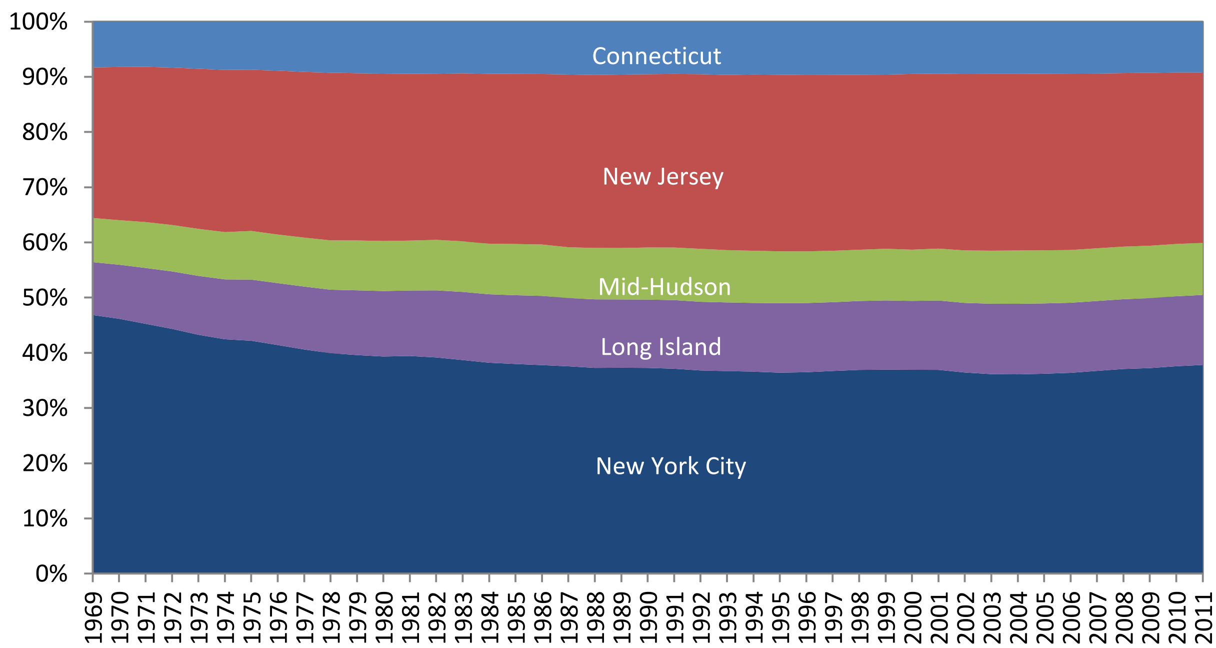
Simulations and sensitivity analyses will be conducted to see what the likely level of commuters would be if the Net Commuter Ratios remained consistent or changed, by margins that would determine our upper and lower bounds for the city and boroughs. An important part of this task will be to achieve a common view of the regional commuting picture; one that melds commuter projections for the city with their projections for other counties in the region.

FIGURE 13 SUB-REGIONAL SHARES OF TOTAL REGIONAL EMPLOYMENT, 1969-2011



Source: BEA REIS

FIGURE 14 SUB-REGIONAL PERCENTAGE OF TOTAL REGIONAL EMPLOYMENT, 1969-2011



Source: BEA REIS

7.2 Economic Method for Determining Total Employment

FIGURE 15 ECONOMIC METHOD FOR DETERMINING TOTAL EMPOLOYMENT

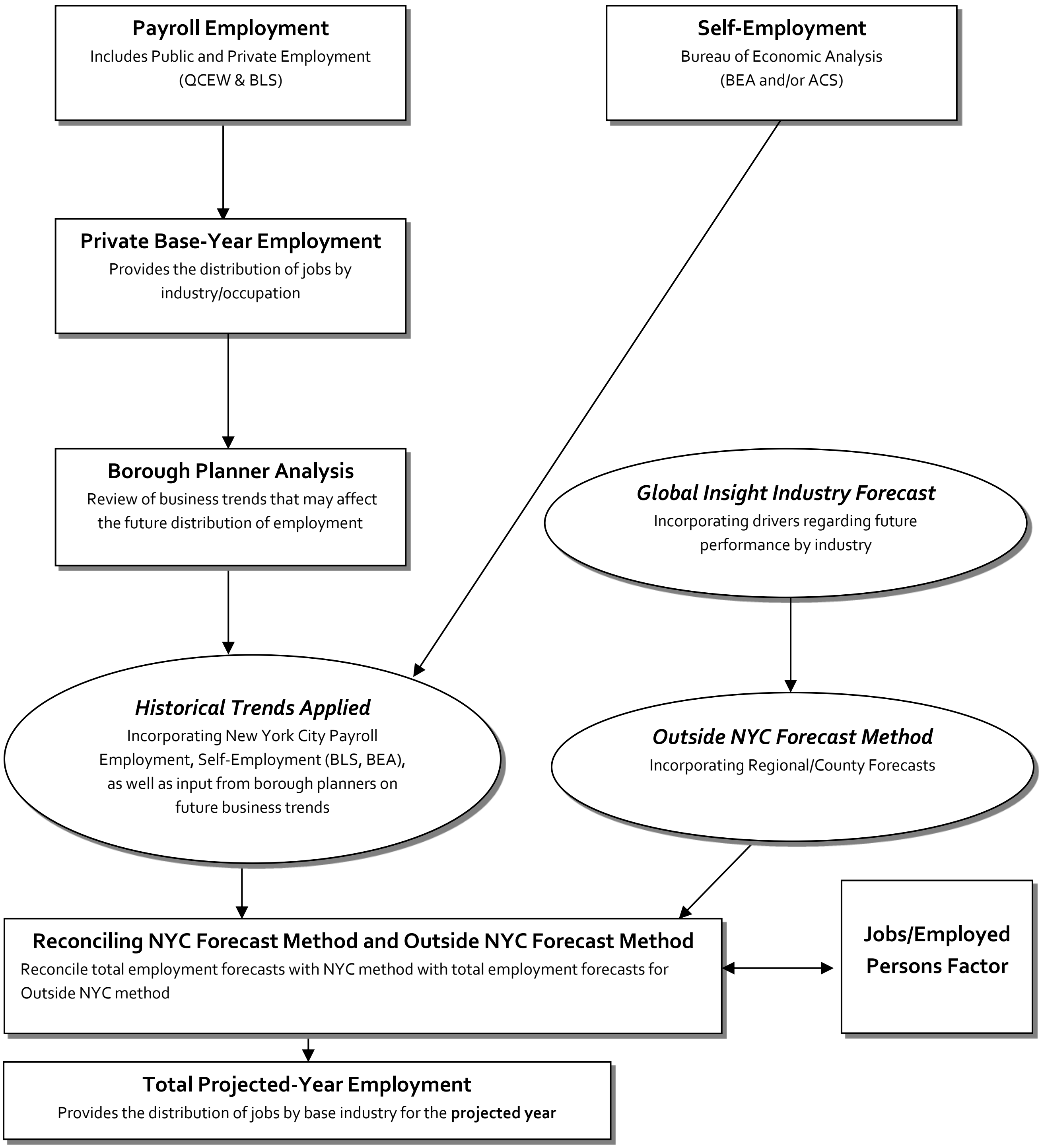
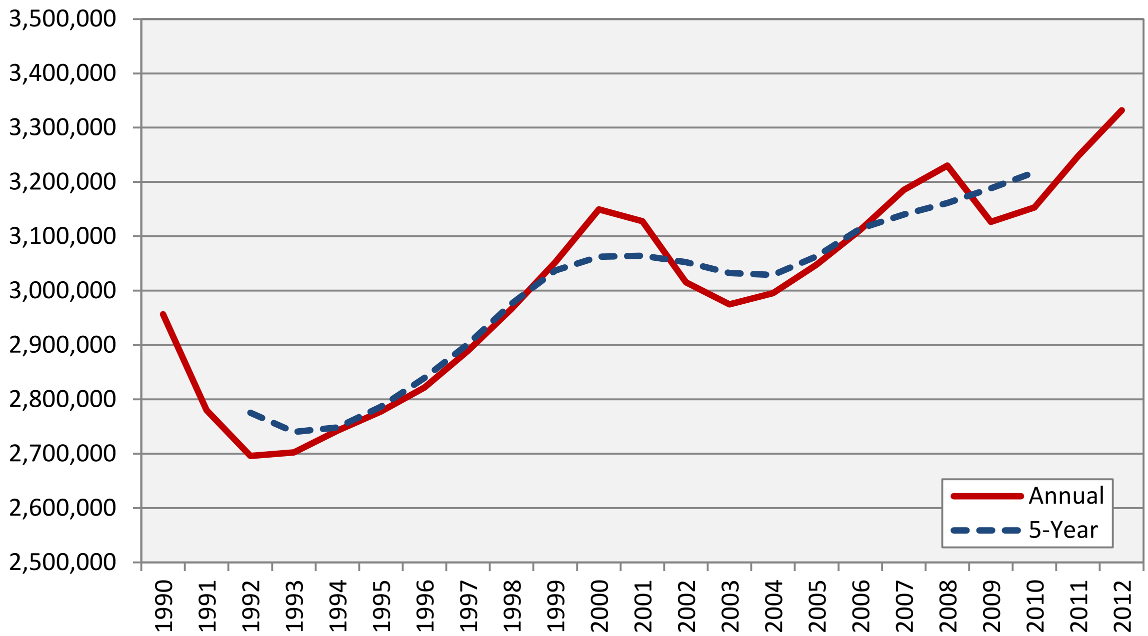
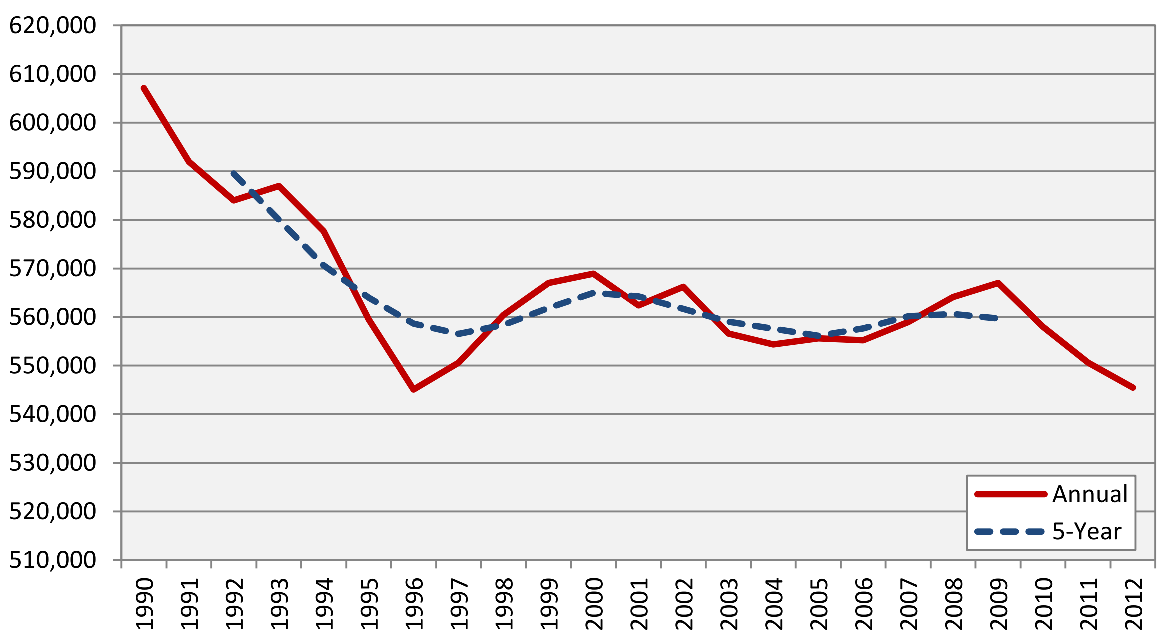


Figure 16 New York City Private Employment, 1990-2012



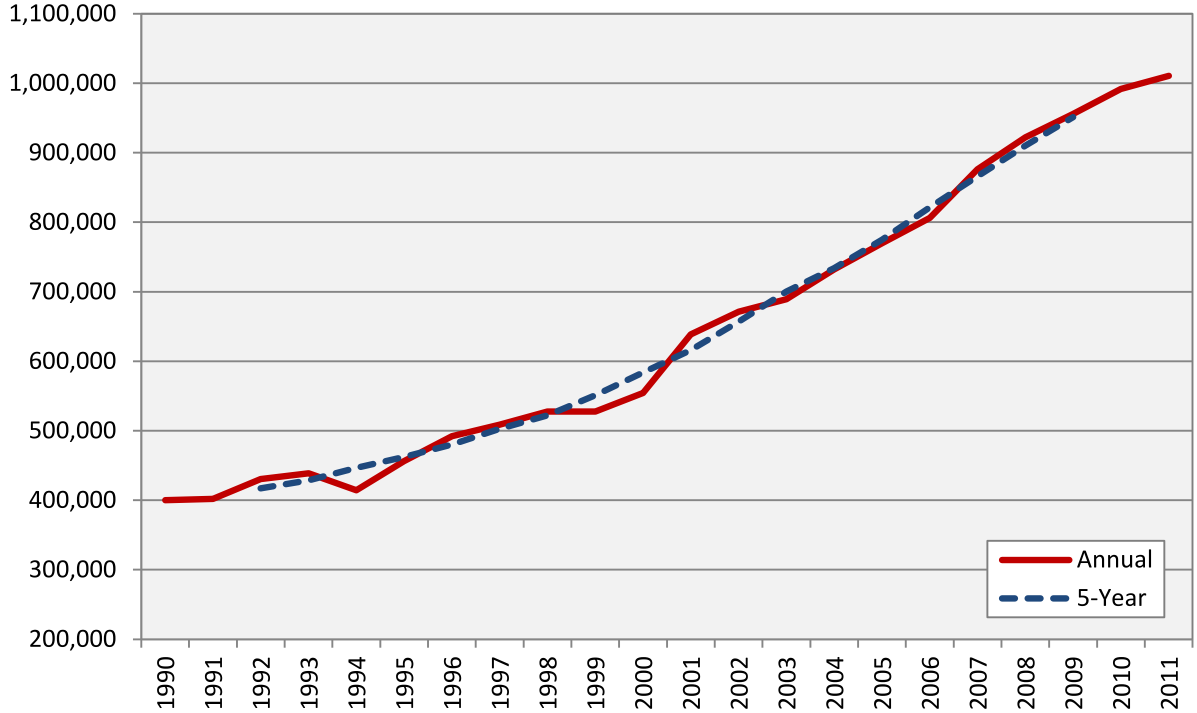
Above is a simple line graph showing the trend in total wage and salary private employment in NYC since 1990, and the five-year moving average since 1995, which smoothes the cyclical fluctuations. There is an obvious upward trend which likely correlates well with population growth. Source: BLS, Current Employment Survey (CES)

Figure 17 New York City Government Employment, 1990-2012



The above chart is for total government employment. Since 1997 the five-year moving average total government employment has fluctuated within a narrow range. Government employment per capita has fallen sharply. Source: BLS, CES

Figure 18 New York City Proprietors' Employment, 1990–2011



This last chart is proprietors' (self-) employment. The trend is exponential, reflecting the changing structure of employment as more and more people are at least part-time free-lancers. The trend is completely detached from the business cycle. Source: BEA, Local Area Personal Income, CA-04 series

7.2.1 Base Year Employment

Current Payroll Employment of wage and salary workers will be compiled by borough from the New York State Department of Labor (NYSDOL) Quarterly Census of Employment and Wages (QCEW). These data provide employment information by place of work, based on quarterly reports from employers covered under New York State's Unemployment Insurance Law.[19] Employment information provided in the QCEW dataset counts the number of jobs at an establishment level and aggregates them in to various geographies. Publically available at the county and New York City levels of geography, this information will be used as a base from which to work off.

NYCDCP has an agreement with NYSDOL to receive un-aggregated QCEW data annually by firm address, of firms located or that have employees who have worked within NYC within the past quarter. The data will be then geo-coded, or tied to an address that is within NYC that is recognized within DCP's addressing system. After a location for firms has been established, the data will be aggregated to sub-borough geographies, such as PUMAs or TAZs. It is important to note that confidentiality agreements exist for use and disclosure regarding this dataset.[20] As a result, employment data cannot be released at a sub-borough level without proper screening, adhering to confidentiality requirements set forth by NYSDOL.

Once the base numbers are gathered, a base-year industry distribution of employees by primary industry for private employment will be established. Using QCEW data, firms with employees will be assigned a 2 digit NAICS code and then aggregated by each major industry, including government employment. The employment data will initially be gathered at a borough level and then later allocated to PUMA then to TAZ in Task 6.

Since proprietor employment is not included in the QCEW dataset, base year self-employment data will be gathered separately based on data from the Bureau of Economic Analysis (BEA), which includes proprietor employment for New York City from 1969 to 2011.

Once the base-year industry distribution is finalized, meetings will be scheduled with borough planners to review current business trends by borough, as well as new trends that may affect the future distribution of employment. This analysis review will be used to update any new employment activity or new industry that incorporates on-the-ground knowledge that is not already included in the data.

7.2.2 Reconciling Forecasts by NYC Method and Outside NYC Method

The approach to forecasting total employment outside NYC is different from the economically driven method for NYC, described earlier. The results of both forecasts will be evaluated before making any conclusions regarding projected total year employment. The methodology for NYC will be more of a bottom up approach, where total employment will be derived at the county level. The forecasts outside NYC will be more of a top-down approach, where total employment will be projected at the regional level and then distributed down to the county level by industry, using a shift-share method of distribution. Therefore, total employment projections for NYC by each method will be evaluated against each other, with a goal of conforming to regional control totals and industry distributions. As will be discussed in Part III, Section 2.4, regional employment projections will be developed using shift/share methods driven by Global Insight economic projections.

7.2.3 Projected Year Employment

Total employment projections for the five boroughs of New York City independent of the resident employed/labor force projections based on the cohort-component population model will be created. Projected Year Employment is a function of several factors including national and local area trends in employment by industry, population, personal income, and the statewide economic outlook. Projections will be done on a borough level by primary industry (2-digit NAICS code) and then aggregated to the city. Projected employment growth will be dependent on a combination of sources which include historical data by industry and earnings for wage and salary workers, incorporating regional and national trends in employment reported by the NYSDOL and BLS.

In addition to projecting payroll wage and salary workers, self-employment projections for the projected year will be derived separately. After examining some historical patterns in proprietor data (Figure 18), the trends beg for separate treatment especially in light of a huge increase in self-employment. Self-employment projections will then be combined with payroll wage and salary employment to calculate total employment.

In addition to applying historical trends to payroll base-year employment, as mentioned above, national drivers from the research consulting company Global Insight will also be incorporated. Several Global Insight resources may provide a better understanding of the future performance of major industries and these drivers will be applied to the 31-county region to derive regional control totals for employment forecasts. Using the shift-share method, long- term economic projections by major industry will be incorporated into projected-year employment. The US County forecasts from Global Insight and county forecasts will further help determine market strength and vulnerability for industries located within NYC and provide an independent economic and financial outlook by industry. This collaborative effort will provide a better perspective regarding the future performance of major industries in NYC for the projected years.

7.2.4 Reconciling Population and Economic Based Models

There are several important points and caveats about the economic model that need to be noted. First, in addition to corroborating estimates of employment and commuters, the economic model will provide important information on changes in employment by industry, allowing the creation of selected industry subdivisions (e.g. Office jobs) required for this task. Second, the economic model provides estimates of "employment" from the perspective of jobs. The unit of analysis in the decennial censuses and in the ACS is employed persons, with those holding multiple jobs represented just once. The employment data focuses on jobs, thus allowing those with more than one job to be represented multiple times in the data. This means that the residual that exists between the economically-driven projection and the population-driven one involves some number of multiple job holders. In order to get a handle on the ratio of census employed persons to total employment from the economic model, historic ratios will be examined to create an adjustment factor that accounts for this difference. In making this determination, one expected approach is to compare census employed persons to total employed from the U.S. BEA.

8.0 HOUSEHOLD INCOME (ITEM 7)

The goal will be to look at the current 2012 distribution of income for all households in New York City by borough (and PUMA for allocation to TAZs). The idea is to construct a model of how median household income is distributed by borough (and PUMA), and apply the population to this distribution for projected years. The best way to measure future changes in household income would be to designate income quintiles by maintaining the current distribution moving forward, but assigning high, medium and low income households to upper, medium and lower fifths of households and apply the projected year population to the income distribution. Since it will not be assumed that income distribution gets more unequal ad infinitum, the quintiles for 2012 will be used going forward.

8.1 A Note about Other Counties in the New York Region

Much of what is done for New York City will need to be integrated into a common demographic and economic picture for the region, especially on the commuting front. This will be accomplished through a thorough examination of the assumptions underlying the models. This is especially true since population and employment shifts in the counties surrounding New York City would have big ramifications for the city's forecasts; establishing a consensus will be essential.

III. OUTSIDE NYC FORECAST

1.0 INTRODUCTION

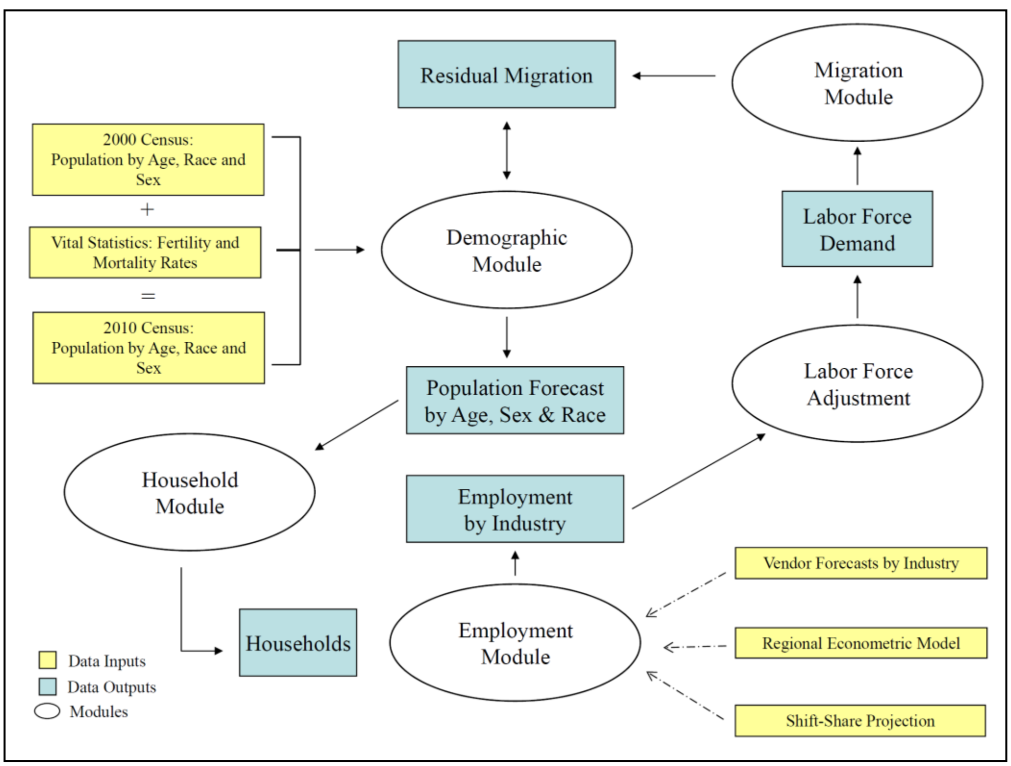
Part III of this paper will describe the proposed forecast methodology for the 26 counties outside of NYC. The key objectives are to provide further elaboration on the key issues to be addressed in the development of the forecast and further detail on methods and best practices in long-range SED forecasting. The next section of this paper provides a summary outline of the proposed forecasting methodology, which reflects the literature review's findings. Subsequently, the paper will discuss additional topics for consideration in the context of enhancing the existing process; these topics include economics, forecast stability, incorporating land use constrains and including input from stakeholders and experts in the field.

2.0 SUMMARY OF PROPOSED METHODOLOGY

2.1 Overall Structure

A new forecasting model drawing will be developed from the current SED demographic model, where appropriate, to incorporate the best available data and effective practices in forecasting as discussed above. A detailed approach to estimation of natural increase, migration, and labor force demand and participation is proposed with calculations conducted in several modules as outlined below and depicted in Figure 19.

FIGURE 19 OVERVIEW OF PROPOSED SED APPROACH FOR OUTSIDE NYC REGION



2.2 Population Model

The methodology used to prepare projections of population and demographic characteristics at the county level will be a disaggregate cohort-component model, similar to the methodology for the NYC region (see Part II, Section 4). The method employs detailed age, sex, and race/ethnicity (if determined to be a relevant contributor to forecasting during testing, a recommendation on approach will be made to the FWG) data from the U.S. Census Bureau and state vital statistics data. The model produces estimates of population change attributable to natural increase (births minus deaths) and net migration (net population flows into or out of the county). The steps involved for estimating each component of population change are as follows:

2.2.1 Natural Increase

The first step in the estimation of county population involves projecting the natural increase in the population from rates of fertility and mortality. This "zero-migration" estimate of population growth of the 26 counties outside NYC will be developed using vital statistics by age, sex and race/ethnicity (as appropriate) groupings. The aging population is tracked in five-year increments by applying annualized cohort-specific survival ratios (the adjustments made for deaths) in each age cohort. Births during the period are derived by tracking females of childbearing age (i.e., 15-49 years) and applying age-specific fertility rates. Calculations are performed for age, sex groupings (to capture observed differences in fertility, mortality, labor force participation and household formation); and to segment the population to make projections based on observed national and regional demographic differences in a manner consistent with the comprehensive forecasting analyses conducted by the U.S. Census Bureau, the Bureau of Labor Statistics and DCP forecasts.

2.2.2 Fertility & Mortality Projections

The demographic model will allow for survival and fertility ratios to remain constant (fixed at the average of the rates observed in each category of the ten-year vital statistics time-series) or to fluctuate over time. Change over time will be keyed to the U.S. Census bureau national forecasts of vital statistics index.

2.2.3 Initial Net Migration Estimate

To derive an initial estimate of net migration, the natural increase in population is calculated by estimating the number of births (to develop a new cohort) and deaths (to determine the survival rate for other cohorts) between the base year (2010) and the previous decennial census (2000). This calculation produces an estimate of the "survived-population" and new population for the initial period. Then, by utilizing Census 2010 control totals, the residual or net migration level is determined for each age cohort. The residual is the difference between historic natural increase or survived-population and Census counts. Dividing the residual figure by the survived population in each age cohort provides a "migration-index" or a propensity to migrate for each age cohort stratified by sex. To promote stability in the forecast and to understand the impact of in-flows and outflows of commuters affect the county-level forecasts, the individual county migration-index estimates will be compared with a region-wide estimate developed through a regional-level CC model.

2.3 Household Formation Model

2.3.1 Household Module

The calculation of households from population will involve a detailed evaluation of household formation patterns. A simple calculation would rely on the average persons per household, but a more sophisticated analysis is necessary to disentangle variations in the propensity to form households amongst the various age and sex cohort categories within the Demographic Model. As such, the Household Module relies on base-year headship rates to forecast household formation. In addition, the Household Module estimates average household income for each age cohort categorized by sex. Headship rates are derived by age and sex based on review of the Public Use Microdata Sample (PUMS 1% Sample, 2000 and as available, 2010 and ACS). A headship rate is the ratio of household-heads (self-identified classification by census respondents) and the population. To forecast the specific household and housing variables required by the BPM, the module will also include historical data for these variable including household size and type as well as housing type and tenure.

2.3.2 Household Income Calculation

Using the PUMS data set, the number of households in each of the census income categories will be estimated for each of the age cohorts by sex and race/ethnicity. Utilizing the income distributions and aggregate income levels from the PUMS data set, the average household income is a function of the changing age-sex-race/ethnicity characteristics of the region. The model will project income-levels in constant dollars consistent with the date of the PUMS data set and will be calibrated to observed household income statistics by county.

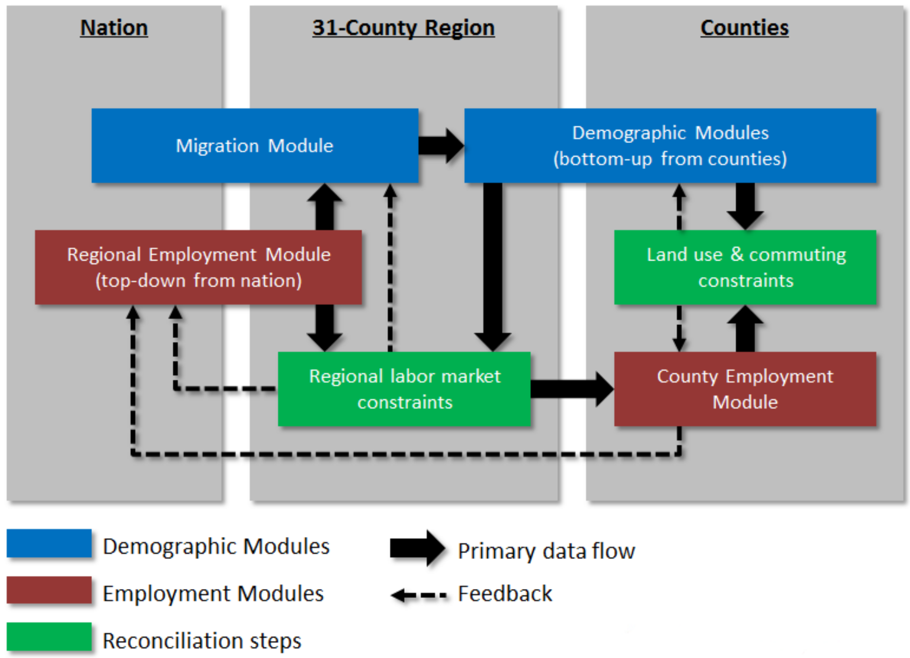
2.4 Employment Model

As discussed in the Executive Summary, the employment methodology for the outside NYC region will be used to forecast employment for the full 31-county region; a reconciliation process through the FWG will reconcile the forecasts for NYC generated by the methodology presented here and the methodology for NYC presented in Part II, Section 7. An employment forecasting methodology will be developed that is stable, transparent and easy to implement/update. Specifically, the approach to employment forecasting includes the development of a spreadsheet-based model (including Graphical User Interface – GUI) that presents the employment forecast assumptions and calculations in a straightforward manner. The proposed technical approach is to apply a top-down allocation of national economic growth that leverages available data subscriptions, but that also allows flexibility to modify forecasts for calibration purposes or based on FWG feedback. The resulting approach – similar to the SCAG approach reviewed in Part I – allocates national growth to county employment growth in four primary steps (see Figure 20):

1. Allocate national employment growth to 31-county region;
2. Reconcile regional employment with population via labor supply and demand constraints;
3. Allocate regional employment to counties;
4. Reconcile county employment with population via land use & commuting constraints.

This four-step process incorporates third-party forecasts together with a series of econometrically-calibrated shift/share models in a framework that allows for multiple points of calibration and feedback. Detailed aspects of these steps are described in the following four sections.

FIGURE 20 OVERVIEW OF EMPLOYMENT FORECASTING APPROACH



2.4.1 Regional Employment Module

To promote stability and consistency in the employment estimates as the forecasts are updated on a period basis, first developing one or more region-wide control totals for employment by industry is recommended. As discussed above, the goal is to produce a stable forecasting platform that can be easily adjusted to (1) reflect input from stakeholders and the FWG and (2) ensure consistency between the top-down employment forecasts and bottom-up demographic forecasts. Introducing the 31-county region as an intermediate target will assist both these goals.

The proposed Regional Employment Module will allocate national economic growth in several sub-tasks. First, a third-party national forecasting model (likely IHS Global Insight) will be used to establish a baseline national economic forecast of employment by 2-digit NIACS sector and components of GDP. As will be discussed in Section 3.1, large-scale national macroeconomic models – such as IHS Global Insight and Moody's – have the greatest predictive power and are most theoretically valid at the national level.

Parallel to this task, an accounting tool will be developed to be able to modify select aspects of the third-party national forecasts (for example, net exports or overall GDP growth). The goal here is twofold. First, this introduces a point of flexibility into the overall forecasting framework, providing national scenario "levers" that can be used to adjust the NYMTC forecasts upwards or downwards for calibration purposes. This is particularly important if the built-in assumptions of the third-party forecast are irreconcilable with the bottom-up demographic forecast. Second, the accounting tool provides a risk-analysis component through the possible development of a range of national forecast scenarios.

The result of the previous sub-task will be one or more national forecasts of employment by 2-digit NAICS sectors. In the next step, a set of econometrically calibrated shift/share models will be developed to allocate national employment to the 31-county region. These will be calibrated sector-by-sector using historical BEA and BLS data for the 31-county region, as well as cross-sectional CBP and LEHD data to determine robust correlations between employment growth and facets of metropolitan activity (similar to the SCAG econometric methods). Here, the goal is to arrive at a small set of robust specifications for each NAICS sector. These specifications will then be incorporated into the forecasting spreadsheet and applied to national employment to determine employment by industry in the 31-county region. It is also proposed that regional totals from the third-party forecasts be incorporated as a separate scenario in the Regional Employment Module. As before, this approach is data-driven, yet provides flexibility to make adjustments for calibration purposes or based on FWG feedback.

2.4.2 Regional Labor Market Adjustment Module

The next main step of the proposed methodology is to perform an explicit comparison of population and employment at the 31-county region level. It is envisioned that this reconciliation be similar to the SCAG process, with the goal of ensuring that the assumptions built into the bottom-up demographic forecasts are commensurate with the top-down employment forecasts. To perform this reconciliation step, several adjustments will be applied to the "raw" population and employment estimates to simulate labor market supply and demand (see Section 2.5):

- labor force participation rates
- self-employment rates
- employment not accounted for in QCEW estimates
- multiple-job holding rates
- net in/out commuting at the region's border

Applying these adjustments will yield values for total employment supply and demand – and implied unemployment. Here, the FWG will review the implied unemployment rate in the context of NYMTC history and third-party national vendor forecasts to determine the level of convergence. If the FWG is satisfied that employment and demographic forecasts are balanced, then the employment forecasting process can proceed with the down-allocation to counties. If not, then the following adjustments are available to bring the forecasts closer in line:

- Adjust the national macroeconomic forecast(s);
- Adjust industry-specific shift/share specification(s);
- Adjust demographic migration rates (or other aspects of the demographic module).

2.4.3 County Employment Module

The previous two steps are designed to ensure broad region-wide consistency between the employment and population models. With this satisfied, the forecasting process can proceed to the next step, which is to down-allocate the regional employment totals to individual counties. As with the regional employment module, the proposed approach is to apply econometrically calibrated shift/share models for each industry.

Although this process will be informed by the specifications for the regional allocation (Step 2), county allocations will undergo a separate calibration process. As before, the goal is to arrive at a narrow set of robust drivers of county employment share for each industry. In this case, more targeted data (such as land use or commuting patterns) may be incorporated with the economic and demographic data sets listed above. These specifications will then be incorporated into the forecasting spreadsheet and used to allocate regional employment control totals. It is also recommended that county-level detail of the third-party forecasts be adopted as a reference scenario.

2.4.4 Land Use and Commuting Constraints

The final step in the proposed employment forecasting model is to perform an explicit reconciliation of the bottom-up demographic module and top-down employment module at the county level. The previous reconciliation step ensured broad agreement between the supply and demand of labor for all forecast years. In the present step, the goal is to ensure that the spatial pattern of employment and population is reasonable in the context of land use forecasts and commuting patterns.

To reconcile county employment to land use constraints, it is recommended that the forecasts be reviewed by appropriate FWG members and local planning agencies to determine if the projected employment (by sector) can be plausibly accommodated within each county. If necessary, county forecasts can be adjusted by modifying county shift/share specifications, regional shift/share specifications, or even national forecast scenarios (on an industry-by-industry basis).

A final step in the reconciliation process is to ensure that the commuting patterns implied by county-level population and employment forecasts are reasonable. To perform this step, county labor force estimates will be derived as above and matched to employment demand based on a trend of historical county-to-county commuting patterns (from CTPP, LEHD and other sources). It is proposed that the FWG and NYMTC staff review the implied county-to-county commuting flows to determine if the results are plausible. As before, necessary adjustments can be made at the county or regional level.

2.5 Labor Force Model

In the 31-county region and other areas with a strong and varied employment base, a primary determinant of net migration is labor force demand. Unlike a pure demographic model, which relies on the residual migration index, the demographic-economic model allows migration patterns to fluctuate depending on the demand for labor in the local economy. Instead of dictating both volume and composition, residual migration in the inter-census period (2000 to 2010) dictates the allocation of future estimates of migration flows – into sex and race/ethnic specific age cohorts – produced by the calculation of local labor force demand. Net migration of the labor force for each county reflects the residual required to balance labor demand (available jobs) and labor supply (labor force) after accounting for unemployment. Labor force net migrants are then converted into total persons by analyzing dependency ratios and trends in elderly migration (i.e., persons 0-15 years and adults over 65 years of age). The demographic composition (i.e., the age, sex, race/ethnicity characteristics) of net migration is allocated in proportion to each cohort's share of observed county migration during the 2000 to 2010 period. Steps in the calculation include:

- **Labor Force Participation Estimate** - Civilian non-institutionalized labor force projections are a critical variable for the development of net migration estimates for an accurate determination of future population projections. For base year estimates, the Labor Force Module will utilize the sex and race/ethnicity specific participation rates as reported by the Bureau of Labor Statistics (BLS) for age cohorts reflecting key periods in the life cycle of the worker. Currently the overall national rate of labor force participation is approximately 63.5 percent.
- **Multiple Job Holdings Adjustment** – An important facet of the labor market is multiple job holding patterns. While employment is typically measured as a head count of filled positions

(including self-employed), the labor force is measured by the number of persons employed or seeking employment. Current and historic data on multiple job holding at the regional level (available from the BLS) will be used to develop an adjustment factor. Currently, the proportion of multiple job holders to total employment at the national level is approximately 4.7 percent.

- **Commutation Adjustment** - A final control factor for local labor force demand and in turn, migration, is the extent of non-resident worker commutation. Often times, non-resident workers will satisfy local labor force demand before migration patterns change because of proximity, better access to information and lower transaction costs. To account for this phenomenon, the relationship between employment (i.e., establishment-based counts or commute destinations) and labor force (i.e., household-based counts or commute originations) is estimated household travel survey data maintained by NYMTC and neighboring MPOs, FHWA ACS CTTP tabulation, LEHD, and other sources. Future year commute flows will either be derived from the BPM or projected from the base year using the Fratar[21] method.
- **Group Quarters Adjustment** – A careful review of decennial census and ACS group quarters estimates as well as input from stakeholders will be sought to confirm these assumptions at the county level.

Migration of Dependent and Senior Populations – Migration patterns of the "dependent population" are primarily a function of labor force demand for household workers. The number of migrants 15 years and younger will be determined based on the prevailing child-dependency ratio in each county population. The prevailing relationship identified for the period 2000-2010 establishes relative shares for the allocation of dependent migrants to age cohorts by sex. Persons 65 years of age and older are estimated based on observed migration patterns for the 2000-2010 period, and as appropriate, previous periods, at the county level. During this analytical stage, regional annual estimates of population 65 years and over are reviewed as an additional reference and method for adjusting net migration rates and patterns by age-sex.

3.0 Additional Considerations for Proposed Methodology

Based on the review of the existing NYMTC forecasting process, a series of additional topics to be considered in the proposed methodological approach have been identified:

- Economic considerations
- Forecast stability
- Land use considerations
- Incorporating expert/stakeholder input
- Coordination and integration with NYC Forecasting Process

A deeper investigation of these topic areas and their relevance to the forecasting process is provided in the following sub-sections. The resulting ideas and recommendations are intended to augment the proposed approach detailed in the scope of work, which is also summarized in the prior section (Section 2) of this paper.

3.1 Economic Considerations

Household and employment growth in the metropolitan region are governed in part by long-term trends related to aging of the population, household composition participation in the workforce, and general economic conditions. Migration into and out of the metropolitan area is a manifestation of these trends, and is directly linked to both, the short-term business cycle, long-term trends in industry composition and to large-scale demographic shifts. In addition, at the sub-regional level, occupation and industry trends, type of housing, and housing affordability all influence inflows and outflows.

3.1.1 Long-Run Structural Change

Over time, economies undergo structural change in ways that affect the demand for labor. Examples include growth and decline in specific industry sectors, changes in import/export activity, and productivity trends. While these trends are difficult to predict in the long-term, select forecasting tools can be used to understand the duration of trends as well determines the sensitivity of structural changes to future employment.

Large-scale macroeconomic models such as those employed by IHS Global Insight and Moody's are perhaps the best available tools to predict near-term changes in economic structure – particularly as they relate to national and international issues. Their long calibration history and multiple feedbacks allow them to project recent trends while also being "mean reverting" to prevent implausibly divergent forecasts.

In many cases, structural changes can be explicitly controlled for in alternative forecasts. For example, when models separately identify sources of demand by industry (for example, domestic versus international demand), these can serve as points of sensitivity to determine the robustness of regional employment forecasts.

It should also be noted that the large-scale macroeconomic models tend to yield more accurate predictions at larger spatial scales. This is because (1) smaller spatial scales diverge more easily from national trends due to specific local factors, and (2) the equilibrating assumptions of macroeconomic models become more valid as the number of interacting agents increases.

Separate from time-series approaches that propagate recent structural changes, several quantitative methods are available to understand the *sensitivity* of regional employment to structural change. One is shift/share analysis, which uses one facet of structural change – industry mix – as an input to the forecasting process. Another approach (previously used by ABAG) is an extension of the shift/share approach that uses a fully-balanced input-output model (for example, RIMS or IMPLAN) to test regional consequences of select structural changes. In this case, changes such as variations in domestic demand, international demand, or household composition by income group can be simulated to determine regional employment implications.

3.1.2 Effects of Business Cycles

As in the previous section, large-scale national macroeconomic models such as those employed by IHS Global Insight and Moody's are perhaps the best available tools for forming economic projections in the context of business cycles. These models incorporate a wide range of variables across leading and trailing indicators that help them steer projected growth back to stable, long run ratios of unemployment, inventories, and investment. However, the present cycle's unprecedented depth (in recent history) creates uncertainty around the exact timing of a full employment recovery. In that context, it may be useful to employ alternative macroeconomic forecasts where available.

These could either be alternative forecast scenarios or ranges from individual forecast services, or comparisons between multiple forecast services (such as Global Insight and Moody's).

3.1.3 Scenario/Risk Analysis

The term "risk and uncertainty" tends to be applied generically to the analysis of situations with unknown outcomes. In essence, risk is a quantity subject to empirical measurement, while uncertainty is non-quantifiable. Thus, in situations involving risk it is possible to indicate the likelihood of a variable falling within stated limits. On the other hand, in situations of uncertainty the fluctuations of a variable are such that they cannot be described by a probability calculus. Thus, risk and uncertainty are best thought of as representing a spectrum of unknown situations which an analyst may face, ranging from perfect knowledge of the likelihood of all the possible outcomes at one end (i.e., risk) to no knowledge of the likelihood of possible outcomes at the other (i.e., uncertainty).

Risk analysis in forecasting is used to provide decision-makers with a means for exploring uncertainty in the inputs and assumptions used to derive future projections. The application of risk analysis methods to the forecasting process is designed to minimize uncertainty by defining the range of potential outcomes, and by providing estimates of the probability of attaining each outcome.

What-if scenarios – based on non-probabilistically determined high, medium and low projections – have been the traditional method of accounting for forecasting uncertainty.[22] The real range of possible outcomes is not revealed through this traditional approach to addressing forecasting uncertainty. The range of outcomes could better be represented through a simulation process (Monte Carlo or Latin Hypercube) that not only incorporates the probability of occurrences of an individual variable, but also takes into account covariance – the likelihood of each occurrence in combination with other related variables.[23]

There are three main approaches to deriving probabilistic forecasts of both population and employment.[24]

- **Time-series extrapolation method.** This process involves fitting time series model on historical data, and subsequent extrapolation based on estimated parameters.
- **Extrapolation of empirical errors.** This process uses observed errors from historical forecasts to ascertain uncertainty.
- **Random scenario approach.** This process uses expert opinions in the definition of probability distributions for the future values of key indicators.

The time series extrapolation method is considered the traditional means of addressing forecasting uncertainty while the random-scenario approach is appealing to forecasters due to its simplicity and flexible framework that better facilitates direct involvement by stakeholders and experts. The Project Team is intimately familiar with both the time series extrapolation and the random scenario methods, and will therefore combine both approaches in the forecasting process used to develop probabilistically determined ranges of potential outcomes. This combined approach dovetails with a number of elements of the proposed forecasting process as described in the steps below, as opposed to the extrapolation of empirical errors process that requires additional computational steps that are less familiar to the Project Team.

3.1.3.1 Input Variables

It is expected that input variables from IHS Global Insight will be used in the forecasting process and the data will include a historical series, as well as three future projections (baseline, optimistic and pessimistic).

The three future projections provided by IHS Global Insight do not include a probabilistic estimate of likelihood and therefore @Risk software will be used to conduct an independent evaluation of probabilistic future outcomes – this process will thereby incorporate simulation techniques into the time series extrapolation method noted above. Given the time series of a particular variable, @Risk software can be used to generate probabilistic ranges of future outcomes either manually or by directly relying on @Risk's internal features.

FIGURE 21 EXAMPLE OF INPUT VARIABLE GROWTH RATE DISTRIBUTIONS

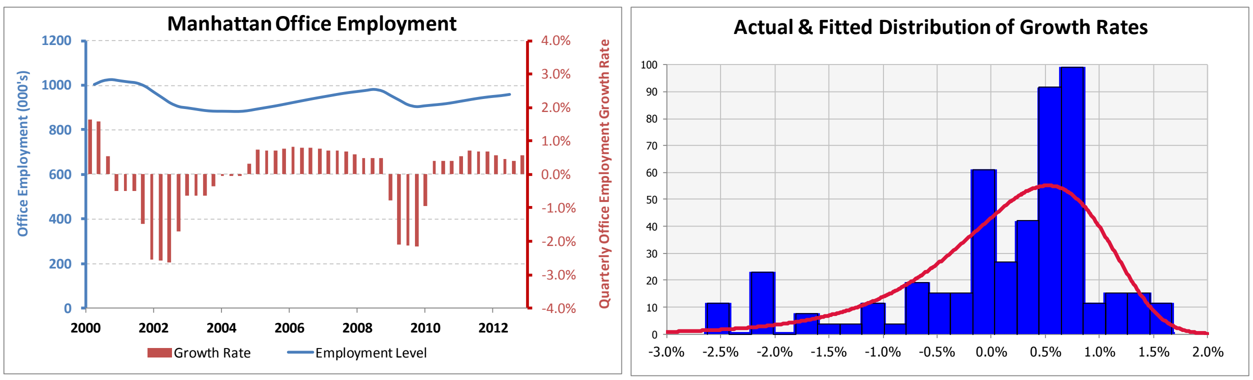
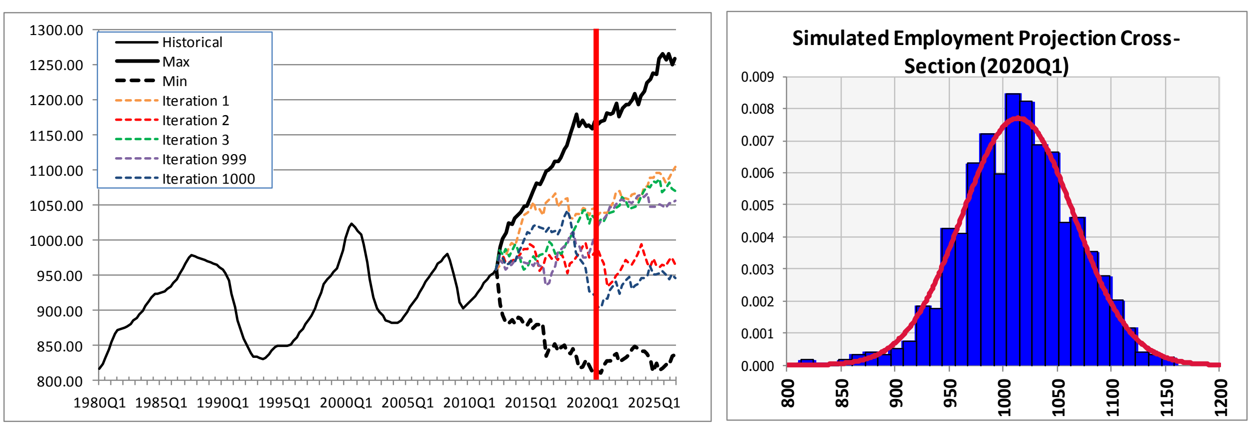


FIGURE 22 EXAMPLE OF SIMULATED FORECAST SCENARIO OUTCOMES

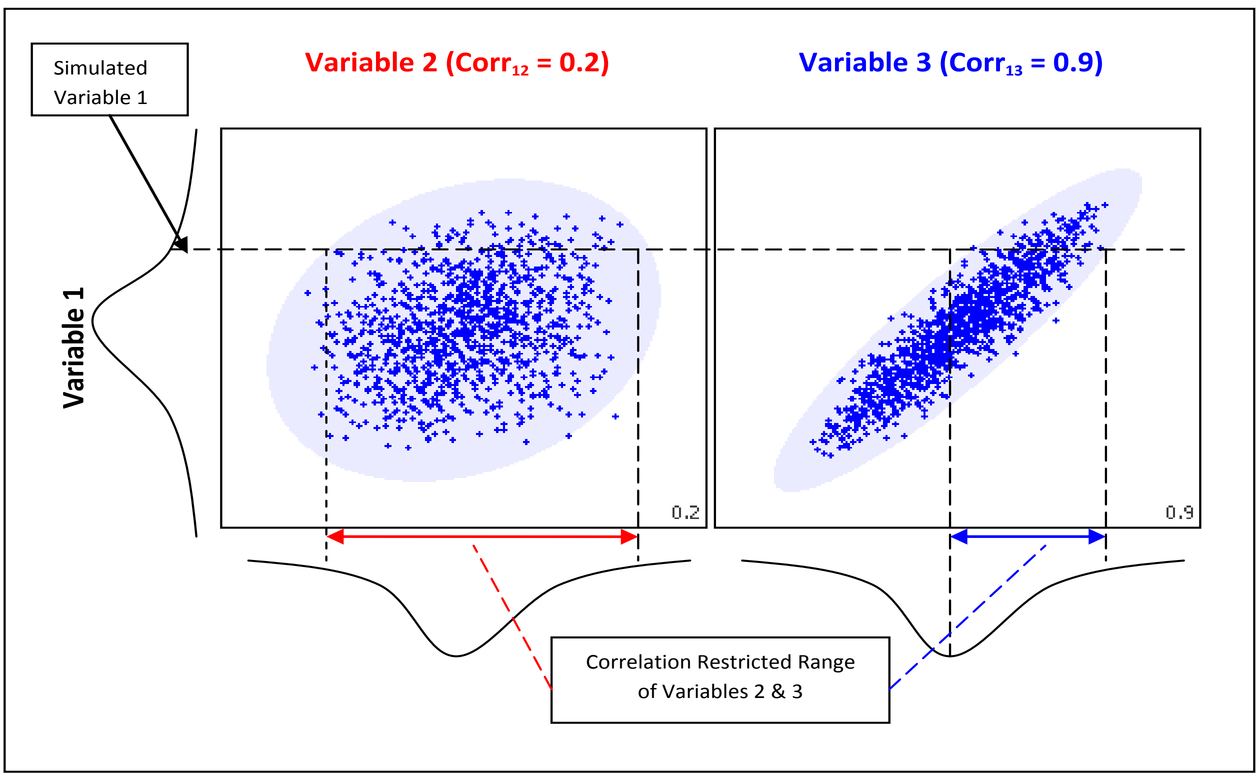


The manual simulation process uses historical growth rates of a given variable to create a sample of potential future growth rates (Figure 21) that will be applied to the variable's most current level and projected forward. Starting from the current level of the given variable, a growth rate from the defined distribution will randomly be picked using Monte Carlo/Latin Hypercube simulation techniques, and used to predict the level of the variable in the next future period – this process is repeated as many times as needed to extend through the duration of the desired forecasting horizon and compounds the effect of the randomly picked growth rates to create a single hypothetical projection of the variable. Each simulation will involve several iterations of this process (1,000 – 10,000 iterations) thereby generating a probabilistic range of future outcomes as shown in Figure 22 that traces the projected growth trajectory of select iterations and that also provides a cross sectional view revealing the distribution of the simulated employment levels in 2020Q1.

@Risk software also provides the capability to incorporate correlations between variables to ensure that the scenarios generated are consistent with economic realities. For instance, one might expect a strong positive correlation between GDP and personal income, and perhaps a weaker correlation between population and personal income. A stylized illustration of the effect of correlations in the simulation process is presented in Figure 23, which represents the scatterplots of variable 1 against 2, and variable 1 against 3. Given a simulated value of variable 1, this figure shows the effect that correlation will have in the range of potentially simulated values of Variables 2 and 3. The lower correlation between variable 1 and 2 effectively means that for a given simulated value of variable 1, there is a wider range of potential values for variable 2 than there is for variable 3.

These types of correlations, where meaningfully identified, will be incorporated through statistical tests to ensure that the range of simulated outcomes in the time-series extrapolation is constrained by real world linkages across variables.

FIGURE 23 EXAMPLE OF INPUT VARIABLE CORRELATION EFFECTS IN SIMULATION



3.1.3.2 Expert and Stakeholder Input

The results of the technical analysis described above will constitute the first part of the risk analysis based scenario development process. The FWG and the TAC will subsequently be invited to review and comment on the scenarios provided by the IHS Global Insight and the Monte Carlo simulations.

Where necessary as a primary input into the Monte Carlo simulation process, expert judgment from appropriate members of the FWG or TAC will be solicited to provide their input on likely scenarios and associated probabilities to be used in the Monte Carlo simulation process (refer to Section 3.4).

3.1.3.3 Integration with Econometric Models

The capabilities of @Risk are not limited only to the evaluation of plausible scenario ranges of input variables; the @Risk simulation procedures will be integrated into spreadsheet-based econometric forecasting tools to produce a range of potential outcomes that includes a probability estimate of likelihood. These estimates will aid in the selection of a reasonable range of high, medium and low employment forecasts.

3.2 Forecast Stability

A bottom-up, county-level forecast process helps ensure that sub-regional and county-level variations and detail within this diverse metropolitan region are accurately incorporated in the forecast results. However, disaggregated forecasting methods can also introduce substantial variability into the forecast from one period to another. This is especially problematic with econometric models for employment (which also have relationships to labor demand/supply and migration/commuting) that rely on input variables that are subject to periodic restatement of the near-term outlook and lagged adjustments based on updated reporting to historic time series data.

3.2.1 Top-Down/Bottom-Up Forecasting

Incorporating both a bottom-up and top-down approach is necessary for arriving at stable long-run forecasts. A bottom-up approach is instrumental in governing local and county-level growth patterns to so called on-the-ground realities of land use and population growth constraints. However, purely bottom-up forecasts can yield implausible regional totals, and frequently do not incorporate seasonal and cyclical effects or predictable changes in economic structure.

In contrast, national macroeconomic models are well-suited to forecasting economic growth in the context of the business cycle, even when used with shift/share analysis or other methods to allocate national growth to the 31-county region. However, the accuracy of these forecasts tends to diminish at finer spatial scales – precisely because they lack more detailed local data.

Given the inherent value and trade-offs of these methods, the following steps will be taken to ensure the proper balance between bottom-up and top-down methods. First, methodology documents accompanying GI forecasts will be reviewed to determine the steps and data involved with down-allocation to counties. Second, the population and employment models will be calibrated and reconciled at multiple spatial scales: the entire 31-county forecasts region, sub-regions, and individual counties.

3.2.2 Variability Patterns, Forecast Smoothing Techniques and Risk Adjustments

A number of techniques will be applied to ensure that the process of reviewing historical patterns in the data provided isolates the effects of seasonal or business cycle variability to reveal clearer trends that broadly reflect the true direction of changes in the data over time. Historical data will go back at least as far as 1990, with some Global Insight data starting in 1980.

- Moving averages and exponential smoothing techniques will be applied as necessary to reveal underlying trends in the data. Simple moving averages are a popular smoothing technique and help provide a relatively simple visual depiction of progression in the data.

  Exponential smoothing techniques are also a popular method used to produce smoothed time series. Whereas moving averages weight each past observation equally, exponential smoothing assigns decreasing weights to older observations in the data. Double or triple exponential smoothing (repeatedly smoothing a data series) is better for handling trends in a given data series.

- Fluctuations in the data resulting from seasonal or business cycle variability can also be addressed through the inclusion of categorical (dummy) variables in the econometric specifications of time series data. There is an added advantage of this approach from the standpoint of understanding forecasting uncertainty – in addition to isolating seasonal or business cycle impacts on the underlying trend, this approach also provides a measure of the magnitude of that impact. This measurement can be used in sensitivity or risk analysis methods discussed in Section 3.1.3 to replicate a future shock in the trend.

3.3 Land Use Considerations

Local land use policies or decision-making can impose constraints on a model's ability to most effectively predict future patterns of population, employment, labor force and household formation. Such constraints are particularly relevant in the New York metropolitan region with established population and employments centers nearing buildout capacity, and large land areas subject to environmental and other protections.

MPOs and local governments have addressed the following issues related to land use: 1) gauging capacity constraints imposed on those who reside and work in different parts of the region, 2) identifying developable and protected land and 3) reconciling county and TAZ forecasts. This paper will address all three, but will only briefly address the third item; a subsequent white paper on TAZ allocation will delve deeper into reconciling county and TAZ forecasts.

3.3.1 Land Use Capacity Issues

In developing SED forecasts, the methodology employed cannot fully account for the realities of the situation on the ground. Land use factors, such as housing availability constrain population, employment and labor force projections.

A key capacity constraint imposed on workers who live and work in different parts of region is the availability of affordable housing. In a region such as New York, where housing costs near to employment centers are significantly higher compared to those further from the CBD, employment forecasts should take into account the limits imposed by available housing. Recently, the Association of Bay Area Governments (ABAG) and the Metropolitan Transportation Commission (MTC) published the Draft Plan Bay Area Forecast of Jobs, Population, & Housing for the nine-county San Francisco Bay Area. The job forecast projected 1.1 million new jobs would be created between 2010 and 2040; this gain, however, would not translate directly into new construction of office, commercial and industrial space. Existing office space and facilities are anticipated to house the new workers. Despite the ability to provide adequate office space for the projected workers, the forecasters had to adjust the job growth forecast "based on the difficulties in supplying sufficient housing in the Bay Area to meet the needs of workforce housing within reasonable commute times."[25] Thus, had the forecasters not taken into account the "historic imbalances in the Bay Area housing market," the job growth forecast would have been overly optimistic and unconstrained by the realities of the Bay Area's housing and employment situation.

The methodology for the NYC forecast contains an approach evaluating residential growth in terms of land use. This approach – detailed earlier in Part II, Section 4 – uses two benchmarks, 2020 and 2050, to conduct the borough projections in three growth categories – high, medium and low. The categories will be determined by the amount of capacity used to account for projected populations. The high scenario, for example, assumes that all the capacity will be developed by 2050. The benchmarks differ in their estimation procedures. The 2020 benchmark will combine actual data and estimated data based on building permits issued from 2005 to 2012, as well as Certificates of Occupancy. The 2050 benchmark is a capacity analysis designed to establish potential housing unit count, based on: NYC Department of City Planning initiatives and re-zonings, large private applications, large public site projects, opportunity areas and background growth. Housing units will be converted into population, through occupancy rate and average household size being applied to housing units to create a projected population based on residential land use capacity. Migration rates will be adjusted to bring the population projections from the cohort-component model in sync with the population generated from residential land use scenario.

At the county level, these considerations provide a framework for imposing constraints on the SED forecasts. NYMTC's forecasts could similarly benefit from the recognition of the interplay between land use and population/employment growth. For population projections, a land-use based methodology will be incorporated to evaluate the results of cohort-component module, and identify any implausible growth projections at the county level. This effort will be dependent upon the level of land use and development data provided by the counties. Similar to the methodology employed for NYC, future development information will be utilized to estimate population.

3.3.2 Identification of Developable and Protected Land

Accurately assessing the extent of land that is available for development is a crucial component for population, household, employment and labor force projections. In the tri-state metropolitan region being considered, there are numerous areas protected from development, or with significant limitations on development. The Highlands Water Protection and Planning Act for instance, preserves roughly 1.1 million acres over seven New Jersey counties. The New York State Department of Environmental Conservation (NYSDEC) has also established numerous conservation easements, where restrictions are placed on properties. Additional protected areas in the region include state and local parks, federally restricted lands and other conservancy efforts.

MPOs from across the United States have employed various methods for identifying vacant and developable land, offering NYMTC key considerations in its evaluation. A GIS-based approach to this analysis is the accepted best practice across most organizations. The Chicago Metropolitan Agency for Planning (CMAP) developed base year population and employment estimates for 2010 for use in a planning horizon out to 2040. In assessing the amount of redevelopment necessary from the household forecast, CMAP used the 2005 CMAP Land Use Inventory and selected polygons of vacant or agricultural land; polygons developed since 2005 were selected from the Development Database "and 'erased' through geo-processing from the 2005 agricultural or vacant land. This yielded an estimate of available developable land in 2010 after aggregating to the township level."[26] These results were further refined to establish a ratio between the "amount of land needed to accommodate new households on available land" and the "amount of vacant land that is actually available for development."[27] A ratio larger than one indicates either larger densities for new development or redevelopment of non-vacant properties must take place. The methodology is unique to CMAP, but reveals how an MPO can use available resources to arrive at an assessment of vacant land for use in its forecast.

The Sacramento Area Council of Governments (SACOG) produced its Metropolitan Transportation Plan/Sustainable Communities Strategy (MTP/SCS) for 2035 in 2012. In each of its updates (every 4 years), SACOG prepares a regional growth forecast and land use pattern to accommodate the forecasted increases in population and employment. Through 2035, SACOG projects the region will need to develop an additional 53,000 acres of land to provide for such growth.[28] SACOG developed a "Community Types" framework to help determine the growth allocations for population, employment and housing. Local land use plans, such as general plans, specific plans, master plans, etc. were divided into these Community Types based on their location. "Developing Communities," for example, are typically situated on vacant land on the periphery of existing urban or suburban development.[29] It is these areas that are most ripe for urban expansion. One aspect, therefore, of SACOG's method to identify vacant, developable land is to use local plans as a proxy.

Some of the land identification methods outlined above will be incorporated as SED forecasts are conducted; this data will be gathered and used extensively in the development of the Zonal Allocation Program (ZAP). Primarily, a GIS-based approach will allow identification of substantial areas of vacant and protected land with the potential to constrain county levels of growth and development. This information will be used in consultations with county planners to determine the overall reasonableness of the county-level forecasts.

Using available shapefiles from departments of environmental protection, county governments and other resources, the quantity and location of both vacant and protected land will be identified. Land use data at the county level will, no doubt, vary based on differing jurisdictions. As such, the recommended approach will be adapted to have a few, broad categories to create a more uniform process. These categories could include vacant land, preserved land, developable land and developable parcels. Where data is unavailable or outdated for an area, a methodology similar to the one employed by CMAP may be appropriate and updated parcel data will serve as a proxy for estimating vacant or protected land. Local plans will be used to fill gaps in identifying vacant or protected areas.

The goal is to create an aggregated shapefile for the 31- county region, with unique layers for each county. This will allow for analysis at the county level and prime the data for the TAZ allocation process (ZAP) in Task 6. This shapefile will be constructed on a best-effort basis. The Project Team will conduct outreach to all counties in the region, specifically requesting land use and environmental data as well as data for residential, commercial and industrial developments. As the counties in New Jersey, Connecticut and mid-Hudson Valley do not have direct representation in the FWG, the Project Team will coordinate with those members who have contacts in the counties to reach individuals who can provide this information.

3.3.3 Reconciling County and TAZ Forecasts

A model's ability to produce an accurate picture of future conditions is only as effective as the realities it incorporates. As the county forecasts for population, employment, labor force and households are distributed to the TAZ and other sub-regional levels, land use constraints will have to be imposed to ensure the forecast reflects conditions on the ground. A subsequent white paper will detail the zonal allocation process and relevant constraints and considerations for reconciling the TAZ distribution process with the county forecasts for population, employment, labor force and households.

3.4 Incorporating Expert/Stakeholder Judgment and Input

The NYMTC forecasting process is a collaborative effort that engages a multitude of regional stakeholders and experts with valuable insights and perspectives. Incorporating stakeholder input into the forecasting process and the development of an effective procedure for reconciliation of quantitative and qualitative forecast inputs are an important focus of this model development process. This section highlights examples from other comparable forecasting efforts that have been reviewed to identify effective practices for soliciting and reconciling various inputs from multiple sources.

The Southern California Association of Governments (SCAG) embarked on a similar socioeconomic and demographic long term forecasts in the middle of the Great Recession (2007 – 2009). Given the uncertainty of economic conditions surrounding that time, SCAG initiated a series of three separate panel workshops in May of 2009, 2010 and 2011. These panels were intended to reduce forecasting errors (particularly in the more uncertain short term) by leveraging the collective opinions of several experts on critical factors and economic/demographic assumptions.[30] Experts were provided with a list of survey questions ahead of each panel meeting focusing on three major aspects of population and employment projections: 1) short term economic outlook; 2) long term economic assumptions (e.g. regional share of national job projections, retirement age, labor force participation rates, etc); 3) long term demographic assumptions (e.g. fertility rates, life expectancy, etc).

SCAG's structured approach to expert panel engagement is similar to the Delphi Method. The Delphi Method is a structured technique that aggregates expert opinions of likely future outcomes. The Delphi method is based on the assumption that group judgments are more accurate/valid than individual judgments, and is used to help to obtain a consensus regarding uncertain future predictions. It is important to note that lack of consensus does not imply failure, but may accurately reflect a situation in which only a wide range of impacts can be foreseen – too many variables are unknown and/or the panelists have defensibly divergent views.[31]

Other characteristics of the Delphi Method include:

- The panel consists of a diverse group of individuals;
- Each panel member carries out their own analysis;
- Each panelist's analysis is shared with rest of the panel (typically anonymously);
- Panelists have an opportunity to revise initial analysis after reviewing and hearing other panelist's findings.

Although the existing NYMTC forecasting process already includes elements of data review and consensus building from regional stakeholders, some of the features of the Delphi Method described provide the opportunity to enhance the efficacy of these efforts during FWG meetings. Given the project schedule constraints, the need for a targeted number of FWG meetings will be identified where elements of the Delphi Method will be adapted to facilitate the review of critical project input data and assumptions in a structured workshop setting as described in SCAG example. This process could be useful in identifying and obtaining consensus in the following areas:

- Future fertility/mortality rates, retirement ages, immigration rates;
- Review of IHS Global Insight data – long term economic trends and scenario ranges;
- Identification of plausible model input scenarios and probability assessments;
- Validity/accuracy of key data sources (commutation patterns etc);

- Future land use assumptions.

3.5 Coordination and Integration with NYC Forecasting Process

As discussed in the Executive Summary, a region-wide employment forecast for all 31 counties will be produced using the outside NYC forecast methodology (detailed earlier in Section 2.4). For variables other than employment, forecasts for the 26 counties outside of New York City will produced using a separate methodology from that used for the five NYC boroughs. The detailed parallel effort for household, population and labor force projections undertaken for New York City will need to be coordinated with the broader regional methods, while the results of the employment forecast using the outside NYC methodology will need to be reconciled with the results produced by the NYC methodology (described in Part II, Section 7). Descriptions of these two tasks follow.

3.5.1 Population and Household Forecasts

For the NYC forecast, a cohort-component model will be used to produce projections of population and households that is corroborated by a parallel housing unit/land use model. The forecast outside NYC will also employ a similar CC approach based on the existing 2035/2040 NYMTC model as the basis of the 2050 demographic projections.

Given regional context, boundaries, interdependencies and linkages, a parallel demographic forecast for New York City will be developed using the methodology employed for the outside NYC forecast. Data sharing and coordination during this parallel effort will provide an additional opportunity to re-evaluate the broad methodological assumptions for the region and to identify potential changes in assumptions to ensure greater consistency across the whole region. This parallel effort will also be useful in evaluating other elements of the overall forecast, such as the employment and labor force models.

Some other key areas and issues which require coordination between the NYC and outside NYC forecasts include but are not limited to the following:

- The proposed cohort-component model approach for the 2050 SED forecasts differs from the current NYMTC practice in that the new approach does not rely on separate cohort-component projections on the basis of race or ethnicity. In addition to the existence of several gaps in compiling vital statistics by race and ethnicity from the Census, NYMTC documentation also notes that race specific fertility and survival rates are not available at subnational, subregional or county levels.[32]

  The use of a single (non-race/ethnicity based) cohort-component model for each of the 31 counties in the region will be used. To derive estimates of the race and Hispanic origin composition of the forecasted population, a proportional allocation of forecast outputs will be developed by race and ethnicity, based on an evaluation of historical localized trends, current conditions, and potential future trends. This simplified approach provides a more transparent and responsive means of arriving at forecast estimates of the race and ethnic composition of the population according to current definitions.

- The methods used to account for undercounts in the 2010 Census require a coordinated approach between the methodologies' for the NYC and outside NYC region forecasts. Appropriately similar methodologies across NYMTC members will be considered to account for potential under/overcounts in the 26 counties outside of NYC to ensure a consistent approach for the entire region.

- As indicated in Section 3.3, the population and household forecasts outside of NYC will be evaluated against broad land use capacity impacts. The development of appropriate land use capacity constraints for the 26 counties outside of five New York City boroughs will be developed, drawing from the land use based methodology used for the NYC region.
- The critical balance between the net commutation assumptions for New York City's five counties and those of the remaining 26 counties in the tri-state region must be maintained. Both the previous NYMTC forecast and the proposed approach for NYC are based on county-county worker flows from the Census. The impact of any potential undercount issues on the implied commuter flows will be considered, using the recently released 2010 CTPP Package.[33]

3.5.2 Employment Forecasts

Employment and labor force forecasts for NYC are based on the reconciled outputs of two separate approaches.

- The population-based method uses the historical record of resident labor force participation rates, resident employment and unemployment and data on net commutation to project total employment levels. Total growth in employment is subsequently subject to the population growth derived from the cohort-component model.
- The economic based method first establishes base year employment using current payroll employment data. Future year projections are derived by growing the base year estimates based on anticipated regional and national trends that reference multiple sources – IHS Global Insight data, input provided from borough planners, and statewide economic outlooks.

The population approach produces estimates of employed persons, while the economic based approach generates estimates of jobs. The reconciliation of the two outputs revolves around explaining the residual of the two outputs by identifying a ratio of jobs per employed persons.

The employment forecast will primarily be based on econometric specifications using data inputs provided by IHS Global Insight. Given that outputs from this effort will serve as a primary input into the economic based projections for NYC, efforts will be coordinated with IHS Global Insight to discuss the impact of economic conditions in the construction of forecast scenarios.

IV. APPENDIX

1.0 APPENDIX A – DETAILED DISCUSSION OF NYC REGION METHODOLOGY

There are several adjustments to the population inputs that will have to be performed before the baseline component rates for NYC can be established.

Adjusting for Undercount in 1990 - The U.S. Census has been unable to count the entire population as far back as the first "modern" census in 1940. The Census Bureau has steadily worked to reduce this phenomenon in its decennial estimates using methods such as the post-enumeration survey, but a 1991 census report estimated that the undercount in 1990 was 4.86 million for the nation, or 1.86 percent. In 2000, there have been a number of estimates from newly implemented Accuracy and Coverage Evaluation (A.C.E), but all show the 2000 population with negligible levels of undercount.[34] Ultimately, the Census Bureau made no adjustments to the 1990 and 2000 population counts; the former for political reasons, the latter because the count was considered accurate.[35]

Using the 1990 and 2000 populations as starting and ending points for calculating component rates will produce inconsistent results. This is because they are invariably two different measures of the New York City population. Since the 1990 census population is actually much lower than it really was, the overall growth that is observed in the 1990's will appear much more significant than it really is. Thus, fertility rates will seem higher, survival rates will seem better, and migration rates will seem more positive. This effect will project a population that is increasing at a greater rate than the "real" components would suggest. In order to determine these "real" components, it is necessary to adjust the 1990 population to correct for the under-enumeration of nearly 250,000 people in New York City.[36]

Adjusting for Undercount in 2010 - In 2010 there was an undercount of the city's population, partly due to the Census reporting an increase in vacant units, a 46 percent rise since 2000. A disproportionate share of this increase was found in southern Brooklyn and northwest Queens. The huge concentration of vacant units in these areas cannot be explained by new construction or foreclosures; nor is it consistent with other surveys or administrative data. There was also an issue with units being deleted from Queens as well. Since this would affect projections moving forward in these two boroughs, steps have been taken to adjust the population to at least compensate for the vacant and deleted unit problem in Brooklyn and Queens.[37]

2010 Census Enumeration and DCP Adjusted Population				
	2010 Population		Difference	
	Census Enumeration	DCP Adjusted	Number	Percent
New York City	**8,175,133**	**8,242,624**	**67,491**	**0.8**
Bronx	1,385,108	1,385,108	-	-
Brooklyn	2,504,700	2,552,911	48,211	1.9
Manhattan	1,585,873	1,585,873	-	-
Queens	2,230,722	2,250,002	19,280	0.9
Staten Island	468,730	468,73		

Group Quarters Population - When considering those that live in a dormitory, nursing home, or even a prison, it is clear that their mortality, fertility, and migration patterns are entirely different with respect to the general population. It is a common practice to remove group quarters from the general population, and place them back into the population figures after the projection is completed. This method allows for those in group quarters to be included into the forecast, but under completely constant conditions. That is, the group quarters population remains the same from base year to final projection because it is subtracted from the population before the projection, and added to the population after the projection, which makes them independent of any components of change.

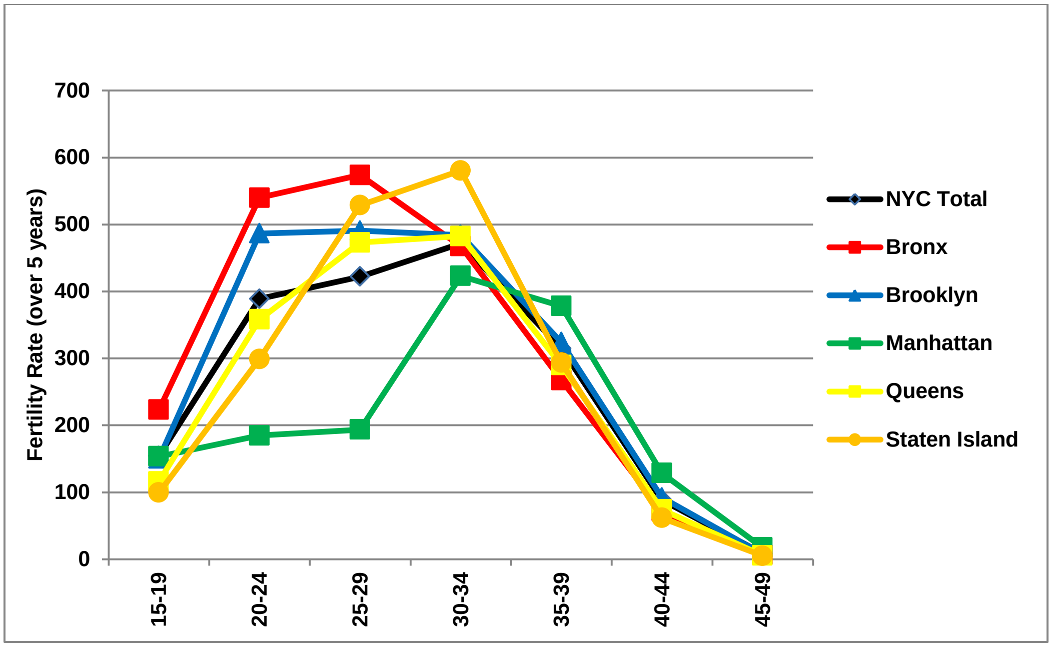
1.1 Cohort-Component Model

1.1.1 Fertility

So as not to be subject to any one-year anomaly, births were averaged for 2008, 2009, 2010, and age-specific rates were calculated based on the 2010 population.[38] Birth data were obtained from the New York City Department of Health and Mental Hygiene, while population data by age were from the decennial census. The age-specific fertility rates determine how births are distributed by age of the mother. Unlike a conventional age pattern of fertility, where child-bearing occurs primarily in the younger age groups, New York City sees a disproportionate number of births to older women. Rates rose at ages 20-24, continued to increase until 30-34, and then fell steeply thereafter (Figure 24). The two boroughs that had the highest overall fertility, the Bronx and Brooklyn, also had the highest age-specific fertility rates for those ages 20-24. Alternatively, Queens and Staten Island followed a more "suburban" fertility pattern, with moderate overall levels, and rates peaking at ages 30-34 for both boroughs. Manhattan had low overall fertility, especially for those under age 30. Unlike the case with other boroughs, women in Manhattan ages 35 to 39 had fertility rates higher than for those in their 20s, usually the prime child-bearing ages.

Since the baby boom ended in the 1960s, New York City has experienced a fairly steady decline in fertility rates that has continued to fall through the 2000s. Given these low fertility rates and the difficulties inherent in projecting fertility, 2008-2010 fertility rates will be held constant for the entire 2010-2050 period. Since the aggregate number of births is the result of both fertility rates and the number of women in the childbearing ages, the number of births may increase in some years, despite unchanging or even declining fertility rates.[39]

FIGURE 24 AGE–SPECIFIC FERTILITY RATES BY BOROUGH, 2010



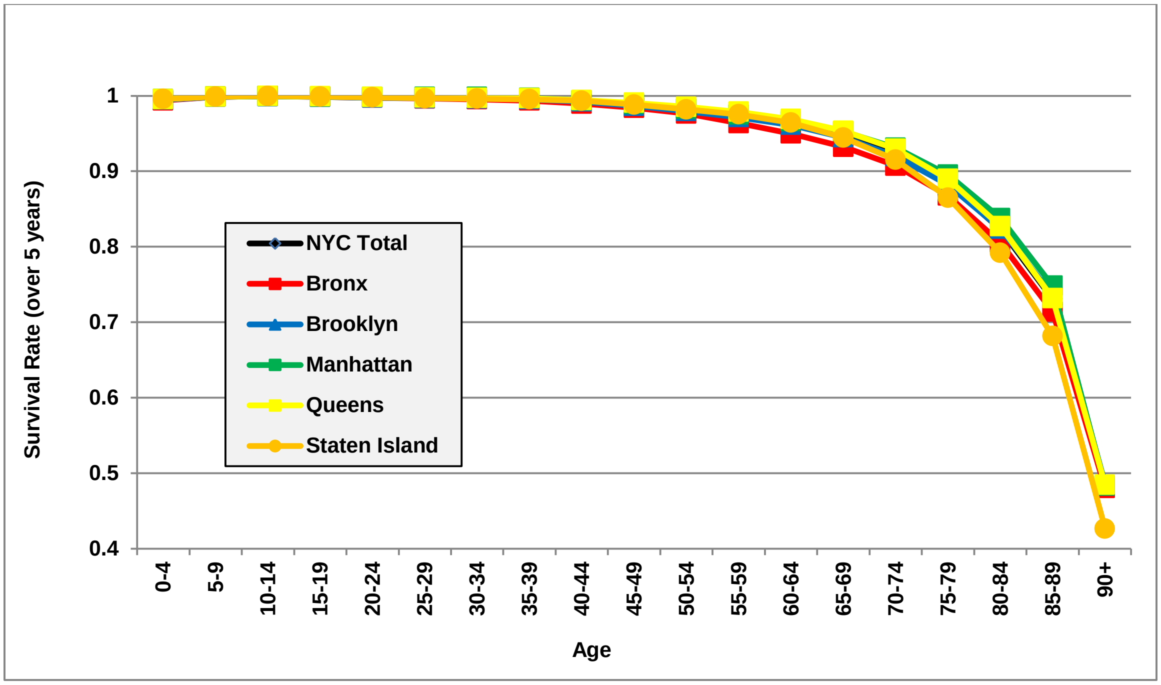
1.1.2 Mortality

As with births, deaths were averaged for 2008, 2009, and 2010 to calculate age-specific death rates based on the 2010 population. Data on deaths were obtained from the New York City Department of Health and Mental Hygiene, while population data by age were from the decennial census. These age-specific death rates were then used as the foundation for a life table that calculated survival rates by age. These rates represent the percentage of persons who are likely to survive to the next five year time point.[40] Younger age groups have much higher survival rates than older ones, but no age group is immune from death over a five year period.

The initial survival rates, which were employed for the 2010-2020 period, follow a very traditional pattern of high probability of survival for the younger ages, with very little attrition until ages 55-59 (Figure 25). Thereafter, the probability of survival begins to fall, declining steeply for the older age groups. Survival rates in the Bronx were minimally lower than those for other boroughs.

Starting in 2020-2025, survival will be increased for each group by applying age-specific improvements anticipated by The United States Social Security Administration. The ratio of increase in survival for the United States will be applied to each borough. Given the fact that the city's life expectancy in 2010 already exceeded the national average, it is unrealistic to assume that New York City's experience would continue upward at the same rate as the nation. Therefore, 50% of the difference between the boroughs and the national survival rate will be added to the borough's rate. This method will repeated at each 10 year time point, while holding the 5 year period in between constant with the decade's survival rates.

FIGURE 25 SURVIVAL RATES BY AGE AND BOROUGH, 2010



1.1.3 Migration

Since migration is the most volatile component, age-specific and crude migration rates (CMRs) were calculated using decennial census data from 1990-2000[41] and 2000-2010.

For all but the youngest age groups, net migration is derived as a residual by applying mortality rates and "surviving" an enumerated population, yielding an "expected" population for each age/sex group five or ten years later. For example, some number of 45-49 year olds in 2000 will survive to ages 55-59 in 2010. This expected population of 55-59 year olds is then compared to the observed or actual population of 55-59 year olds in the 2010 census; the difference or residual is net migration. Net migrants were divided by the initial population to create age-specific migration rates for each 5 year period, and then averaged to arrive at the rate for the entire decade.[42] Like all measures, however, this calculation is subject to error. Since the reporting of deaths by age/sex is largely complete in New York City, most of the error is likely related to coverage—undercount or overcount—of the population in 2000 or 2010.

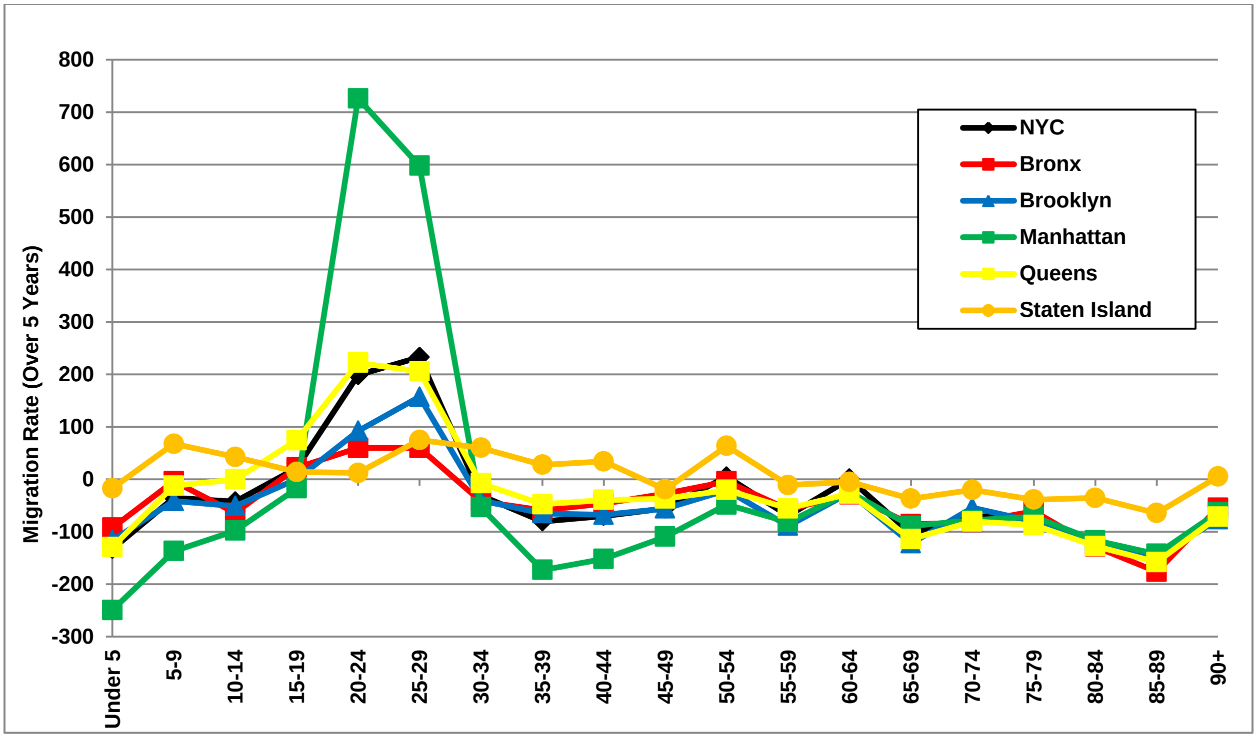
It can be very difficult to detect the impact of coverage errors. The Census Bureau evaluates coverage after every census, but rarely are estimates provided in a form that is useful for performing population adjustments. Given the recognition of substantial undercount in the 1990 Census, the Census Bureau did provide some estimates by age/sex that were applied in these projections. With some exceptions (see earlier discussion), the systematic undercount problems identified in 1990 were likely not present in the 2000 or 2010 Censuses. Still, all censuses contain errors of coverage and it is prudent to be aware of this when analyzing the rates that form the foundation of these projections.

Since it is calculated as a residual, any estimate of net migration will include errors associated with population coverage. This can manifest itself in the form of unusual changes in net migration rates among successive age/sex groups. In Brooklyn and Queens, volatility was observed in net migration rates among persons between the ages of 45 and 60, which was characterized by big shifts in the direction of rates for successive age groups. As a result, age cohorts that historically had net outflows and declined as they moved across the age spectrum were erroneously seeing net inflows and were increasing in size as they aged. These effects were dampened by averaging the rates for adjacent age groups. This has the effect of smoothing rates in the middle/upper ages, making for a scenario that was likely less sensitive to anomalies in the coverage of age/sex groups in the census enumeration.

With the exception of a small inflow for ages 15-29, the Bronx had negative migration for all of its age groups (Figure 26). The rates in Brooklyn and Queens for the young age groups were more than twice that of the Bronx. In comparison, Manhattan had an extraordinarily positive net migration rate for those 20-29. Manhattan also had among the highest rates of out-migration for most age groups thereafter. Staten Island, unlike the other boroughs, had positive migration rates for a majority of its age groups, especially those groups of people younger than 55 years old. In all of the boroughs, migration rates in the older age groups showed greater out-migration than in the younger age groups.

While overall rates of migration vary by borough, there are age-specific patterns that hold across boroughs. For example, with the exception of Staten Island, all boroughs have a net outflow for those under the age of 5, as new parents often leave for the suburbs to raise their children. In contrast, migration rates are positive for those ages 20-29 in each borough. Among those ages 30 to 54, migration rates tend to be close to zero, with the notable exception of Manhattan, which has a sizable outflow, especially for those ages 30-49. For those ages 55 to 89, migration rates are negative for all boroughs.[43]

FIGURE 26 AGE-SPECIFIC MIGRATION RATES BY BOROUGH, 1990-2010



The overall migration dynamic is captured by the CMR, which was calculated by totaling net migrants in each 5 year period and dividing that by the mid-decade population.[44] A positive CMR means that those who move into the city outnumber those who leave, while the reverse is true if the rate is negative. Migration trends for New York City from 1990-2010 are slightly negative overall, but vary by borough. The Bronx, Brooklyn, and Manhattan all had negative CMRs. While Queens also had a negative overall rate, it was not to the same degree as the other boroughs. Staten Island was the only borough to have a positive CMR.

2.0 APPENDIX B – DETAILED DISCUSSION OF NYC REGION HOUSING UNIT AND PLANNING METHODS

2.1 Establishing a 2020 Population Driven Housing Unit Baseline

The baseline population that will be produced initially by population projections will be converted into housing units by applying an average household size to the output from the cohort-component model. In addition, 2010 Group Quarters (GQ) population and vacancy rates will also be considered.

The GQ population will be removed from the total population to derive a household population number. The household population will then be divided by the average household size to obtain the number of occupied units. A vacant-to-occupied ratio will then be multiplied by the occupied units to get the number of vacant units, and then those vacant units will be added to the occupied units to derive a total housing unit count. The vacant-to-occupied ratio will then be obtained by dividing the number of vacant units by occupied units.

2.2 2010-2020 Population Projections from Housing Permits and Certificates of Occupancy

Change in the number of housing units will be separated into the following components:

A. Building Permits Issued – Provide the number of housing units listed on building permits issued by the New York City Department of Buildings, from 2005 to 2012;

B. Certificates of Occupancy – Or C of Os, issued by the New York City Department of Buildings, are linked to the building permits above, and are issued when units are completed and ready to be occupied;

C. Demolitions – estimate of the number of units in structures that were demolished, from the New York City Department of Buildings; and

D. Re-zonings & Public/Private Applications – have "build years" associated with them that can be used as a guide to compare to our building permits & C of Os. Build years are estimated completion dates, or dates when developments can be expected to come to fruition.

A 2020 housing target will be created using three inputs : 1) Certificates of Occupancy (C of Os) for 2010 to 2013[45] ; 2) demolitions for the years 2005 to 2012; and 3) the average number of building permits issued between 2005 and 2012 that was used as the estimated number of building permits for each of the years between 2014 to 2020. Permits issued between 2005 and 2012 will be examined and a net value will be created by subtracting the demolitions from the permits for each year. For each year, C of Os that are tied to the permits issued in each of the above years will be tracked, thus enabling a determination of how many units will be completed and occupied each year (for permits issued in a given year, it takes an average of six years for them to be built and occupied). The annual rate of completion/occupancy will be applied to the estimated number of building permits for 2014-2020, to obtain the estimated number of C of Os for each of these years. The actual number of C of Os for 2010 to 2013 will be summed, along with the estimated number of completions for 2014 to 2020, to obtain the estimated number of housing units built between 2010 and 2020.

The housing number will then be converted into population by using the average household size and occupancy rates for each borough. The baseline CMRs will be adjusted for 2010 to 2020 in order to bring the cohort-component population projection for each of the boroughs in line with the population based on the 2020 housing target.

2.3 Using a Land Use Analysis to Inform the 2020-2050 Population Projections

2.3.1 Land Use Scenario Used in Concert with the Cohort-Component Model

Once created, the cohort-component projections need to be evaluated against the opportunities and constraints related to changes in each borough's land use, zoning, and the production of housing. This is referred to as the planning component or a check of the reasonability of the demographic model, to ensure that the demographic projections can be accommodated by the city's housing. In general, it is this planning component that may dampen the tendency for demographic projections to show linear increases as the lack of capacity for new residential development acts to constrain growth.

The planning component, done with local knowledge from planners for each borough, consists of five categories of opportunities for future growth.

1. *DCP Initiatives and Re-zonings*: Certified/Adopted re-zonings will be taken into account since each rezoning is required to provide a number of housing units that the change will generate. The number of housing units that is generated by a rezoning is listed for each rezoning in a borough's land use scenario. These housing units will be a combination of projected and potential increases. Projected housing units have attached "build years" – units expected to be completed within the next ten years from the date of certification, which are used in the2020 Benchmark. However, the creation of potential units, which tends to occur beyond the "build year", will be factored into the 2050 benchmark.
2. *Large Private Applications*: These applications will also be accounted for, since each project has a number of units tied to its application. Examples of private applications are Atlantic Yards in Brooklyn and Flushing Commons in Queens.
3. *Large Public Site Projects*: These include Department of Housing Preservation and Development-sponsored affordable housing, New York City Economic Development Corporation projects, as well as New York City Housing Authority properties that are new or are being rehabilitated
4. *Opportunity Areas*: These are areas that could see future housing growth or may be rezoned in future years. They are mostly located along major thoroughfares and are in proximity to mass transit. While the previous three categories are more firm in their accounting of new units, Opportunity Areas are less certain regarding when and how many units will be developed. In meeting with borough planners, DCP will be able to get more insight into these areas and assign a number of potential units that may be created in future years.
5. *Background Growth*: This is growth and development of new housing irrespective of the previous four categories. It mainly consists of underutilized lots and the development of these lots through as-of-right development. Sites that fall into this category are considered "soft" since they are built to less than 50 percent of their maximum allowable floor area. Land area is converted into housing units by multiplying the square footage by 1,000.

Categories 1, 2 and 3 of the residential land use scenario require few assumptions, since all of the rezoning areas and sites are required to provide unit counts for public review as well as on their site plans, as they go through the approval process and/or become adopted and certified. These housing units are used as a guide to potential new housing units to be developed by 2050.

Categories 4 and 5 are more uncertain and assumptions need to be made in order to determine new housing unit potential, especially for areas that haven't experienced a change in zoning or public/private applications.

Opportunity Areas will be selected and examined by DCP's borough planners as areas that could see future housing development. New housing units in these areas, while not as definite as the first three categories, will be assumed to be developed based on current and future broad agency initiatives that support transit oriented development. They will also selected because of the likelihood that future rezoning will likely occur.

Background Growth[46] takes into account underutilized or vacant land located in residential zoning districts, irrespective of the other four categories. In New York City, much of the development that occurs is considered "as-of-right" development, or development that does not have to provide an application or become adopted/certified by in public review. Background Growth should not, however be seen as a block-by-block analysis, but rather a way of estimating zoned capacity in gross terms over large areas. Although zoning changes could be made that would modify future capacity, it is too speculative to make large-scale assumptions about such changes and their effects in the present analysis.

Each development or opportunity identified in the planning component will be accompanied by the projected number of housing units that could be built when the development will be completed. The number of housing units will then summed for each borough and, using current occupancy rates and average household size, a population associated with the projected number of housing units in 2050 will be derived. A check will then be performed to evaluate the population produced from the cohort component model to evaluate whether each borough will have enough housing to accommodate the projected growth in population between 2020 and 2050.

3.0 APPENDIX C – PUBLIC COMMENTS MAY 8-JUNE 6, 2014

TABLE 5. PUBLIC COMMENTS ON NYMTC 2050 WHITE PAPER ON METHODOLOGY: COUNTY LEVEL FORECASTS

	Commenter			
No.	Name	Organization	Comment	Response
1	Jack McGreevy		Kindly keep the Peconic Bay, five town region in your plans – especially the North Fork region. We could certainly use light-rails out here.	While the comment is outside of the framework of the development of the 2050 SED forecasts, it will be considered as an input in the transportation planning process.
2	Dennis Terry	161 Street BID	Does the NYMTC public notification mailing list include utility companies?	Yes, the NYMTC mailing list does include individuals at several power companies serving the region.

V. End Notes

[1] See http://rtpscs.scag.ca.gov/Documents/2008/fGrowthForecast.pdf

[2] See http://rtpscs.scag.ca.gov/Documents/2012/final/SR/2012fRTP_GrowthForecast.pdf

[3] See http://rtpscs.scag.ca.gov/Documents/2012/final/SR/2012fRTP_GrowthForecast.pdf (p. 23)

[4] See http://rtpscs.scag.ca.gov/Documents/2008/fGrowthForecast.pdf (p. 5)

[5] It is important to note that the model used to prepare the population projections for the five boroughs by age/sex through 2050 will also be used to create population projections for the 55 PUMA subareas of the city. This is in anticipation of Task 6, where population projections will be provided at the Transportation Analysis Zone (TAZ) level. While these 55 areas were designed to encompass many of the neighborhoods of New York and to approximate New York City's 59 Community Districts, they are not intended to be definitive regarding neighborhood boundaries.

[6] A cohort is simply a group of people sharing the same demographic characteristics, in this case age and sex (ex. 30-34 year old males). Those in the same cohort are subject to the same components of change, and will ultimately move through the projection together.

[7] Despite acknowledging a substantial undercount of the population, in 1991 the U.S. Commerce Department decided not to adjust the 1990 Census results.

[8] See Salvo, J.J. and A.P. Lobo (2013). "Misclassifying New York's Hidden Units as Vacant in 2010: Lessons Gleaned for the 2020 Census." *Population Research and Policy Review*, August.

[9] See Voss, P. and K, Marton (2012). "Small Populations, Large Effects: Improving the Measurement of Group Quarters Population in the American Community Survey." Washington, DC: National Academies Press.

[10] For 2013, data were available for only the first six months; this figure was doubled to obtain the total for the entire year.

[11] For permits issued in a given year, it takes an average of six years for them to be built and occupied.

[12] It is important to note that the quantification of housing unit increase is not an exact science, but an approximation of the number of housing units that may be created as a result of a borough's current land use and zoning.

[13] Swanson, David A. (2010). "The Methods and Materials Used to Generate Two Key Elements of Housing Unit Method of Population Estimation: Vacancy Rates (VR) and Persons Per Household (PPH). Department of Sociology, University of California Riverside. In addition, the ACS is not a reliable source used to predict the future trends in household size, since the data it generates are not suitable as inputs into a model. "Substantial massaging needs to be done to iron out the temporal instabilities and large variances found for person per household values in many areas."

[14] For example, each change of .01 in average household size (e.g. from 2.58 to 2.59) on a base of 3.4 million households can alter the population by 34,000 people.

[15] Grip, Richard (2008). "Does Projecting Enrollments by Race Produce More Accurate Results in New Jersey School Districts?" Paper Presented at the Annual Meeting of the Southern Demographic Association, Greenville, South Carolina, December.

[16] Keyfitz, N. (1977). Applied Mathematical Demography. New York: John Wiley & Sons.

[17] See http://www.bls.gov/emp/emplab1.htm

[18] Henderson, Richard (2012). Industry employment and output projections to 2020. *Monthly Labor Review*, January 2012, http://www.bls.gov/opub/mlr/2012/01/art4full.pdf;
DiCecio, R., Engemann, K. M., Owyang, M.T., & Wheeler, C.H. (2008). Changing trends in the labor force: A survey. *Federal Reserve Bank of St. Louis Review*, 90(1), 47-62 - http://research.stlouisfed.org/publications/review/08/01/DiCecio.pdf;
Gendell, M. (2008). Older workers: Increasing their labor force participation and hours of work. *Monthly Labor Review*, 131(1), 41-54 - http://www.bls.gov/opub/mlr/2008/01/art3full.pdf;
Hotchkiss, J.L. (2009). Changes in the aggregate labor force participation rates. *Federal Reserve Bank of Atlanta Review Economic Review*, 94(4), 1-6 - http://www.frbatlanta.org/documents/pubs/economicreview/er09no4_Hotchkiss.pdf;
Kwok, J. Daly, M., & Hobijn, B. (2010). Labor force participation and the future path of unemployment. *Federal Reserve Bank of San Francisco Economic Letter*, 2010-27, 1-5- http://www.frbsf.org/publications/economics/letter/2010/el2010-27.pdf;
Lee, M.A. & Mather, M. (2008). U.S. labor force trends. *Population Bulletin*, 63(2), 1-16 - http://www.prb.org/pdf08/63.2uslabor.pdf;
Toossi, Mitra (2012). Labor force projections to 2020: a more slowly growing workforce. Monthly *Labor Review*, January 2012, http://www.bls.gov/opub/mlr/2012/01/art3full.pdf.

[19] Quarterly Census of Employment and Wages, developed through a cooperative program between the State of New York and the US Bureau of Labor Statistics.

[20] Aggregate Confidential Data for purposes of this agreement shall mean Confidential Data sets consisting of no fewer than three (3) employers with (1) employer comprising no more than eighty percent (80%) of the total.

[21] The Fratar method is a mathematical formula used for trip distribution, using a structure that extrapolates a base year trip table to the future based on growth

[22] Lee R.D., Tuljapurkar S., Stochastic Population Forecasts for the United States: Beyond High, Medium, and Low, Statist. Ass, 89, (1994)

[23] National Cooperative Highway Research Program, Synthesis 364: Estimating Toll Road Demand and Revenue, Transportation Research Board, (2006)

[24] Billari F.C., Graziani R., Melilli E., Stochastic Population Forecasts Based on Conditional Expert Opinions, Journal of the Royal Statistical Society, (2011)

[25] Association of Bay Area Governments & Metropolitan Transportation Commission. "Bay Area Plan. Strategy for a Sustainable Region. Draft Forecast of Jobs, Population and Housing." March 2013.

[26] CMAP. "Socioeconomic Inventory Validation and Forecasting Method." June 2010 (revised January 2011).

[27] Ibid

[28] Sacramento Area Council of Governments. "Metropolitan Transportation Plan/Sustainable Communities Strategy (MTP/SCS)." April 2012.

[29] Ibid

[30] Southern California Association of Governments, Growth Forecast Draft, (December 2011).

[31] Seskin S.N., Still K.G., Boroski J., *The Use of Expert Panels in Analyzing Transportation and Land Use Alternatives*, AASHTO Standing Committee on Planning, (2002)

[32] New York Metropolitan Transportation Council, *Technical Memorandum 2040 Forecasts Modeling Methodology County/Subregional Level*, (2012)

[33] The NYMTC 2040 Forecast Methodology Technical Memorandum highlighted deficiencies in the enumeration of resident labor force that resulted in the inaccurate representation of gross in- and out-commutation flows.

[34] Among the A.C.E., the A.C.E. II, and a revised Demographic Analysis approach that accounted for a better measurement of immigration, the total undercount rates were 1.15%, -0.48%, and 0.12% respectively.

[35] As far as the projection is concerned, the 2000 count is correct. Although it is believed that there are some inconsistencies by race within the 2000 count, this is acceptable since the projection only evaluates the population by age and sex.

[36] The undercount was adjusted by the age and sex rates in 1990 PL-471 file (revised 1999) that was based in methods of Demographic Analysis, not by the Post-Enumeration Survey. All age groups with any overcount were set to zero, and redistributed so that all other age groups experienced a decrease in their percentage of undercount. The result was the same total adjustment for each borough, with only a slight modification in the age structure of the undercount. The percentages were multiplied by the population of the age/sex group, and that figure was added to the initial unadjusted population number.

[37] For more information, see Salvo, J.J. and A.P. Lobo (2013). "Misclassifying New York's Hidden Units as Vacant in 2010: Lessons Gleaned for the 2020 Census." *Population Research and Policy Review*, August.

[38] Age-specific fertility rates are based on five year age groups beginning with 15-19 year olds and ending with 45-49 year olds. All these age groups represent the "child-bearing" population. In order to calculate the actual rates, the number of births an age group produced is divided by that age group's female population. Births were based on a three year average (2008, 2009, and 2010) so that the yearly estimate of births is not subject to any one-year anomalies. All the age-specific fertility rates are multiplied by five to obtain a rate for the five year projection interval.

These age-specific rates can be summed to produce an overall measure of fertility for women in each borough, called the Total Fertility Rate (TFR). Adjusting the TFR permits a change in overall rates for women in all age groups (moving them up or down, but maintaining the overall pattern by age), while adjustments to the age-specific fertility rates permits an alteration of fertility levels for particular age groups.

[39] For example, despite the post baby-boom decline in rates, births did increase in the 1980s and peaked in 1990 at 135,000. This was still well below the level achieved in 1961, when the number of births exceeded 168,000.

[40] Survival rates are not precisely the proportion surviving to the next age group. Instead, they are calculated through a life table, which determines survival in terms of life expectancy and person years lived within an interval against all remaining intervals. This is because each age group is not subject to the exact same chances of dying, since people are distributed evenly throughout the age group. For instance, the 0-4 year old age group is not merely comprised of infants at the beginning of the projection. Rather, it includes infants, 1, 2, 3, and 4 year olds alike. This is why a simple proportion cannot be applied to each age group -- people are moving to the next age group before the five year period has concluded and are thus subject to a different probability of dying.

[41] For 1990-2000, an adjusted population was used for 1990 and the Census population was used for 2000. The undercount for 1990 was high, compared to a negligible number for 2000. The age/sex distribution of the undercounted population in New York City was not available. At the national level, undercount rates by age/sex were available through demographic analysis, so this distribution was used to make adjustments to the city's population in 1990.

[42] For each decade, migration rates were calculated for a 5 year period and then averaged to create the rate for the entire decade. The rates for 1990-2000 and 2000-2010 were then averaged to create a rate for the entire 1990-2010 period.

[43] Males and females in each borough tend to have similar patterns of migration, differing only in terms of magnitude. While this section focused on age-specific patterns of migration, age/sex-specific migration rates will be used for the migration component.

[44] The crude migration rate (CMR) is calculated by adding the estimated number of net migrants in the first portion of the relevant historical period to the net migrants in the second portion of the period, dividing by 2, then dividing that figure by the mid year population of the same period. For example, to obtain the CMR for 1990-2000, 1990-1995 net migrants would be added to the 1995-2000 net migrants, then divided by two, then divided once again by the 1995 population. A rate for 2000-2010 would be calculated in a similar manner and the rates for the two decades would then be averaged. While the age-structure of migration can be altered by changing age-specific rates, adjusting the overall CMR changes migration rates across all age groups while maintaining the overall pattern of migration by age.

[45] For 2013, since data is only for the first six months, it was doubled and assumed that total for the year.

[46] Excluded from background growth are lots that are built over 50% of its maximum capacity under current zoning. In addition, lots less than 2,500 sq ft, publically own land, lots with rent stabilized buildings, landmark buildings, historical districts, parkland and cemeteries, were all excluded from the background growth analysis. Also, in the case of Brooklyn and Queens, one and two family homes were removed, since many one and two family homes are owner-occupied and located in low-density areas, it is unlikely that these homes will be redeveloped into larger buildings with more units.