Regional Science and Urban Economics 43 (2013) 635–648

Contents lists available at SciVerse ScienceDirect

Regional Science and Urban Economics

journal homepage: www.elsevier.com/locate/regec

Interjurisdictional housing prices and spatial amenities: Which measures of housing prices reflect local public goods?☆

H. Spencer Banzhaf [a,*], Omar Farooque [b]

[a] Georgia State University, NBER, PERC, United States
[b] Northwestern University, United States

ARTICLE INFO

Article history:
Received 27 January 2012
Received in revised form 10 March 2013
Accepted 24 March 2013
Available online 4 April 2013

Keywords:
Housing prices
Price index
Capitalization

ABSTRACT

Understanding the spatial variation in housing prices plays a crucial role in topics ranging from the cost of living to quality-of-life indices to studies of public goods and household mobility. Yet analysts have not reached a consensus on the best source of such data, variously using transaction values, self-reported values from the census, and rental values. Additionally, while most studies use micro-level data, some have used summary statistics such as the median housing value.

Assessing community housing price indices in Los Angeles, we find that indices based on transaction prices are highly correlated with indices based on self-reported values, but that the former are better correlated with public goods. Moreover, rental values have a higher correlation with public goods and income levels than either asset-value measure. Finally, indices based on median values are poorly correlated with the other indices, public goods, and income.

1. Introduction

Housing is the most important asset and largest expenditure category in most households' budgets. Accordingly, accurate data on the value of homes is a lynchpin in many economic studies. For example, because housing accounts for about 30% of households' expenditures, housing costs play a key role in computing geographic comparisons in the cost-of-living, such as the US ACCRA index, as well as intertemporal indices of inflation.

An accurate representation of home values is also a critical step in many empirical studies of local public goods and household mobility. Through the process of capitalization, home values are deeply intertwined with spatial public goods such as school quality, crime, air quality, hazardous waste sites, and green spaces, and the taxes that pay for them. The differences in housing values associated with differences in these public goods and taxes have long been used by economists to infer people's demand for such goods. Prominent examples of such "hedonic" methods include applications to intercity quality-of-life measures (Albouy, 2012; Blomquist et al., 1988; Winters, 2009), education (Bayer et al., 2007; Figlio and Lucas, 2004), crime (Bishop and Murphy, 2011), racial segregation (Bajari and Kahn, 2005), air quality (Chay and Greenstone, 2005; Grainger, 2012; Smith and Huang, 1995), superfund sites (Gamper-Rabindran et al., 2011; Greenstone and Gallagher, 2008), cancer risks (Davis, 2004), and property taxes (Palmon and Smith, 1998).

Over the last twelve years or so, economists have overcome two related challenges in this literature. The first is the endogeneity problem in teasing out the demands for public goods from the sorting process, when unobserved heterogeneity leads to a simultaneous choice of both the levels of those goods and their implicit cost. Secondly, economists have begun to account for general equilibrium effects in evaluating households' demand for public goods and the effects of public policies (Kuminoff et al., 2013). Importantly, changing the geographic distribution of local public goods will induce households to move, with attending effects on local real estate prices, local peer groups, and tax revenues. For example, Sieg et al. (2004) show that improving air quality in some communities can cause what they call "environmental gentrification," which benefits landlords at the expense of poor renters, leading to very different distributional welfare effects in general equilibrium (see also Tra, 2010). Calabrese et al. (2006) consider the implications of household sorting on endogenous neighborhood demographics for voting on public goods. Bayer et al. (2007) emphasize the importance of accounting for endogenously formed neighborhood demographics and their interactions with school quality and unobserved locational goods. And Walsh (2007) illustrates the importance of endogenous development patterns when understanding the effect of open space policies.

To model the sorting process and incorporate these feedbacks in an equilibrium model, an emerging strategy in the public economics literature is to use a discrete-continuous model, in which households first choose a community in which to live and then a continuous quantity

☆ This paper was begun under the Andrew Young Summer Policy Internship program, under NSF award 0452702. We thank Kelly Bishop, James Marton, and especially Nicolai Kuminoff for comments and Christopher Mothorpe for assistance.

* Corresponding author at: Dept. of Economics, Georgia State University, PO Box 3992, Atlanta, GA, 30302, United States. Tel.: +1 404 413 0252.
E-mail addresses: hsbanzhaf@gsu.edu (H.S. Banzhaf), omarfarooque2015@u.northwestern.edu (O. Farooque).

0166-0462/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.regsciurbeco.2013.03.006

of housing in that community. There are several variants on this strategy,[1] but in all cases a crucial step is the estimation of community-level price indices. Usually this is done by regressing micro-level housing prices on housing structural characteristics and community fixed effects. These fixed effects become the price indices. Once community-level housing prices are estimated, they enter the model as one community-level attribute, along with community-level amenities, determining how households sort across communities.

Despite the importance of this step, economists have not yet settled on best practices for estimating those price indices. In particular, researchers have used a variety of data sources on housing values, including recorded transaction prices, survey data of owner-reported housing values (available from the US Census), tax assessments, and rental prices. Transaction prices often are considered the gold standard, well worth the additional expense of obtaining them, but there has been little actual evidence to justify this presumption.

A related issue is whether to use micro-level data or aggregate summary statistics. Most studies use micro-level housing data from one of the above sources, but others have relied on aggregate statistics, such as the median housing value in the county or census tract conditioned on community-level housing variables (e.g. Chay and Greenstone, 2005; Grainger, 2012; Greenstone and Gallagher, 2008). Again, there has been little evaluation of these judgments.[2]

Each of these data sources has inherent benefits and shortcomings. In this paper, we evaluate them in two ways. Our first strategy is simple and direct: we construct community-level price indices using a variety of data sources, all of which have been used in the peer-reviewed literature, and compare them to one another. This strategy is one way to gauge how sensitive spatial housing price indices are to data sources, but cannot indicate which data are preferable.

Our second strategy employs an "ascending bundles" criterion suggested by Sieg et al. (2002). Using the insights from equilibrium models of locational choice, they note that more expensive communities should be those with better public goods and other locational amenities. Accordingly, those community price indices that best correlate with observed amenities would appear to be most sensible. By this criterion, the "best" index is one that has the best fit when regressed on local amenities.[3] By a similar logic, these communities will also be inhabited by the richest households. By this criterion, the "best" index is one that has the tightest correlation with average income in the community.

We find that indices based on self-reported values in the US Census are highly correlated with indices based on transaction prices, though the latter are somewhat better correlated with public goods. This finding suggests that while households may provide accurate estimates of the value of their structures, they do a poorer job of recognizing components of their property values capitalizing local amenities. Indices based on rental prices perform even better when judged by the criterion of correlation with public goods, as well as income. This may be because rents represent current conditions, whereas asset values reflect expected future conditions. Finally, whether based on self-reported asset values or rental values, indices that rely on the community's median price, rather than micro data, are not as strongly correlated either with the other indices or with amenities or local incomes. The median price does not appear to aptly represent community-wide conditions.

Our approach to assessing housing price data is firmly rooted in the logic of discrete-continuous sorting models. Nevertheless, our results also speak to other applications where similar data questions arise, including inter-city quality-of-life indices and more standard hedonic price regressions, in which prices are regressed on structural characteristics as well as public goods in one step.

2. Approaches to estimating community housing price indices

Perhaps the most fundamental question when it comes to constructing a community-level housing price index is whether to use asset values or rental prices. Often, this choice is determined by the research context, and sometimes researchers make use of both types of data (e.g. Calabrese et al., 2006; Greenstone and Gallagher, 2008). Rental prices may be most relevant for short-term service flows, but only if rental contracts adjust rapidly to changes in public goods and taxes. If rents are sticky, they may be out of equilibrium over long periods. Observed asset prices may be more likely to be in equilibrium at a point in time, but reflect anticipated future service flows.[4]

Even when asset values are desired theoretically, in many cases they must be converted to annual user costs (see e.g. Poterba, 1992). Concluding that this conversion is too sensitive to assumptions about the opportunity cost of capital and expected capital gains, the US Bureau of Labor Statistics has for many years used rental properties as a proxy for the cost of owner-occupied housing in the Consumer Price Index (see e.g. Gillingham, 1983). The BLS's approach has the advantage of side-stepping the imputation of user costs from asset prices. When survey-based data must be used, this approach also has the advantage that households presumably are much more likely to know their monthly rent with accuracy than the market value of their home.[5] But it has the disadvantage of relying only on rental properties, which may be very different from the owner-occupied housing stock (Glaeser and Gyourko, 2007).

When asset values are to be used, they generally come from one of three sources: surveys of owners (e.g. the American Community Survey or American Housing Survey), actual transaction prices, and tax assessments. Each of these sources also has its characteristic advantages and disadvantages. Survey-based data is readily available from the US Census and has been used by Albouy (2012), Bajari and Kahn (2005), Bayer et al. (2009), Greenstone and Gallagher (2008), and many others. This data source is by far the most convenient, if publically available data are used. It also has the advantage of being a representative sample of all homes.

However, one disadvantage with publically available census housing data, at least in the US, is that they are available as micro data only at large geographies, such as a public-use micro area (PUMA), an area with about 100,000 people.[6] One solution to this problem is to use the restricted data, but of course this requires sacrificing the convenience of public data. An alternative approach is to forego the micro data and use aggregate, community-level statistics such as the median home value in a county or census tract (e.g. Chay and Greenstone, 2005; Grainger, 2012; Greenstone and Gallagher, 2008). Most economists would probably agree that, if available, the micro-level data is preferable, for two reasons. First, it makes use of a broader sample of houses.

[1] Broadly speaking, there are two branches to the literature. One approach leverages insights from hierarchical models, in which the demand for high-quality communities makes them more expensive, to estimate community-level quality indices based on these price indices. These quality indices, in turn, can then be decomposed into a function of observed public goods (see e.g. Calabrese et al., 2006; Epple and Sieg, 1999; Kuminoff, 2011, and Sieg et al., 2004). Another approach applies insights from logit models with horizontal differentiation to locational choice, estimating a preference for communities that can be decomposed into the utility for public goods and disutility for high housing costs (see e.g. Bayer et al., 2007, 2009, and Tra, 2010). See Kuminoff et al. (2013) for discussion.

[2] See Gamper-Rabindran et al. (2011) for one recent discussion of this issue.

[3] Sieg et al. (2002) apply their approach only to functional form and related issues using a single data source. They do not address the question of the type of price information (surveys, transactions, appraisals, or rents), which has been a larger question in the literature. We apply their basic insight to this question.

[4] For additional discussion of these issues, see Winters (2012). See Greenstone and Gallagher (2008) for an excellent example of the comparative use that can be made of both types of data.

[5] Calabrese et al. (2006) find that though tax assessments correlate well with self-reports of values at the community level, the same is not true for rental units. However, this may be a consequence of biases in tax assessments or in the user cost formula they use in their exercise.

[6] Previously, another concern was that the data were only available through the annual census. Annual micro data for PUMAs are now available through the American Community Survey.