Methods

UNITED STATES CENSUS SOURCES

A number of the indicators presented in the *State of New York City's Housing and Neighborhoods* are derived from four data sources collected by the United States Census Bureau. These sources are described below along with a discussion of issues of comparability across sources.

Decennial Census (Census)

In recent decades, the Census has consisted of two parts: the 100% "short form" that collects information from every person and about every housing unit in the country, and the "long form" of additional questions asked of a sample of people and households. The "short form" collects information on age, race, Hispanic or Latino origin, household relationship, sex, tenure, and vacancy status. The "long form" provides more in-depth information about personal and housing characteristics such as income, employment status, and housing costs. In this edition of the *State of the City*, we use data from the 2000 Census short and long forms to derive demographic, economic, and housing measures for the year 2000. To create most of these indicators, we use summary census data reported at the City, borough and sub-borough area levels.

American Community Survey (ACS)

The American Community Survey is a relatively new annual survey that collects data similar to that collected by the Census "long form" described above. As with the long form, the ACS covers only a sample of individuals and housing units. However, the ACS uses a smaller sample: the Census "long form" covered 1-in-6 housing unit addresses while the ACS only covers 1-in-40 housing unit addresses each year. The Census Bureau began to work on developing the ACS in 1996, but reliable annual estimates for geographic areas with a population of 65,000 or more only became available in 2005. In December 2008, the Census Bureau began releasing 3-year rolling estimates for all geographic areas with populations of 20,000 or more. We use ACS data to generate the same statistics we obtained from the 2000 Census, but for the years 2007 and 2008. Going forward, the ACS is intended to replace the Census "long form," providing annual data that were previously available only at ten-year intervals. Most of the indicators in this edition are derived from summary-level data reported by the Census for Public Use Microdata Areas (PUMAs). A PUMA contains at least 100,000 people, and the geographic boundaries of PUMAs are almost identical to those of New York City's sub-borough areas. Summary-level data is also reported at the borough and City levels. Because each PUMA in New York City has at least 100,000 residents, reliable annual estimates are available for each PUMA from the ACS. In this edition of the *State of the City* we use annual estimates for almost all of the data we get from the ACS. One exception is the rental vacancy rate, for which we use a 3-year estimate (see the section below for more details). We also use a 3-year estimate to describe the racial composition in CDs 104, 105, 215, and 405 because 2008 data was not available for those areas.

Census and American Community Survey Microdata

In order to calculate our income diversity ratios and several indicators in our State of New Yorkers section, we determine the quintile distribution of incomes for the various geographies in New York City; to do this we use microdata (data reported at the individual- and household-level). The smallest level of geography reported in the microdata is the PUMA.

New York City Housing and Vacancy Survey (HVS)

The Housing and Vacancy Survey (HVS) is conducted every three years by the U.S. Census Bureau under contract with the City of New York. The New York City Department of Housing Preservation and Development (HPD) sponsors and supervises the HVS. The primary purpose of the HVS is to satisfy the City's statutory requirement to measure the rental vacancy rate in order to determine if rent regulations should be continued. In addition to the housing unit information, a limited set of data also is collected about the household and the individual answering the questionnaire, presumably the head of the household.

In this edition of the *State of the City*, we use HVS data to construct two indicators that are specific to New York City and therefore not captured in the ACS—the percentage of rental units that are subsidized and the percentage of rental units that are rent-regulated.

Notes on Sampling

Because both the ACS and HVS are sample surveys, not censuses, all data derived from the surveys are estimates, not exact counts. The ACS sample includes approximately 3 million housing unit addresses nationwide including about 66,000 in New York City; 18,000 housing unit addresses are sampled for the HVS. The sample for the HVS is designed primarily to achieve acceptable reliability in estimating the "vacant available for rent" rate for the entire City, so estimates for smaller geographic units such as sub-borough areas are subject to potentially large sampling errors. This report uses the convention established by HPD in cautioning the reader about any estimates that are based on 3,000 or fewer weighted observations. Readers should treat these estimates with some skepticism and be aware that the true value may differ significantly from the reported estimate.

Comparisons between the Decennial Census and American Community Survey Years

The U.S. Census Bureau makes continual adjustments to the Decennial Census and the American Community Survey to improve the coverage of the surveys and accuracy of the results. These adjustments often make cross-year comparisons difficult. Below is a discussion of the key areas where changes in sampling, question construction, or other methodology might affect the comparability of indicators that we report in the *State of the City* over time.

Income

The question construction and data collection for income information differs between the Decennial Census and the ACS. The 2000 Census asked for the respondent's 1999 income; thus incomes reported in 2000 are all for one fixed period of time (calendar year 1999). The ACS, by contrast, asks for the respondent's income over the "past 12 months" and this information is collected on an on-going monthly basis. Therefore these figures are not directly comparable. The Census Bureau notes that a comparison study of the 2000 Census income data and the 2000 ACS data found that incomes reported in the Census were about 4% higher than the incomes reported in the ACS.

Because of these data collection methods, adjacent years of ACS data may have reference months in common; thus comparisons of income data between ACS years (2007 and 2008) should not be interpreted as precise comparisons of economic conditions in those years.

Indicators affected by the income methodology issues are: Income Diversity Ratio, Median Household Income, Poverty Rate, and Poverty Rate by Age.

Note that for comparison purposes, we adjust all dollar amounts reported in this book to 2009 dollars using the Consumer Price Index for All Urban Consumers (Current Series) from the Bureau of Labor Statistics for all major expenditure classes.

Rental Vacancy Rate

Nearly two-thirds of the sub-borough areas in New York City lacked enough sample observations to calculate a rental vacancy rate for at least one year of ACS data. However, all but two of the SBAs had sufficient observations to calculate a 3-year average of the rental vacancy rate. Thus, on the community district pages, for the rental vacancy rate only, we report a 3-year average rental vacancy rate for 2006–2008. We are still reporting annual rental vacancy rates on the borough and City pages, however, the 2006–2008 average for community districts cannot be directly compared to any one year of borough or City data.

More information about comparability between U.S. Census data sources is available at: http://www.census.gov/acs/www/UseData/compACS.htm.

HOUSING PRICE APPRECIATION INDICES

The index of housing price appreciation, also called the repeat sales index, is a measure of relative change in property values over time. We construct housing price appreciation indices for four different property types (condominiums, single-family homes, 2–4 family homes, and 5+ unit apartment buildings) for New York City as a whole and for each borough. Estimating price indices separately for different types of properties allows for different market valuations and fluctuations within each property type. Due to insufficient data, we report the price indices only for the most representative building type at the community district level.

The primary data set used to construct the price index was obtained under an exclusive arrangement with the New York City Department of Finance. This data set contains information on address, price, and date of sale for all transactions involving sales of apartment buildings, condominium apartments and single- and multi-family homes in New York City between 1974 and 2009. We used roughly 239,000 pairs of sales in the estimation.

The repeat sales price indices are created using statistical regression techniques. Economists use two basic approaches to estimate housing price indices: the hedonic regression and the repeat sales methods. Both of these approaches estimate temporal price movement controlling for the variation in the types of homes sold from period to period. Each method has its own strengths and weaknesses.

The repeat-sales methodology controls for housing characteristics by using data on properties that have sold more than once. An attractive feature of this method is that, unlike the hedonic approach, it does not require the measurement of house quality; it only requires that the quality of individual houses in the sample not vary over time. The most important drawback of the repeat sales method is that it fails to use the full information available in the data. In most data sets, only a small proportion of the housing stock is sold more than once; the data on single sales cannot be used. Moreover, properties that transact more than once may not be representative of all properties in the market, raising concerns about sample selection bias. However, as the index period lengthens, more properties have changed hands more than once. This reduces sample selection bias but exacerbates a heteroskedasticity problem; Case and Shiller (1989) show evidence that price change variability is positively related to the interval of time between sales.

Most of the problems associated with the repeat sales method are overcome in this report. Specifically, the data set used here is quite large so we lose little precision by eliminating properties that sold only once. Moreover, the time period of 35 years is long enough that we capture a fairly large proportion of the housing stock. Finally, we use the three-step procedure suggested by Case and Shiller[1] (1989) and modified by Quigley and Van Order[2] (1995) to account for the possibility of time-dependent error variances.

In the first stage, the difference between the log price of the second sale and the log price of the first sale is regressed on a set of dummy variables, one for each time period in the sample (a year, in this case) except for the first. The dummy variables have values of +1 for the year of the second sale, -1 for the year of the first sale, and zeros otherwise.

In the second stage, the squared residuals from the first stage are regressed on a constant term, the time interval between sales, and the time interval squared. The fitted value in the stage-two regression is a consistent estimate of the error variance in the stage-one regression. In the third stage, the stage-one regression is re-estimated by generalized least squares, using the inverses of the square root of the fitted values from the stage-two regression as weights.

HMDA

The Federal Home Mortgage Disclosure Act (HMDA) requires financial institutions with assets totaling $31 million or more to report information on loan applications and originations. Thus, the HMDA data capture most, but not all, residential mortgage lending activity.

1 Case, K.E. and R.J. Shiller. 1989. "The Efficiency of the Market for Single Family Homes." American Economic Review, 79, p.125-37.

2 Quigley, J.M. and R. Van Order, 1995. "Explicit Tests of Contingent Claims Models of Mortgage Default." Journal of Real Estate Finance and Economics, 11(2), p.99-117.

All figures in our analysis are based on conventional, owner-occupied, 1–4 family, non business-related loans. We exclude from our analysis any government sponsored loans (such as FHA insured or VA guaranteed), any loans for properties that the owner acknowledged he or she did not occupy as a principle dwelling, any loans for manufactured or multifamily housing (5 or more families), and any loans deemed to be business related (classified as those loans for which a lender reports an applicant's ethnicity, race and sex all as "not applicable"). Conventional, owner occupied, 1–4 family, and non business-related loans constituted more than 89% of all loan applications in New York City in 2008.

HMDA requires lenders to report when the spread between the annual percentage rate (APR) of a loan and the rate of Treasury securities of comparable maturity is greater than three percentage points for first lien loans and five percentage points for junior lien loans. In this report, all loans with APRs above this threshold were referred to as high-cost loans.

Loan applicants are assigned to a racial/ethnic group for purposes of our research based on the first reported race of the primary applicant. However, if the applicant reported his or her ethnicity as "Hispanic" the applicant is classified as Hispanic, regardless of the applicant's reported race. If an applicant provided information to the lender via mail, internet or telephone and did not provide information on their race we assign those loans to the "not reported" racial category. These loans are included in our national, City, and borough level analyses, but are not included in our calculation of the racial share of new home purchase borrowers for the *State of New York City's Mortgage Lending* chapter or the *State of New Yorkers* section.

LIS PENDENS DATA (NOTICES OF FORECLOSURE)

The Furman Center collects data on *lis pendens* filings from a private vendor, Public Data Corporation (PDC). A *lis pendens* may be filed for a host of reasons unrelated to a mortgage foreclosure. The Furman Center uses a variety of screening techniques to identify only those *lis pendens* related to a mortgage. Further, if the same property received any additional *lis pendens* within 90 days of the initial *lis pendens*, the additional *lis pendens* are not included in our rate to avoid double-counting the same foreclosure.

PROPERTIES THAT ENTERED REO

The data for this indicator came from three sources—*lis pendens* from PDC, residential sales data from the New York City Department of Finance (DOF) and the Real Property Assessment Data (RPAD) from DOF. Each of these datasets identifies properties using a unique borough, block and lot number (BBL). Starting with the set of all lis pendens, we use BBLs to match each *lis pendens* (LP) issued since 1993 with the most recent sale of that property prior to the LP (if the sale happened in 1974 or later). We then match the LP to any sales that occurred within three years after the LP, and assume that the first such sale was undertaken in response to the foreclosure filing. Finally to identify transfers into REO, we search the grantee name field of the first sale after the LP for the word "bank" or the name of any large bank or subsidiary. Further, we check if the name of the grantee matches the name of the LP servicer, and if so, we classify the sale as a transfer into REO. All such transfers are included in the calculation of this indicator.

POPULATION WEIGHTING FORMULA

Several indicators included in this book are provided to us at geographic levels other than the community district level, such as police precincts or school districts. In order to make comparisons at the community district level, the Furman Center uses a population weighting formula.

For instance, when aggregating the felony crime rate from the 76 police precincts to the 59 community districts, we first calculate the rate for each of the 76 police precincts. If a community district only contains one police precinct then that rate is directly used for the community district. If a community district contains more than one police precinct, we weight the rates for each precinct based on the number of housing units within the community district that are in each precinct.

For example, if community district 1 contains three precincts A, B, and C and of the 100 housing units in community district 1, 50 are in precinct A, 30 and in

precinct B, and 20 are in precinct C the resulting formula would be:

*rate(CD1) = rate(A) * .5 + rate(B) * .3 + rate(C) * .2*

Since police precincts and community districts are not co-terminus, it is possible that the same precinct would be included in the calculation of two or more community districts. However, it would be weighted accordingly each time.

CALCULATING DISTANCE TO AMENITIES

This book presents several indicators that show the percentage of housing units within a given walking distance to amenities, such as parks and subway stations.

To determine walking distances to amenities, we use the NYC Deptartment of City Planning's LION shapefile to create a network walking buffer of streets with pedestrian right-of-ways. This method is an improvement on traditional "as the crow flies" buffers because it accounts for actual walking distance down streets and around corners. Then, for each subway entrance or park perimeter, we create a network walking buffer of every possible combination of routes emanating from each amenity. Finally, we select all lots that fall within one half-mile walking distance of any amenity.

Subway/Rail Entrances: For calculating walking distance to a subway or rail entrance, we use a database of entrances to MTA subway stations in the Bronx, Brooklyn, Manhattan, and Queens from the NYC Department of Transportation (DOT). While DOT has already geocoded most of these entrances, we supplement their work by assigning geographies for non-geocoded entrances. For the Staten Island Railway, Long Island Rail Road, Metro-North Railroad, and Amtrak, we interpolate station entrances using a variety of GIS techniques, including current satellite imagery.

Parks: Because our data on parks do not contain information on their entrances, we calculate walking distances from points along their perimeters. For parks with areas of 2.5 acres or less, we base our analysis on perimeter points of each park, generally the corners of the park. For parks larger than 2.5 acres, perimeter points (corners) are generally too far apart. Instead, we use the intersections of pedestrian right-of-ways within 150 feet to approximate their perimeters. These generally include all of the street intersections bordering the park.