A Panel of Price Indices for Housing, Other Goods, and All Goods for All Areas in the United States 1982-2008

Paul E. Carrillo
Department of Economics
George Washington University
2115 G Street, NW
Washington, DC 20052
pcarrill@gwu.edu

Dirk W. Early
Department of Economics and Business
Southwestern University
Georgetown, TX 78627
earlyd@southwestern.edu

Edgar O. Olsen*
Department of Economics
University of Virginia
Charlottesville, VA 22901
Phone: 434-924-3443
Fax: 434-982-2904
eoo@virginia.edu
*Corresponding author

The perfect is the enemy of the good – Voltaire

December 8, 2010

The authors are grateful to Bettina Aten, Brad Case, Jim Follain, Dean Frutiger, Thesia Garner, Richard Green, Joseph Gyourko, Steve Malpezzi, Frank Ptacek, Steve Reed, Kathy Short, and Randall Verbrugge for helpful conversations, information, and comments, Abby Adams, Nia Harrison, Abigail Lacker, Tony Zhang, and Yiyi Zhou for excellent research assistance, and the Bankard Fund for Political Economy at the University of Virginia for financial support.

Abstract

This paper produces a panel of price indices for housing, other produced goods, and all produced goods for each metropolitan area in the United States and the non-metropolitan part of each state from 1982 through 2008 that can be used for estimating behavioral relationships, studying the workings of markets, and assessing differences in the economic circumstances of people living in different areas. Our general approach is to first produce cross-sectional price indices for a single year 2000 and then use BLS time-series price indices to create the panel. Our geographic housing price index for 2000 is based on a large data set with detailed information about the characteristics of dwelling units and their neighborhoods throughout the United States that enables us to overcome many shortcomings of existing interarea housing price indices. For most areas, our price index for all goods other than housing is calculated from the price indices for categories of non-housing goods produced each quarter by the Council for Community and Economic Research. In order to produce a non-housing price index for areas of the United States not covered by their index, we estimate a theoretically-based regression model explaining differences in the composite price index for non-housing goods for areas where it is available and use it to predict a price of other goods for the uncovered areas. The overall consumer price index for all areas is based on the preceding estimates of the price of housing and other goods. The paper also discusses existing interarea price indices available to researchers, and it compares the new housing price index with housing price indices based on alternative methods using the same data and price indices based on alternative data sets. Electronic versions of the price indices are available online.

Keywords: Interarea price indices, interarea housing price indices, geographic cost-of-living differences, geographic price differences

JEL Codes: C8, R1, R2, R3

1. Introduction

Empirical estimates of behavioral relationships explaining how individuals will respond to changes in their circumstances are often based on data for households living in different geographic areas. In economic theory, prices play a central role in explaining individual behavior. Despite the obvious large differences in prices that prevail in different areas, few empirical studies based on data for households living in different areas include price indices for consumer goods as explanatory variables. The majority of studies make no attempt to account for geographic price differences, and most of the rest attempt to control for such differences by adding to their empirical models location fixed effects or location specific characteristics such as total population.

Recent studies have shown that the failure to account for price differences can have large effects on the conclusions of empirical studies. Moretti (2008) finds that half of the apparent increase in the return to college between 1980 and 2000 disappears when account is taken of geographic price differences. Slesnick (2005) shows that the failure to account for geographic price differences leads to severely biased estimators of the parameters of systems of demand equations. Effects on descriptive statistics are equally large. For example, Citro and Michael (1995), Short, Garner, Johnson, and Doyle (1999), Slesnick (2002), Nelson and Short (2003), Dalaker (2005), and Jolliffe (2006) find that accounting for geographic prices differences have large effects on poverty rates in different locations and noticeable effects for different demographic groups. For example, Jouliffe (2006, Table 1) finds that when poverty thresholds are not adjusted for geographic price differences, the poverty rate in non-metropolitan areas is 28 percent higher than in metropolitan areas, but when it is adjusted for them, the poverty rate is 12 percent lower in non-metropolitan areas. Dalaker (2005, Table 4) finds that the poverty rate for Hispanics is about 11 percent higher when geographic price differences are accounted for.

An important reason for the failure to account for price differences in the United States has been the absence of reliable cross-sectional price indices covering all areas of the country. The absence of a good housing price index is particularly important because housing is a large fraction of consumer spending and previous research suggests that housing prices vary much more across locations than the prices of other goods [Kokoski, Cardiff, Moulton, 1994; Citro and Michael, 1995; Aten, 2006]. The U.S. government has not produced official cross-sectional price indices since 1981 when the Bureau of Labor Statistics discontinued its series. Despite some exploratory studies on this topic since then [Primont and Kokoski, 1990; Kokoski, Cardiff, and Moulton, 1994; Moulton, 1995; Aten, 2005, 2006, 2008], the publication of an official index is not imminent. Due to ignorance of their existence and concerns about their reliability, the best privately-produced consumer price indices have been little used in economic research.

The purpose of this paper is to produce a panel of price indices for housing, other goods, and all goods for all areas of the United States from 1982 through 2008. Our general approach is to first produce cross-sectional price indices for a single year 2000 and then use BLS time-series price indices to create the panel. Unlike many previous papers, ours is not intended to contribute to the methodology for constructing consumer price indices. Instead, it uses well established methods, a new data set that is especially well suited to producing housing price indices throughout the country, and the best existing non-housing price indices to produce price indices whose use in empirical research would be much better than current practices in accounting for geographic price differences. These price indices will be useful for estimating behavioral relationships, studying the workings of markets, and assessing differences in the economic circumstances of people living in different areas.

Our geographic housing price index for 2000 is based on data on the gross rent (rent received by the landlord plus any tenant paid utilities) and numerous housing, neighborhood, and location characteristics of about 173,000 units throughout the United States. Information on the census tract of each dwelling unit makes it possible to append detailed information on its immediate neighborhood from the Decennial Census to each observation. For most areas, our price index for all goods other than housing is calculated from the price indices for categories of non-housing goods produced each quarter by the Council for Community and Economic Research, formerly the American Chambers of Commerce Research Association (ACCRA). In order to produce a non-housing price index for areas of the United States not covered by their index, we estimate a theoretically-based regression model explaining differences in the composite price index for non-housing goods for areas where it is available and use it to predict a price of other goods for the uncovered areas. The overall consumer price index for all areas is based on the preceding estimates of the price of housing and other goods.

Given the geographic information in the CSS, many alternative levels of geographic aggregation are possible. In this paper, we report separate price indices for each metropolitan area in the United States and the non-metropolitan part of each state. This paper provides a detailed account of the methods used to produce the panel of price indices and links to Excel and Stata files containing them.

The next section discusses the existing price indices available to researchers. Section 3 documents the data and methods used to produce the new cross-sectional housing price index for 2000. Section 4 reports the results. Section 5 compares this housing price index with housing price indices based on alternative methods using the same data, and Section 6 compares it with price indices based on alternative data sets and with existing indices that are often used to approximate the rental price of identical housing in different locations. Section 7 describes the methodology used to construct the price indices for other goods and all produced goods for 2000 and reports these estimates. Section 8 describes how the BLS's CPI time-series price indices for particular areas together with these cross-sectional price indices are used to produce a panel of price indices for all years between 1982 and 2008, and it discusses selected results. The paper concludes with a summary of the findings.

2. **Existing Cross-Sectional Price Indices**

Since the BLS discontinued its geographic price index in 1982, a small number of cross-sectional price indices have been produced.[1] Some of these studies produce price indices for many broad categories of goods (including housing) and an overall consumer price index. Others are devoted exclusively to producing a housing price index. None produces cross-sectional price indices covering all locations in the United States. All housing price indices suffer from either very limited geographic coverage, failure to account for many characteristics of the housing units and their neighborhoods, or errors in the prediction of the rental value of owner-occupied units.

This section describes briefly the best overall consumer price indices and housing price indices, and it discusses their advantages and disadvantages. Since our overall price index depends in part on the ACCRA price indices for non-housing goods and this is the best existing

[1] See Johnson, Rogers, and Tan (2001, pp. 32-33) for an account of the development and demise of the BLS price indices. The BLS used the data that underlies the time-series CPI to produce cross-sectional indices for 6 broad categories of goods and an overall consumer price index across 39 metropolitan areas and the non-metropolitan urban areas in four regions.

publicly-available price index produced for many places on a regular schedule, it receives special attention.

Overall Consumer Price Indices

For many years, the Council for Community and Economic Research and its predecessor ACCRA have published an overall consumer price index and six composite price indices that are expenditure-weighted averages of the price indices of 59 categories of goods.[2] ACCRA picks one narrowly defined good, for example, 160-count Kleenex brand facial tissue, to represent price differences for all goods in a category. That is, it assumes that if the particular good priced is 10 percent more expensive in one location than in another, all goods in its group are 10 percent more expensive. Price indices are produced quarterly for urban areas that account for about 70 percent of the U.S. urban population. In recent years, price indices have been produced for more than 300 urban areas.

The primary concerns about the ACCRA price indices have been the small number of price quotes in each area (5 per quarter for each good), volunteer data collectors, and expenditure weights applicable to households in the top quintile of the income distribution with a member in a professional or managerial occupation. Except for housing, the narrow definition of the goods involved ameliorates the objection based on small sample size. There is surely less variation in the prices of narrowly defined goods than more heterogeneous goods. No direct evidence shows that ACCRA's volunteers are less accurate than professionals in recording price data. ACCRA provides its volunteers with detailed instructions, and it reviews their reported prices carefully for seeming anomalies (Council for Community and Economic Research, 2006, pp. 1.4-1.5). Because ACCRA reports the individual prices that underlie its overall consumer price index, alternative expenditure weights can be used to produce an overall price index and price indices for composite commodities such as food. Koo, Phillips, and Sigalla (2000, pp. 130-131) find that replacing ACCRA's expenditure weights with weights reflecting average expenditure shares

[2] The Statistical Abstract of the United States has reported these price indices since 1990. They are a series of cross-sectional price indices rather than a panel. Since 1990, the Council has also provided the prices of the individual goods underlying the indices. Council for Community and Economic Research (2006) documents their data collection procedures and price index construction.

has very little effect on the overall price index. In section 7, we report similar results for our new price indices.

The ACCRA housing price index leaves much more room for improvement than its other price indices. The main problems are accounting for differences in housing and neighborhood characteristics and predicting the rental value of owner-occupied units. The data set underlying our housing price is much superior to the ACCRA data set in these regards, and our sample size is much larger (170,000 versus 3,000).

Accounting for the many differences in the characteristics of the dwelling unit and its neighborhood is a perennial problem in constructing an accurate cross-sectional housing price index. Differences in these characteristics lead to enormous differences in the rental value of dwelling units within a given market, and the average values of these characteristics are not the same across markets. ACCRA does account for many such differences. Its housing price index is a weighted average of a price index for homeowners and renters. For both, ACCRA controls accurately for the size of the unit and (for homeowners) the size of the parcel. To control for the condition of the unit, ACCRA prices apartments less than 10 years old whenever possible. For homeowners, it prices newly built units. A much greater attempt is made to account for amenities for homeowners than renters. For renters, ACCRA makes no direct attempt to account for amenities beyond the provision of a stove and refrigerator. To account for the many differences between units that are not directly specified, ACCRA attempts to price units occupied by managerial and professional couples in the top fifth of the income distribution. The range of differences in the overall desirability of units among this group is certainly much less than for the entire population. Nevertheless, the remaining differences in the characteristics of the structures and their neighborhoods among the units in the ACCRA sample might be significant. This makes ACCRA's small sample size in each area (5 rental and 5 owner) more problematic for housing than for other goods.

A second problem with the ACCRA housing index is prediction errors in the price index for homeowners. This is particularly important because the housing price index for homeowners accounts for 80 percent of the overall housing price index and 24 percent of the overall consumer price index. Our purpose is to produce a housing price index that compares the cost of occupying an identical unit during a year in different locations. The rents of apartments correspond exactly to this concept. For homeowners, the ideal is how much their units would rent for. ACCRA's homeownership price index can be viewed as an approximation of this ideal. ACCRA attempts to determine the sales price of very similar new houses in all locations. It then determines the level payment on a 30 year mortgage with a 25 percent downpayment at the average local mortgage interest rate. The average of these level payments across all units in a locality scaled to have a mean of 100 across all localities is the housing price index for homeowners. The question is how well these level payments reflect the rental price of the unit during its first year. Identical houses in the same location sell for the same amount whether they are purchased by a person who wants to live in the unit or rent it to someone else. The sales price of a house depends not only on its rental value during the current year (net of depreciation and operating expenses) but also its expected net rental value in future years. The ratio of current rent to sales price for identical units can be different in different locations due to different expectations about the future. For example, suppose that it is announced that a large plant will be constructed in a small community in several years. This would have an immediate effect on the sales prices of existing houses and vacant land, but it would not affect current rents.

From time to time, analysts at the BLS and BEA have used the data set underlying the time-series CPI to produce exploratory cross-sectional price indices for broad categories of goods. The studies that have produced price indices for broad categories of goods accounting for the bulk of consumer expenditure have usually developed overall consumer price indices.

Unlike ACCRA, the CPI data set is collected by professionals. It also has more individual price observations each year than ACCRA (about 1,000,000 versus 360,000) and prices many more goods (about 370 versus 59).[3] Like ACCRA, the CPI data set covers only urban areas. However, the CPI collects data from many fewer urban areas than ACCRA (88 versus more than 300), albeit selected by stratified random sampling to represent all urban areas [Moulton, 1995, pp. 183-184]. The housing information in the CPI comes from a survey of about 50,000 dwelling units. This is much larger than the ACCRA survey (about 3,000) and much smaller than ours (about 170,000). In some years, the BLS housing survey has contained owner-occupied as well as rental units. In other years, it has been limited to rental units.

[3] The ACCRA sample size is now smaller. Since 2007, it has collect data for only the first three quarters of the year. The results reported in the fourth quarter are averages of the previous three quarters.

However, in all years since 1982, its housing price index has been based in part on estimates of the market rental value of owner-occupied units [Ptacek and Baskin, 1996]. The CPI Housing Survey contains only a few rudimentary housing characteristics. BLS and BEA analysts who have used it to produce cross-sectional price indices have typically supplemented it with neighborhood characteristics from the Decennial Census. Unlike the ACCRA data set, the CPI data is not available to independent researchers. Therefore, it could not be used for this study.

The most important BLS and BEA studies are Kokoski, Cardiff, and Moulton (1994) and Aten (2005, 2006, 2008). Based on CPI data from July 1988 through June 1989, Kokoski, Cardiff, and Moulton produced price indices for 11 categories of goods for 44 areas (32 specific urban areas and all other urban areas divided into 12 categories), but did not produce an overall consumer price index. Aten (2005) used 2003 CPI data to produce price indices for eight composite commodities and an overall consumer price index for 38 areas (31 large urban areas and all other urban areas divided into 7 categories). Aten (2006) used simpler procedures to produce price indices for 2003 and 2004 at the same level of geography. Finally, Aten (2008) used the 2005 CPI data to produce an overall consumer price index for 363 metropolitan areas and 51 states. Although the methods used to construct these price indices are not described in detail, it is clear that many assumptions are involved in getting from the data used to the results. This is necessitated in large part by the limited geographic coverage of the CPI data set.

Most BLS and BEA studies use more refined price index formulas but produce price indices for many fewer separate urban areas than the ACCRA index. For the specific years and urban areas identified, the BLS and BEA price indices are almost surely better. However, with one exception, the BLS and BEA studies assign the same value for each price index to all urban areas in the same region and broad size class and hence may be less accurate than the ACCRA price indices for specific urban areas not separately identified. Aten (2008) produces price indices for all metropolitan areas. However, her assumptions and approximations are not well documented.

Koo, Phillips, and Sigalla (2000) shed light on the reliability of the ACCRA index compared with an overall price index based on CPI data, albeit in a comparison limited to 23 metropolitan areas. Specifically, they compare ACCRA's cost-of-living index with a cost-of-living index based on Kokoski, Cardiff, and Moulton (KCM)'s price indices. The KCM price indices are based on larger samples, better data collection procedures, and more refined price index formulas than the ACCRA index. When the same simple formula and expenditure weights are used to produce the cost-of-living indices and the two indices are rescaled to have the same mean, the mean of the absolute percentage deviations between the cost-of-living indices is 5.8 percent.

In addition to the preceding studies whose primary purpose is to produce price indices, some studies such as Citro and Michael (1995), Early and Olsen (2002), and Moretti (2008) have constructed crude overall consumer price indices to study particular questions. A common approach is to use an expenditure-weighted average of a crude index of housing prices combined with the assumption of no geographic variation in the prices of other goods. Others such as Slesnick (2003) have produced more refined measures at a much higher level of geography.

Housing Price Indices

The most reliable housing price indices have been produced with data from the metropolitan sample of the American Housing Survey (AHS). In the most recent detailed study, Thibodeau (1995) used these data to produce a cross-sectional housing price index that accounted for many housing and neighborhood characteristics and paid careful attention to model specification. Two major shortcomings of this price index for many purposes are its vintage and geographic coverage. It is only available for about 44 metropolitan areas (about 11 per year in each year between 1984 through 1992 with each area represented in several years). Blackley and Follain (1986), Follain and Ozanne (1979), Follain and Malpezzi (1980), Malpezzi, Ozanne, and Thibodeau (1980) and Thibodeau (1989) used AHS data and similar methods to produce housing price indices for selected metropolitan areas in earlier years. In an analysis of the usefulness of the AHS for creating house price indices, Kiel and Zabel (1997) concluded that its biggest drawback is its lack of objective information on neighborhood quality. Almost all of the information about neighborhood conditions comes from asking the respondent, and no AHS reports location for an area smaller than 100,000 units.

Normally, the BLS and BEA studies that produce price indices for many broad categories of goods or an overall consumer price index do not carefully document the methods used to create their housing price index.[4] However, in a methodological paper devoted to comparing housing price indices based on different statistical models, Moulton (1995) describes in some detail the CPI housing data and the general approach used to create the price indices in most BLS and BEA studies. Like KCM (1994), this paper produces housing price indices for 32 specific urban areas and all other urban areas divided into 12 categories. The CPI housing data has the same shortcomings as the ACCRA data, namely, limited information about housing characteristics and prediction errors in estimating the market rents of owner-occupied units. A comparison of Moulton's Table 1 with ours makes clear that the CPI data set contains many fewer housing characteristics than the CSS data set underlying our housing price index. It also contains data for many fewer geographical areas. Construction of housing price indices from the CPI data set has always involved estimating the rental value of owner-occupied units. At some times, this has been the owner's guess. At other times, it has been based on estimating a simple statistical model [Ptacek and Baskin, 1996, p. 34]. In contrast, our data set is limited to rental units.

Malpezzi, Chun, and Green (1998) have produced a housing price index for 1990 for 272 MSAs and the nonmetropolitan areas within each state based on the limited set of housing characteristics in the Decennial Census. Their hedonic equation explaining rent has 19 regressors representing only 11 rudimentary characteristics such as the number of rooms and bedrooms, the existence of complete plumbing and kitchen facilities, and the age of the structure. Dwelling units that are the same with respect to these characteristics can differ enormously in their condition, amenities, neighborhoods, and convenience to jobs, shopping, and recreation facilities. If there were differences in the mean values of these omitted characteristics across areas among units with the same values of the included characteristics, their housing price index would be biased on that account.

In addition to the preceding studies whose primary purpose is to produce cross-sectional housing price indices, some studies such as Gabriel and Rosenthal (2004), Chen and Rosenthal (2008), Moretti (2008) and Albouy (2008, 2009) have produced such indices for specific years

[4] This is understandable. Although the hundreds of goods in the CPI survey are very narrowly defined, they are not completely homogeneous. The survey collects data on differences in at least a few characteristics of each good, and the BLS and BEA analysts estimate hedonic equations for each to produce a price index for that good. So the housing hedonic equation is only one of hundreds involved in their analysis.

and places to study other issues. Not surprisingly, these are much cruder than those developed in this paper.

3. **Data and Methodology for Constructing Housing Price Indices**

Our geographic housing price index for 2000 is based on data on the gross rent (rent received by the landlord plus any tenant paid utilities) and numerous housing, neighborhood, and location characteristics of about 173,000 units throughout the United States. The primary data set is HUD's 2000 Section 8 Customer Satisfaction Survey (CSS). The CSS provides detailed information on the characteristics of the dwelling unit and tenant perceptions of its neighborhood.[5] The CSS was mailed to 280,000 families in HUD's voucher or certificate program. Families were instructed to fill out the survey and return it to HUD. The response rate was roughly 62 percent [Gray, et al., 2002]. The questionnaire asks 60 questions about the unit, the building, and the neighborhood. The pilot study indicated a very high agreement between residents and trained inspectors in answering these questions [Building Research Council, 1998]. Because the data set identifies the census tract of each dwelling unit, we are able to append data on its immediate neighborhood from the 2000 Decennial Census to each observation. HUD appended information on the gross rent of the unit (that is, the sum of the tenant's and government's payment to the landlord plus an allowance for tenant-paid utilities), the number of persons in the unit, and its location from its administrative records.

All units in the data set were occupied by families with HUD's Section 8 housing vouchers. Voucher recipients are free to occupy any unit that meets the program's standards and they can afford with the help of the voucher subsidy. Previous research has indicated that the rents paid to landlords of tenant-based voucher units are very close to the rents of unsubsidized units *with the same characteristics* [Wallace et al., 1981; Weinberg, 1982; Leger and Kennedy, 1990; ORC/Macro, 2001, Chapter V].

The joint distribution of housing and neighborhood characteristics is different for units occupied by voucher recipients and all households. Due to the program's minimum housing standards, voucher recipients do not live in the worst housing units, and the generosity of the

[5] Building Research Council (1998) describes the pilot studies that led up to the CSS survey.

voucher subsidy is not sufficient to induce them to live in the best housing. The average unit occupied by a voucher recipient is similar to the average unsubsidized rental unit in terms of its overall desirability and hence less desirable than the average dwelling unit because owner-occupied units tend to be better. On average, voucher units rent for amounts about equal to the program's Fair Market Rent [Leger and Kennedy, 1990, p. 28], the average two-bedroom FMR in April 2000 was $625 a month, and the median gross rent of all two-bedroom rental units in this year was $620 a month. Mast (2009, Exhibit 7) reports that the mean values of the answers to two broad questions about the desirability of the housing and its neighborhood are virtually identical for voucher recipients and other renters in the 2001 National American Housing Survey (AHS). Although the average desirability of owner-occupied units is greater than renter-occupied units, the differences between rental and all units with respect to the AHS measures of the overall desirability is not large. The mean values of the two measures for rental units are 25 and 35 percent of one standard deviation below the means of these measures for all units (U.S. Census Bureau, 2002, Table 2-7, 2-8). Unlike households in subsidized housing projects, voucher recipients are widely dispersed. More than 80 percent of all census tracts in the 50 largest metropolitan areas have at least one voucher recipient (Devine and others, 2003, p. 10). Voucher recipients account for more than 10 percent of all households in only 3 percent of these census tracts and more than 25 percent in almost none (Devine and others, 2003, p. ix).

The difference in the joint distribution of housing and neighborhood characteristics between voucher recipients and all households is of little consequence for our purposes, namely, to produce a single housing price index to characterize differences in housing prices across areas. Obviously, units with different combinations of characteristics have at least somewhat different relative rents across areas. A separate price index for units with each different combination of characteristics would more accurately characterize differences in housing prices across areas. Producing accurate price indices for units with each combination of characteristics would require a much larger sample. The best available evidence suggests modest differences in relative rents for units at very different points in the quality spectrum (Appendix A). Our price indices are intended for users who seek a single housing price index to characterize differences in housing prices across areas. To the extent that it is viewed as applying to a particular sector of the housing market, it is arguably most applicable to rental housing of average quality and owner-occupied housing of somewhat below average quality.

To construct rental housing price indices, we follow the well-established literature that uses hedonic regressions. Hedonic regression models explain the rent of a unit as a function of its attributes. The housing price index produced in this study is based on data on the gross rent and numerous housing, neighborhood, and location characteristics of units occupied by families with tenant-based housing vouchers throughout the United States.

The hedonic specification used to produce housing price indices in this study is:

$$\ln RENT_i = \beta_0 + \beta_1 X_{1i} + ... + \beta_{ni} X_{ni} + \alpha_1 Z_{1i} + ... + \alpha_m Z_{mi} + \nu_i \qquad (1)$$

where $RENT_i$ is the gross rent of unit *i*, the *X*s represent characteristics of the rental unit and its neighborhood broadly conceived to include proximity to jobs, the *Z*s are dummies representing m different geographic areas (one area is omitted), the α's and β's are coefficients, and ν_i is an error term with Gauss-Markov properties. Previous studies have found that this functional form fits the data particularly well, and it is the most widely used specification.[6] It is also consistent with our intention to produce a single housing price index to characterize differences in housing prices across areas.

Estimates of the α's are used to produce estimates of the price index for rental housing across areas. Specifically, if the price index is normalized to 1 in the area not represented by a *Z* in the hedonic equation, the price index for any other area *j* is e^{α_j}, the ratio of the median rent in region j to the median in the base region, conditional on any set of housing and neighborhood characteristics *X*.

In total, 122 regressors are included to describe the attributes of the unit, its neighborhood, and contract conditions. Table 1 provides descriptive statistics. Only a few regressors require explanation.

Previous research indicates that the gross rent of a dwelling unit depends importantly on contract conditions. Length of tenure is included to capture discounts normally available to long

[6] Thibodeau (1989, pp. 102-103) argues for the semi-log form based on empirical results that indicate less heteroskedasticity in the error term than from a linear form.

term tenants. Landlords may offer lower rents to tenants who create fewer problems, or the tenure discount may reflect the worse condition in unobserved respects of units that have been continuously occupied for many years. Some maintenance is normally delayed until a unit is vacant. Because housing units depreciate faster as the number of persons in a unit increases, landlords may charge more for additional persons being added to the lease to compensate for the additional depreciation. Since square footage and the total number of rooms are not available in the CSS, the number of persons divided by one plus the number of bedrooms is used as a proxy for the level of crowding in the unit.

Although the CSS contains many variables describing the unit and its neighborhood, some determinants of market rents are surely omitted. To help capture the effects of omitted characteristics, we append to the CSS data characteristics of each unit's census tract from the 2000 Decennial Census. The census tract mean travel time to work for those not working at home is included as an explanatory variable in order to capture the convenience of the location to jobs. Since low income neighborhoods may offer fewer amenities, two measures of the income distribution of the census tract are included, namely, the poverty rate and median household income. Neighborhood amenities may also vary with the racial and ethnic composition of the area. For this reason, we include the fraction of the census tract population African-American and the fraction Hispanic.[7] Because a substantial number of vacant units in an area might indicate that it is less desirable, we also include in the hedonic regression the fraction of housing units in the census tract that are vacant. The CSS does not ask about the age of the structure. Measures of the age distribution of rental units in the census tract are included as regressors to capture unobserved characteristics of the dwelling unit as well as its neighborhood. Finally, we include population density to account for the net effect of the unobserved neighborhood amenities that attract people to particular locations and the unobserved disamenities resulting from high density.

[7] It is also possible that people of different races or ethnicities may pay different prices for identical housing and the magnitudes of these differences might depend on the racial and ethnic composition of the neighborhood. We are using the CSS and Census data to explore this issue in a separate paper. For present purposes, it suffices to say that the inclusion of these two variables in the hedonic equation has virtually no effect on the resulting housing price index. The two price indices are almost proportional with a correlation coefficient of .999.

It has been suggested that we include as explanatory variables amenities and disamenities such as climate, pollution levels, and the existence of a symphony orchestra or professional football team that are common to a broad area. Rosen (1979) and Roback (1982) have shown how these factors affect land prices and wage rates and how to produce quality-of-life indices based on their models. Bloomquist, Berger, and Hoehn (1988) produce better quality-of-life indices for a much larger number of urban areas by accounting for amenity variation within large urban areas and using better data. Gyourko and Tracy (1991) recognize that local tax revenue is used to produce some local amenities and show the importance of fiscal differentials in explaining quality-of-life differentials. Albouy (2008) makes a number of important improvements in methodology that lead to more credible quality-of-life rankings. The differences in land prices and wage rates across areas that result from differences in amenities and disamenities affect the production or distribution cost of all goods consumed in the area, albeit to different extents for different goods. Therefore, these amenities and disamenities affect the prices of housing and other goods.

Our housing price index is based on a narrower definition of housing services. In our approach, an area that has a higher market rent for otherwise identical units on account of amenities that are common to a broad area is said to have a higher price per unit of housing service rather than to provide a higher quantity of housing services for each unit. This approach seems better suited to explaining differences in consumption patterns across areas such as differences in the size of housing units.

Differences in common amenities and disamenities should be taken into account in assessing differences in standards of living across areas. Nominal income divided by a standard price index is inadequate for this purpose because it reflects only differences in consumption of goods that are purchased directly. The quality-of-life literature fills this gap.

As with other surveys, some questions either are not answered or do not contain a valid response. Although few variables had missing information for more than 5 percent of the observations, roughly 50 percent of observations had missing data for at least one variable. In all analyses reported, we omit from the estimation of the hedonic regressions observations with missing data for more than 20 of the underlying variables. This removed 2,733 observations, less than 2 percent of the total. In addition, observations with unrealistic rents (less than $200 a month) were excluded in estimating the hedonic regressions. This resulted in 194 observations being dropped.[8] A common method for handling missing data is to restrict the analysis to observations with complete data, normally referred to as complete case analysis (CCA), and we report a housing price index based on this approach. However, since CCA required the omission of about half of the sample, we also produced a price index based on the estimation of a hedonic regression with omitted variable indicators in which we excluded only the 2,927 [=2,733+194] observations mentioned above. This is the housing price index against which all others are compared. To implement it, a new variable was constructed for each underlying variable with missing values that is coded 0 if the data exists, and 1 otherwise, and the variable itself is assigned a value of 0 if its value is not reported and the reported value otherwise. With the addition of the missing values indicators, the hedonic specification is:

$$\ln RENT_i = \beta_0 + \beta_1 X_{1i} + ... + \beta_{ni} X_{ni} + \gamma_1 M_{1i} + ... + \gamma_{ni} M_{ki} + \alpha_1 Z_{1i} + ... + \alpha_m Z_{mi} + \nu_i \qquad (2)$$

where the *M* represents the missing indicator variables and other variables are as defined above.

In our judgment, twenty-four metropolitan areas had insufficient sample size to estimate with much precision a separate rental housing price index. If an area had fewer than 50 observations, those observations were combined with another area. This procedure is based on the assumption that the price of housing differs little between the areas combined. The smallest metropolitan areas were combined with observations on the nonmetropolitan part of the same state. The observations for other metropolitan areas with less than 50 observations were combined with another nearby metropolitan area of similar size. The estimate of the rent of a unit for the combined areas is then used as an approximation of the price for those metropolitan areas. Delaware and Connecticut had insufficient numbers of observations for their nonmetropolitan areas to allow precise estimates for those areas. Instead, nonmetropolitan observations for Delaware were combined with those for Maryland and nonmetropolitan observations for Connecticut were combined with those for Massachusetts. The price indices derived from the combined samples are used as an approximation of the price of rental housing

[8] Including these observations had little effect on the housing price indices.

in those areas. The nonmetropolitan areas of Alaska had only 36 observations. Since no area is in close proximity to the nonmetropolitan areas of Alaska, the price of rental housing was estimated for Alaska using the small number of observations.

Units occupied by households in the voucher program are not a random sample of the population of rental units in each area whose rents are market rents.[9] As mentioned earlier, they tend to be somewhat below average with respect to their overall desirability. However, this does not necessarily lead to estimation problems. For example, since a housing voucher recipient must occupy a unit meeting minimum quality standards in order to receive a voucher subsidy, the sample will largely exclude units with certain deficiencies. Only in cases where deficiencies emerge between annual inspections will voucher units have them. However, since the CSS data include essentially all the variables involved in the housing standards, the program's minimum housing standards merely lead to a difference in the joint distribution of the observed explanatory variables included in the hedonic equation between voucher units and all units. This creates no bias in OLS estimation of the hedonic equation specified or the price indices based on this specification. The hedonic specification assumes that the percentage difference in median rent between two areas is the same for any combination of housing and neighborhood characteristics. To the extent that these percentage differences are different for units with different combinations of characteristics, no single housing price index can represent differences in housing prices across areas.

More problematic is the upper limit on the rents of the units that could be occupied under two of the three voucher programs that existed at the time of our data 2000. To the extent that these limits are binding constraints for some voucher recipients, they imply that voucher recipients, especially those who occupy housing that is the best with respect to observed characteristics, will occupy units that are worse on average than other units with the same observed characteristics. This leads to bias in OLS estimators of the hedonic equation. The extensive set of housing and neighborhood characteristics included in the hedonic regression reduces the variance in its error term and hence ameliorates this problem. However, in addition to price indices based on the standard OLS estimation of the hedonic equation, we produce a

[9] A small fraction of the rental units in each area are in subsidized housing projects. Neither the rents paid by their tenants nor the payments received by their owners are market rents.

price index based on maximum likelihood estimation of a stochastic model that accounts for this truncation, and we compare this price index with the price index based on the standard OLS estimation.

4. **Basic Housing Price Index**

Our basic cross-sectional housing price index that will be compared with many alternatives and used to construct the overall consumer price index is based on the hedonic regression model (2). The first column of Table A-1 reports the coefficient estimates and standard errors from OLS estimation of this regression model.[10] The coefficient estimates for the missing value indicators and geographic dummy variables are omitted. The first column of Table A-2 provides housing price indices for all metropolitan areas and the nonmetropolitan parts of each state, scaled so that the price is 1 in Washington, D.C. The first column of Table 2 gives the values of the rental housing price index for the ten areas with the highest, lowest, and middle housing prices based on the results of this regression. Section 7 describes the construction of the two other price indices in Table 2.

Since our purpose is the (asymptotically) unbiased prediction of the price indices for different locations, we include in the regression all variables at our disposal that are expected to affect the market rent of a dwelling unit. Failure to include these variables risks biasing the estimators of coefficients of the area dummy variables due to correlation between them and the omitted variables. Among units that are the same with respect to the variables included, the mean values of the omitted variables may be different in different locations. Because we have access to many variables that are likely to have small effects on market rent, it is not surprising that some coefficients have unexpected signs and others with the expected sign are statistically insignificant at the standard levels. Good econometric practice argues for the inclusion of all relevant variables.

Given the reason for estimating the hedonic equation and its large number of regressors, we limit our discussion of the results to a few variables. Among units that are the same in other respects, one-bedroom apartments rent for about 19 percent more than efficiencies, two-bedroom

[10] Due to their length, Tables A-1, A-2, and A-3 are posted along with the price indices and other supplementary material under the heading Price Indices at http://artsandsciences.virginia.edu/economics/facultystaff/eoo.html.

apartments for 35 percent more than efficiencies, three-bedroom apartments for 53 percent more, and each additional bedroom adds about 10 percent to rent. Living in a census tract where the mean travel time to work is 30 minutes longer reduces rent by about 10 percent. Households with an additional person per bedroom plus one pay about 14 percent more for an identical unit.

The fit of the hedonic equation was excellent ($R^2 > .8$), and the coefficients used to create the price indices were estimated with considerable precision. The estimated price indexes were consistent with popular views about differences in housing prices. Among the most expensive places to rent an apartment were San Francisco, San Jose, and Santa Cruz, California; Stamford, Connecticut; Boston, Massachusetts; and New York City and its suburbs. The least expensive places to rent tended to be nonmetropolitan parts of states and small metropolitan areas in the South. The most expensive place to rent (San Francisco) was somewhat more than three times as expensive as the least expensive (nonmetropolitan Missouri).

5. Alternative Housing Price Indices Based on CSS Data

To check the robustness of the results, alternative methods were employed to produce housing price indices based on the CSS data. This section describes these methods, and it compares the alternative price indices with the basic index. Table A-1 reports the results of the hedonic regressions and Table A-2 the corresponding price indices. For each alternative housing price index, Table 3 reports the results of OLS estimation of a linear regression of the alternative price index on the basic index, after scaling each so that its mean is 1. It also reports the mean and maximum absolute percentage difference between alternative price indices across all areas.

If the price indices were identical, the slope coefficient and coefficient of determination would be 1. The null hypothesis for testing the proportionality of the price indices on average is that the slope coefficient is 1. Because the price indices are scaled so that their means are one, the estimated constant term is one minus the estimated slope, and the test of the hypothesis that the slope is equal to one yields the same conclusion as the test of the hypothesis that the intercept is zero. For this reason, we report only the estimated slope coefficient and its standard error. Although we can reject the proportionality hypothesis at the usual levels of significance in most cases, the magnitudes of the deviations from proportionality are minuscule in all cases.

Our first alternative housing price index is based on estimating the coefficients in the regression model (2) by minimizing the sum of absolute deviations, the usual estimator of the median regression model. This tests the sensitivity of our price index to a relaxation of our assumptions about the conditional distribution of the error term in the hedonic regression and provides a reasonable alternative estimator of its parameters under the initial assumptions. Table 3 shows that the slope and coefficient of determination deviate only slightly from one, the mean absolute percentage difference between the price indices across all areas is only one percent, and the maximum absolute percentage difference is less than six percent.

Including missing value indicators allowed nearly all observations to be used in estimating the hedonic regression and constructing the basic housing price index. An alternative method is to base the housing price index on the estimation of equation (1), omitting observations with any missing values. This requires the omission of roughly half of all observations. In addition to a CCA based on the full set of variables, a CCA based on a shorter list of variables, omitting those variables with the most missing values, was also employed. The second and third rows of Table 3 report the comparisons of the price indices based on these regressions with our basic housing price index. In both comparisons, the slope and coefficient of determination deviate only slightly from one, the mean absolute percentage difference between the price indices is less than one percent, and the maximum absolute percentage difference is less than six percent.

As mentioned earlier, two of the three HUD programs of tenant-based housing assistance in 2000, the year of our data, had ceiling rents. On October 1, 1999, HUD began to phase out its old Section 8 certificate and voucher programs in favor of the new Section 8 Housing Choice Voucher Program.[11] This transition continued into 2002. About 90 percent of the households in our sample received assistance under the old housing certificate or new housing voucher program. Unlike the old voucher program, these programs have upper limits on the rent of the unit that can be occupied. To the extent that these limits are binding constraints for some voucher recipients, they imply that voucher recipients, especially those who occupy housing that is the best with respect to observed characteristics, will occupy units that are worse on average than other units with the same observed characteristics. This leads to bias in OLS estimators of

[11] Olsen (2003, pp. 400-404) describes the main features of these programs and how they affect the budget spaces of families offered these subsidies.

the hedonic equation. Since the dependent variable in the hedonic regression is a monotonic transformation of gross rent, a truncated regression specification is arguably more appropriate than the Gauss-Markov specification. Estimation of this model requires information on the ceiling rent that faced each certificate and voucher recipient in our sample. The CSS data does not include these ceiling rents. Using the approximations described in Appendix B, we estimated a hedonic regression model based on standard truncated regression assumptions [Maddala, 1983]. The fourth row of Table 3 shows that the resulting housing price index differed little from the basic index. The slope and coefficient of determination deviate by only slightly from one, the mean absolute percentage difference between the price indices across all areas is less than two percent, and the largest absolute percentage difference is fifteen percent.

Finally, Early (2006) has produced housing price indices with the CSS data based on the estimation of separate hedonic equations in each location. Precise estimation of the mean or median rent of units with specified characteristics in an area requires a substantial sample size relative to the number of characteristics involved in the hedonic regression. [12] The CSS data set has a relatively small number of observations in some metropolitan areas and the nonmetropolitan parts of some states. To retain the maximum number of observations, Early imputed the missing values of explanatory variables using Stata's imputation procedure. To increase the sample size in each area, he combined CSS data for three years and included year dummy variables in the hedonic regression. Even with these methods for expanding the sample size, 21 areas did not meet his low cutoff of 110 observations for estimation of the hedonic equation. Eighteen small metropolitan areas were combined with the nonmetropolitan part of their states and price indices were not produced for the nonmetropolitan parts of three states. Early used the resulting hedonic equations to predict median market rents of units with sample mean values of the regressors. The results reported in the last row of Table 3 indicate that this price index is highly correlated with our basic index and the indices are very close to

[12] In an earlier study, Moulton (1995) found that estimating separate regressions for different areas with the CPI sample led to poor out-of-sample predictions compared with a single regression that imposed the same coefficients on housing and neighborhood characteristics across areas. He attributed this result to small sample sizes in some areas.

proportional on average. The mean and maximum absolute percentage differences are larger than in the previous comparisons, but still small.

In summary, the results reported in this section indicate that reasonable alternative methods for producing housing price indices with the CSS data yield indices that are very similar.

6. Comparisons with Housing Price Indices Based on Different Data

Our detailed information about the housing, neighborhood, and location characteristics of a large sample of rental units throughout the country has enabled us to produce a housing price index better than existing indices. This section compares our basic housing price index with price indices based on different data (the AHS and Decennial Census) and with existing indices that have often been used to compare the rental price of identical housing in different locations (HUD's Fair Market Rent, median gross rent, and the ACCRA housing price index). Table 4 reports statistics similar to those in Table 3.

The most accurate existing housing price indices are based on the AHS because it contains by far the best information on housing characteristics among public-use data sets available on a regular basis. However, these price indices have been limited to large metropolitan areas in the AHS metropolitan samples to insure sufficient sample sizes for the estimation of separate hedonic equations in each area. We explore the accuracy of a housing price index for all areas that can be produced with the national AHS.

To create an AHS-based price index for all areas in 2000, we estimate a single hedonic regression with data from the 1999 and 2001 National AHS, combining data for the two years in order to create a sufficiently large sample for a reasonable number of metropolitan areas (64 of the 133 identified in the data set), and including in the hedonic equation dummy variables for these areas and the two years. To account for the locations of households that did not live in metropolitan areas with a sufficiently large sample size in the national AHS, we included in the hedonic regression dummy variables for all combinations of region and metropolitan status. This yields housing price indices that are the same for the non-metropolitan parts of all states in the same region and all metropolitan areas in a region that are not separately represented by a dummy variable in the hedonic equation. Otherwise, our hedonic specification follows closely the work of Thibodeau (1989, 1995). Unlike Thibodeau, we use monthly housing cost, which includes the cost of utilities, instead of contract rent. Using monthly housing costs more closely resembles the gross rent measure used in the housing price indices constructed from the CSS. The hedonic regression based on the AHS data contained 45 regressors (in addition to the location and time dummy variables) compared with the 122 in the hedonic based on our combination of CSS and decennial census data. The coefficient of determination in the AHS regression was .57 compared with .81 in the CSS regression.

The first row of Table 4 compares the price index based on the AHS data with the basic price index based on the CSS data. The AHS housing price index differs greatly from our basic housing price index. First, the results indicate significant deviation from proportionality. Specifically, the AHS index tends to be much lower than the CSS index for metropolitan areas with the highest CSS index. Second, deviations between the indices tend to be large. The mean absolute percentage difference between the price indices across all areas is about 10 percent, and the largest absolute percentage difference is more than 40 percent. The most plausible explanation for these large deviations is the necessity of combining all metropolitan areas in a region not separately represented by a dummy variable in the hedonic equation and the non-metropolitan part of all states in the same region. In some cases, the areas combined have very different housing prices for identical housing. The mean absolute percentage deviation between the price indices for the 64 metropolitan areas separately identified in the hedonic equation was less than 5 percent compared with more than 10 percent for all areas.

The second row in Table 4 compares housing price indices for the 64 metropolitan areas based on hedonic equations estimated with AHS and CSS data for these metropolitan areas alone. Because the AHS contains information for a random sample of dwelling units and considerable information about housing and neighborhood characteristics, it has been suggested that this comparison would shed light on bias in our price indices due to the non-representative nature of the CSS sample. Since our data set contains better information about the housing and its neighborhood than the AHS, a difference between these price indices would not necessarily indicate a bias in our price indices on this account. However, this comparison perhaps gives some additional reason to believe that the non-representative nature of the CSS sample has not caused any significant bias in our price indices. The results indicate that these two price indices are much more highly correlated and closer to proportional than the previous comparison.

Another important cross-sectional housing price index has been produced using the 1990 Decennial Census PUMS [Malpezzi, Chun, and Green, 1998]. Unlike the AHS, the Decennial Census PUMS provides a sufficiently large sample to estimate a hedonic equation for each area. The American Community Survey now provides the same information about housing on an annual basis for more than a million units each year. The primary shortcoming of these data sets for constructing a housing price index is their very limited information about the dwelling unit and its neighborhood. Dwelling units that are the same with respect to the characteristics available can differ enormously in their condition, amenities, neighborhoods, and convenience to jobs, shopping, and recreation facilities. Following closely Malpezzi, Chun, and Green's hedonic specification but estimating a single hedonic equation for the entire country with dummy variables for different areas, we construct a housing price index using data from the 2000 Decennial Census PUMS and compare it with our basic housing price index. The results reported in Table 4 indicate that on average these price indices are very close to proportional to each other. However, the correlation between them is less than between the alternative price indices based on the CSS, and the mean and maximum absolute percentage differences between the price indices are much larger.

Because HUD describes the Fair Market Rent in the Section 8 Housing Voucher Program as the cost of renting decent and safe housing in the private market and FMRs are available in all locations in each year, FMR is often used as a housing price index in empirical research. However, it is clear that the procedures used to produce them are not attempting to estimate the rent of identical units in different locations. At the time of our data with some exceptions, FMRs in each locality were to be set at the fortieth percentile of the rents of unsubsidized rental housing units of standard quality that were not built within the last two years and were occupied within this period.[13] The standards used to calculate FMR refer to only a few housing characteristics. Dwelling units of standard quality differ greatly in many respects. Three decades ago, Follain (1979) compared the FMR with an AHS-based housing price index for 39 large cities. We

[13] Since FY 1996, HUD has established higher FMR in many nonmetropolitan counties than would result from the application of this rule [HUD, 2007, p. 10]. These places are not included in our analysis. More recently, HUD has implemented a policy of using rents at the fiftieth percentile for areas that meet specified criteria (HUD 2000). These affect 39 metropolitan areas that account for about 27 percent of all program participants.

compare it with our housing price index across 331 metropolitan areas in 2000. Although the two price indices are highly correlated, the results in Table 4 indicate significant deviation from proportionality. Specifically, the FMR index tends to be higher than the CSS index for metropolitan areas with the highest CSS index. The mean absolute percentage deviation from our basic price index is similar to the deviation for the index based on the 2000 Decennial Census PUMS, but the largest absolute percentage deviation is much larger.

Median rent is the most widely used measure of differences in rental housing prices, especially in the popular press. This measure takes no account of differences in the average values of housing and neighborhood characteristics across areas. Table 4 indicates that the deviations of this measure from our basic housing price index are similar to the deviations for the housing price index based on the 2000 Decennial Census PUMS that accounts for a few rudimentary housing characteristics, except that there is a much greater deviation from proportionality. As might be expected, median rent understates housing prices in places where housing prices are greatest. As will be shown later, the price of housing relative to other goods is greatest on average in these places. This will lead consumers at these locations to economize at least on the space dimension of the housing bundle.

Finally, we compare our basic housing price index with the ACCRA index for the 226 metropolitan areas where it was available in 2000. As mentioned earlier, the primary deficiencies of the ACCRA index are accounting for differences in housing and neighborhood characteristics and predicting the rental value of owner-occupied units. The sixth row of Table 4 suggests that the ACCRA index is nearly proportional with our basic housing price index on average, but the correlation between the two indices is much lower than in any of the previous comparisons. The mean absolute percentage deviation from our basic price index is also much larger (except compared with the AHS), and the largest absolute percentage deviation is almost five times as large as in any of these comparisons. The maximum absolute percentage deviation is for the New York metropolitan area. According to the ACCRA index, housing prices are almost five times higher in this metropolitan area than the mean of the 226 metropolitan areas covered. According to our basic price index, the New York PMSA is 71 percent more expensive than the mean of these areas. One explanation for the difference is obvious. The people who collect the ACCRA data sometimes limit their pricing to units in certain parts of the urban area.

In 2000, ACCRA data for the New York metropolitan area was limited to Manhattan. Deleting the New York PMSA from the sample yields a price index that is far from proportional to our basic index and not highly correlated with it. This suggests that the previous finding that the ACCRA index is roughly proportional to our basic index is an artifact of an extremely implausible value of the ACCRA index for one locality.

In short, all widely used housing price indices differ from ours to some extent. For many, the differences are substantial.

7. **Construction of Price Indices for Other Goods and All Produced Goods**

Most research questions require price indices for other produced goods or all produced goods, in addition to or instead of a housing price index. For example, the demand for housing services depends not only on its price but also on the prices of other goods. Labor supply depends on the wage rate divided by an index of the prices of produced goods rather than the nominal wage rate.

Each quarter, ACCRA provides an overall cross-sectional consumer price index and price indices for most privately produced goods grouped into six categories for many areas.[14] However, its indices are not available for many other areas, and our housing price index is better than the ACCRA index in accounting for differences in housing and neighborhood characteristics and avoiding errors in predicting the rental value of owner-occupied units. Our index is also based on a much larger sample of dwelling units. This section describes how we create a price index for all non-housing goods and an overall consumer price index for 2000 based on ACCRA non-housing price indices, our housing price index, and other data.

The first step is to calculate an index of the price of all goods except housing and utilities for the places where the ACCRA index exists. To do it, we use ACCRA price indices for the four broad categories of other goods and average expenditure shares for all consumers from the Consumer Expenditure Survey (CEX). Table 5 reports our judgment about which CEX categories correspond to the four ACCRA non-housing composite commodities based on an examination of the specific goods that ACCRA prices in each broad category. It also reports the expenditure share for each broad category used by ACCRA to create its overall consumer price

[14] It also publishes the prices of the individual items used to create these price indices.

index and the CEX expenditure share for all consumers. Our price index for non-housing goods for the areas covered by ACCRA is the weighted mean of the ACCRA price indices for grocery items, transportation, health care, and miscellaneous goods using the CEX expenditure shares for all consumers as weights.

Our estimate of the price of non-housing goods for areas not covered by the ACCRA index can be justified by a simple theoretical model that recognizes that each good consumed in a locality involves some local labor and land and some imported inputs, often semi-finished or finished products. We assume that the production functions for housing services H and other goods X are Cobb-Douglas with constant returns to scale, where output depends on the quantities of local labor L, local land K, imported inputs I, and inputs F whose prices are the same at all locations. Specifically,

$$Q_H = A_H Q_L^{\alpha_{LH}} Q_K^{\alpha_{KH}} Q_I^{\alpha_{IH}} Q_F^{\alpha_{FH}} \tag{3}$$

$$Q_X = A_X Q_L^{\alpha_{LX}} Q_K^{\alpha_{KX}} Q_I^{\alpha_{IX}} Q_F^{\alpha_{FX}} \tag{4}$$

where the A's and the α's are constants.

These production functions imply the following minimum long-run average cost of production.

$$LRAC_H = (1/A_H)(P_L/\alpha_{LH})^{\alpha_{LH}}(P_K/\alpha_{KH})^{\alpha_{KH}}(P_I/\alpha_{IH})^{\alpha_{IH}}(P_F/\alpha_{FH})^{\alpha_{FH}} \tag{5}$$

$$LRAC_X = (1/A_X)(P_L/\alpha_{LX})^{\alpha_{LX}}(P_K/\alpha_{KX})^{\alpha_{KX}}(P_I/\alpha_{IX})^{\alpha_{IX}}(P_F/\alpha_{FX})^{\alpha_{FX}} \tag{6}$$

In the absence of government action, the long-run equilibrium prices of the two goods would be equal to these minimum long-run average costs.

Local government policies might affect output prices only through their effects on input prices. To account for the possibility that local government policies also create gaps between long-run equilibrium prices and minimum long-run average production cost at prevailing input prices, we assume that

$$P_H = E_H LRAC_H \tag{7}$$

$$P_X = E_X LRAC_X \tag{8}$$

where E_H and E_X are expected to be greater than or equal to 1.

Substituting (5) into (7) and (6) into (8) and taking the logarithm of both sides yields

$$\ln P_H = K_H + \ln E_H - \ln A_H + \alpha_{LH} \ln P_L + \alpha_{KH} \ln P_K + \alpha_{IH} \ln P_I + \alpha_{FH} \ln P_F \tag{9}$$

$$\ln P_X = K_X + \ln E_X - \ln A_X + \alpha_{LX} \ln P_L + \alpha_{KX} \ln P_K + \alpha_{IX} \ln P_I + \alpha_{FX} \ln P_F \tag{10}$$

where K_H and K_X are constants that depend on the α's in the respective equations. Since P_F is the same everywhere, we can rewrite (9) and (10) as

$$\ln P_H = C_H + \ln E_H - \ln A_H + \alpha_{LH} \ln P_L + \alpha_{KH} \ln P_K + \alpha_{IH} \ln P_I \tag{11}$$

$$\ln P_X = C_X + \ln E_X - \ln A_X + \alpha_{LX} \ln P_L + \alpha_{KX} \ln P_K + \alpha_{IX} \ln P_I \tag{12}$$

where C_H and C_X are constants.

If data were available on the three composite input prices and determinants of E_X and A_X, it would be possible to estimate (12) using data for locations where P_X is reported and use this estimated regression equation to predict this variable for other locations.[15] Equation (11) would be irrelevant. However, data on land prices P_K and the prices of imported inputs P_I are not readily available. To account for these unobserved input prices, we first solve (9) for P_K and

[15] The error term in this regression model stems from error terms in equations explaining $\ln E_H$ and $\ln A_H$.

substitute into (10). This yields

$$\ln P_X = [C_X - (\alpha_{KX}/\alpha_{KH})C_H] + [\ln E_X - (\alpha_{KX}/\alpha_{KH})\ln E_H] + [(\alpha_{KX}/\alpha_{KH})\ln A_H - \ln A_X]$$
$$+ [(\alpha_{LX}\alpha_{KH} - \alpha_{LH}\alpha_{KX})/\alpha_{KH}]\ln P_L + (\alpha_{KX}/\alpha_{KH})\ln P_H + [(\alpha_{IX}\alpha_{KH} - \alpha_{LH}\alpha_{KX})/\alpha_{KH})]\ln P_I \quad (13)$$

If local government policies affect output prices only through their effect on input prices $(E_H = E_X = 1)$, the terms in square brackets reflect differences in the parameters of the production functions for housing services and other goods. If there were no such differences, these terms would be zero, the coefficient of $\ln P_H$ would be 1, and the prices of the two goods would be the same in each location. In this case, our housing price index would also be a price index of non-housing goods and all goods. If there are differences in production functions for the two goods, the inclusion of the price of housing services in a regression model explaining differences in the price of other goods is useful because it captures the effect of unobserved input prices, especially land prices.

To complete the regression model, we write $\ln E_H$ and $\ln E_X$ as functions of an index of land use regulation (*regindex*), $\ln A_H$ and $\ln A_X$ as functions of climate variables (*coolingdays, heatingdays, precip*), and $\ln P_I$ as a function of the distance to the nearest metropolitan area with a population in excess of 1.5 million (*dist*), with additive error terms in each equation. Regulations might create a deviation between price and production cost, and weather might affect the output that can be produced with a given input bundle. Appendix C provides the definitions and sources of these variables. Substituting these equations into (13) and reparameterizing yields the regression model used to explain differences in the price index for other goods across areas where it was available.

$$\ln P_X = \beta_0 + \beta_1 regindex + \beta_2 \ln(coolingdays + 1) + \beta_3 \ln(heatingdays + 1)$$
$$+ \beta_4 precip + \beta_5 precip^2 + \beta_6 \ln P_L + \beta_7 \ln P_H + \beta_8 dist + \varepsilon \quad (14)$$

Table 6 reports the OLS estimates of the parameters of this model.[16] Under plausible assumptions about the underlying error terms, the error term in (14) would be correlated with $\ln P_H$ and hence OLS estimators of the β would be biased. However, since the purpose of this estimated equation is to predict the index of non-housing prices where it is not reported based on data available, this is not problem with OLS estimation.

Figure 1 depicts the predicted values and corresponding residuals for the observations used to estimate the prediction equation. It suggests no misspecification of functional form, heteroskedasticity, or outliers. If the functional form is correct, we expect the mean of the residuals to be about zero at all predicted values of $\ln P_X$. If the error term in the regression model is homoskedastic, the residuals should have about the same variance at all predicted values of $\ln P_X$. Figure 1 reveals that both are true to a remarkable extent. It also shows that essentially no deviations between predicted and observed values of the price index for other goods exceed 10 percent and relatively few exceed 5 percent.

INSERT FIGURE 1

Table 2 reports the price indices for housing, other goods, and all goods in 2000 for the ten areas with the highest, lowest, and middle housing price index. Each price index is scaled to have a mean of 1 across all locations. The price index for non-housing goods is the rescaled ACCRA index for the areas where it was available and the predicted index for other areas. The overall price index is the weighted average of the price indices for housing and other goods where the weights are the CEX expenditure shares for all consumers. Table A-3 reports these price indices for all locations. Table 2 suggests what is generally true. On average, non-housing prices are higher in areas where housing prices are higher, and the ratio of housing prices to the prices of non-housing goods are higher in areas with the highest overall CPI. The highest housing price index is three times as large as the smallest. The highest price index for other goods is only 39 percent greater than the smallest.

Since some researchers will want to use the overall consumer price index to study subsets

[16] Because the ACCRA prices for New York City refer to Manhattan, an unusually expensive part of the NYC PMSA, we treat these prices as not reported in estimating the model and predict the non-housing price index for it.

of the population, it is worthwhile to determine its sensitivity to the weights used to construct it. The ACCRA price indices are based on expenditure weights that reflect the consumption patterns of a very special subset of the population. One reason that economists have been reluctant to use the ACCRA index is that they are studying different populations with different expenditure patterns and they believe that price indices would be sensitive to these differences. As mentioned earlier, Koo, Phillips, and Sigalla (2000, pp. 130-131) have found that replacing ACCRA's expenditure weights with weights reflecting average expenditure shares has very little effect on the overall price index, albeit in a study limited to 23 metropolitan areas. The results of our study based on 380 areas supports their conclusion. When we compare an overall price index using the ACCRA expenditure shares in Table 5 with our price index based on the very different expenditure shares of all consumers from the CEX, the resulting indices are virtually identical. The correlation coefficient between the two price indices exceeds .99, the largest percentage difference between the two is less than 7 percent, and the mean absolute percentage difference is less than 2 percent.

The simple formula used to calculate our overall price index is not ideal from the viewpoint of measuring differences or changes in well-being. The ratio of an individual's income to this price index is not an index of the individual's well-being for any preferences with the standard general properties [Deaton and Muellbauer, 1980, Chapter 7; Pollak, 1989, Chapter 1].[17] In a simple world in which income is not subject to choice and there are no differences in amenities across locations and individuals face budget frontiers that are hyperplanes, an ideal price index for an individual exists if and only if the individual's indirect utility function can be written as the ratio of the individual's income to an expression involving only the prices of goods and constants. The expression in the denominator is an ideal price index. If there are differences in amenities across locations and amenities are separable from other goods in the person's utility function, then the ratio of income to the price index is an index of the well-being that results from consumption of the goods priced. An index of overall well-being would require accounting for differences in amenities across locations.

[17] Furthermore, when it exists at all, an ideal price index is different for each person. We ignore this complication.

To get some sense of whether moving towards an ideal price index would yield very different results, we develop an ideal overall price index based on a simple assumption about preferences and compare it with our price index. The ideal price index is based on the assumption that all people have a Cobb-Douglas utility function involving two goods housing and non-housing with exponents equal to the expenditure shares that underlie the previous overall price index. The formula for this price index is:

$$CPI = (PH/.252)^{.252}(PX/.748)^{.748} \tag{3}$$

After rescaling this ideal price index to have the same mean as the simple expenditure weighted average of the housing price index *PH* and the price index of other goods *PX* in Table A-3, the price indices are almost identical. The correlation coefficient exceeds .999, the mean absolute percentage difference is less than three-tenths of a percent, and the maximum absolute percentage difference is 2.3 across the 380 locations.

8. **Construction of Price Indices for Other Years**

To this point, we have described how we developed interarea price indices for a single year. Most applications require cross-sectional price indices for some other year or a panel of price indices. This section describes how we use the best available time-series price indices for different areas to generate a panel of price indices for 1982 through 2008 from our cross-sectional price indices. A major advantage of this approach is that the panel can be easily expanded forward and backward in time. The entire panel of prices is available as an Excel and a Stata file at http://artsandsciences.virginia.edu/economics/facultystaff/eoo.html under the heading Price Indices.[18]

Like Moretti (2008) and Slesnick (2002, 2005), we use BLS time-series price indices to create a panel of price indices from our cross-sectional prices. For quite some time, the BLS has produced time-series price indices for groups of goods and all goods combined for specific

[18] Our suggested citation is CEOPricesPanel02.

metropolitan areas and groups of urban areas based on region and population. [19] Almost all of our metropolitan areas fit unambiguously into one of these categories. Seventy nine of our MSA or PMSA are within the 27 BLS metropolitan areas. For our remaining metropolitan areas, we use the BLS price indices for the relevant population size category in its region. Finally, we use the BLS price indices for the smallest population size category in a region for the nonmetropolitan part of each state in that region, except for Alaska and Hawaii. For non-metropolitan Alaska, we use the BLS indices for Anchorage. For non-metropolitan Hawaii, we use their indices for Honolulu. The BLS does not provide time-series price indices for Phoenix from 1982 through 2001, Tampa from 1982 through 1986, or Washington-Baltimore from 1982 through 1996, for the urban areas in each region with populations between 50,000 and 1,500,000 that are not specifically identified prior to 1998, or for rural areas. Table 7 describes how we handled these cases.

The BLS does not produce time-series price indices for our categories of goods, namely, shelter and utilities combined and all other goods as a group. With a minor exception, we apply their methods and weights to produce these indices [U.S. Bureau of Labor Statistics, 2010, Chapter 17].[20] With a trivial exception, our time-series price indices are exactly the same as theirs would be if they had produced indices for these composites. First, we use BLS methods and time-series price indices for shelter and utilities to create a time-series price index for housing in each area. The BLS reports a composite housing price index that includes household furnishing and operations as well as shelter, fuel and utilities. Our housing index does not include household furnishing and operations. Second, we use this price index and the BLS price index for all goods to create a time series price index for goods other than housing. Third, we use these two time-series price indices and the overall CPI to inflate and deflate our three cross-sectional price indices.

Table 8 provides illustrative results, namely, price indices for housing and all goods in the first and last year of the panel for the ten areas with the highest, lowest and middle overall

[19] Details about the geographic sample of the CPI can be found in U.S. Bureau of Labor Statistics (2010) and Williams (1996).
[20] The BLS does not collect prices every month in all areas. To obtain an annual price index for these areas, they interpolate to obtain price indices for those months where prices are not collected before averaging over the year. We take a simple average of the reported price indices.

CPI in 2008. The percentage increase in the CPI between 1982 and 2008 tended to be much higher in the places with the highest cost of living than in places where it is average or extremely low. The mean increase was 138 percent in the highest group and only 113 percent in each of the other two groups. The mean increase in housing prices was about the same in the middle and lowest group (122 and 116 percent respectively) and only slightly higher than the increase in the CPI. For the ten areas with the highest CPI in 2008, the mean percentage increase in housing prices (175 percent) was much greater than mean increase in their CPI and the increase in housing prices in the other groups.

The complete panel of price indices contains some seemingly anomalous results. Because our overall consumer price index in 2000 is a weighted mean of the price indices for housing services and other goods with weights between 0 and 1, it is always between these two price indices. Applying the BLS time-series price indices to our cross-sectional does not necessarily maintain this property. This is not due to the difference between the weights used to create the cross-sectional CPI and those used by the BLS in constructing its time-series price indices. Even with the same weights, it can occur. It is a general mathematical phenomenon. In our panel, the CPI was outside the range of the two composite prices in about 3 percent of the cases, always by small amounts. Among these few cases, the mean absolute percentage deviation of the CPI from the nearest price was 0.3 percent and the maximum deviation was 2.9 percent. This is not a problem unless both the overall CPI and the individual price indices are used in the same analysis. In this case, users may prefer to create their own CPI as a weighted mean of the prices indices for housing services and other goods.

The price indices reported in this paper are for the metropolitan areas announced by the Census Bureau on June 30, 1999 and in effect until June 6, 2003.[21] They are based on the standards for defining metropolitan areas adopted by the U.S. Office of Management and Budget (OMB) in 1990 and the Census Bureau's analysis of data from the 1990 Decennial Census and other sources.

The basic concept of a metro area has stayed the same since its inception, namely, a densely populated core area and adjacent areas that have a high degree of economic and social

[21] The list is at http://www.census.gov/population/www/metroareas/pastmetro.html.

integration with it. However, the boundaries of the metropolitan areas can change over time due to changes in the locations of households and the official definition of a metropolitan area. Around the time of each decennial census, OMB adopts a new definition of a metropolitan area. About three years later, the Census Bureau creates specific metro areas based on the general definition and an analysis of data from the decennial census. In the ten years between these major modifications, some additions, deletions, and modifications occur based on other data.

Each time a new set of metropolitan areas is announced at least some new metropolitan codes are added and others are abolished. Prior to June 6, 2003, the codes were four digits. (0040 is considered a four-digit code.) Since this time, the codes have been five digits, all between 10,000 and 50,000. Therefore, potential users of our price indices may encounter a set of metropolitan codes somewhat or entirely different from the 331 codes that existed on June 30, 1999.

To assist users of our panel of prices, we have produced price indices for all metropolitan areas that have existed between 1982 and 2008.[22] To each metro area that existed prior to or after June 30, 1999, we assign the prices of the 1999 area that has the greatest population overlap with it based on the 2000 populations of the counties (or cities and towns in New England prior to June 6, 2003) in the non-1999 metro area. Specifically, it is assigned the prices of the 1999 area that accounts for the largest fraction of the 2000 population of its counties in the year closest to 2000.

Few public-use data sets contain information on the location of observations at this detailed level of geography. That is, few contain the exact metropolitan area of each observation in a metro area and the state of each observation in a non-metro area. Some report specific metro codes only for the largest metropolitan areas. Others report whether an observation is in a metro area but not the specific metro area. Some report only region rather than specific state. The Excel and Stata files CEOPricesPanel02 contain the information needed to produce good price indices at the lowest level of geography possible with the geographic information that is available in a wide range of public-use data sets, and its user's guide suggests how to do it.

[22] These are in the files entitled CEOPricesPanel02 under the heading Price Indices at http://artsandsciences.virginia.edu/economics/facultystaff/eoo.html . This site also contains a user's guide to the data.

9. Conclusion

Data on differences in prices in different locations are important for economic research and private and public decision making. Recent studies have shown that the failure to account for price differences can have large effects on the conclusions of empirical studies. Despite the importance of this information, the United States government does not produce official cross-sectional price indices. BLS and BEA analysts have produced such price indices on a few occasions at least for the largest urban areas and other urban areas divided into about 12 categories by region and population. Since 1968, the Council for Community and Economic Research has produced the ACCRA price indices for six broad categories of goods and an overall consumer price index for many urban areas. However, these price indices have rarely been used in economic research, and this paper indicates that its housing price index is problematic.

This paper estimates cross-sectional housing price indices for each metropolitan area and the nonmetropolitan part of each state based on a large data set with detailed information about the characteristics of dwelling units and their neighborhoods that overcomes many shortcomings of existing housing price indices. The fit of the hedonic equation was excellent, and the estimated price indexes were consistent with popular views about differences in housing prices. Alternative housing price indices based on alternative methods and the same data are highly correlated. All housing price indices based on inferior data and methods differ from the preceding housing price indices in some important respects. In some cases, the differences are very substantial.

The paper then combines the housing price index for all areas of the United States with the ACCRA price indices for other goods that exist for many places and data from the Consumer Expenditure Survey to produce price indices for non-housing goods and all consumer goods for all areas of the United States. It is shown that the resulting overall consumer price index is not sensitive to the expenditure weights used, and it differs little from a simple ideal consumer price index. Finally, the best available time-series price indices are used to produce a panel of price indices for housing, other goods, and all goods from 1982 through 2008 from the cross-sectional price indices. The panel can be easily expanded forward and backward in time.

In assessing whether to use the new panel of price indices, the following questions seem relevant. Is it better to completely ignore geographical price differences in economic research? Is it better to account for geographic price differences in alternative ways such as including dummy variables for different types of areas as explanatory variables in behavioral relationships? Is it better to use other available price indices rather than these price indices in empirical research? Does it make any sense for each economist who does an empirical study that would benefit from cross-sectional or panel price indices to construct his or her own? It will be interesting to see what will happen when old issues are revisited with good price indices and what new issues will be explored with them.

Appendix A

The housing price indices produced by Thibodeau (1989, 1995) shed light on the differences between housing price indices for units of different qualities. Thibodeau produced separate rental housing price indices for units built in the last three years, older standard units, and older substandard units. Less than 4 percent of rental units were built in the last 3 years and less than 7 percent were severely or moderately inadequate according to the AHS's definitions of these terms. Because he used data from the Metropolitan AHS, he had sufficient observations to estimate separate hedonic equations for each place and year. He used these estimated hedonics to predict the market rent of units in each time and place at the national average values of the regressors for the three categories of units. To compare the extent to which these predicted market rents indicate the same percentage differences in housing prices, we first rescale the three price indices in each study to have a mean of 1.[23] This led to 163 observations for each price index in the 1989 study and 103 in the 1995 study. For the observations in the 1989 study, the correlation between the price index for new units (PNew) and the price index for older standard units (PStand) is .94 and the correlation between PStand and the price index for older substandard units (PSub) is .98. The mean of the absolute percentage deviations between PNew and PStand is 7.6 percent and between PSub and PStand is 4.8 percent. For the observations in the 1995 study, the correlation between PNew and PStand is .90 and the correlation between PStand and PSub is .96. The mean of the absolute percentage deviations between PNew and PStand is 9.3 percent and between PESub and PStand is 6.5 percent.

[23] We delete the obviously erroneous result reported for Indianapolis's first survey in the 1989 study.

Appendix B

This appendix describes the approximations of the rent ceilings that faced the recipients of housing certificates and vouchers in our sample. These ceilings were used in the estimation of the truncated regression model discussed in section 5.

Roughly 25 percent of the sample members were in the old certificate program. Most recipients under this program faced a rent ceiling equal to the local Fair Market Rent (FMR) for a unit with the number of bedrooms deemed appropriate for a family of its size and composition. Local housing authorities were allowed to approve rents up to 10 percent greater than the relevant FMR for up to 20 percent of recipients, and with HUD field office approval, they could allow rents up to 20 percent greater than the relevant FMR for these recipients. About 29 percent of certificate recipients were served by housing authorities that used exception rents at the time of our data [Devine et al., 2000, Table IV-8, Table A-2]. Recipients had a substantial incentive to find the best unit available renting for no more than their ceiling since occupying a more expensive unit, within that constraint, did not require them to sacrifice consumption of other goods. Our data contains information on the FMR that applied to each recipient, but not specific ceiling rents faced by recipients granted exceptions under the certificate program.

To approximate the preceding reality, we made the following assumptions. If the gross rent was less than or equal to the relevant FMR, the FMR was the ceiling rent. If the gross rent was greater than the FMR but less than or equal to $1.1 \cdot FMR$, the ceiling rent was $1.1 \cdot FMR$. If the gross rent was greater than $1.1 \cdot FMR$ but less than or equal to $1.2 \cdot FMR$, the ceiling rent was $1.2 \cdot FMR$. Finally, if the gross rent exceeded $1.2 \cdot FMR$, the ceiling rent was the gross rent.

The remaining 75 percent of sample members participated in the old or new voucher program. The CSS data does not distinguish between these programs. This distinction is important for our purposes because the old voucher program did not, and the new voucher program does, have a ceiling on the rent of the unit occupied. Based on other information, we conclude that about 10 percent of the CSS sample members were under the old voucher program and 65 percent under the new program. Since we could not determine which units were under the old program and these units were a distinct minority of all voucher units in the sample, we assumed that all voucher units were covered by the rules of the new voucher program.

When a family enters the new voucher program or when it moves to a new unit under the program, it faces a ceiling on the rent of its unit equal to a local payment standard applicable to families of its type plus 10 percent of the family's adjusted income. Unlike the certificate program, these recipients do not have a strong incentive to occupy a dwelling unit renting for the ceiling rent. They bear the full marginal cost of more expensive housing for units renting for more than the local payment standard and less than the ceiling rent. Beyond the first year in a given unit, the rent can exceed this amount provided that the housing authority certifies that the rent does not exceed the market rent of similar units.

Since implementation of the 1998 Quality Housing and Work Responsibility Act, housing authorities have been allowed to establish local payment standards within 10 percent of the relevant FMR for some or all types of voucher recipients without HUD approval.[24] Devine et al. (2000, p. 48) reports that 90 percent of housing authorities had adopted a uniform percentage of the FMR for all recipients at the time of our data.[25] Among these housing authorities, about 64 percent used the FMR themselves as payment standards and about 21 percent used payment standards above FMR. The CSS data contains information on the FMR applicable to each recipient under the new voucher program, but it does not contain the applicable local payment standard. We approximated them based on a data file from HUD's Office of Public Housing and Voucher Programs that contains the payment standard applicable to each voucher recipient and other relevant information during our time period.[26] It is not possible to match the households in this file with those in the CSS data. To approximate each housing authority's payment standard, we calculated separately for each housing authority the median payment standard among households living in the same zip code, with the same number of bedrooms on the voucher, and with and without a disabled member of the household. (Using the mean and mode payment standards produced similar results.) This allowed us to link an estimate of the payment standard at this level of specificity to more than 90 percent of voucher recipients in the CSS sample.

[24] With HUD approval, they could establish payment standards outside this range. However, few exceptions had been granted at the time of our data [Devine et al., 2000. p. 48].
[25] Most of the rest established percentages that differed for families of different sizes and compositions and in different areas within their jurisdiction.
[26] We are grateful to Milan Ozdinec and Juan Garcia for providing this information.

To approximate the ceiling rent for each household in the CSS data, we made the following assumptions. If the gross rent was less than or equal to $PS + .1 \cdot AINC$ (where $AINC$ is adjusted income), the ceiling rent was $PS + .1 \cdot AINC$. If the gross rent is greater than $PS + .1 \cdot AINC$, the ceiling rent was the gross rent. For the cases where the CSS did not report one of the variables needed to use the estimated payment standards for its locality (less than 10 percent of the cases), we assumed that the payment standard was $1.1 \cdot FMR$.

Appendix C

This appendix documents the sources of the explanatory variables in the regression explaining differences in the non-housing price index and how values were imputed when they were not reported.

Land use regulation index (*regindex*)

We estimated a regulatory index for our areas using the Wharton Residential Land Use Regulatory Index (WRI) developed by Gyourko, Saiz and Summers (2007). This index is based on a nationwide survey of local land use controls. The survey was sent to about 6,900 municipalities across the U.S. About 38 percent responded, representing about 60 percent of the surveyed population. The survey data together with information on state land use policies and other measures of community pressure (using information from environmental and open space ballot initiatives) are used to create eleven subindexes that summarize different aspects of land use regulation. Higher values of these indices indicate more restrictive regulations. An aggregate index, the WRI, is created using factor analysis. The WRI is standardized so that its sample mean is zero and standard deviation equals one. We use their municipal-level WRI index and weights (which are available online) to compute (weighted) average regulatory indices for most of our areas. Forty of our areas contain no subareas for which the WRI is reported. To impute the regulation index for 37 of these areas, we use a state-level average WRI provided by the authors in the paper. For three areas whose boundaries cross state lines (Cumberland, MD-WV MSA, Grand Forks, ND-MN MSA, and Clarksville-Hopkinsville, TN-KY MSA), we use simple averages of the corresponding state level regulation indices.

Climate (*coolingdays, heatingdays, precip*)

Census Bureau (2007, Table C-6) provides average annual precipitation and the total number of cooling and heating days between 1970 and 2000 for many cities. The level of geography in our study is the metropolitan area (MSA or PMSA, hereafter MSA), and the non-metropolitan part of each state. Metropolitan areas often contain multiple cities, but MSA names usually include the name of its largest city. For these MSA, our values of the climate variables are the values for the largest city. In 18 cases, the source did not contain data for the cities mentioned in the MSA name or the MSA name contains only counties. In these cases, our imputed values of the climate variables were for the closest MSA. The median distance from the center of these 18 MSA to the closest MSA whose climate data were reported was 25 miles; the maximum was only 53 miles. The imputed values for the non-metropolitan part of each state are the mean values of the variables for the MSA in the state.

Wage rate (P_L)

Using U.S. Census data from the 2000 Integrated Public-Use Microdata Series (IPUMS), David Albouy (2009) computed wage differentials across 290 areas of the U.S. The wage differentials are computed for full time workers (working at least 30 hours a week, 26 weeks a year) ages 25 to 55. To estimate the wage differential, a log-wage regression is estimated. Covariates include educational attainment, potential experience, industry, gender, English proficiency, and marital, veteran, minority, and immigrant status, their interactions, and MSA dummy variables. The regression model is estimated using weighted OLS. Albouy's index is an index of ln P_L, whose value is zero in Reading, PA.

Some assumptions are necessary to predict wage indices for all of our areas. Albouy computes only one wage index for each Consolidated Metropolitan Statistical Area (CMSA). We assume that all PMSAs within the CMSA have the same wage index. Wage data are unavailable for 35 small MSAs. The average population in these areas is about 120,000; the largest is Huntington-Ashland with about 315,000 residents. We use the wage index computed for the non-metropolitan part of the corresponding state to impute the wage index of these small areas. In the eight cases where the MSA spans several states, we compute a simple average of the corresponding non-MSA state wage indices.

Distance to nearest large metropolitan area (*mindist*)

This variable is the 'as-the-crow-flies' distance between the center of each area and the center of the nearest MSA with at least 1.5 million residents. For the 41 large metropolitan areas, it is zero. The center of the non-metropolitan part of each state is assumed to be the center of the state. The longitude and latitude of the center of each area were obtained from Google Maps.

References

Albouy, David. 2008. “Are Big Cities Really Bad Places to Live? Improving Quality-of-life Estimates across Cities.” National Bureau of Economic Research Working Paper 14472.

Albouy, David. 2009. “The Unequal Geographic Burden of Federal Taxation.” *Journal of Political Economy* 117(4): 635-667.

Aten, Bettina H. 2005. “Report on Interarea Price Levels.” U.S. Department of Commerce, Bureau of Economic Affairs Working Paper 2005-11.

Aten, Bettina H. 2006. “Interarea Price Levels: An Experimental Methodology.” *Monthly Labor Review* 129(9): 47-61.

Aten, Bettina H. 2008. “Estimates of State and Metropolitan Price Parities for Consumption Goods and Services in the United States, 2005.” http://www.bea.gov/papers/pdf/aten_estimates_state_metro_2005.pdf

Blackley, Dixie M., and James R. Follain, with Haeduck Lee. 1986. “An Evaluation of Hedonic Price Indexes for 34 Large SMSAs.” *AREUEA Journal* 14(2): 179-205.

Bloomquist, Glenn C., Mark C. Berger, and John P. Hoehn. 1988. “New Estimates of Quality of Life in Urban Areas.” *American Economic Review* 78(1): 89-107.

Building Research Council, School of Architecture, University of Illinois at Urbana-Champaign. 1998. *Resident Assessment of Housing Quality: Lessons from Pilot Surveys*. Washington, DC: U.S. Department of Housing and Urban Development.

Chen, Yong, and Stuart S. Rosenthal. 2008. “Local Amenities and Life-Cycle Migration: Do People Move for Jobs or Fun?” *Journal of Urban Economics* 64(3): 519-537.

Citro, Constance F., and Robert T. Michael, ed. 1995. *Measuring Poverty: A New Approach.* Washington, D.C.: National Academy Press.

Council for Community and Economic Research. 2006. *ACCRA Cost of Living Index Manual*. Arlington, VA:

Dalaker, Joe. 2005. *Alternative Poverty Estimates in the United States: 2003.* U.S, Bureau of the Census Current Population Report P60-227.

Deaton, Angus, and John Muellbauer. 1980. *Economics and Consumer Behavior.* Cambridge: Cambridge University Press, 1980.

Devine, Deborah J., Barbara A. Haley, Lester Rubin, and Robert W. Gray. 2000. *The Uses of Discretionary Authority in the Tenant-Based Section 8 Program.* Washington, DC: U.S. Department of Housing and Urban Development.

Devine, Deborah J., Robert W. Gray, Lester Rubin and Lydia B. Taghavi. 2003. *Housing Choice Voucher Location Patterns.* Washington, D.C.: U.S. Department of Housing and Urban Development.

Early, Dirk W. 2006. "Metropolitan Area Cost of Rental Housing Indices Using the Customer Satisfaction Survey," HUD Report, Research for Assisted Housing II Grants Program, Grant Number GS-10F-0405M, US Department of Housing and Urban Development, Washington, D.C.

Early, Dirk W., and Edgar O. Olsen. 2002. "Subsidized Housing, Emergency Shelters, and Homelessness: An Empirical Investigation Using Data from the 1990 Census." *Advances in Economic Analysis & Policy*, 2 (1): 1-34.

Follain, James R. 1979. *How Well D o Section 8 FMRs Match the Cost of Rental Housing?: Data from 39 Large Cities.* Contract Report 249-15. Washington, D.C.: The Urban Institute.

Follain, James R., and Larry Ozanne, with Verna Alberger. 1979. *Place to Place Indexes of the Price of Housing*. Washington, DC: The Urban Institute.

Follain, James R., and Stephen Malpezzi. 1980. *Dissecting Housing Value and Rent*. Washington, DC: The Urban Institute.

Gabriel, Stuart, and Stuart S. Rosenthal. 2004. "Quality of the Business Environment versus Quality of Life: Do Firms and Households Like the Same Cities?" *Review of Economics and Statistics* 86(1): 438-444.

Gray, Robert W., Barbara A. Haley, Seth R. Marcus, Huda M. Elhassan, and Vivienne Liu. 2002. *Tell Us About Your Home: First Year Results of the Survey of Households Assisted by the Housing Choice Voucher Program, Draft Report*. Program Monitoring and Research Division, Office of Policy Development and Research, Washington, DC: U.S. Department of Housing and Urban Development.

Gyourko, Joseph, and Joseph Tracy. 1991. "The Structure of Local Public Finance and the Quality of Life." *Journal of Political Economy* 99(4): 774-806.

Gyourko, Joseph, Albert Saiz and Anita Summers. 2007. "A New Measure of the Local Regulatory Environment for Housing Markets: The Wharton Residential Land Use Regulatory Index." *Urban Studies,* forthcoming.

Johnson, David S., John M. Rogers, and Lucilla Tan. 2001. "A Century of Family Budgets in the United States." *Monthly Labor Review* 124(5): 28-45.

Jolliffe, Dean. 2006. "Poverty, Prices, and Place: How Sensitive Is the Spatial Distribution of Poverty to Cost of Living Adjustments." *Economic Inquiry* 44(2): 296-310.

Kiel, Katherine, and Jeffrey E. Zabel. 1997. "Evaluating the Usefulness of the American Housing Survey for Creating House Price Indices." *Journal of Real Estate Finance and Economics* 14: 189-202.

Kokoski, Mary F., Patrick Cardiff, and Brent R. Moulton. 1994. "Interarea Price Indices for Consumer Goods and Services: A Hedonic Approach Using CPI Data." U.S. Department of Labor, Bureau of Labor Statistics, Working Paper 256.

Koo, Jahyeong, Keith R. Phillips, and Fiona D. Sigalla. 2000. "Measuring Regional Cost of Living." *Journal of Business & Economic Statistics,* 18(1): 127-136.

Leger, Mireille L., and Stephen D. Kennedy. 1990. *Final Comprehensive Report of the Freestanding Housing Voucher Demonstration*. Volume 1 & 2. Cambridge, MA: Abt Associates Inc.

Maddala, G. S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Econometric Society Monographs, Cambridge, UK: Cambridge University Press.

Malpezzi, Stephen, Gregory Chun and Richard Green. 1998. "New Place to Place Housing Price Indexes for U.S. Metropolitan Areas, and Their Determinants: An Application of Housing Indicators," *Real Estate Economics* 26(2): 235-75.

Malpezzi, Stephen, Larry Ozanne and Thomas Thibodeau. 1980. *Characteristic Prices of Housing in 59 SMSAs*. Washington, DC: The Urban Institute.

Mast, Brent D. 2009. "Measuring Housing Quality in the Housing Choice Voucher Program with Customer Satisfaction Program Data," *Cityscape* 11(2): 101-112.

Moretti, Enrico. 2008. "Real Wage Inequality." National Bureau of Economic Research Working Paper 14370.

Moulton, Brent R. 1995. "Interarea Indexes of the Cost of Shelter Using Hedonic Quality

Adjustment Techniques," *Journal of Econometrics* 68(1): 181-204.

Nelson, Charles, and Kathleen Short. 2003. "The Distributional Implications of Geographic Adjustment of Poverty Thresholds." http://www.census.gov/hhes/www/povmeas/papers/geopaper.pdf

Olsen, Edgar. 2003. "Housing Programs for Low-Income Households." In *Means-Tested Transfer Programs in the United States,* ed., Robert Moffitt. Chicago: University of Chicago Press.

ORC/Macro. 2001. *Quality Control for Rental Assistance Subsidies Determination.* Washington, D.C.: U.S. Department of Housing and Urban Development, Office of Policy Development and Research.

Pollak, Robert A. 1989. *The Theory of the Cost-of-Living Index.* Oxford: Oxford University Press.

Primont, Diane F., and Mary F. Kokoski. 1990. "Comparing Prices across Cities: A Hedonic Approach." U.S. Department of Labor, Bureau of Labor Statistics Working Paper 204.

Ptacek, Frank, and Robert M. Baskin. 1996. "Revision of the CPI Housing Sample and Estimators." *Monthly Labor Review* 119(12): 31-39.

Roback, Jennifer. 1982. "Wages, Rents and the Quality of Life." *Journal of Political Economy,* 90(6): 1257-78

Rosen, Sherwin. 1979. "Wage-Based Indexes of Urban Quality of Life." In *Current Issues in Urban Economics,* ed., Peter Mieskowski and Mahlon Straszheim. Baltimore: Johns Hopkins University Press.

Short, Kathleen, Thesia Garner, David Johnson, and Patricia Doyle. 1999. *Experimental Poverty Measures: 1990 to 1997.* U.S. Bureau of the Census, Current Population Report P60-205.

Slesnick, Daniel T. 2002. "Prices and Regional Variation in Welfare." *Journal of Urban Economics* 51(3): 446-468.

Slesnick, Daniel T. 2005. "Prices and Demand: New Evidence from Micro Data." *Economic Letters* 89(3): 269-274.

Thibodeau, Thomas G. 1989. "Housing Price Indexes from the 1974-83 SMSA Annual Housing Surveys." *AREUEA Journal* 17(1): 100-17.

Thibodeau, Thomas G. 1995. "House Price Indices from the 1984-1992 MSA American Housing

Surveys," *Journal of Housing Research* 6(3): 439-479.

U.S. Bureau of Labor Statistics. 2010. *BLS Handbook of Methods*. http://www.bls.gov/opub/hom/pdf/homch17.pdf.

U.S. Census Bureau. 2002. *American Housing Survey for the United States: 2001*. Current Housing Reports, Series H150/01. Washington, DC: U.S. Government Printing Office.

U.S. Census Bureau. 2007. *County and City Data Book: 2007.* 14th edition. Washington, D.C.

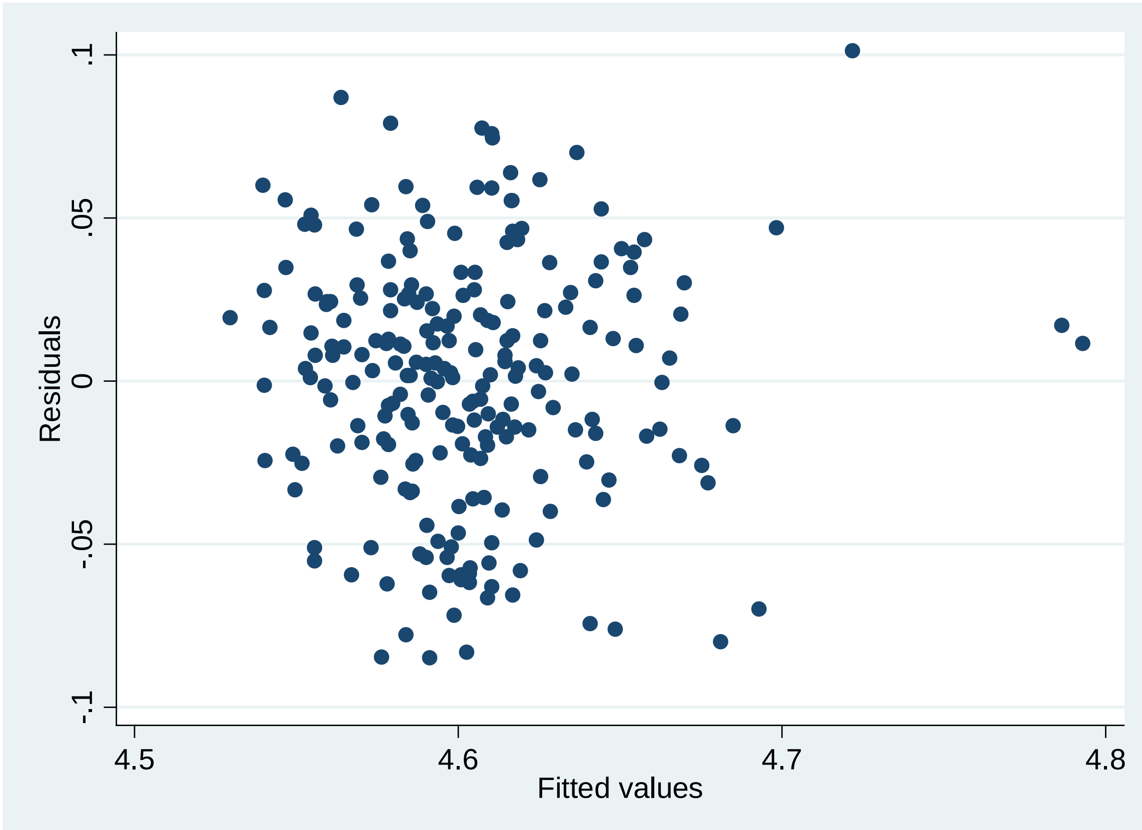
U.S. Department of Housing and Urban Development. 2000. "Fair Market Rents: Increased Fair Market Rents and Higher Payment Standards for Certain Areas; Interim Rule" F*ederal Register* **65**(191) (Oct. 2), 58870.

U.S. Department of Housing and Urban Development. 2007. "Fair Market Rents for the Section 8 Housing Assistance Payments Program." http://www.huduser.org/datasets/fmr/fmrover_071707R2.doc

Wallace, James E.; Bloom, Susan Philipson; Holshouser, William L.; Mansfield, Shirley; and Weinberg, Daniel H. 1981. *Participation and Benefits in the Urban Section 8 Program: New Construction and Existing Housing.* Volume 1 & 2. Cambridge, MA: Abt Associates Inc.

Weinberg, Daniel H. 1982 "Housing Benefits from the Section 8 Program." *Evaluation Review* 6: 5-24.

Williams, Janet L. 1996, "The redesign of the CPI geographic sample." *Monthly Labor Review*. 119(12): 10-17.



.1
.05
0
-.05
-.1
Residuals
4.5
4.6
4.7
4.8
Fitted values

Table 1. Variables Used in the Hedonic Regressions, their Definitions, and Summary Statistics

Variable	Definition	Mean	Std. Dev.
Dependent Variable[1]			
LNRENT	Log of gross rent (contract rent to owner + utility allowance)	6.274	0.346
Explanatory Variables			
Bedrooms[1]			
BDRMS0	unit has no bedrooms (omitted)		
BDRMS1	unit has 1 bedroom	0.289	0.453
BDRMS2	unit has 2 bedrooms	0.407	0.491
BDRMS3	unit has 3 or more bedrooms	0.290	0.454
BDRMS4P	number of bedrooms - 3, if number of bedrooms > 3	0.040	0.220
Units in the structure			
UNITS1	single-family detached housing unit	0.375	0.484
UNITS2-4	two to four units in building (omitted)		
UNITS4-8	four to eight units in building	0.119	0.324
UNITS8P	eight or more units in building	0.225	0.417
Length of time in the unit			
LT1YR	lived in the unit less than 1 year	0.303	0.460
Kitchens and bathrooms			
STOVE	all stove burners work	0.928	0.259
OVEN	working oven	0.967	0.178
REFRIG	refrigerator keeps food cold enough that food does not spoil	0.954	0.210
WATER1	tap water has a problem with color or odor	0.098	0.297
WATER2	tap water sometimes has a problem with color or odor	0.087	0.281
KLIGHT	kitchen has a working light fixture	0.962	0.191
KOUT1	one working outlet in the kitchen	0.054	0.227
KOUT2	two or more working outlets in the kitchen	0.913	0.282
HOTCOLD	hot and cold running water in kitchen and bathroom, tub, shower, and sink	0.975	0.158
WLEAK	water is leaking from any kitchen or bathroom sink, pipe, or drain	0.138	0.345
CLOG1	any kitchen or bathroom sink, pipe, or drain is clogged	0.031	0.173
CLOG2	any kitchen or bathroom sink, pipe, or drain is slow	0.342	0.475
BATHVENT	bathroom has either a window that opens or a ventilation system that works	0.913	0.282
TOILETS	all toilets are working	0.969	0.174
BADTOILET13	in the last three months, toilets did not work for more than 6 hours at least once, but fewer than 4 times	0.075	0.264
BADTOILET4P	in the last three months, toilets did not work for more than 6 hours more than 3 times	0.019	0.137

WETFLOOR	bathroom floor was covered by water due to plumbing problem	0.113	0.317
Electrical wiring			
ENCLOSED	all wiring enclosed in walls or metal coverings	0.939	0.240
COVERS	all outlets and switches have cover plates	0.935	0.246
OUTLETS	each room has at least one working outlet (excluding the bathroom)	0.975	0.156
FIXWORK	all ceiling and wall mounted light fixtures work	0.934	0.249
NOFIX	no ceiling or wall mounted light fixtures	0.010	0.100
BLOWN13	fuses blown or circuits tripped 1 to 3 times in last three months	0.139	0.346
BLOWN4P	fuses blown or circuits tripped 4 or more times in last three months	0.031	0.172
Heating and cooling			
HEATOK	heating system provides enough heat in every room	0.784	0.412
HEATDN	do not know whether heating system provides enough heat in every room	0.062	0.242
OVENHEAT1	use oven to heat the unit	0.078	0.268
OVENHEAT2	sometimes use oven to heat the unit	0.073	0.260
NOAC	no air conditioning	0.370	0.483
BADAC	air conditioning is not working	0.062	0.242
ADJHEAT1	can adjust heat when too hot or too cold	0.853	0.354
ADJHEAT2	can partially adjust heat when too hot or too cold	0.051	0.221
NOWINTER	did not live in the unit last winter	0.203	0.402
HEATOFF13	lived in the unit last winter and heating broke down for more than 6 hours at least once, but fewer than 4 times	0.073	0.260
HEATOFF4P	lived in the unit last winter and heating broke down for more than 6 hours more than 3 times	0.013	0.112
COLDHOME	lived in the unit last winter and unit was cold for more than 24 hours	0.089	0.285
Sanitation and safety			
RATS	observed rats in the building or outside around the grounds	0.067	0.250
ROACHES	observed many cockroaches in the unit this week	0.122	0.327
SMELL1	bad odor (sewer, natural gas, etc.) is present in the unit	0.040	0.195
SMELL2	bad odor (sewer, natural gas, etc.) is sometimes present in the unit	0.089	0.284
LOCKS	all doors have working locks	0.930	0.255
WINLOCK	all windows have locks that work	0.886	0.317
BWINDOW	all bedrooms have a window that can open	0.922	0.268
MAILGONE	mail has been stolen	0.058	0.233
DETECTOR	working smoke detector exists	0.932	0.251
DETECTORDK	do not know if a working smoke detector exists	0.020	0.140
EXITS	at least two exits out of the unit to be used in case of a fire	0.931	0.254

GARBAGE	weekly garbage pickup	0.943	0.232
DUMPSTER	covered dumpsters or cans for garbage and trash	0.866	0.341
Dwelling quality			
RAIN	holes or cracks allow outdoor air or rain to enter unit	0.114	0.317
CHIPPING	paint is easily chipped or peeled	0.155	0.362
PEELING	large areas of peeling paint or broken plaster	0.047	0.211
WALLSBAD	walls, ceilings, or floors with serious problems	0.120	0.325
MILDEW	mildew, mold, or water damage on any wall, floor, or ceiling	0.175	0.380
FLOORMISS	flooring material missing, curled, or loose	0.175	0.380
TRIP	floor problems can cause you to trip	0.068	0.252
BADRAILS	secure handrails are not present on all stairs and landings in the unit	0.075	0.263
BADRAILSNA	handrails in unit does not apply	0.325	0.469
BROKENW	any window with broken glass	0.044	0.205
BADPORCH	dangerous porch or balcony	0.059	0.236
BADPORCHNA	porch or balcony condition not applicable	0.178	0.383
BADSTEPS	unsafe handrails, steps, or stairs outside unit	0.076	0.265
BADSTEPSNA	condition of handrails, steps, or stairs outside unit not applicable	0.187	0.390
SIDEWALK	sidewalk, driveway, or parking lot damaged	0.097	0.296
NOLIGHT	not enough exterior light for safety	0.116	0.320
BADFENCE	problems with the fences or gates in bad repair	0.063	0.243
NOFENCE	no fence	0.441	0.497
EXWALLS	exterior walls have serious problems	0.045	0.208
BADROOF	roof sagging, holes, or missing roofing	0.046	0.210
ROOFDK	cannot see roof	0.182	0.386
SAFEYARD	agree or strongly agree that yards, playgrounds, and off-street parking are safe	0.703	0.457
UNSAFEYARD	disagree or strongly disagree that yards, playgrounds, and off-street parking are safe	0.153	0.360
OUT_SAME	lived in unit for one year and condition of building same as a year ago	0.372	0.483
OUT_WORSE	lived in unit for one year and condition of building worse than a year ago	0.046	0.209
SUPER_SAME	lived in unit for one year and landlord’s supervision of vacant units is the same as a year ago	0.420	0.494
SUPER_WORSE	lived in unit for one year and landlord’s supervision of vacant units is worse than a year ago	0.019	0.137
REPAIR_SAME	lived in unit for one year and repair of problems the same as a year ago	0.408	0.492
REPAIR_WORSE	lived in unit for one year and repair of problems is worse than a year ago	0.047	0.211

Apartment complex amenities			
LAUNDRY1	live in an apartment complex with a laundry room in working condition	0.326	0.469
LAUNDRY2	live in an apartment complex with a non-working laundry room	0.008	0.089
PLAYAREA1	live in an apartment complex with a useable play area	0.265	0.441
PLAYAREA2	live in an apartment complex with a play area, but it is not usable	0.019	0.138
ELEVATOR1	live in an apartment complex with a working elevator	0.041	0.199
ELEVATOR2	live in an apartment complex with an elevator, but it is not in working condition	0.003	0.058
Neighborhood quality			
CRIMEOK	crime or drugs not a problem	0.521	0.500
CRIMEBAD	crime or drugs big problem	0.068	0.252
CRIMEDK	do not know whether crime is a problem	0.207	0.405
TRASHOK	trash or junk nearby not a problem	0.709	0.454
TRASHBAD	trash or junk nearby big problem	0.054	0.226
TRASHDK	do not know whether trash is a problem	0.051	0.219
VACANTOK	vacant or run-down homes or stores not a problem	0.764	0.425
VACANTBAD	vacant or run-down homes or stores big problem	0.022	0.147
VACANTDK	do not know whether vacant or run-down buildings are a problem	0.075	0.264
NBHDGRT	scale from 1-10 (10 being best) rated neighborhood 9 or 10	0.379	0.485
NBHDOK	scale from 1-10 (10 being best) rated neighborhood 6 - 8	0.394	0.489
General opinion of home (rental unit) as a place to live			
HOMEGRT	scale from 1-10 (10 being best) rated home as a place to live 9 or 10	0.440	0.496
HOMEOK	scale from 1-10 (10 being best) rated home as a place to live 6 - 8	0.354	0.478
Contract conditions			
CROWDED[1]	number of persons in the unit divided by 1 + number of bedrooms	0.747	0.349
Census tract variables[2]			
BLT95_98	fraction of rental units built between 1995-1998	0.049	0.072
BLT90_94	fraction of rental units built between 1990-1994	0.050	0.059
BLT80_89	fraction of rental units built between 1980-1989	0.136	0.113
BLT70_79	fraction of rental units built between 1970-1979	0.196	0.122
BLT60_69	fraction of rental units built between 1960-1969	0.136	0.090
BLT50_59	fraction of rental units built between 1950-1959	0.114	0.082
BLT40_49	fraction of rental units built between 1940-1949	0.082	0.070
BLT39	fraction of rental units built between before 1940	0.177	0.179
VACRATE	vacancy rate	8.364	6.768
TRAVELTIME	mean travel time to work, minutes	21.054	7.494

MEDINC	median household income, in $1,000s	32.762	13.864
POV_RATE	poverty rate	0.158	0.104
BLACK	fraction African-American	0.137	0.227
HISP	fraction hispanic	0.113	0.190
DENSITY	population density 1,000s of persons per square kilometer	1.390	2.308

Notes:

[1] From Form HUD-50058, Family Report

[2] From the 2000 Decennial Census.

All other data from HUD Customer Satisfaction Survey "Tell us About Your Home."

Unless otherwise noted, all variables are coded 1 if the condition exists, 0 otherwise.

Table 2. Price Indices for Housing, Other Goods, and All Produced Goods across Areas (2000)

Geographical Area	Housing	Other Goods	All Goods
Areas with Ten Highest Housing Price Levels			
San Francisco, CA PMSA	2.043	1.155	1.379
Stamford-Norwalk, CT PMSA	1.969	1.124	1.337
San Jose, CA PMSA	1.963	1.124	1.336
Nassau-Suffolk, NY PMSA	1.814	1.233	1.379
Santa Cruz-Watsonville, CA PMSA	1.789	1.134	1.299
Boston, MA-NH PMSA	1.658	1.141	1.271
Middlesex-Somerset-Hunterdon, NJ PMSA	1.634	1.087	1.224
New York, NY PMSA	1.626	1.087	1.223
Bergen-Passaic, NJ PMSA	1.587	1.092	1.216
Monmouth-Ocean, NJ PMSA	1.569	1.089	1.210
Areas with Ten Middle Housing Price Levels			
Springfield, IL MSA	0.949	0.927	0.932
Corpus Christi, TX MSA	0.948	0.969	0.964
Jacksonville, FL MSA	0.947	0.972	0.966
Gainesville, FL MSA	0.945	0.974	0.967
Tallahassee, FL MSA	0.945	1.058	1.029
Toledo, OH MSA	0.943	0.996	0.983
Racine, WI PMSA	0.943	1.002	0.987
Sheboygan, WI MSA	0.943	0.959	0.955
Grand Junction, CO MSA	0.943	0.988	0.976
Corvallis, OR MSA	0.943	1.069	1.037
Areas with Ten Lowest Housing Price Levels			
Nonmetro ND	0.704	0.940	0.880
Dothan, AL MSA	0.702	0.957	0.892
Nonmetro TN	0.702	0.945	0.883
Gadsden, AL MSA	0.682	0.945	0.879
Hattiesburg, MS MSA	0.681	0.951	0.883
Nonmetro MS	0.681	0.951	0.883
Nonmetro LA	0.674	0.942	0.875
Nonmetro AL	0.672	0.916	0.855
Nonmetro AR	0.664	0.955	0.882
Nonmetro MO	0.660	0.937	0.867

Notes: Housing price index is based on specification using missing value indicators. Other goods price index is based on ACCRA indices for goods other than housing and utilities when available and fitted values otherwise weighted by expenditure shares from CES. Overall consumer price index applies average expenditure shares from CES to the price indices for housing and other goods. Each index is scaled so that the mean across all 380 areas is 1.

Table 3. Comparison of Housing Price Indices Based on Alternative Methods Using CSS Data (2000)

Alternative Methodology	Regression Results			Sample	Absolute Percent Difference	
	Slope	Std. Error	R2	Size	Mean	Maximum
Median regression	1.031	0.003	0.996	380	1.145	5.710
CCA -- All variables	0.977	0.003	0.997	380	0.995	5.427
CCA --Selected variables	0.982	0.002	0.998	380	0.730	3.639
Truncated regression	0.983	0.006	0.986	380	1.880	15.411
Separate hedonic each area	1.017	0.013	0.942	360	4.182	21.091

Table 4. Comparisons with Housing Price Indices Based on Different Data

Alternative Price Index	Regression Results Slope	Std. Error	R2	Sample Size	Absolute Percent Difference Mean	Maximum
American Housing Survey (all areas)	0.637	0.029	0.567	380	10.577	42.342
American Housing Survey (64 large metropolitan areas)	1.111	0.038	0.934	64	4.929	27.051
Decennial Census PUMS	1.016	0.025	0.832	343	7.870	26.836
HUD Fair Market Rents	1.217	0.024	0.891	331	7.063	37.748
Median Gross Rent	0.880	0.022	0.832	331	6.977	28.978
ACCRA (with NYC)	0.968	0.090	0.343	226	11.382	181.957
ACCRA (without NYC)	0.676	0.048	0.472	225	10.610	58.407

Table 5. CEX Average Expenditure Shares for ACCRA Composite Commodities

Goods Categories		Expenditure Shares	
ACCRA	CES	ACCRA	All Consumers
Housing	Shelter	28.0	18.7
Utilities	Utilities, fuels, and public services	8.0	6.5
Grocery Items	Food at home Tobacco products and smoking supplies Housekeeping supplies	16.0	10.0
Transportation	Gasoline and motor oil Other vehicle expenses Public transportation	10.0	10.5
Health Care	Health Care	5.0	5.4
Miscellaneous	All other items	33.0	48.9

Source: http://www.bls.gov/cex/2000/share/quintile.pdf

Table 6. Regression Explaining Differences in Non-Housing Prices

Regressors	Coefficient	Standard Error	t-score	P>t
regindex	0.00314	0.00525	0.60	0.550
ln(coolingdays+1)	-0.01561	0.00522	-2.99	0.003
ln(heatingdays+1)	-0.00094	0.00662	-0.14	0.887
precip	-0.00157	0.00086	-1.82	0.070
precipsq	0.00002	0.00001	1.94	0.053
lnPL	0.08589	0.04108	2.09	0.038
lnPH	0.12777	0.02758	4.63	0.000
dist (in hundreds of miles)	0.00371	0.00253	1.47	0.144
constant	4.75178	0.08540	55.64	0.000

Notes. Dependent variable is natural logarithm of the price index for non-housing goods with sample mean 4.60. Number of observations is 225, $F(8,216)$ is 26.95, and $R2$ is .50.

Table 7: Interpolation of BLS Missing Data

Area	Missing data period	Assumptions
Phoenix-Mesa, AZ	1982 - 2001	Missing prices are estimated using price changes in the overall West - Size Class B/C region.
Tampa-St. Petersburg-Clearwater, FL	1982 - 1986	Missing prices are estimated using price changes in the Miami-Fort Lauderdale, FL, region.
Washington-Baltimore, DC-MD-VA-WV	1982 - 1996	Missing prices are estimated using changes in the New York-Northern New Jersey-Long Island, NY-NJ-CT-PA metro area.
Midwest metro- Urban Size Class B/C	1982 - 1997	Missing prices are estimated using price changes in the overall Midwest region.
Northeast metro - Urban Size Class B/C	1982 - 1997	Missing prices are estimated using price changes in the overall Northeast region.
South metro - Urban Size Class B/C	1982 - 1997	Missing prices are estimated using price changes in the overall South region.
West metro - Urban Size Class B/C	1982 - 1997	Missing prices are estimated using price changes in the overall West region.
Non-metro parts of each state	1982 - 2008	Missing prices are estimated using smallest urban size class in region except that the Anchorage index was used for non-metro Alaska and Honolulu for non-metro Hawaii

Table 8. Price Indices for Housing and All Produced Goods

Geographical Area	Housing 1982	Housing 2008	Housing % inc.	All Goods 1982	All Goods 2008	All Goods % inc.
Areas with Ten Highest CPI (2008)						
Nassau-Suffolk, NY PMSA	0.874	2.573	194%	0.720	1.782	147%
Stamford-Norwalk, CT PMSA	0.949	2.793	194%	0.698	1.728	147%
San Francisco, CA PMSA	0.941	2.568	173%	0.747	1.704	128%
San Jose, CA PMSA	0.904	2.468	173%	0.723	1.651	128%
Boston, MA-NH PMSA	0.821	2.265	176%	0.661	1.629	146%
Santa Cruz-Watsonville, CA PMSA	0.973	2.306	137%	0.738	1.606	118%
Middlesex-Somerset-Hunterdon, NJ PMSA	0.787	2.317	194%	0.639	1.582	147%
New York, NY PMSA	0.784	2.306	194%	0.639	1.580	147%
Bergen-Passaic, NJ PMSA	0.765	2.250	194%	0.635	1.572	147%
Nonmetro AK	0.963	1.736	80%	0.806	1.567	95%
Areas with Ten Middle CPI (2008)						
Houston, TX PMSA	0.686	1.314	92%	0.619	1.209	95%
Grand Junction, CO MSA	0.513	1.215	137%	0.555	1.208	118%
Redding, CA MSA	0.524	1.241	137%	0.554	1.207	118%
Nonmetro MT	0.441	1.045	137%	0.554	1.207	118%
Asheville, NC MSA	0.513	1.113	117%	0.566	1.206	113%
Milwaukee-Waukesha, WI PMSA	0.566	1.317	133%	0.569	1.205	112%
Fort Myers-Cape Coral, FL MSA	0.571	1.240	117%	0.565	1.204	113%
Nashville, TN MSA	0.624	1.354	117%	0.565	1.203	113%
Louisville, KY-IN MSA	0.555	1.205	117%	0.564	1.202	113%
Biloxi-Gulfport-Pascagoula, MS MSA	0.542	1.178	117%	0.564	1.202	113%
Areas with Ten Lowest CPI (2008)						
Gadsden, AL MSA	0.405	0.878	117%	0.509	1.085	113%
McAllen-Edinburg-Mission, TX MSA	0.432	0.938	117%	0.509	1.084	113%
Nonmetro TX	0.420	0.915	118%	0.496	1.082	118%
Nonmetro AL	0.391	0.852	118%	0.492	1.075	118%
Nonmetro MO	0.405	0.838	107%	0.518	1.074	107%
Joplin, MO MSA	0.428	0.931	118%	0.506	1.073	112%
Johnson City-Kingsport-Bristol, TN-VA MSA	0.479	1.040	117%	0.502	1.069	113%
Jonesboro, AR MSA	0.455	0.987	117%	0.499	1.063	113%
Texarkana, TX-Texarkana, AR MSA	0.444	0.965	117%	0.498	1.061	113%
Fort Smith, AR-OK MSA	0.424	0.920	117%	0.498	1.060	113%

Notes: Each index is scaled so that the mean across all 380 areas in 2000 is 1.