Appendix: Data and Methods

New York City Housing and Vacancy Survey (HVS)

The HVS is conducted every two to 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. A wealth of other information about housing and households in New York is collected during the administration of this survey, and this information is used extensively throughout this report. In addition to the housing unit and household information, a limited set of data is also collected regarding each person in the household. This person-level data is used to calculate employment statistics such as labor force participation rate.

Because the HVS is a sample survey, not a census, all data derived from the HVS are estimates, not exact counts. These estimates are subject to statistical sampling error, which can be substantial, particularly for small areas. 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. In technical reports, estimates derived from samples are often presented in conjunction with standard errors, confidence intervals, or other indicators of statistical reliability. In order to keep this data readable, however, this type of technical information has been omitted. Instead, cautionary notes have been provided whenever an estimate is based on a small number of observations. 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.

Since the 1999 and 2002 surveys were derived from separate samples (the 1990 census and the 2000 census, respectively), it is very difficult to compare the two sets of data in a reliable manner for various reasons. First, the 2000 census included over 370,000 housing unit addresses that were missed during or added after the 1990 census was completed. Second, the weighting for the 1999 HVS was based on the 1990 Census, while the weighting for the 2002 HVS was based on the 2000 Census. Therefore, none of the 370,000 extra units were reflected in the 1999 HVS. As a result of this difference, the number of residential housing units in the 2002 HVS is substantially more than the number of units in the 1999 HVS, an increase which does not truly reflect the actual 3-year increase in units. In order to make these 2 datasets comparable, the 1999 HVS needs to be reweighted applying the weights used for the 2002 HVS. While this data is not available at this time, the Census Bureau is in the process of reweighting the data, which should be available in 2005. *Un-*

til the reweighted data is available, the State of New York City's Housing and Neighborhoods report and NYCHANIS are using the original 1999 HVS data. Treat differences between the two data sets with caution.

Another area of caution when comparing the 2002 and 1999 HVS data is that statistical tests have not been performed on these differences. Differences between two sample estimates are even more subject to sampling error than are single point estimates. Users should therefore be especially cautious when interpreting these changes over time. In particular, apparently small changes may in fact reflect no change at all. In this report, cautions have been provided whenever the difference is calculated from one or two estimates that are based on 2,000 or fewer weighted observations.

Readers who wish to learn more about the sampling errors associated with the HVS, to calculate confidence intervals associated with the estimates published in this report, or to calculate statistical test for changes over time can consult the "Sample Design, Estimation Procedure and Accuracy Statement" published by the U.S. Census Bureau for the 1999 or 2002 HVS. These documents contain information on the construction of the sampling frame for the HVS and formulas for calculating confidence intervals, and are available from the Census Bureau at http://www.census.gov/hhes/www/nychvs.html.

The sampling unit in the HVS is the housing unit, not the household or person. Both occupied and vacant housing units are included in the sample. For occupied units, information is collected from the household member that owns or rents the housing unit. For vacant units, respondents can include building superintendents, building managers, realtors, and knowledgeable neighbors.

Data from the 1999 and 2002 HVS are available throughout this report. For 1999, the HVS sample included roughly 18,000 housing units. The sample was selected from four sampling frames: 1) housing units included in the 1990 census; 2) housing units constructed since the 1990 census; 3) housing units that were converted to residential use since the 1990 census; and 4) housing units in buildings owned by the City of New York (*in rem*). Updated information on the last three sampling frames was provided by the City to the Census Bureau prior to each survey administration. The 2002 HVS, based on the 2000 census, was selected from three sampling frames: 1) housing units included in the 2000 census; 2) housing units constructed since the 2000 census; 3) housing units that were converted to residential use since the 2000 census; and 4) housing units in buildings owned by the City of New York (*in rem*).

The HVS sample is designed to be representative of the entire City of New York. Housing units are sampled throughout the City, although the exact location of each housing unit is kept confidential by the Census Bureau. The most detailed information released about the location of each unit is the sub-borough area in which it is located. Sub-borough areas are geographic units created by the Census Bureau. They are designed to approximate New York's community districts. How-

ever, since sub-borough areas are constructed from census tracts, their boundaries do not precisely coincide with community district boundaries. In addition, there are 59 community districts in New York but only 55 sub-borough areas. The (approximate) areas of four pairs of community districts were combined by the Census Bureau in creating the sub-borough areas to improve sampling and enhance the confidentiality of respondents. These pairs are Mott Haven/Melrose and Hunts Point/Longwood in the Bronx, Morrisania/Crotona and Belmont/East Tremont in the Bronx, the Financial District and Greenwich Village/Soho in Manhattan, and Clinton/Chelsea and Midtown in Manhattan.

For purposes of rent and income, 1999 dollar values from the HVS have been inflated to 2002 dollars, so that change is expressed in real dollars. To calculate the inflated value, we multiplied the 1999 value by (178.9/166.2).

Additional information about the HVS can be obtained from the Census Bureau through their HVS web site: http://www.census.gov/hhes/www/nychvs.html. HPD posts summary tables of information from the HVS on their web site: http://www.ci.nyc.ny.us/html/hpd/home.html. HPD also publishes a printed report summarizing findings from the HVS, entitled *Housing New York City*. The most recent report currently available is for the 1999 HVS.

New York City Department of City Planning

The Department of City Planning provided data on new Certificates of Occupancy issued each year from 1990 to the present. A Certificate of Occupancy is required for all newly constructed housing units. Rehabilitated housing units generally do not require a Certificate of Occupancy, unless the rehabilitation is so significant that the floor plan of the unit is changed. These data originate in the New York City Department of Buildings.

New York City Department of Housing Preservation and Development (HPD)

HPD provided two sets of data for this database that are not generally available to the public. The first data set provides information on housing rehabilitation and new construction from 1987 to the present through New York City's Capital Programs (also known as the Ten-Year Plan). These data cover over 100 separate programs that subsidize the rehabilitation and construction of affordable housing. This data covers a wide variety of housing programs, including those that assist the rehabilitation of occupied, privately owned units (most notably the Article 8-A loan program); rehabilitate vacant buildings (including *in rem* buildings) either for rental- or owner- occupancy; and construction of new affordable housing units, primarily under the Nehemiah Plan and the New York City Housing Partnership New Homes Program.

HPD also provided data on housing code complaints and violations from their internal database. The data are counts of new complaints received from the public and violations issued by HPD

housing inspectors. Any particular building or housing unit may have received multiple complaints or been issued multiple violations for any year available in this database.

Repeat Sales Indices

Repeat sales indices are a relative measure of changes in property values over time. The indices have been constructed 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. Due to insufficient data, at the community district level the price indices have been created only for the most representative building type. Estimating price indices separately for different types of properties allows for different market valuations and fluctuations within each property type.

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 family homes in New York City between 1974 and the present. A total of roughly 200,000 pairs of sales were used in the estimation.

The repeat sales price indices are created using statistical regression techniques. There are two basic approaches that are used by economists to estimate housing price indices: the hedonic regression and the repeat sales method. Both of these approaches estimate the temporal price movement controlling for the variation in the types of homes sold over time. Each method has its own strengths and weaknesses.

In the hedonic regression approach, the house sales price is regressed on a set of house characteristics (e.g., number of rooms, square footage, number of bathrooms, quality of construction, condition, number of apartments - a multi-family building characteristic, etc.), and a constant term for each time period. These intercepts – essentially the coefficients on a set of time dummy variables – reflect the movement in prices over time controlling for house characteristics. The hedonic approach requires a large quantity of data on each individual house sold. If, as usually is the case, there are variables that cannot be included, either because they are unmeasured or because data are unavailable, the coefficients on the included variables may be biased – including both the coefficients on the other house characteristics and the estimated price indices. The extent of the bias will depend upon the extent to which the omitted variables are correlated with the included variables.

The repeat-sales methodology controls for hedonic 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 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, and none of the data on single sales can 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. As the index period lengthens, more houses fit the multiple sales category. This attenuates sample selection bias but exacerbates a heteroskedasticity problem; Case and Schiller (1989) show evidence that price change variability is positively related to the interval of time between sales.

This report uses the repeat-sales method to estimate price indices. Most of the problems associated with this method are overcome in this report. Specifically, the data set used here is quite large, so that we lose little precision by eliminating observations. Moreover, the time period–27 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 Schiller (1989) and modified by Quigley and Van Order (1995) to account for the possibility of time-dependent error variances.

In the first stage, the log price per unit of the second sale minus the log price per unit of the first sale is regressed on a set of dummy variables, one for each time period (year, in this case) in the sample 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 GLS, using as weights the inverse of the square root of the fitted values in the stage-two regression.

Case, K.E. and R.J. Schiller, 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.

New York City Department of Finance (DOF)

The DOF generously provided several important data sets regarding housing In New York City. One of these is the data set of real estate transactions used to create the repeat sales price indices.

Another important data set provided by the DOF is the Real Property Assessment Data (RPAD). Although this data set is publicly available, it can be difficult to use. Each year of RPAD contains nearly 1 million records. Each record represents a unique tax lot, and provides data on geographic location, land use, building type, building size, number of residential units, property tax as-

sessment, and estimated market value, among other information. RPAD is used in this report for information on the land use of lots, and was used in conjunction with other DOF data sets to create various indicators.

The DOF also supplied data from their Open Balance File (OBF). The OBF is a continuously updated database of information concerning outstanding property taxes and other monies owed to the City related to properties (e.g., sewer charges, fines for violations, etc.). New data is added to the OBF every day, as old charges are paid and new charges are added. Different 'snapshots' from the OBF were used to create the information on property tax delinquencies. Each year of OBF data that is available in this database was obtained through an extract of OBF records at the end of the year to provide updated tax delinquency information. Each of these data sets contains information on every tax lot in the city that has delinquent property taxes, including the amounts and dates due.

Data on tax lien sales were also provided by DOF. This data was derived from a specially created data file containing information on the tax lot, amount sold, and date of sale for every tax lien sold by the city since 1996.

New York City Housing Authority (NYCHA)

NYCHA provided current counts of public housing units and section 8 vouchers. The public housing unit counts include Federal-, State-, and City- funded units.

Home Mortgage Disclosure Act (HMDA)

This annual, national data set of information on home mortgage loan applications and originations is collected by the Federal Financial Institutions Examination Council (FFIEC) under the authority of the Federal Home Mortgage Disclosure Act (HMDA). The requirements of HMDA have been amended several times over the years; currently, financial institutions with assets totaling $29 million or more are required to report data on all loan applications they process. Information contained in the data set includes lender identification, the location of the property (census tract), loan type, loan purpose, loan amount, owner-occupancy, action taken (e.g., denial of loan, origination of loan, etc.), applicant's race, applicant's sex, reason for denial, as well as demographic data regarding the property's census tract taken from the U.S. census. Because not all mortgage lenders are required to report HMDA data, the data set covers most, but not all, mortgage loan applications and originations. HMDA data is available to the public for a fee from the FFIEC. More information can be found on the FFIEC web site at http://www.ffiec.gov/hmda/default.htm.

For this report, HMDA records for census tracts were extracted from the national data. To calculate rates of mortgage lending (loans per 1,000 properties), the HMDA data were merged with estimates of the numbers of properties obtained from the RPAD data.

In order to identify loans made by subprime lenders, the HMDA data were matched to a list

of subprime lenders published by the U.S. Department of Housing and Urban Development (HUD). (This list also include lenders that specialize in manufactured homes, an area of mortgage lending that is not particularly relevant to the New York City housing market.) HUD compiled the list by examining the lending practices of lenders. Most lenders identify themselves as primarily a prime or subprime lender. For those that offer both prime and subprime loans, HUD classified the lender as a subprime lender if at least 50% of the conventional originations were subprime loans or manufactured home loans. It is not possible to identify specific subprime loans in the HMDA data, since terms of the loans are not disclosed. More information on the HUD subprime lender list can be found on the HUD web site at http://www.huduser.org/datasets/manu.html.

Public Data Corporation (PDC)

PDC is a private firm that provides searchable access to real estate legal filings in New York City (legal records from Staten Island are not available). PDC's services are most often used by real estate professionals and attorneys. Under contract with the Furman Center for Real Estate and Urban Policy, PDC conducted specialized searches of these filings in order to create several data sets for this project. *Lis pendens* filings since 1996 were extracted and matched to RPAD records (see Department of Finance information) to create the indicators of mortgage foreclosure. Information about PDC's services can be found on their web site: http://www.pdcny.com.

Lepercq, Inc.

Lepercq, Inc. is a private firm that manages tax liens sold by the City on behalf of the trust that holds the liens. Lepercq generously agreed to provide data that is not available to the public so that aggregated information about the disposition of lien sales could be included in this database.

United States Census

Census data were derived from the summary files produced by the U.S. Census Bureau from the 2000 and 1990 censuses of the United States population. Data presented in this report are 100% percent count data unadjusted for undercounting. Census data for the five boroughs were extracted from Summary File 3 (SF3, known as STF3 for the 1990 census), which presents detailed population and housing data collected from a 1-in-6 sample and weighted to represent the total population. Census data are widely available to the public in a number of formats. For more information see the Census Bureau web site at http://www.census.gov.

New York City Department of Education (DOE)

The education data related to elementary and middle school students and teachers was obtained from the DOE. This information is presented at the school district and borough levels. The

data is freely available to the public at the DOE's website: http://www.nycenet.edu.

Zip Code Business Patterns

The department of commerce provides data on business establishments at the zip code level. This data is available to the public for a fee from the United States Department of Commerce.

Tax Classes

For the purposes of this report, tax class 1 and tax class 2 may not correspond precisely with the City's definitions of class 1 and class 2 properties, in part because they do not include vacant land parcels. Using the building codes from the New York City Department of Finance, we derived the following classifications.

Class 1 buildings comprise the following building types: all single family dwellings; all 2 family dwellings; walk-up apartments with only 3 families; residential units in 1-3 story buildings that were originally built as condominiums; residential units of a 1-3 unit condominium that were originally in tax class 1; commercial units of a 1-3 unit condominium that were originally in tax class 1; primarily 1 family dwellings with two stores or offices; primarily 1 family dwellings with one store or offices; and primarily 2 family dwellings with one store or office.

Class 2 buildings comprise the following building types: all walk up apartments except those with 3 families; all elevator apartments; residential condominiums of 2-10 units that are primarily residential condominiums; residential condominiums in walk-up buildings; residential condominiums in elevator apartment buildings; commercial units of a 2-10 unit, primarily residential condominium; coops; primarily 3 family dwellings with one store or office; primarily 4 family dwellings with one store or office; primarily 5-6 family dwellings with one store or office; and primarily 1-6 family dwellings with stores or offices.

Treatment of Dollar Values

Unless specifically mentioned, the dollar values in this report are expressed in current year dollars.