NBER WORKING PAPER SERIES

USING LOCATIONAL EQUILIBRIUM MODELS
TO EVALUATE HOUSING PRICE INDEXES

Holger Sieg
V. Kerry Smith
H. Spencer Banzhaf
Randy Walsh

Using Locational Equilibrium Models to Evaluate Housing Price Indexes
Holger Sieg, V. Kerry Smith, H. Spencer Banzhaf, and Randy Walsh
NBER Working Paper No. 7934
October 2000
JEL No. C43, H73, R21

ABSTRACT

This paper analyses how the properties of locational equilibrium models can be used to evaluate approaches for constructing price indexes for heterogeneous houses. Housing markets play a key role in locational equilibrium models. Prices for houses determine that implicit costs that households bear when locating in a given community. We evaluate a variety of price indexes all relying on hedonic models for predicting interjurisdictional housing prices. The application uses a unique panel data set of housing transactions in Southern California. The rank predictions of different models are robust with respect to the hedonic model and the composite commodity definition used in aggregation. They do not depend significantly on the spatial or temporal definitions used to define the choice set of local housing markets. Finally, housing price estimates are strongly correlated with education and environmental amenities.

Holger Sieg
Department of Economics
Duke University
Box 90097
Durham, N.C. 27708
and NBER
hs25@econ.duke.edu

V. Kerry Smith
Department of Agr. and Resource Economics
North Carolina State University
Box 8109
Raleigh, N.C. 27695

Spencer Banzhaf
Department of Economics
Duke University
Box 90097
Durham, N.C. 27708

Randy Walsh
Department of Economics
Duke University
Box 90097
Durham, N.C. 27708

1   Introduction

The main purpose of this paper is to explore different approaches of estimating interjuris-
dictional price indexes and to analyze the properties of these estimates. In particular,
we investigate whether the estimated interjurisdictional price indexes have properties pre-
dicted by a large class of general equilibrium models which are increasingly used in urban
economics and local public finance.[1] These locational equilibrium models provide condi-
tions for a sorting of heterogeneous households among a finite set of communities. Invoking
plausible single-crossing assumptions on the underlying preferences of households, an equi-
librium in these types of models will satisfy stratification and ascending bundles properties.
Stratification implies that households with similar demands for local public goods will tend
to live together. Ascending bundles implies that higher amenity communities will have
higher housing prices in equilibrium. These models give, therefore, rise to a strict hierar-
chy of communities in equilibrium, similar to vertical product differentiation models in the
industrial organization literature.

Housing prices not only determine the rank of a community within the hierarchical sys-
tem, but also the implicit price that households pay for locating in a particular community
within the metropolitan area. One of the main challenges encountered in using these mod-
els in an empirical analysis is to estimate interjurisdictional housing prices since housing
prices per unit of housing services are not directly observed by researchers. In this paper we
explore different regression techniques for predicting interjurisdictional housing prices using
a unique panel data set of housing transactions in Southern California. We also investigate
whether housing price indexes provide robust rank predictions since ranking communities
is an important part of matching the theoretical models into the data.

Hierarchical equilibrium models in local public finance are built around the assumption

---

[1]Tiebout (1956) inspired research on inter-jurisdictional mobility which gave rise to a number of theoreti-
cal models including Ellickson (1971), Westhoff (1977), Epple, Filimon, and Romer (1984, 1993), Goodspeed
(1989), Epple and Romer (1991), Fernandez and Rogerson (1996) Nechyba (1997a, 1997b) and Epple and
Platt (1998). Epple and Sieg (1999) and Epple, Romer, and Sieg (1999) provide a structural empirical
analysis of location equilibrium models.

that households are mobile and make locational choices based on the mix of site specific public goods and the housing market conditions in a finite set of communities. These models replace a continuum of choices of tax and expenditure packages with a discrete set and hence differ significantly from earlier work on hedonic equilibrium (Rosen, 1974). The community choice set parallels the horizontal extent of the market in the industrial organization literature on differentiated products. It defines the extent of substitution in these models.[2] From an empirical perspective, researchers have some freedom of defining the relevant choice set. This raises the question how sensitive our price indexes and their rank predictions are to temporal and the spatial aggregation technique used to construct the community choice set. We therefore need to investigate the relationship between the community choice set and the housing price indexes.

Once we have established the rank properties of the different housing price estimates, we analyze whether housing price estimates satisfy the ascending bundles property. An empirical analysis of the ascending bundles property is complicated by the fact that the public good provision is multi-dimension, i.e. local public good provision is not determined by a single factor like school quality, but also depends on a multitude of other factors like protection from crime and environmental quality. It may also depend on factors which are either unobserved by the researcher or measured imprecisely. We would therefore expect that housing prices are not perfectly correlated with each amenity. Nevertheless high correlations between housing prices and amenities or the ranks can be interpreted as evidence in support of the predictions of hierarchical Tiebout models.

The rest of the paper is organized as follows: Section 2 presents an extended discussion of our data set. Section 3 discusses how to construct interjurisdictional price indexes based on our data. Section 4 reports the price index estimates and discusses their properties with special reference to locational equilibrium models. Section 5 concludes the analysis.

---

[2]The analogy between the set of differentiated products in a standard IO model and the community choice sets is only partial, since at least some characteristics of communities (like expenditures and tax rates) are endogenous and depend on the composition of the population within each community which are jointly determined in equilibrium.

2   Data

In this paper we analyze housing markets in Southern California. We assembled for our analysis a comprehensive data base on housing markets in the LA metropolitan area. We focus on the Los Angeles Metropolitan Area which consists of five counties: Los Angeles, Orange, Riverside, San Bernardino and Ventura. We chose housing units in all school districts to include the entire greater L.A. metropolitan area west of the San Gabriel Mountains. A map of our housing sample is given in Figure 1.

Figure 1: Map of LA Metropolitan Area and Housing Sample
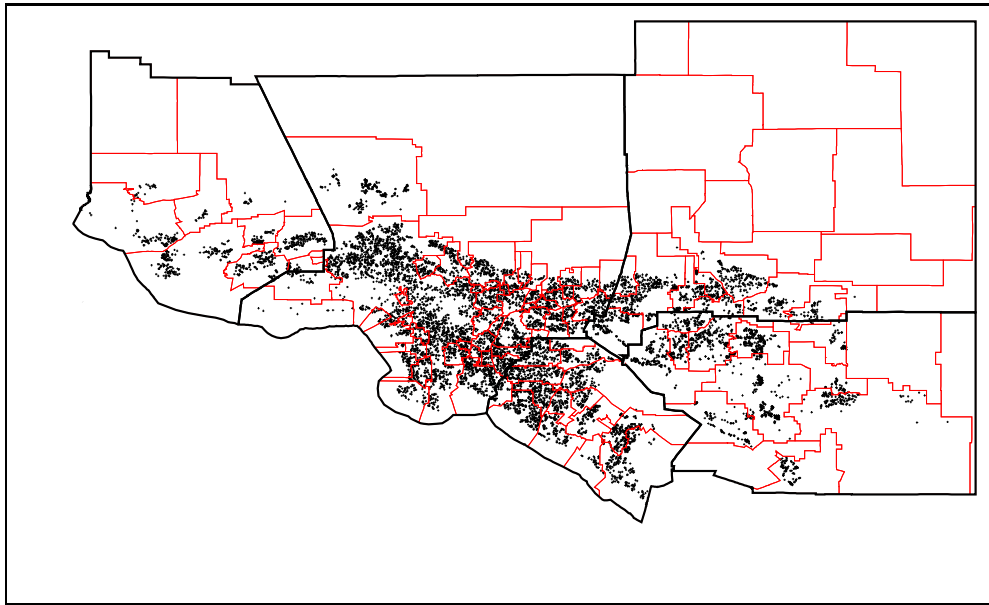
Our data set is based on housing transactions which were collected by Transamerica Intellitech, a market research firm. This data set contains housing characteristics and transaction prices for virtually all housing transactions in Southern California between 1985 and 1995.[3] Table 1 reports means of the main variables in the housing sample by counties.

---

[3]In contrast, other work has relied on home-owners' subjective assessments of their homes' value (Zabel and Kiel., 1995) or on samples subject to selection effects due to eligibility criteria such as the FHA houses (Chattopadhyay, 1998).

Table 1: Descriptive Statistics of the Housing Sample

Variable	Orange	Riverside	San Bernardino	Ventura	Los Angeles
Number of Observations	40894	33132	24493	14817	109529
Price	253315	139771	151313	244888	243889
Number of Bathrooms	2.16	2.07	2.10	2.24	1.94
Number of Bedrooms	3.33	3.26	3.27	3.49	3.05
Lot Size	0.16	0.24	0.21	0.22	0.19
Square foot Building	1748	1627	1615	1838	1591
Pool	0.16	0.12	0.13	0.15	0.17
Fireplaces	0.26	0.84	0.79	0.79	0.54
Age	23.8	9.7	16.8	17.4	37.0

Means of housing values and structural characteristics by county.

In general, these data are quite consistent across counties. Significant differences across counties are present only for price, the variable to be explained in our research, in age of the house, and in the average percentage of fire places. Most of these differences are consistent with a priori expectations. Mean prices are higher for the LA, Orange, and Ventura counties. Unlike the other two counties, these are coastal counties. In addition, two of the counties (LA and Orange) are more urban than the others. Being further from the central city, it is not surprising that Ventura, San Bernardino, and Riverside counties would have the youngest houses on average. Of the five counties, these three experienced the largest population growth from 1980 to 1990, with Riverside and San Bernardino having the largest population growth in the state at 67 and 54 percent respectively (California Department of Finance 1990). Consequently, it is not surprising that they would have younger houses on average.

One potential drawback associated with using California data relates to Proposition 13. It has been argued in the literature that Proposition 13 created a lock-in effect on house

owners. A household faces a tax on mobility because property taxes on newly purchased property are based on the market value. If the market value exceeds the assessed value, the revaluation creates additional mobility costs. O'Sullivan, Sexton, and Sheffrin (1995) provide a detailed analysis of the quantitative importance of these lock-in effects. They conclude that "the effects [of Proposition 13] on mobility and loss of economic welfare from this lock-in effect are small ... We estimate that for the average household a 13 % inflation rate will lengthen the average time between moves by only approximately two months" (p.138).

Closely related to the lock-in effect is the question relating to turn-over in housing markets and the representativeness of our housing transaction sample. As mentioned above, our sample contains a large number of housing transactions for each school district. So there is no question that there is a large degree of turn-over in the housing market, which is also indirect evidence for the fact that mobility costs are likely to be small. Nevertheless, we need to establish that our sample is a representative sample of the underlying housing stock of each school districts. Some information pertinent to this question is available from the 1990 US Census. In particular, the US census reports histograms on housing expenditures, the number of bedrooms, and the age of houses.

The pattern of housing expenditures across communities is quite close in the two data sets, with a correlation of 0.99. However, prices tend to be uniformly higher in the US Census. Across our 92 school districts, prices are 6 to 12 percent higher in the census (inter-quartile range). This might suggest smaller homes are over-sampled in the transactions data, but the data do not confirm this hypothesis. In fact, 34 percent of houses in the US census have two or fewer bedrooms, compared to only 19 percent in our sample. In addition, homes are much younger in our data set. Fifteen percent of our houses are younger than 1 year, whereas in the 1990 census only three percent of homes were built in 1989-1990. The difference is particularly striking in Ventura, San Bernardino, and Riverside counties. There, 36 percent of our sample is younger than 1 year, compared to 6 percent built in 1989-1990 in the census. The over-sampling of new homes is not surprising, because newly built homes will automatically show up in a data set of housing transactions, whereas older

homes will only show up when they turn over.

In any case, since our hedonic regressions suggest that younger homes are more expensive, ceteris paribus, it is not clear why homes appear more expensive in the US census despite being older and having fewer bedrooms on average. Three possibilities suggest themselves: (1) The US Census, which does not rely on transactions data, systematically overestimates housing prices. (2) Other characteristics which affect housing prices, included in our hedonic regressions but not available in the US census for comparison, are systematically different in the stock of houses and the set of housing transactions. (3) Unobserved variables in our hedonic regression are systematically higher in the stock of houses than in the set of housing transactions. Unfortunately, we can not determine which of these three variables is driving the result. The fact that price differences are only about 10 percent, however, and consistently ranked across communities, is encouraging.

Table 2: Housing Tenure

	< 1 year	1-5 years	5-10 years	11-20 years	21-30 years	> 30 years
Los Angeles	24.1	30.9	13.3	17.2	8.0	6.4
Orange	25.9	33.5	12.7	18.0	7.2	2.8
Riverside	29.9	34.8	13.2	15.1	4.5	2.4
San Bernadino	28.9	35.0	12.4	14.5	5.4	3.8
Ventura	22.5	34.3	14.3	19.0	7.1	2.8

Housing Tenure Data from the 1990 U.S. Census.

Table 2 reports the empirical distribution of housing tenure for the five counties of interest based on data collected by the 1990 U.S. Census. We find that between 70 and 80 percent of all households change houses within 10 years. Given the scope of our housing data, we expect to capture most of these housing transactions in our sample. Approximately 20 to 30 percent of the houses in the US Census have a housing tenure which is greater than 10 years. By construction our sample only contains a fraction of these house. This analysis

suggests that our sample is likely to oversamples houses with a tenure less than 10 years and under-samples houses with a higher tenure.

3 Interjurisdictional Price Indexes

In order to test some of the predictions of the equilibrium models of local jurisdictions, we need interjurisdictional housing prices. Unfortunately, housing prices per unit of housing services are not observed by the researcher, and hence need to be estimated. The most frequent approach is to estimate housing prices based on housing values or rental expenditures. This approach implicitly relies on the assumption that housing is a homogeneous good in all communities and can only be consumed in a fixed quantity. In our model, housing is treated as a normal good in economic analysis and hence subject to a strong income effect. Hence both median rents or median housing values are not good candidates to measure relative housing prices among a set of communities. They fail to account for differences in the quantity of consumed housing services.[4]

One of the main advantages of this study is that we have access to a panel data set on housing transaction (prices and housing characteristics) for each community in the sample. We can therefore rely on hedonic regressions to estimate housing prices for the set of communities. Following Poterba (1992), we can easily convert the transaction price of a house n located in community j at time t into the imputed rent which is denoted by r_{jnt}. We also observe a vector of housing specific characteristics denoted by z_{jnt}. Note that the housing characteristics are indexed by t indicating that they may change over time time due to renovations and additions. Let u_{jnt} denote the unobserved housing characteristics.

There is a substantial literature on the development of quality adjusted quantity indexes for heterogeneous commodities. Indeed, one interpretation of the hedonic model has been in these terms (Triplett, 1990). Our analysis draws a distinction between the effect of

---

[4]We therefore need to correct for the income effect in the housing expenditures in order to obtain reliable housing price estimates. Epple and Sieg (1999) provide such an approach which exploits the structure of the demand for housing.

structural characteristics and the effects conveyed through location. For the former we assume a spectrum of choices such that within any community it is possible to select any of the continuum of houses available. This is equivalent to unbundling the local public goods and amenities from the effects of structural and lot characteristics. Thus the quality adjusted units of housing is given by the following equation:

$$h_{jnt} \quad = \quad h_j(z_{jnt}, u_{jnt}) \tag{3.1}$$

By definition, rent measured for a quality adjusted unit is the product of the adjusted housing price and the number of quality adjusted housing units, i.e. $r_{jnt} = p_{jt} \ h_{jnt}$. Substituting equation (3.1) into the identity above and taking logarithms, we obtain the following regression model:

$$\ln(r_{jnt}) \quad = \quad \ln(p_{jt}) \ + \ \ln(h_j(z_{jnt}, u_{jnt})) \tag{3.2}$$

Additionally, we assume that the intercept in equation (3.2) can be decomposed as follows:

$$\ln(p_{jt}) \quad = \quad \ln(p_j) \ + \ \ln(p_t) \tag{3.3}$$

and we can interpreted $\ln(p_j)$ as the housing price index of community j. In order to estimate the housing price indexes we need to impose a parametric assumption on the functional form on $h_j(z_{jnt}, u_{jnt})$. There are a number of functional forms which have been used in the related literature on hedonic regression, one of them is the log-linear specification, which is obtained by assuming that:

$$h_{jnt} \quad = \quad e^{\delta'_j \ z_{jnt} \ + \ u_{jnt}} \tag{3.4}$$

which implies the following regression model:

$$\ln(r_{jnt}) \;\; = \;\; \ln(p_{jt}) \; + \; \delta'_j \, z_{jnt} \; + \; u_{jnt} \tag{3.5}$$

Provided that $E(u_{jnt}|z_{jnt}) = 0$, we can estimate the regression above using least squares and obtain a consistent estimate of the price of an efficiency unit of housing in community j. A special case of equation (3.5) is the simple fixed effects estimator which assumes that $\delta_j = \delta$ for all communities.

Thus far we have given the regression model above a structural interpretation which allows us to interpret the community specific fixed effect as a housing price index. We can also interpret equation (3.5) in slightly less structural way and exploit its predictive properties. The basic idea behind this procedure is to use the regression above to predict the housing prices of a fixed set of houses for each community. The average predicted price for these houses can be interpreted as a price index since it holds the quantity of housing consumed in each community constant. This approach only exploits the predictive powers of our regression model. In particular, it is less important that the underlying coefficients of the model can be given a structural interpretation. the key concern in this approach is that the model has good predictive powers. This is much easier to establish based on out-of-sample predictions.

4 Empirical Results

4.1 Estimates of Interjurisdictional Price Indexes

Table 3 reports summary statistics of the price estimates obtained from five different hedonic regressions approaches. Communities are defined as school districts. Columns 1 and 2 report estimates from variable intercept regressions. Column 1 (2) uses log-log (log-linear)

specification. Column 3 reports estimates based on a linear specification.[5] Column 4 is the same as column 1, but uses quantile regression techniques instead of least squares. Column 5 reports estimates based on variable coefficient regressions and allows for county specific coefficients.[6]

Table 3: Descriptive Statistics of Price Index Estimates by S.D.

	1	2	3	4	5
	F.E. ln-ln	F.E. ln-lin	Round Robin lin-lin	Quantile F.E. ln-ln	Round Robin (by county)
Mean	0.99	1.01	1.07	1.01	0.99
Std. Deviation	0.42	0.43	0.59	0.44	0.41
Min	0.44	0.46	0.49	0.44	0.43
Max	2.70	2.96	4.97	2.90	2.65
Correlation:					
mean income	0.683	0.687	0.685	0.675	0.682
median income	0.512	0.503	0.456	0.496	0.513

In general we find that the summary statistics reported in Table 3 are robust to the choice of functional form and estimation technique.[7] The highest price school districts have price levels between 6 and 6.5 times that of the lowest. Correlations with mean school district income are clustered near .68. The largest discrepancies are with the linear specification reported in column 3. The primary source of this discrepancy is the Beverly Hills School District which is the highest price school district under all models and an extreme outlier. The next highest price under the linear specification is 2.95, a difference of almost 4 standard

---

[5]To develop a price index based on a linear in prices specification, a sample of 10,000 house types were drawn from the entire data-set. The average price of these 10,000 houses was then predicted for each school district.

[6]This price index was simulated using the same approach as in column 3.

[7]Several additional models were also estimated. A subset is reported here for ease of discussion. The reported results effectively characterize the results from the larger set of models.

deviations from the level of 4.97 estimated for Beverly Hills.

Table 4: Correlations of Price Index Estimates

	Fixed Effect ln-ln	Fixed Effect ln-lin	Round Robin linear	Quantile F.E. ln-ln	Round Robin by county
Fixed Effect ln-ln	1.00	.	.	.	.
Fixed Effect ln-linear	0.99	1.00	.	.	.
Round Robin linear	0.93	0.95	1.00	.	.
Quantile F.E. ln-ln	0.99	0.99	0.94	1.00	.
Round Robin (by County)	0.99	0.99	0.92	0.99	1.00

The high degree of similarity among the different regressions is also reflected in the correlation matrix which is reported in Table 4. As was the case in the discussion above, the largest differences across price index models result from the inability of the linear model to account for the extremely high prices in Beverly Hills. If we remove Beverly Hills from the sample, the correlation between the linear model's price predictions and those of the other models all exceed .97.

4.2   Rank Analysis

The price estimates ultimately determine the rank of a given community in the hierarchy of communities. We would like to know how robust the rank predictions are based on the different estimates of the price indexes. Table 5 provides some results. For each model, the table reports the number of communities whose ranking by price index differs by more than 0, 3, or 5 places when compared with each of the other price index models. It also reports the mean difference for each model comparison. The rankings are relatively stable across price indices. In all but one of the comparisons, less than 10% of the communities absolute ranking varies by more than 5 places (relative to a total of 92 communities).

12

Table 5: Rank Differences Across Price Indices

	Fixed Effect ln-ln				Fixed Effect ln-linear				Round Robin County				Quantile Fixed Effect			
	# com. w/ dif.			mean	# com. w/ dif.			mean	# com. w/ dif.			mean	# com. w/ dif.			mean
	> 0	> 3	> 5	diff.	> 0	> 3	> 5	diff.	> 0	> 3	> 5	diff.	> 0	> 3	> 5	diff.
F.E. ln-ln	0	0	0	0.00	.	.	.	.	.	.	.	.	.	.	.	.
F.E.ln-lin	44	2	0	0.78	0	0	0	0.00	.	.	.	.	.	.	.	.
RRCnty	42	2	1	0.72	48	6	3	1.00	0	0	0	0	.	.	.	.
Qnt F.E.	68	15	7	2.00	70	19	7	2.20	64	17	9	2.04	0	0	0	0.00
RR linear	79	21	8	2.50	81	21	7	2.57	78	20	9	2.50	82	26	15	2.93

The necessary conditions for locational equilibrium imply that equilibrium allocations must satisfy the "ascending bundles" property: housing prices and levels of public provision should be monotonically increasing in the price rank of the community. Since the level of public good provision determines the desirability of a community, higher amenity communities must have higher housing prices which gives rise to a strict hierarchy of communities. It would be surprising if estimated housing prices and observed amenities satisfied this condition perfectly. One of the main problems with testing the ascending bundles property is that the level of public good provision is a function of many amenities, some of them are unobserved to the econometrician.

As a first step of the empirical analysis of the ascending bundles property , it is useful to analyze the correlation structure between housing prices and measures of the quality of public good provision. We compute correlations between each of the price index models and different measures of public good provision: expenditure per student, students per teacher, particulate matter, ozone, and average 10th grade math scores for reading, writing and math. In all cases, the correlation is negative for bads (students/teacher and air pollution) and positive for goods (expenditure/student and average test scores).

The previous correlation analysis lends support for the ascending bundles property. An alternative test of the ascending bundles property directly focuses on the ranks of the variables. We would expect that the price rank of a community is very highly correlated with the amenity rank of a community. There are $J(J-1)/2$ pairwise comparisons between communities. With 92 communities in the sample, there are 4186 possible rank violations. We focus on the median math score within the school district as a measure for school quality. We find that there are 1271 rank violations, or roughly 30.3 percent. This suggests that the rank predictions based on housing prices are quite similar to those based on median math score.

Table 6: Correlations of Price Indexes and Amenities

	Expenditure per Student	Students per Teacher	5 Year PM Average	5 Year Ozone Average	Reading Test Average	Writing Test Average	Math Test Average
F.E. ln-ln	0.2050	-0.5602	-0.2592	-0.3113	0.5033	0.4065	0.6017
F.E.ln-linear	0.2182	-0.5780	-0.2419	-0.2936	0.4856	0.4014	0.5906
Round Robin by County	0.2057	-0.5516	-0.2692	-0.3191	0.5092	0.4081	0.6031
Quantile F.E. ln-ln	0.2257	-0.5733	-0.2393	-0.2902	0.4865	0.3975	0.5934
Round Robin Linear	0.2939	-0.6345	-0.1815	-0.2449	0.3852	0.3583	0.5267
Expenditure Per Student	1.0000	.	.	.	.	.	.
Students per Teacher	-0.2922	1.0000	.	.	.	.	.
5 Year PM Average	-0.0188	0.1112	1.0000	.	.	.	.
5 Year Ozone Average	-0.1200	0.1299	0.7460	1.0000	.	.	.
Reading Test Average	-0.0947	-0.3794	-0.2570	-0.1713	1.0000	.	.
Writing Test Average	-0.0734	-0.3390	-0.0055	-0.0716	0.6899	1.0000	.
Math Test Average	-0.0837	-0.4657	-0.1457	-0.0611	0.7804	0.6911	1.0000

4.3 Interjurisdictional Price Estimates and the Community Choice Set

A key component of testing hierarchical equilibrium model is choosing the temporal and geographical boundaries which delineate a market in equilibrium. We compare geographic and temporal price index estimates based on three different market definitions over the greater Los Angeles area. The first market definition includes all residential housing sales between 1988 and 1992 for 92 school districts in Ventura, Los Angeles, Orange, Riverside, and San Bernardino Counties. This market definition includes the entire greater L.A. metropolitan area west of the San Gabriel Mountains. The second market definition includes the same years (1988-1992), but is restricted to 36 school districts west of the San Gabriel Mountains in Orange, Riverside and San Bernardino Counties. The third market definitions includes the same 36 school districts as the second, but use a longer time period (1985-1995). Table

Table 7: Correlation of Price Indices Across Market Definitions

	36 districts 11 years	36 districts 5 years	92 districts 5 years
92 districts by 5 years	0.9977	0.9999	1.0000
36 districts by 5 years	0.9983	1.0000	0.9999
36 districts by 11 years	1.0000	0.9983	0.9977

7 reports correlations between price indices under the 3 different temporal/spatial market definitions (estimated using the fixed effects ln-ln specification). The high correlations suggest that at least for the greater Los Angeles area relative price indices are stable across spatial market definitions. The results appear to be only slightly less invariant to temporal definitions.

16

5 Conclusions

In this paper, we have developed a number of simple techniques which can be used to esti-
mate interjurisdictional price indexes based on a cross-sectional or panel sample of housing
transactions. We find that the different techniques yield price estimates which are aston-
ishingly similar. The correlation between the different price indexes often exceeds 0.95. We
also find that these price indexes are reasonably robust with respect to the rank predictions
for the communities in the choice set. Differences for rank predictions are small and often
do not significantly alter the hierarchy among communities. We then investigate whether
housing prices and local amenities are highly correlated as predicted by the underlying hi-
erarchical equilibrium models. Since we only observe proxies of public good provisions, we
would not expect that the ascending bundles prediction holds perfectly in the data. We find
that all correlations at least have the expected sign and are often large in magnitude pro-
viding some support for the ascending bundles property. Finally, we investigate how robust
our findings are to different spatial and temporal aggregation schemes used to construct the
community choice set. Our data does not provide any evidence which would suggest that
these aggregation schemes are driving any of our empirical findings. While this analysis
cannot replace a more careful structural analysis of the underlying properties of the class of
hierarchical equilibrium models, they nevertheless provide some solid evidence in favor of
some of the strong qualitative predictions which are shared by a broad class of equilibrium
models.

A Housing Data Appendix

To conduct the analysis described in this paper, we obtained a large data set of housing characteristics and actual prices from virtually all housing transactions in five counties of southern California from Transamerica Intellitech, a market research firm. The five counties are Los Angeles and the four surrounding it: Ventura, San Bernardino, Riverside, and Orange.

In general, the data contain a rich set of independent variables for a hedonic price regressions. These variables include the age of the house; the number of rooms, bathrooms, and bedrooms; the size of the building in square feet; the size of the lot; the number of stories; information about the garage; the presence of a swimming pool, a fire place, air conditioning, and central heating; the availability of public sewers, and connections to natural gas. Unfortunately, however, the collection of these variables is quite uneven across counties so that only a sub-set may be included in pooled regressions, creating an unobserved variable problem. Variables available for all counties include age, the number of bedrooms and bathrooms, the size of the building and lot, and the presence of a swimming pool and a fireplace.

These data were limited to a time frame and geographic extent of the market judged to be appropriate for the task of estimating hedonic price indexes for the equilibrium model. Temporally, we restricted the data to the years 1988-1992 (a five-year window surrounding 1990, the point at which we observe the income distributions). As described in the body of the paper, to test the sensitivity of this restriction on the price estimates we also obtained data from 1985-1995 for the four counties surrounding LA. Geographically, we restricted the data to lie inside a ring surrounding the city of LA roughly defined by a series of national forests (specifically, Los Padres National Forest, Angeles National Forest North, Angeles National Forest, and San Bernardino National Forest). Much of the omitted area in the east is sparsely populated desert. In terms of the number of school districts, the restriction involves omitting one school district in northern Ventura County, one in northern Los Angeles County, all but nine in the southwestern corner of San Bernardino County, and

all but fourteen in western Riverside County. Preliminary hedonic regressions including some of these omitted areas indicated that they tended to have prices much lower than the immediate LA area. Finally, one additional school district, Murrieta Valley Unified in western Riverside County, had to be dropped because we were unable to obtain adequate schooling data.

After dropping houses with missing or incomplete sales prices (close to 50 percent of observed housing transactions), confining the data set to houses that sold between 1988 and 1992, and confining the data set to the 92 school districts surrounding Los Angeles, there were 247,454 observations available in our data set. Of these, 24,589 observations, or close to 10criteria designed to screen the data of unreasonable observations or extreme outliers that were not likely to be in the choice sets of most Californians. These criteria, and the number of observations successively dropped under each criterion, are as follows:

- 3,145 were dropped because their age was less than -1. (In addition, 5,352 houses with an age between zero and -1 years were re-coded to an age of zero years under the assumption that they were new houses sold slightly before completion.)

- 4,643 were dropped using a 1 percent-trim on the nominal sale price.

- 9,397 were dropped because the number of bedrooms was zero or greater than 20.

- 643 were dropped because the number of bathrooms was less than 1 or greater than 15.

- 141 were dropped because the area of the house was greater than 4 times the area of the lot. Since there were no houses with greater than 4 stories, this would be a logical impossibility.

- 59 were dropped because the area of the building was less than 250 square feet or greater than 10,000 square feet.

- 6,591 observations were dropped because the size of the lot was less than 500 square feet or greater than 100,000 square feet.

19

Note that these numbers are successive, so that, for example, 6,591 observations were still outside the criteria for lot size even after the a-trim on price and the other steps. The result of these steps is a rich data set of 222,865 observations representing the types of homes likely to be available to most people in the LA area.

References

Chattopadhyay, S. (1998). An Empirical investigation in the Performance of Ellickson's Random Bidding Model, with an Application to Air Quality Valuation. *Journal of Urban Economics*, *43*, 292–314.

Ellickson, B. (1971). Jurisdictional Fragmentation and Residential Choice. *American Economic Review: Papers and Proceedings*, *61*, 334–339.

Epple, D., Filimon, R., and Romer, T. (1984). Equilibrium Among Local Jurisdictions: Towards an Integrated Approach of Voting and Residential Choice. *Journal of Public Economics*, *24*, 281–304.

Epple, D., Filimon, R., and Romer, T. (1993). Existence of Voting and Housing Equilibrium in a System of Communities with Property Taxes. *Regional Science and Urban Economics*, *23*, 585–610.

Epple, D. and Platt, G. (1998). Equilibrium and Local Redistribution in an Urban Economy when Households Differ in Preferences and Incomes. *Journal of Urban Economics*, *43(1)*, 23–51.

Epple, D. and Romer, T. (1991). An Econometric Model of Exit and Voice in Systems of Local Jurisdictions. Mimeo, Carnegie Mellon University.

Epple, D., Romer, T., and Sieg, H. (1999). The Tiebout Hypothesis and Majority Rule: An Empirical Analysis. NBER Working Paper 6977.

Epple, D. and Sieg, H. (1999). Estimating Equilibrium Models of Local Jurisdictions. *Journal of Political Economy*, *107(4)*, 645–681.

Fernandez, R. and Rogerson, R. (1996). Income Distribution, Communities, and the Quality of Public Education. *Quarterly Journal of Economics*, *111*, 135–164.

Goodspeed, T. (1989). A Reexamination of the Use of Ability-to-Pay Taxes by Local Governments. *Journal of Public Economics*, *38*, 319–342.

Nechyba, T. (1997a). Existence of Equilibrium and Stratification in Hierarchical Public Good Economies with Voting. *Economic Theory*, *10(2)*, 277–304.

Nechyba, T. (1997b). Local Property and State Income Taxes: the Role of Interjurisdictional Competition and Collusion. *Journal of Political Economy*, *105(2), April 1997*, 351–84.

O'Sullivan, A., Sexton, T., and Sheffrin, S. (1995). *Property Taxes and Tax Revolts*. Cambridge University Press.

Poterba, J. (1992). Taxation and Housing: Old Questions, New Answers. *American Economic Review*, *82 (2)*, 237–242.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, *82*, 34–55.

Tiebout, C. (1956). A Pure Theory of Local Expenditures. *Journal of Political Economy*, *64*, 416–424.

Triplett, J. E. (1990). *Hedonic Methods in Statistical Agency Environments: An Intellectual Biopsy*, pp. 207–233. Fifty Years of Economic Measurement: The Jubilee of the Conference on Research in Income and Wealth. Chicago University Press.

Westhoff, F. (1977). Existence of Equilibrium in Economies with a Local Public Good. *Journal of Economic Theory, 14*, 84–112.

Zabel, J. and Kiel., K. (1995). Estimating the Demand for Clean Air in the United States. Working Paper.