ESTIMATING INTER-CITY DIFFERENCES IN THE PRICE OF HOUSING SERVICES*

Harvey S. Rosen

Econometric Research Program
Research Memorandum No. 231

September 1978

* I would like to thank Richard Quandt, A. Mitchell Polinsky, Roger Gordon, and two referees for useful comments, and N. Greg Mankiw for assistance with the computations. This project was funded in part by a grant from the Hoover Institution on War, Revolution and Peace. A portion of the research was completed while I was a Visiting Fellow at The Institute for Advanced Studies, Hebrew University of Jerusalem.

Econometric Research Program
Princeton University
207 Dickinson Hall
Princeton, New Jersey

I. Introduction

In an important article published several years ago, Muth (11) used the economic theory of production to develop a procedure for estimating inter-city differences in the price of housing services. Since Muth's goal was other than to develop a price index for housing, he did not show the ranking of cities implied by their prices of housing services. The purposes of the present paper are to:

a) extend Muth's method by employing a more general model of housing technology,

b) rank a group of American cities according to their prices of housing services, and

c) compare the ranking with that implied by the well-known housing price figures published by the Bureau of Labor Statistics (BLS).

In Section II are described several different approaches to measuring inter-city differences in the price of housing services, including Muth's. Section III outlines the method used in this paper, and contains the empirical results. A brief concluding section lists some caveats and suggestions for future research.

II. Conceptual Problems in Measuring the Price of Housing

The fundamental source of difficulties in calculating the price of 'housing' is its heterogeneous nature. A house is a bundle of many 'characteristics' such as number of rooms, amount of floor space, etc., and there is enormous variation in the ways these characteristics are combined. In this section, we discuss how the heterogeneity problem is handled in three different approaches to computing the price of housing services: the Bureau of Labor Statistics approach, the hedonic approach, and finally, the Muth approach, which forms the basis for the analysis of this paper.

BUREAU OF LABOR STATISTICS METHOD

The BLS deals with the heterogeneity problem by specifying *a priori* a particular set of housing characteristics, and computing for each city the annual cost of living in a house with those characteristics. (More precisely, the BLS calculates the annual expenses of dwelling in several types of houses - the standard house for a 'moderate' income family involves less expense than that for a 'high' income family.) According to the BLS,

> For homeowner families, the costs of maintaining the shelter standards were calculated for a five - or six-room house, with 1 - or 1½ baths for the moderate, and 1 or more baths for the higher standard. Both standards called for a fully equipped kitchen, hot and cold running water, electricity, and central or other installed heating.[1]

An important problem is inherent in this approach to measuring housing prices. By examining the price of the same standard unit across cities, the BLS implicitly rules out the possibility of factor substitution in the production of housing across cities. In effect, a fixed coefficient production technology is assumed. Thus, the structure whose price is used in the BLS calculation may not in fact be 'typical' of a given community. Moreover, although all units are the same with respect to the few characteristics used by the BLS, they are not necessarily the same with respect to a much larger number of other characteristics. There may be systematic differences in excluded characteristics across cities.

HEDONIC PRICE EQUATION METHOD

In light of these considerations, it would seem desirable to construct a housing price index using a procedure which explicitly allows for the fact that the characteristics of the housing stock may vary from city to city.

The tack taken by Gillingham (6) improves upon the BLS approach by taking account of more characteristics. The first step of the Gillingham procedure is to estimate a hedonic price equation for housing. The regression coefficient of a given characteristic is an estimate of its implicit price (see (7) or (10)).

Gillingham uses the results from hedonic price equations to construct a price index for rental housing[2] as follows: Let x_s denote the vector of average characteristics of housing in city s. An hedonic price equation estimated for city t can be used to estimate what the price of the unit x_s would be if it were situated in city t. The ratio of this figure to the price of x_s in city s is Gillingham's price index.

Although more systematic than the BLS's 'typical dwelling' approach, this technique suffers from several problems. In theory, the ordering of cities may depend upon which city is used as the point of reference; in fact, Gillingham's rankings do vary considerably as the reference city changes. Moreoever, as Alexander (1) has pointed out, the choice of characteristics for the hedonic price equation may be quite arbitrary.[3]

MUTH'S APPROACH

Rather than specifying a list of housing characteristics, Muth deals with the heterogeneity problem by introducing the concept of 'housing services'. Structures with different characteristics are viewed as producing different quantities of the <u>homogeneous</u> commodity 'housing services'.[4] The housing market, then, is one in which a homogeneous product is traded.

For purposes of developing a price index for housing, the advantage of Muth's approach is that it indicates precisely the number that is needed for each city - the average price per unit of housing services. Unlike the hedonic or BLS techniques, there is no need to establish what the 'standard' housing unit looks like. As Olsen (12) notes, "... in this theory there is no distinction between the quantity and quality of a dwelling unit as these terms are customarily used." (p.613) A better than average house is simply one that yields more units of housing services than average, but the price per unit of housing services at a given location is the same for all dwellings in long run competitive equilibrium ((12), p. 614).

The main problem is that the price of a unit of housing services is not directly observable. However, Muth (11) indicates how it can be estimated. The most important assumption required for estimation is the existence of a constant returns to scale housing production function. Specifically, Muth assumes that land (L) and built structures (S) are inputs into the production of housing (Q),where Q is the capacity to provide a flow of housing services ((11), p. 244). Polinsky and Ellwood (13) have suggested that it is more convenient to think of the problem in terms of the dual of the production function, the unit cost function,[5] and our analysis will be conducted within that framework.

Let the prices of structures and land be P_S and P_L, respectively. Following Muth (11) and Polinsky and Ellwood (13), assume constant returns to scale. Then the unit cost function, $C(\cdot)$, gives the minimum average cost of producing housing services as a function of those prices.

(1) $C = C(P_S, P_L; \rho)$;

where ρ is the vector of parameters that determine the shape of the cost function.[6]

Given a specific functional form for $C(\cdot)$ and the value of ρ, it is possible to determine the unit cost of housing for a particular city by substituting into (1) the values of P_S and P_L for that city. Assuming long run equilibrium, the price of housing services is equal to their unit cost. (Many individuals would prefer a procedure which does not require this equilibrium assumption. Unfortunately, little is known about price determination outside of equilibrium.) Repeating the process for each city, we can generate a series of housing service prices, and hence, our index.

Of course, we come to the problem knowing either $C(\cdot)$ nor ρ. Some 'reasonable' specification for $C(\cdot)$ must be taken as a maintained hypothesis, and ρ estimated conditional on that hypothesis. Deferring for the moment the specification of $C(\cdot)$, consider the problem of estimating ρ conditional on a specific functional form. To do so, it is useful to take advantage of Shepherd's lemma, which states that the derivative of the cost function with respect to its ith argument is the demand for the ith factor:

$$(2) \quad S = \frac{\partial C}{\partial P_S}(P_S, P_L; \rho) = f(P_S, P_L; \rho)$$

$$(3) \quad L = \frac{\partial C}{\partial P_L}(P_S, P_L; \rho) = g(P_S, P_L; \rho)$$

If cross-sectional observations on the variables S, L, P_S and P_L can be obtained, then the vector ρ is amenable in principle to econometric estimation.[7]

In short, given estimates of the parameters of the factor demand equations, one can infer the parameters of the underlying cost function.[8] Similar procedures are common in econometric work on consumer demand theory, where estimates of commodity demand equations are used to infer the parameters of the underlying utility function. (See, e.g., (16).)

Muth (11) and Polinsky and Ellwood (13)[9] compute unit prices of housing services on the assumption that $C(\cdot)$ can be characterized by a constant elasticity of substitution (CES) specification.[10] Although the CES function is more general than the popular Cobb-Douglas form, it constrains the elasticity of substitution in production to be constant regardless of the ratio of structures to land. The procedure used below draws upon the basic framework which has been described in this section, but it uses a more general characterization of technology.

III. Model and Results

In this section, we specify the housing unit cost function to be estimated, discuss the data, and present the parameter estimates. The econometric results are then used to generate estimates of cities' unit costs of housing.

A. The Model

Assume that the relationship between the unit cost of housing and the prices of structures (P_S) and land (P_L) can be represented by the translog function:

$$(4) \quad \ln C = \alpha_0 + \alpha_S \ln P_S + \alpha_L \ln P_L + \frac{1}{2}\delta_{SS}(\ln P_S)^2 + \delta_{SL}\ln P_L \ln P_S + \frac{1}{2}\delta_{LL}(\ln P_L)^2 ,$$

where the α's and δ's are parameters, and the other variables are defined above. The translog function has been used in a number of studies of both consumer and production behavior (e.g., (3), (4)). It is a flexible functional form which does not constrain the elasticity of substitution to be a constant independent of factor ratios.

As Burgess (3) has emphasized, since the translog cost and production functions are not self-dual, there is some arbitrariness in selecting the cost rather than the production function as the starting point for analysis. Because our ultimate goal is to estimate unit costs, the cost function is chosen for the sake of convenience.

Taking the derivative of (4) with respect to the ith price (i = S,L) and rearranging yields

$$(5) \quad M_S = \alpha_S + \delta_{SS}\ln P_S + \delta_{SL}\ln P_L$$

$$(6) \quad M_L = \alpha_L + \delta_{SL}\ln P_S + \delta_{LL}\ln P_L$$

where M_i is the cost share of the ith factor (i = S,L). Assuming linear homogeneity,

$$(7) \quad \alpha_S + \alpha_L = 1$$

$$\delta_{SS} + \delta_{SL} = 0$$

$$\delta_{LL} + \delta_{SL} = 0.$$

Because of the adding up constraint and the symmetry of (δ_{SL}), it is necessary to estimate only one of the equations (5) and (6); i.e. the estimates are invariant to the equation dropped. (See (3).) Selecting (6) as the equation to be estimated and imposing the constraints (7),

(8) $M_L = \alpha_L + \delta_{LL} \ln(P_L/P_S) + \varepsilon$

where ε is a random error term which represents errors in optimizing behavior.

B. Data[11]

The data for this study consist of a subsample of households which purchased single family homes in 1969 and whose mortgages were insured by the Federal Housing Administration (FHA) under its Section 203 program. There were 10,054 households from 31 metropolitan areas. Only frame houses were included in the sample because it was only for these that the full set of price data was available.[12] Although these data are the best currently available for a study of this kind, it should be noted that the sample is not entirely representative, because homes purchased under FHA mortgages may not be typical of the entire housing stock. (See (11) on this.) It is not obvious how the production of FHA insured homes might differ from the rest of the housing stock. This issue should be investigated when more complete data become available.

The variables of equation (8) are defined as follows:[13] M_L is the value of the lot upon which the house stands divided by the value of the house. The value of the house is the sales price minus any closing costs, and the value of the lot is the FHA's estimate of the price for an equivalent site. The price of land, P_L, is the lot value divided by the lot size in square feet.[14] P_S, the price of structures, is the Boeckh building cost index number for frame residential structures. P_S is computed on a metropolitan area basis. The Boeckh index is not ideal for our purposes, because it suffers from problems similar to the BLS index - it is based on the cost of building a house with certain characteristics in different cities. Although attempts are made to cost identical structures, it is clear that the particular results generated below must be regarded with caution due to possible errors in the measurement of P_S.

C. Results

When (9) is estimated by ordinary least squares[15] we find

$$\text{(9)}\quad M_L = \underset{(.00579)}{.854} + \underset{(.000864)}{.0959} \ln(P_L/P_S)$$

$$R^2 = .55 \qquad SSR = 23.8 \qquad N = 10{,}054$$

where the numbers in parentheses are standard errors. The R^2 of 0.55 is comparable in magnitude to those of other studies which analyze micro observations.

According to (9) and the constraints (7), $\delta_{SS} = \delta_{LL} = .0959$, $\delta_{SL} = -.0959$, $\alpha_L = .854$, and $\alpha_S = .146$. Using these results and the fact that the elasticity of substitution σ is given by[16]

$$\text{(10)}\quad \sigma = (\delta_{SL} + M_S M_L)/M_S M_L \text{ ,}$$

we find that evaluated at mean values of the cost shares, $\sigma = .429$ (s.e. = .0051). This is comparable to Polinsky and Ellwood's result of .45 generated from a CES function, and Field's (5) estimate of .5 from the analysis of grouped data with a translog function. Note, however, that our estimate of σ varies with the ratio in which land and structures are used, while Polinsky and Ellwood's does not. If, for example, we evaluate (10) at one standard deviation from the mean of M_L, σ changes to 0.56, an increase of almost 25 percent.[17]

Given the estimates of the cost function parameters, the price index for housing services can be calculated as follows:

a) For each household in the sample calculate the unit cost of housing by substituting its price of structures and land into (4) and exponentiating.[18]

b) For each city, compute the average of the unit costs so calculated.[19]

c) Arbitrarily assign some city a value of 1.0 in the index, and compute index values for all the other cities by dividing their respective unit costs by unit cost for the base city.

The results are shown in Column 1 of Table III.1. (The base area is Chicago-N.W. Indiana.) In order to facilitate comparison with the BLS indices, they have been adjusted so that the figures for Chicago-N.W. Indiana are also 1.0; i.e., for both the high and moderate income levels, the vector of BLS annual housing expenses was divided by the figure for Chicago. Casual inspection indicates that there is some correlation in the three indices. For example, regardless of the choice of index, Anchorage, Honolulu and New York rank near the top. In some cases, however, the correspondence is weaker. The translog index gives Nashville the lowest unit cost of housing, while in the BLS High Income Index, there are five cities less expensive than Nashville.

There is no obvious or unambiguous way systematically to compare the translog index with its BLS counterpart. One interesting issue is the degree to which the indices agree on how the cities rank with respect to housing prices. A convenient summary measure of the extent to which the rankings of two vectors differ is the Spearmann rank correlation r_s:

Table III.1

Alternative Indices of the Cost of Owner-Occupied Housing

City	(1) Translog Index	(2) BLS High Income Index	(3) BLS Moderate Income Index
Anchorage	1.307	1.434	1.417
LA-Long Beach	1.331	.973	.889
San Diego	1.099	.961	.871
SF-Oakland	1.219	1.015	.964
Denver	.893	.875	.859
Hartford	.867	1.067	1.044
Washington, D.C.	.920	.939	.908
Atlanta	.741	.744	.700
Honolulu	1.582	1.219	1.095
Chicago-N.W. Indiana	1.000	1.000	1.000
Indianapolis	.833	1.013	.955
Wichita, KS	.892	.899	.847
Baton Rouge	.800	.860	.736
Portland, ME	.732	.892	.893
Baltimore	.924	.832	.762
Boston	.977	1.203	1.144
Detroit	1.044	.902	.837
Minneapolis-St. Paul	.880	.887	.857
Kansas City	.838	.931	.854
St. Louis	.964	.885	.874
Buffalo	.998	.987	.976
New York-N.E. New Jersey	1.164	1.152	1.112
Cincinnati	.970	.884	.881
Dayton	.994	.885	.798
Philadelphia	.935	.910	.896
Pittsburgh	.950	.880	.798
Nashville	.714	.862	.797
Dallas	.833	.826	.748
Houston	.856	.789	.722
Seattle-Everett	.933	.978	.923
Milwaukee	1.003	1.036	1.054

$$(11) \quad r_s = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)}$$

where d denotes the differences between the ranks of the vectors, and n is the number of elements in the vector ((17), p.435). Applying (11) to the numbers of Table III.1, we find that the rank correlation between the translog cost index and the BLS high income index is 0.629; between the translog index and the BLS moderate income index it is 0.579.[20] It appears that the ranking implied by the BLS figures is an inadequate substitute for that generated by the theoretically preferred translog cost function.[21]

Another way to compare the indeces is via the usual product-moment correlation. The simple correlation between the translog cost index and the BLS high income index is 0.674; between the translog index and the moderate income index it is 0.589. Although there is a positive relation between the translog index and each of the BLS indices, the correlations are certainly not perfect.

IV. Concluding Remarks

The theory of cost functions has been used to estimate an inter-city index of the price of owner-occupied housing. Comparisons with the Bureau of Labor Statistics indices yield both rank and product-moment correlations of roughly 0.6. There is some similarity in the indices, but the correspondence is inexact.

Whether or not the difference in rankings matters depends upon the purpose for which the index is to be employed. From an econometrician's point of view, the issue is the extent to which estimates of housing demand equations are sensitive to which index is used for the price of housing variable. Using their results from a CES cost function, Polinsky and Ellwood (13) have shown that in certain cases, parameter estimates can differ substantially depending upon whether cost function or BLS indices are employed. From the point of view of the government or of a firm with plants in a number of cities, it seems clear that the amounts of income required to compensate individuals for differences in intercity housing costs will vary considerably with the index chosen. Generally speaking, urban economists' inability to observe prices in a market which is of prime importance for them indicates the need to develop adequate measures for the price of housing services.

The search for a better index, then, appears worth pursuing. In addition to the ways suggested above, there are a number of directions in which the current research could be extended. The input 'structures' is really generated by combining 'capital' and 'labor.' Therefore, it would be useful to estimate a cost function with three arguments, the wage, cost of capital, and cost of land. Also, the analysis should be applied to dwelling units other than owner-occupied homes. In particular, study of the unit cost of rental units would contribute significantly to understanding inter-city differences in the choice between rental and owner-occupied housing.

Footnotes

[1] See (15), p.42, for further details.

[2] This approach could also be used for owner-occupied housing.

[3] Some other problems in the interpretation of hedonic price indices are discussed in (14).

[4] This view of housing is quite similar to that of capital in the neo-classical theory of production. Just as capital services (which are generated by the capital stock) are an input in the firm's production function (8), housing services (which are proportional to the housing stock) are an argument in the household's utility function.

[5] For a discussion of cost functions, see (2), pp. 295-296.

[6] It is assumed that C is increasing, linear homogeneous, and concave in P_S and P_L. These assumptions correspond to the usual conditions for "well-behaved" production functions.

[7] This assumes, of course, that an appropriate stochastic specification is selected.

[8] Depending upon the form of $C(\cdot)$, it may be more convenient to estimate share equations or expenditure equations rather than the factor demand equations *per se*.

[9] In both these studies, the unit costs are estimated for the purpose of generating price variables for housing demand equations. The authors do not investigate the implications of their estimates for ranking cities according to the price of housing services.

[10] For the case of the CES function, both the production and cost functions have the same functional form, but in general this is not the case. See (3).

[11] I am most grateful to M. Polinsky and D. Ellwood for making available to me the data described in this section.

[12] In addition, only households whose principal earner was a male between 21 and 60 were included, because the full set of data was available only for these families.

[13] More detailed definitions are available in (13), Appendix B.

[14] Observe that the value of neighborhood 'amenities' (positive and negative) is incorporated in the price of land. A major problem with this approach is its inability explicitly to integrate these amenities into the analysis.

[15] Since the left hand side variable M_L is bounded in the interval (0,1), ordinary least squares estimates suffer from heteroskedasticity. However, the estimates are consistent. Given the large number of observations, the use of generalized least squares to increase efficiency did not seem worth the increased computational costs.

[16] See (3), p.118.

[17] The implications of a non-constant elasticity of substitution upon the steepness of rent gradients have been explored by Field (5).

[18] Note that if the right hand side of (4) gives a consistent estimate of $\ln C$, then $\exp(\ln C)$ gives a consistent estimate of C.

[19] As noted above, since the price of land varies within a city, the unit price of housing services will not be the same at different locations. This fact has been stressed especially by Polinsky and Ellwood (13).

[20] The rank correlation between the two BLS indices is .940.

[21] The estimated standard error of r_s is given by $1/\sqrt{n-1}$, which is equal to .18 for n = 31. However, we cannot perform a formal statistical test of the null hypothesis $r_s = 1$ because not enough is known about the distribution of r_s. See (9).

Bibliography

(1) Alexander, W.E., "Comment," Annals of Economic and Social Measurement, vol. 4, no. 1 (1975), pp. 175-178.

(2) Baumol, W.J., Economic Theory and Operations Analysis (4th ed.), Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1977.

(3) Burgess, D.F., "Duality Theory and Pitfalls in the Specification of Technologies," Journal of Econometrics, vol. 3 (1975), pp. 105-121.

(4) Christensen, L.R. and W.H. Greene, "Economies of Scale in U.S. Electric Power Generation," Journal of Political Economy. vol. 84, no. 4 (August 1976), pp. 655-676.

(5) Field, B.C., "Estimating Substitution Elasticities in Housing with a Trans-Log Cost Function," mimeo, University of Massachusetts at Amherst, undated.

(6) Gillingham, R., "Places to Rent Comparisons," Annals of Economic and Social Measurement, vol. 4, no. 1 (1975), pp. 153-173.

(7) Griliches, Z. (ed.), Price Indexes and Quality Change, Cambridge: Harvard University Press, 1971.

(8) Jorgenson, D.W., "Capital Theory and Investment Behavior," American Economic Review, vol. 53 (1963), pp. 242-259.

(9) Kendall, M.G., Rank Correlation Methods, Hafner Publishing Company, New York, 1962.

(10) King, A.T., Property Taxes, Amenities, and Residential Land Values, Cambridge, Mass.: Ballinger Publishing Company, 1973.

(11) Muth, R.F., "The Derived Demand for Urban Residential Land," Urban Studies, vol. 8, no. 3 (October 1971), pp. 243-254.

(12) Olsen, E.O., "A Competitive Theory of the Housing Market," American Economic Review, vol. 59, no. 4 (September 1969), pp. 612-622.

(13) Polinsky, A.M. and D. Ellwood, "An Empirical Reconciliation of Micro and Grouped Estimates of the Demand for Housing," Harvard Institute of Economic Research, no. 567, August, 1977.

(14) Rosen, S., "Hedonic Prices and Implicit Markets," Journal of Political Economy, vol. 82 (1974), pp. 34-55.

(15) U.S. Department of Labor, Bureau of Labor Statistics, 3 Standards of Living for an Urban Family of Four Persons 1969-1970. Bulletin No. 1570-5, Washington, D.C.: U.S. Government Printing Office, 1972.

(16) Wales, T.J. and A.D. Woodland, "Estimation of Household Utility Functions and Labour Supply Response," International Economic Review, vol. 17 (June 1976), pp. 397-410.

(17) Yamane, T., Statistics. An Introductory Analysis, New York: Harper and Row, 1964.