A Price Index for New Multifamily Housing

By Frank de Leeuw

The author was formerly BEA's Chief Statistician; this article covers work that he did while at BEA.

One feature of the December 1991 comprehensive revision of the national income and product accounts (NIPA's) was the introduction of a price index designed specifically for new multifamily structures. This index, which extends back to 1978, is based on the sales prices and characteristics of these structures. Before the comprehensive revision, expenditures on new multifamily housing were converted to constant dollars using the U.S. Census Bureau's price index for new one-family houses under construction.[1]

Measuring price changes for construction is a longstanding problem. Because structures do not come in mass-produced models sold at fixed locations, the standard approach of tracking the average price of a narrowly specified commodity is not practicable. In recent years, the measurement of structure prices has been improved by the use of "hedonic" (regression) methods, which estimate the price of a standard structure composed of a given set of characteristics.

Building on work done by the Census Bureau in developing hedonic indexes for single-family houses, the Bureau of Economic Analysis (BEA) has developed a hedonic index for measuring the prices of multifamily structures. To deal with the enormous diversity of housing units, the new index uses the results of annual multiple regressions of structure prices in relation to structure characteristics, such as floor space or the presence or absence of air-conditioning, and location characteristics, such as State wage rates for construction workers. The price differences associated with these characteristics are used to adjust actual structure prices in each period so that the adjusted prices refer to some standard set of characteristics. Thus, the adjusted structure prices resemble the price quotations for a standard specification that are normally used to construct consumer and producer price indexes.[2]

This article has four sections. The first describes the data underlying the indexes. The second sets out a framework for analysis. The third presents the form and results of regression analysis, and the final section explains the procedure for using the regression results to obtain a price index.

The Value and Characteristics of New Multifamily Housing

Primary source data

The Census Bureau collects data on new multifamily housing in two surveys, one of them a sample of entire projects and the other a sample of buildings in those projects. The project survey, known as the value-put-in-place survey, provides data on dollar values (excluding land) of a sample of new construction projects. The building survey, known as the survey of construction, provides data on number of units, square feet, number of bathrooms, and other basic characteristics. The Census Bureau has combined the results of the two surveys into a file of projects with information on the total value of each project and on the characteristics of some or all of the buildings in that project. For the present study, some additional information by State—on income, construction industry

1. The one-family index is presented and described in U.S. Department of Commerce, Bureau of the Census, *Price Index of New One-Family Houses Sold*, Construction Reports Series C-27 (Washington, DC: U.S. Government Printing Office).

2. An earlier attempt to construct a multifamily index by the Census Bureau group that prepares the one-family index is described in Jesse Pollock, Donald M. Luery, and Armando Levinson, "Research on Residential Construction Price Indexes," presented to the Census Advisory Committee of the American Economic Association, October 1987. Paul J. Pieper, "The Measurement of Construction Prices: Retrospect and Prospect," in E.R. Berndt and J.E. Triplett, eds., *Fifty Years of Economic Measurement: The Jubilee of the Conference on Research in Income and Wealth*, University of Chicago Press, 1990, discusses hedonic and other approaches to measuring construction prices. For two recent appraisals of the state of hedonic methods by practitioners, see the articles by Zvi Griliches and Jack Triplett in the same volume.

Paul Armknecht, Larry Ozanne, and Paul Pieper made helpful comments on earlier drafts. Participants at a BEA seminar and at a National Bureau of Economic Research summer workshop in 1990 also made useful suggestions. Staff members of the U.S. Census Bureau's Construction Statistics Division were extremely helpful in providing the data, making them easy to use, and explaining their pitfalls.

wage rates, and climate—was incorporated into the Census Bureau file.

The present study observed the following stipulations in using the Census Bureau file:

- The unit of observation is an entire project of five or more dwellings rather than a single dwelling unit.
- The regression analysis covers the period 1978–89, one year at a time, with all of the sample projects under construction during a given year included in the analysis for that year; thus, some projects are included in the analysis of more than one year.

Distribution of values

A preliminary look at the data reveals a feature that has an important bearing on the index. The distribution of value per dwelling unit becomes increasingly skewed to the right during the sample period: In 1978, there were 8 observations in the sample with value per unit 5 times the median or greater; in 1989, in contrast, there were 25 such observations, even though the sample size was only two-thirds as large as in 1978.

Table 1.—Value Per Dwelling Unit, 1978 and 1989: Summary Statistics

	1978	1989	Ratio, 1989/ 1978
Value per dwelling unit (thousands of dollars):			
Arithmetic mean	26.5	61.3	2.31
Standard deviation	22.6	104.4	4.62
Median	21.8	43.4	1.99
Ratio, median to mean	.82	.71	....
Dwelling characteristics:			
Square feet per unit:			
Arithmetic mean	895	956	1.07
Standard deviation	227	202	.89
Median	850	924	1.09
Ratio, median to mean	.95	.97	....
Units per project:			
Arithmetic mean	91	113	1.24
Standard deviation	122	119	.98
Median	56	68	1.21
Ratio, median to mean	.62	.60	....
Bathrooms per unit:			
Arithmetic mean	1.28	1.55	1.21
Standard deviation	.35	.36	1.03
Median	1.11	1.53	1.38
Ratio, median to mean	.87	.99	....
Proportion of units [1]:			
With central air conditioning:			
Arithmetic mean	.59	.76	1.29
Standard deviation	.49	.43	.88
With covered parking space:			
Arithmetic mean	.30	.39	1.30
Standard deviation	.46	.48	1.04
In buildings of more than three stories:			
Arithmetic mean	.11	.09	.82
Standard deviation	.31	.29	.94
Intended for condominium sale:			
Arithmetic mean	.20	.17	.85
Standard deviation	.39	.37	.95

1. Median values of the proportion variables are all either zero or 1 and do not change over the period.

The first panel of table 1 illustrates the change in the shape of the distribution of value per unit in terms of conventional summary statistics. From 1978 to 1989, the standard deviation increased far more than the mean, reflecting the growing frequency of dwellings with values several times as large as the mean. The median increased less than the mean because it is less sensitive to these observations.

The lower panels of table 1 provide summary statistics for seven of the characteristics variables. The data measure "basic" rather than "luxury" characteristics; that is, they cover the organization of space into size of units, units per building, and height of buildings, and they cover a few features of that space, including one—air-conditioning—that was considered a luxury a few decades ago. The data do not cover many "luxury" features of units and projects that can have a substantial effect on value—such as quality of cabinetry, flooring, and appliances and the presence or absence of landscaping, swimming pools, or tennis courts.

In contrast to the value data, the seven characteristics variables show no tendency for standard deviations to increase relative to means; in fact, standard deviations declined while means increased for three of them. Nor is there a tendency for medians to grow less than means; in fact, medians grew more than means for two variables.

The growing skewness of the value data from 1978 to 1989, the absence of any such tendency in the measured characteristics, and the absence of "luxury" characteristics suggest a *growing importance of high-value units with substantial value attributable to unmeasured luxury characteristics*. Data to determine the extent, or even the existence, of such growth is not available. In the absence of such data, the goal is to construct a measure of price change that will not be distorted by such growth. This search for a distortion-free measure of price change is addressed in the next two sections.

Framework for Analysis

The hedonic approach starts from the premise that the value of a product reflects the values of its characteristics—either their utility to users or the cost of supplying them or both.[3] In applying the hedonic approach to the measurement of multifamily housing prices, two complications

3. The theory is set out in Sherwin Rosen, "Hedonic Price Indexes and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy* 82 (February 1974): 34–55.

arise. The first is the absence of measurements of luxury characteristics—a missing set of variables in the analysis. The second is that structures with the same physical characteristics often vary widely in price from one geographical area to another.

Basic and luxury characteristics

As just described, the Census Bureau data are consistent with the hypothesis that unmeasured luxury characteristics of multifamily housing—dwelling characteristics such as high-quality cabinetry and appliances, and project amenities such as landscaping and swimming pools—grew in value during 1978–89. There may be other explanations for the increasing skewness of the value data, but growth of unmeasured luxury characteristics is a plausible one. With that assumption, it is important to develop estimation strategies that reduce the influence that dwellings with these characteristics have on the aggregates used to construct price indexes. The procedures adopted by BEA make use of two such strategies:

(1) Reliance on medians or mean logarithms of values rather than arithmetic means, because medians and mean logarithms give less weight to the right tail of the value distribution, where luxury dwellings tend to be concentrated; and
(2) disaggregation of the sample to find a subsample with a relatively small shift to luxury dwellings.

Geographic differences

Because the "market" for structures is a collection of local markets rather than a single national market, prices for similar structures, unlike those for most other goods and services, differ widely among geographic areas for long periods of time. Ignoring geography at the very least increases the difficulty of uncovering relationships between structure characteristics and dollar values, and it often can distort results in more serious ways.

With a large enough data sample, a satisfactory approach to handling geographic differences is to estimate the relationship between value and characteristics for each market area. The revised Census Bureau price index for one-family houses takes a step in this direction by utilizing separate annual regressions for detached houses in each of four Census regions. For multifamily housing (like attached housing in the one-family index), the available data sample is not large enough for this option. Regression results for geographic areas (not shown in this paper) are much less stable than national regression results.

Another approach to handling geographic differences is to use "one-zero dummy" variables—that is, variables that equal one for observations in one region and zero for observations in other regions. The main shortcoming of this approach is uncertainty about whether the coefficients of the dummy variables represent price differences or quantity differences or both. The coefficient of a regional dummy variable can grow because housing in the region is improving in quality or because it is commanding a higher price (or both). If changes in the coefficient reflect changes in quality, they should not affect the estimation of a price index. If they represent changes in price, then they should.

A third approach, the one adopted by BEA, introduces explicit variables, rather than one-zero dummy variables, to represent cost and income differences among geographic areas. Regressors that measure cost differences—for example, State wage rates—reflect regional variation in supply (or offer) functions. Regressors that measure income differences—for example, State per capita disposable income—reflect regional variation in demand functions. The use of such regressors represents a modification of the hedonic theoretical framework, in which the only regressors are characteristics. The modification seems necessary in the case of housing in order to remove the influence of geographic "noise."

BEA used average weekly wages in the construction industry by State (based on records of the unemployment insurance system) as the principal variable to represent regional differences. However, the correlation of weekly construction wages, a cost variable, and State per capita income, an income variable, is quite high, and using State per capita income instead of weekly wages gave quite similar results. A third variable, which was derived using a set of regional cost factors published by a construction-cost estimating service, also gave similar results. In light of these similarities, it would be a mistake to interpret the resulting regression coefficients narrowly as reflecting only demand or only costs.

The other regional variable used is "cooling degree-days," which measures the frequency and extent of temperatures above 65° F in each State. The complementary variable, "heating degree-days," gave poorer results and added nothing to the overall fit when both variables were included. The coefficients of these climate variables could represent additional quantities of insulation and roofing strength of structures in cold climates (a separate air-conditioning variable reflects the additional costs of air-conditioning in warm climates). By adopting this interpretation, BEA treats cooling degree-days, unlike wage rates, as a quality variable.

The two geographic variables used in the regression surely capture only a part of value differences due to location. However, these variables were introduced not to represent locational factors fully in the regression, but rather to reduce bias in, and increase the statistical significance of, the coefficients of the characteristics variables.

Regression Form and Results

This section first briefly discusses the functional form of the relationship between value and characteristics. Next, it analyzes various ways of disaggregating the observations. It then presents some regression results for 1978–89.

Functional form

A fundamental question about functional form is whether to use value per unit or some transformation of value per unit as the dependent variable. Two considerations, one statistical and one economic, led to the adoption of a transformation (specifically, the logarithm) of value per unit.

The statistical reason for using a logarithmic dependent variable is that it gives much less weight to extremely high values than does an untransformed variable. If unmeasured luxury characteristics have been creating a growing number of high-value outliers, using logarithms instead of untransformed values will reduce the likelihood that such growth will be interpreted as an increase in prices.

The economic reason for choosing a logarithmic value per unit as the dependent variable is that it, unlike the untransformed value per unit, implies interaction among the independent variables. For example, the cost of air-conditioning would increase with unit size rather than staying constant. To take another example, an increase in the wage rate would increase the cost more for a large unit than for a small one and more for a unit with two bathrooms than for a unit with one.

Three of the independent variables—square feet per unit, number of units, and relative wages—also appear in logarithmic form. Using linear rather than logarithmic forms for these variables gives very similar results, but the logarithmic form has the convenient property that the coefficients of these variables can be interpreted as elasticities.

Most of the other independent variables appear as proportions of total units. Their regression coefficients can be interpreted as the percent increase in value of a typical unit as a result of having a particular attribute. Thus, a coefficient of 0.09 for the proportion of units having central air-conditioning means that central air-conditioning typically raises the value of a unit by 9 percent. The simplicity of this interpretation is one reason to have these independent variables appear as proportions; another is that a logarithmic form is not possible for those characteristic variables that take on values of zero (for example, the absence of central air-conditioning) for some observations.

Disaggregation

The major changes over time in the distribution of value per unit (shown in table 1) apply to some subgroups of dwellings much more than to others. Regression results for subgroups with relatively stable distributions should be less biased by unmeasured luxury characteristics than results for subgroups whose distributions have changed.

Table 2 disaggregates the observations in three different ways: High-rise buildings versus low-rise buildings, California and Florida versus all

Table 2.—Disaggregated Value Per Dwelling Unit, 1978 and 1989: Means and Standard Deviations

	Number of observations		Mean		Standard deviation	
	1978	1989	1978	1989	1978	1989
All observations	1,453	1,018				
Value per unit			26.50	61.30	22.60	104.40
Log of value per unit			3.14	3.86	.46	.52
High-rise	162	95				
Value per unit			33.70	63.10	13.80	46.10
Log of value per unit			3.45	4.05	.36	.37
Low-rise	1,291	923				
Value per unit			25.60	61.10	23.30	108.60
Log of value per unit			3.10	3.85	.46	.53
California and Florida	558	484				
Value per unit			29.10	55.00	21.80	85.30
Log of value per unit			3.25	3.77	.46	.51
All other locations	895	534				
Value per unit			24.90	68.20	22.90	121.70
Log of value per unit			3.08	3.97	.45	.52
Condominium units	284	174				
Value per unit			40.60	118.70	32.70	205.50
Log of value per unit			3.57	4.30	.48	.79
Rental units	1,169	844				
Value per unit			23.10	49.50	17.80	60.50
Log of value per unit			3.04	3.78	.40	.40

NOTE.—Values per unit in thousands of dollars.

other locations, and condominium units[4] versus rental units. A critical statistic in the table is the standard deviation of the logarithm, which measures the percentage dispersion of the dependent variable in hedonic regressions. Groupings with little or no change in this measure are more promising candidates for hedonic analysis than groupings with substantial increases in the measure, especially because the percentage dispersion of the main independent variables has changed little.

The last grouping in the table, rental units and condominium units, is the most promising. Rental units show no increase in the standard deviation of the logarithm, but condominium units show a substantial increase. This result suggests that a price index based on rental unit values might be relatively free of distortion due to unmeasured luxury characteristics. The first grouping in the table contains a subgroup, high-rise units, that also has very little change in the standard deviation of the logarithm; however, the sample of high-rise units is less desirable for regression analysis because it is considerably smaller than the sample of rental units.

4. For this article, "condominiums" refers to multifamily units that are built for sale, including co-operatives.

Rental projects

The regression results for rental projects appear in table 3. Three features of the coefficients are of special interest: Their plausibility, their statistical significance, and their stability from year to year. The overall degree of correlation (R^2) is also of interest, because declining correlation over time indicates that the independent variables explain less of the variation in value per unit.

For rental units, the most significant coefficient in every regression is that of the logarithm of square feet per unit. The coefficients range from 0.648 to 0.920, with most of them significantly different from 1.0 as well as from zero. Coefficients less than 1.0 imply that there are economies of scale in building large units: A large unit typically has a lower value per square foot than a small unit having the same other characteristics. Such results suggest that aggregate value per square foot can be a misleading indicator of price.

The coefficient of the logarithm of the number of units per project measures economies (or diseconomies) of scale in the size of entire projects. The negative (usually quite statistically significant) coefficients for this characteristic imply that there are small economies of scale in building large projects.

Table 3.—Regression Results, Rental Projects

[Coefficients, with t-ratios in parentheses]

Variables	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
Intercept	–1.898	–1.927	–2.410	–2.748	–1.750	–1.694	–1.240	–0.914	–1.210	–1.296	–1.073	–1.305
	(5.3)	(5.6)	(6.4)	(6.0)	(3.5)	(4.2)	(3.2)	(2.5)	(3.7)	(3.6)	(2.9)	(2.8)
Square feet per unit (logarithm)	.730	.766	.876	.920	.801	.787	.716	.648	.696	.731	.732	.775
	(12.7)	(13.9)	(14.5)	(12.5)	(10.0)	(12.4)	(11.6)	(11.3)	(13.3)	(12.7)	(12.4)	(10.6)
Units per project (logarithm)	–.024	–.036	–.045	–.026	–.048	–.044	–.034	–.022	–.011	–.012	–.039	–.039
	(2.9)	(4.3)	(4.8)	(2.5)	(4.2)	(4.5)	(3.9)	(2.6)	(1.4)	(1.4)	(4.4)	(4.0)
Bathrooms per unit (number)	.146	.157	.102	.052	.026	.109	.146	.175	.124	.105	.129	.009
	(3.8)	(4.1)	(2.3)	(1.0)	(0.5)	(2.6)	(3.9)	(5.0)	(3.7)	(2.9)	(3.3)	(0.2)
Bedrooms per unit (number)	–.040	–.076	–.110	–.077	–.037	–.062	–.076	–.059	–.039	–.073	–.118	–.074
	(1.5)	(2.9)	(3.6)	(2.3)	(1.1)	(2.1)	(2.6)	(2.1)	(1.5)	(2.5)	(3.7)	(2.1)
Central air-conditioning (percent of units)	.018	–.003	.014	.042	.029	.040	.045	.072	.104	.054	.013	.019
	(0.8)	(0.1)	(0.6)	(1.4)	(0.9)	(1.4)	(1.6)	(3.0)	(4.6)	(2.2)	(0.5)	(0.7)
Parking structure (percent of units)	.117	.078	.046	.097	.074	.069	.002	–.009	.055	.053	.023	.124
	(5.0)	(3.4)	(1.7)	(2.8)	(2.0)	(2.3)	(0.1)	(0.4)	(2.4)	(2.1)	(0.9)	(4.0)
Building greater than three stories (percent of units)	.428	.369	.330	.365	.394	.407	.389	.304	.218	.205	.228	.186
	(13.3)	(12.9)	(11.2)	(10.3)	(10.3)	(10.9)	(9.6)	(7.4)	(5.8)	(5.7)	(6.3)	(4.6)
By State in which project located:												
Construction weekly wage (log of relative wage)	.234	.089	.177	.344	.524	.359	.192	.336	.339	.496	.563	.538
	(2.7)	(1.1)	(2.1)	(3.5)	(5.0)	(5.0)	(3.1)	(4.2)	(5.6)	(7.0)	(8.1)	(6.9)
Cooling degree-days (thousands)	–.054	–.058	–.053	–.040	–.056	–.066	–.061	–.027	–.044	–.033	–.009	.002
	(4.0)	(4.8)	(4.3)	(2.7)	(3.8)	(5.6)	(5.7)	(2.3)	(4.6)	(3.0)	(0.8)	(0.2)
Summary statistics:												
R^2 (corrected)	.40	.37	.41	.42	.41	.36	.26	.22	.29	.31	.34	.34
Standard error of the equation	.31	.32	.31	.32	.32	.31	.34	.35	.33	.32	.31	.32
Number of observations	1,169	1,373	1,044	796	802	1,142	1,347	1,569	1,568	1,187	1,003	844

NOTE.—The dependent variable is the natural logarithm of dollar value per dwelling unit. The unit of observation is a project of five or more dwelling units under construction during the year of the regression (projects under construction during more than one calendar year may serve as observations in more than one regression).

The coefficient of bathrooms per unit measures the percentage increase in value associated with an additional bathroom in a unit. The coefficients are implausibly low (and not significant) in a few years; however, they are positive throughout, and in most years they imply that an additional bathroom adds 10–15 percent to the value of a unit.

The negative and statistically significant coefficients for bedrooms per unit may appear puzzling, because adding a bedroom to a unit clearly increases its cost. However, this coefficient measures the cost differential of adding a bedroom while holding constant the size of the unit; it can therefore indicate whether bedrooms cost more or less per square foot than other uses of dwelling space. The negative coefficients indicate, plausibly, that they cost less.

The coefficients of central air-conditioning and of a parking structure measure the percentage increase in value associated with the presence of these two features. The coefficients are generally positive, but they vary somewhat erratically. There appears to be some tendency for one of the two coefficients to be high when the other is low.

The coefficients of "building greater than three stories" measures the percentage increase in value associated with a unit's being in a high-rise rather than a low-rise building. The coefficients are high and statistically significant in all years, but they tend to decrease over time, perhaps signifying technological improvement in high-rise construction.

The wage variable is measured as the logarithm of the average weekly wage of construction workers in the State in which each observation is located, *minus* its mean for all observations. (Subtracting the mean affects neither its coefficient nor any of the other coefficients except the intercept, but it facilitates the conversion of the regression results into an index.) In the first 3 years of the sample, the weekly wage refers to the entire construction industry; in the last 9, to residential construction only.

The wage coefficient measures the percent increase in value associated with a 1-percent increase in average wages. A plausible estimate of the coefficient would be the share of onsite labor costs in construction value. The coefficients in table 3 are lower than that share, possibly because the production function is complex. A more likely reason, however, is that the wage vari-

Table 4.—Regression Results, Condominium Units

[Coefficients, with t-ratios in parentheses]

Variables	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
Intercept	–1.387	–2.197	–3.304	–3.724	–2.428	–1.010	–1.573	–0.263	–0.385	–0.010	2.228	4.840
	(1.5)	(2.8)	(5.0)	(5.7)	(2.5)	(1.1)	(1.9)	(0.2)	(0.3)	(0.0)	(1.2)	(2.0)
Square feet per unit (logarithm)	.705	.795	.999	1.077	.895	.694	.771	.626	.582	.557	.238	–.112
	(4.8)	(6.4)	(9.5)	(10.5)	(6.0)	(4.7)	(5.8)	(3.5)	(2.7)	(2.4)	(0.8)	(0.3)
Units per project (logarithm)	–.075	–.075	–.034	–.025	–.065	–.070	–.041	–.076	–.033	–.070	–.106	–.108
	(3.3)	(4.1)	(2.2)	(1.6)	(2.8)	(2.9)	(1.7)	(2.3)	(0.8)	(1.5)	(1.8)	(1.7)
Bathrooms per unit (number)	.269	.303	.108	.048	.084	.232	.081	.031	.086	.127	.018	.174
	(3.1)	(4.1)	(1.7)	(0.8)	(0.9)	(2.5)	(0.9)	(0.3)	(0.6)	(0.7)	(0.1)	(0.7)
Bedrooms per unit (number)	–.103	.007	–.021	–.045	.044	–.014	.089	.063	.125	.085	.278	.057
	(1.5)	(0.1)	(0.4)	(0.9)	(0.6)	(0.2)	(1.3)	(0.7)	(1.1)	(0.7)	(1.7)	(0.3)
Central air-conditioning (percent of units)	.030	.054	.098	.102	.206	.180	.143	.236	.348	.232	.126	.152
	(0.5)	(1.1)	(2.3)	(2.2)	(2.8)	(2.5)	(2.1)	(2.5)	(3.3)	(2.0)	(0.8)	(0.8)
Parking structure (percent of units)	.078	.045	.040	.020	–.012	.019	–.016	–.051	–.065	–.159	–.082	–.054
	(1.4)	(1.0)	(1.0)	(0.5)	(0.2)	(0.3)	(0.3)	(0.7)	(0.7)	(1.5)	(0.6)	(0.4)
Building greater than three stories (percent of units)	.158	.153	.065	.118	.146	.109	.050	.060	–.015	.086	.071	.216
	(1.8)	(2.3)	(1.2)	(2.2)	(1.9)	(1.4)	(0.6)	(0.6)	(0.1)	(0.6)	(0.3)	(0.8)
By State in which project located:												
Construction weekly wage (log of relative wage)	.705	.897	.788	.971	.585	.219	.244	.307	.143	.203	.267	.856
	(2.7)	(4.7)	(5.0)	(6.8)	(2.7)	(1.3)	(1.5)	(1.1)	(0.5)	(0.7)	(0.7)	(2.1)
Cooling degree-days (thousands)	–.019	–.005	–.030	–.008	–.037	–.082	–.038	–.057	–.027	.036	.071	.034
	(0.5)	(0.2)	(1.4)	(0.4)	(1.2)	(3.1)	(1.6)	(1.6)	(0.7)	(0.7)	(1.0)	(0.5)
Summary statistics:												
R^2 (corrected)	.33	.34	.33	.34	.23	.17	.15	.08	.09	.07	.03	.00
Standard error of the equation	.39	.42	.40	.43	.53	.57	.59	.71	.73	.76	.79	.79
Number of observations	284	490	622	680	511	601	666	504	362	273	207	174

NOTE.—The dependent variable is the natural logarithm of dollar value per dwelling unit. The unit of observation is a project of five or more dwelling units under construction during the year of the regression (projects under construction during more than one calendar year may serve as observations in more than one regression).

able is only an approximation of the labor cost conditions facing each project. Among the shortcomings of the variable are its omission of fringe benefits and its failure to reflect wage differences within a State.

The negative coefficients of cooling degree-days for all years except 1989 indicate that, holding other characteristics (including air-conditioning) constant, costs of construction are lower in warmer climates.

Several available and possibly relevant variables do not appear in the regressions. Two of these omitted variables—dummy variables for statistical metropolitan area and for central city location—consistently failed to have significant coefficients of the expected sign. Three others were mentioned earlier: Per capita income by State and a set of State cost factors published by a construction cost firm were closely correlated with the income variable, State construction wage rates, and heating degree-days consistently gave poorer results than the included variable, cooling degree-days.

Values of R^2 in table 3 fall moderately, implying that the included variables explain less of the variation in (logarithms of) values in the late 1980's than in the late 1970's. Equation standard errors, however, do not fall, indicating that R^2 values are falling not because of increasing uncertainty about the effects of the included variables, but rather because of growing dispersion in the distribution of values resulting from other factors.

Condominium projects

Regression results for condominium projects, which appear in table 4, are much more erratic than the results for rental projects. Coefficients of the logarithm of square feet, for example, range from −0.112 to 1.077 for condominium projects, compared with 0.648 to 0.920 for rental projects. Values of R^2 plummet for condominiums, falling to zero in 1989, whereas they drop only moderately for rentals. Furthermore, equation standard errors double for condominiums, whereas they remain stable for rentals.

The results for condominium projects are too weak to warrant any confidence in price indexes derived from them. The following section discusses such indexes mainly to support the hypothesis that missing luxury variables are an important influence on the multifamily data.

Converting Regression Results Into Price Indexes

A general procedure

Two steps are necessary to convert hedonic regression results into a price index, though sometimes the two can be combined into a single computation. The first step is to use regression coefficients for each year to transform the dollar value of each observation that year into an estimate of the dollar value of some specified "standard" unit. The second step is to construct an annual index based on the average, or some other measure of central tendency, of these transformed values. The index will vary depending on the selection of a "standard" unit and on the selection of a measure of central tendency.

The algebra of this procedure can best be described with a simple example. Suppose that the value, V_{it}, of the ith dwelling in the sample of units in year t depends on a constant term, a_t; on a coefficient, b_t, times the number of square feet of the dwelling, SF_{it}; and on an error term, u_{it}, uncorrelated with SF_{it}:

$$(1) \qquad V_{it} = a_t + b_t * SF_{it} + u_{it}.$$

In this example, a price index will be estimated for a unit of Z square feet (the average size in some base year).

The first step is to transform V_{it} into an estimate of the price, P_{it}, of a Z-square-foot unit by adjusting V_{it} for the difference between its actual size and Z:

$$(2) \qquad P_{it} = V_{it} - b_t * (SF_{it} - Z).$$

Each P_{it} corresponds to the observed price of a narrowly specified good from various outlets in a standard price index calculation.

The second step is to calculate some measure of the central tendency of the P_{it} for each year and base a price index on these measures, where I_t refers to the index for year t, $CT(P_{it})$ for the central tendency of observed prices in t, and $CT(P_{ib})$ for the central tendency of observed prices in the base year:

$$(3) \qquad I_t = 100 * [CT(P_{it})/CT(P_{ib})].$$

In this simple linear example, the two steps can be accomplished in one set of computations if the measure of central tendency is the arithmetic mean. Taking the mean of each term in equation 2 and substituting for V from the means of equation 1 gives the following, where P_t and u_t are the means of P_{it} and u_{it}:

$$(4) \qquad P_t = a_t + b_t * Z + u_t.$$

Because the regression procedure is based on the assumption that u_t is zero each year, a price index can be calculated using equation 4 for the current year t and the base year b:

$$(5) I_t = 100 * (a_t + b_t * Z)/(a_b + b_b * Z).$$

However, in more complex cases, these two steps cannot be combined into one.

Changing the standard set of characteristics (Z in the simple example) used in the first step is similar to changing the base period for a price index. The usual propositions about base period changes apply: Generally, shifting from a set typical of some early year to a set typical of a later year involves giving greater weight to characteristics whose relative prices have fallen and therefore lowers the rate of growth of a price index.

The choice of a measure of central tendency used in the second step can significantly affect the results of the regression procedure if there is a change in the underlying distribution of price estimates.[5] Changes in distribution may reflect changes in price or in quality. The measure of central tendency chosen should, as far as possible, reflect price changes and exclude quality changes. Especially relevant to this study is a change in the distribution of estimated prices that is likely to have been caused by changes in the unmeasured characteristics that affect the right tail. Because such a change in distribution reflects changes in quality rather than in price, the best choice of a measure of central tendency in this case is one that is relatively insensitive to the behavior of the right tail.

Application to multifamily housing

The first step in applying the procedure outlined above to multifamily housing is to convert each observation for each year into the estimated price of a "standard" unit. In developing the BEA index, the procedure was applied for two standards: Average values of characteristics in 1982 and in 1987. These standards yielded 1982-weighted and 1987-weighted price indexes.

Because of the logarithmic form of the regressions, it is necessary to calculate the logarithm of estimated prices before calculating the estimated prices themselves. The logarithm of the estimated price of each is equal to the logarithm of the actual value per unit corrected for differences between the characteristics of the observed unit and the characteristics of the standard unit (see equation 2). The estimated price is then simply the antilog of this calculated logarithm.

The second step in applying the procedure is to calculate indexes based on some measure of the central tendency of these estimated prices in each year (see equation 3). Three such measures were tested: The mean, the median, and the antilog of the mean logarithm. Of the three measures, an index based on means gives high-value units the greatest weight, and an index based on medians gives them the least weight; an index based on the antilog of the mean logarithm is intermediate in this regard. Differences among these indexes indicate how sensitive the results are to the choice of central tendency.[6]

This procedure produces six indexes: Two "standard" units (1982 characteristics and 1987 characteristics), each using three measures of central tendency. Calculating these 6 indexes for both rental and condominium units produces the 12 indexes shown in table 5. The last two lines of the table, which show average annual rates of price increase, conveniently summarize three important features of the data.

- For rental projects, differences due to choice of central tendency are fairly small. The median-based and logarithm-based growth rates are almost identical (5.97 percent and 5.94 percent per year, for the 1987 standard), and the mean-based growth rates are only a few tenths of a percentage point higher.
- For condominium projects, differences due to choice of central tendency are large. The median-based growth rates are 1.2 percentage points per year less than the logarithm-based growth rates (5.31 percent and 6.47 percent for the 1987 standard), and the logarithm-based rates are, in turn, 3.8 percentage points less than the mean-based growth rates.
- Differences due to weight year are in the usual direction (lower growth for the 1987 standard than for the 1982 standard) but

5. Clearly, the choice does not matter if the distribution of estimated prices does not change (in the sense that each percentile of the distribution in year 2 is a constant multiple of the corresponding percentile in year 1). In this case, different measures of central tendency may differ in level, but they will all change by the same percentage and will therefore show the same index value relative to a base year.

6. The third measure, the antilog of the mean logarithm, has the statistical advantage that the unexplained variation in the logarithms has been assumed, in the logarithmic regression specification, to average zero. The other two measures, the mean and the median, are based on values for which the unexplained variation has not been assumed to average zero.

amount to only a few tenths of a percentage point per year.

These and other results are shown in charts 1 and 2. Chart 1 shows five indexes of multifamily prices for 1978–89. In addition to 1987-weighted hedonic indexes based on the mean, on the median, and on logarithms, the chart shows two cruder price indexes that are uncorrected for changes in quality. One is simply an index of value per unit, and the other is an index of value per square foot (calculated as average value per unit divided by average square feet per unit). For rental units, shown in the top panel of the chart, the index of value per unit rises faster than the other indexes beginning in 1984. The other indexes have mostly similar rates of growth.

Differences among the indexes are far greater for condominiums, shown in the bottom panel, than for rental units. Three of the indexes—value per unit, value per square foot, and hedonic based on means—rise much faster than the hedonic indexes based on medians and on logarithms.

Table 5.—Experimental Price Indexes, Multifamily Housing

[1978=100]

	1982 standard			1987 standard		
	Mean	Median	Antilog [1]	Mean	Median	Antilog [1]
Rental units:						
1978	100.0	100.0	100.0	100.0	100.0	100.0
1979	109.4	109.9	108.7	109.1	109.5	108.4
1980	121.9	123.8	121.3	119.6	121.5	119.0
1981	133.4	134.8	132.0	130.5	131.9	129.2
1982	146.1	146.5	144.7	142.6	143.0	141.2
1983	150.9	151.1	150.1	149.8	149.9	149.0
1984	161.6	159.1	157.5	157.2	154.8	153.2
1985	172.2	166.0	165.6	165.4	159.4	159.0
1986	174.9	171.5	171.2	171.7	168.4	168.0
1987	184.5	181.9	180.0	178.8	176.3	174.5
1988	197.1	196.6	195.5	188.2	187.6	186.7
1989	204.2	197.2	196.6	196.1	189.2	188.7
Condominium units:						
1978	100.0	100.0	100.0	100.0	100.0	100.0
1979	120.2	108.2	110.6	120.8	108.7	111.2
1980	126.2	117.6	119.6	122.3	114.0	115.9
1981	134.7	126.0	126.5	127.4	119.2	119.6
1982	169.9	138.2	144.4	164.4	133.7	139.7
1983	184.5	141.2	150.4	185.3	141.8	151.0
1984	196.2	145.4	158.0	190.6	141.3	153.5
1985	249.3	148.8	171.1	240.7	143.6	165.2
1986	272.6	158.9	182.4	258.0	150.5	172.7
1987	299.4	168.4	191.7	279.1	157.0	178.7
1988	303.9	175.3	204.2	291.2	168.0	195.7
1989	303.8	182.8	206.1	293.8	176.8	199.3
Average annual percentage increases, 1978–89:						
Rental	6.71	6.37	6.34	6.31	5.97	5.94
Condominium	10.63	5.64	6.79	10.29	5.31	6.47

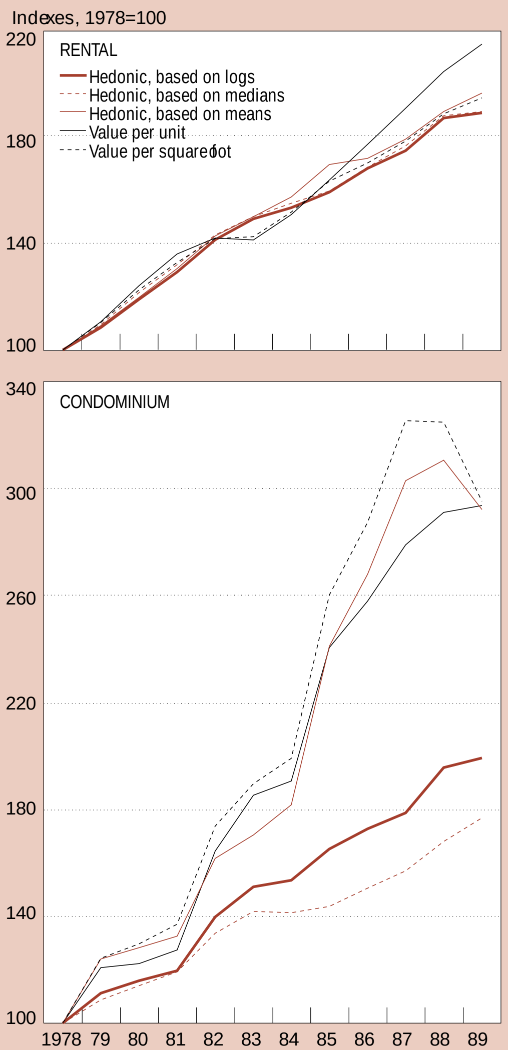
1. Antilog of the mean logarithm.

Of the latter indexes, the logarithm-based index rises a bit faster than the corresponding rental indexes, and the median-based index rises a bit slower.

Chart 2 compares changes in the Census Bureau index for one-family housing with changes in the logarithm-based multifamily indexes for rental housing and for condominium housing. All three are based on 1987 standard dwellings and show some deceleration during the 11-year period. The condominium index is the most volatile, showing sharp accelerations in 1982, 1985, and 1988; the other two indexes are usually within 2 percentage points of one another, but with a few conspicuous exceptions.

CHART 1

Multifamily Price Indexes, 1978-89



Note.--The hedonic indexes are based on 1987 quantity weights

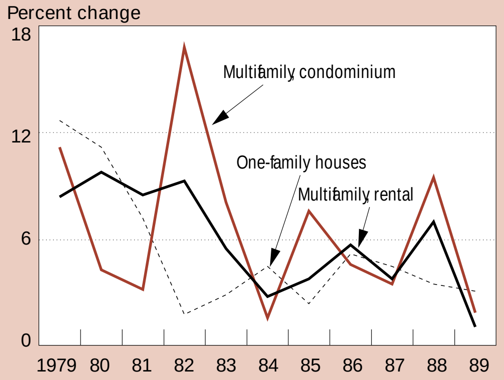
U.S. Department of Commerce, Bureau of Economic Analysis

The condominium index's volatility and extreme sensitivity to the choice of central tendency suggest that growth of unmeasured luxury characteristics may have been substantial for condominium projects. These results strengthen the case against relying on the condominium results.

In light of these results, BEA now uses the logarithm-based index for rental units (shown in table 5), converted to a 1987 base, to deflate expenditures for multifamily housing—both condominiums and rental units. The rental unit index is both interpolated quarterly and extrapolated outside of the period of fit using the Census Bureau index for one-family units (excluding land). BEA is testing a procedure to develop annual and quarterly Paasche indexes for rental units, using information from the regression equation. The regression equation will be updated in preparation for the next comprehensive revision of the NIPA's.

CHART 2

Residential Price Indexes, 1979-89



U.S. Department of Commerce, Bureau of Economic Analysis

A note of caution about the new multifamily price index is warranted: Although this index seems reliable for the present, it may not continue to be so. If more items that are now considered luxuries become standard in future years—as air-conditioning has become in recent decades and as central heating and indoor plumbing became in the past—then it will become necessary to collect data on those characteristics rather than to use a procedure that relies on medians or logarithms and on the omission of condominiums. For example, collecting actual data on swimming pools will become necessary if the proportion of units with access to a swimming pool rises in the future to 40 or 50 percent; in this case, relying on medians or logarithms rather than on arithmetic means will no longer yield an unbiased index.