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 three 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 been comprised 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 1990 and 2000 Census short and long forms to derive demographic, economic, and housing measures. To create these indicators, we use both aggregate census data (reported at the Census tract level), and micro data (available at the individual or housing unit level).

American Community Survey (ACS)

The American Community Survey is a 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. The Census Bureau began to work on developing the ACS in 1996, with reliable estimates for geographic areas with a population of 65,000 or more becoming available in 2006. In this edition of the *State of the City*, we use ACS data to generate the same statistics as the Decennial Census data, but for the years 2005 and 2006. Going forward, the ACS is intended to replace the Census "long form," providing annual data that were previously available only at ten-year intervals. Some of the indicators in this edition come from summary-level data reported by the Census for Public Use Microdata Areas (PUMAs)[1], and some are calculated by aggregating individual- and household-level micro data to the PUMA-level. A PUMA contains at least 100,000 people. The geographic boundaries of PUMAs are almost identical to those of New York City's sub-borough areas.

Comparisons between the Decennial Census and the American Community Survey

The 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. In 2000, for example, the Census Bureau included people living in group quarters in the population covered by the survey. Group quarters are places that are not classified by the Census Bureau as housing units, such as correctional facilities, nursing homes, hospitals, and other types of institutions. The group quarters population was not surveyed in either the 1990 Census or in the 2005 ACS, but it was included in the sampling frame for the 2000 Census and the 2006 ACS. Due to this difference in the sampling frame, we are limited in the comparisons we can make between these years.

Changes in the phrasing of some questions also make comparisons difficult. The income question, for example, differs on the ACS and on the decennial census. In Census years, survey respondents were asked to report their income in the previous year (i.e., 1989 and 1999), whereas the ACS asked respondents their income over the past 12 months. This slight difference appears to have led to a four percent decrease nationally in reported income during the same survey year when using the ACS method.[2] **These differences make it unadvisable to compare income data across the two data sources.**

Estimates of poverty rates are likely to be affected both by the inclusion or exclusion of the group quarters population and by the treatment of income. Therefore, the poverty rate is only presented for 2006 in the community district profiles, and comparisons between years on the borough and city pages are discouraged.

Note that for comparison purposes, we adjust all dollar amounts reported in this book to 2006 dollars, the most recent year for which income data exists from the ACS.

The calculation of some data items has also changed over time. For instance, in 2000 the rent burden was calculated from a universe of "all specified" renter-occupied housing units paying cash rent, while in the 2006 ACS, the rent burden included "all" renter-occupied housing units paying cash rent. As a result, the Census Bureau states that comparisons cannot be made.

As a result of these limitations, comparisons across years are discouraged for:

- **Population**
- **Median Household Income**
- **Median Rent Burden**
- **Poverty Rate**

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

New York City Housing and Vacancy Survey (HVS)

The 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 and household information, a limited set of data also is collected for each person in 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 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 subborough 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 2,000 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.

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 2006. Roughly 239,000 pairs of sales were used 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 method. 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

[1] Steven Ruggles, Matthew Sobek, Trent Alexander, Catherine A. Fitch, Ronald Goeken, Patricia Kelly Hall, Miriam King, and Chad Ronnander. Integrated Public Use Microdata Series: Version 3.0 [Machine-readable database]. Minneapolis, MN: Minnesota Population Center [producer and distributor], 2004. http://usa.ipums.org/usa/

2 Posey, Kirby G., Welniak, Edward, and Nelson, Charles. (2003) "Income in the American Community Survey: Comparisons to Census 2000." U.S. Census Bureau: Washington, D.C. http://www.census.gov/acs/www/Downloads/ACS/ASA_nelson.pdf

method is that, unlike the hedonic approach, it does not require the measurement of house quality; it only requires time invariance of the quality of individual houses in the sample. 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 houses 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 that we lose little precision by eliminating properties that sold only once. Moreover, the time period of 30 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 (1989) and modified by Quigley and Van Order (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 (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.

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

Goodman, A.C. and J.L. Goodman, Jr., 1997. "The Co-op Discount." *Journal of Real Estate Finance and Economics*, 14(1-2), p.223-33.

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.