Journal of Housing Economics 19 (2010) 75–93

Contents lists available at ScienceDirect

Journal of Housing Economics

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

Hedonic versus repeat-sales housing price indexes for measuring the recent boom-bust cycle

Robert E. Dorsey [b], Haixin Hu [a], Walter J. Mayer [a,*], Hui-chen Wang [a]

[a] University of Mississippi, Economics Dept. University, MS 38677, USA
[b] FNC, Inc., 1214 Office Park Drive, Oxford, MS 38655, USA

ARTICLE INFO

Article history:
Received 3 August 2009
Available online 10 April 2010

JEL classification:
C43
R21
C21

Keywords:
Hedonic housing price indexes
Repeat-sales indexes

ABSTRACT

Standard housing price indexes rely on strong constant-quality assumptions and often conflict. Hedonic price indexes overcome limitations of median price and repeat-sales indexes but their implementation has been limited by a lack of data. This paper constructs hedonic indexes at the zip code level for the Los Angeles and San Diego metropolitan areas using considerably more detailed data than previously available. Our sample was collected by a mortgage technology firm, and consists of almost 1.1 million transactions during the boom-bust cycle since 2000. Our hedonic regressions include new spatial models that capture correlations within submarkets (using zip codes as proxies) and allow temporal asymmetry. Compared to a repeat-sales price index constructed from the same data, the hedonic indexes indicate that the market peaked about 11 months later in Los Angeles and about 2 months earlier in San Diego, show less pre-peak appreciation and post-peak depreciation in low-tier housing and more pre-peak appreciation in high-tier housing. We also find that the intensity of the cycle varies greatly across zip codes and price-tiers in a pattern consistent with foreclosure activity.

1. Introduction

This paper uses a new extensive data set to construct constant-quality housing price indexes. During the boom-bust housing cycle of the last several years, releases of price indexes by the National Association of Realtors (NAR), the S&P/Case-Shiller™ and Federal Housing Finance Agency (FHFA) (formerly Office of Federal Housing Enterprise Oversight) have been and continue to be followed closely. Interpretation of these indexes is often difficult, however, because they are computed from samples of houses that are transacted infrequently and have unique attributes. In each period, the houses sampled represent only a small fraction of the housing stock. Comparisons of index values at different dates can be misleading if not adjusted for changes in the quality of the sold houses. For example, greater index values might merely reflect sales of larger houses rather than an increase in the price of a standard unit of housing.

Given a lack of data on property attributes, available housing price indexes have relied on strong assumptions to account for changes in quality. The median-price indexes published by the NAR and the Census Bureau implicitly assume that changes in the composition and quality of houses sold are negligible over time. This amounts to assuming that the distribution of housing prices in a given period would be unaffected if the houses sold were replaced with the houses sold in a different period. This is clearly problematic since the quality of the houses sold can change significantly over time.

The S&P/Case-Shiller™ and the FHFA indexes account for changes in quality through a repeat-sales methodology (Baily et al., 1963; Case and Shiller, 1987). For each sampled house, the repeat-sales approach assumes that the quality attributes and their coefficients are constant

* Corresponding author.
E-mail address: wmayer@olemiss.edu (W.J. Mayer).

1051-1377/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jhe.2010.04.001

between two sales dates. A regression of individual price changes between sales dates on time dummies then yields an estimate of the change in aggregate house prices holding quality constant. Arguably, the repeat-sales method is a considerable improvement over median sales prices. However, the assumption that quality is constant between sales dates is not likely to hold for every sampled house. Attributes are subject to change from a mix of improvements and deterioration due to age, and their coefficients reflect, in part, consumer preferences that change over time.

Repeat-sales indexes also omit a large number of transactions since they are confined to houses for which at least two sales are available in public records. For example, in Los Angeles County approximately 80% of the total transactions are not repeat sales for the three-year period 2003–2006. Clapp and Giaccotto (1999) and Clapham et al. (2006) argue that this causes repeat-sales indexes to be subject to substantial revisions as new data becomes available. Clapham et al. (2006) further argue that the levels of revision are large enough to hamper markets for home equity insurance and housing price futures. Case and Quigley (1991), Case et al. (1991), Quigley (1995) and Englund et al. (1998) highlight and address this shortcoming of repeat-sales indexes by proposing hybrid methodologies that combine samples of single sales and multiple sales.

Unfortunately for users, the three main indexes often conflict. As "The State of the Nation's Housing 2008" report by the Joint Center for Housing Studies of Harvard University (JCHS) (2008, p. 8) notes: "It is difficult to gauge with certainty how far prices have fallen to date. Each of the three major measures paints a different picture of the magnitude of declines to date." While the NAR median price index showed only a 1.8% decline in the national home price for 2007, the S&P/Case-Shiller™ shows a decline of 8.9% and the FHFA index shows an increase of 1.9%.

Hedonic regression can potentially overcome the limitations of the median price and repeat-sales methods. Since the pioneering work of Griliches (1961), hedonic regression has been used to construct constant-quality price indexes for automobiles, computers, consumer durables and other products. For housing, the hedonic approach entails regressing sale prices on a vector of property characteristics. A constant-quality price index is then constructed by using the regression to impute a series of prices for a reference set of houses. In contrast to repeat-sales, the estimates incorporate both single and repeat-sales transactions. The coefficients represent the contributory values of the characteristics. Consequently, unlike repeat-sales methods, changes in these values can be accommodated by allowing coefficients to change over time.

Although theoretically appealing, the main limitation of hedonic regression has been the data requirement (Case, 2006, p. 232). As Rappaport (2007, p. 44) notes, "...properly implementing the hedonic approach requires more detailed attribute data than are typically available." U.S. data are particularly sparse. As Clapham et al. (2006) point out, because property taxes in the U.S. are only local, there are no administrative reasons to develop extensive data sets of housing attributes and sale prices. As a result, many studies have focused on other countries that have better public data. For example, Clapham et al. (2006) use a sample of 600,000 transactions in Sweden, while Hill and Melser (2007) use a sample of 170,000 transactions in Sydney, Australia. In contrast, U.S. studies have relied on samples of fewer than 50,000 transactions (often considerably less) to construct indexes from city or county level hedonic models that are often pooled over 15–20 year periods (Meese and Wallace, 1997; Clapp, 2004). Such data might not be sufficient to adequately allow heterogeneity across region and over time. Economic conditions and the rate at which they change often vary greatly across zip codes within a metropolitan area. Moreover, there appear to be no studies that construct hedonic indexes for the recent U.S. boom-bust housing cycle.

In this paper we construct hedonic housing price indexes using considerably more detailed and comprehensive data than has been previously available. The data are from the National Collateral Database™ (NCD) operated by a mortgage technology firm, FNC, Inc. The NCD combines transaction and physical property characteristic data from appraisals used in loan originations with public records to create detailed records of residential properties. Our sample consists of almost 1.1 million transactions with detailed property characteristics that occurred in Los Angeles and San Diego metropolitan areas from 1999 to 2008. The extensive data allows us to account for regional and temporal heterogeneity by constructing monthly indexes at the zip code level from coefficients that vary over time.

Another contribution is that we use spatial autoregressive versions of hedonic models to construct the price indexes. The models are estimated with the two-stage least squares estimator proposed by Lee (2003). There are two advantages to addressing the spatial dependence of sale prices. First, the characteristic coefficients can be estimated more precisely because a potential source of omitted variable bias is eliminated. Second, the price of a house can be imputed holding constant the characteristics of the house as well as those of other houses in the neighborhood. The latter is important since neighborhood houses can enhance or detract the value of a house and, thus, are a key component of quality. In addition to the conventional fixed-distance spatial models, we propose and estimate new spatial models that capture correlation between houses in the same submarket (using zip codes as proxies) and allow temporal asymmetry in the spatial parameter. Conventional spatial models used in the housing literature are based on correlation between houses within a specified distance, and do not allow the correlation between the sale price of a house and other houses to depend on whether the other houses were sold at an earlier or later date. These models overlook that buyers typically search similar neighborhoods throughout a fairly large area, and that the correlation between sale prices of different houses is likely to be greater the closer the sale dates.

Our results include comparisons of six hedonic-based indexes to repeat-sales indexes. To identify differences that arise purely from differences in methodologies, we constructed a repeat-sales index by applying the Case-Shiller methodology to the same data used to construct the hedonic indexes. The comparisons cover estimates of peaks and price changes for the recent boom-bust housing cycle for three different price-tiers in the Los Angeles and San Diego metropolitan areas. Among our findings: (1) San Diego metro peaked earlier than Los Angeles and experienced less pre-peak appreciation and greater post-peak depreciation; (2) the hedonic indexes indicate that the market peaked about 11 months later in Los Angeles and about 2 months earlier in San Diego than the repeat-sales price index; (3) the hedonic indexes show less pre-peak appreciation and post-peak depreciation than the repeat-sales index overall in both Los Angeles and San Diego, the greatest differences being in low-tier housing; (4) an exception is high-tier housing where the hedonic indexes show greater pre-peak appreciation than the repeat-sales index; (5) the intensity of the cycle varies greatly across zip codes and price-tiers; (6) consistent with the collapse of the subprime market, there is strong negative correlation between distressed sales and appreciation at the zip code level.

Section 2 describes the hedonic regression models used to construct the indexes. Section 3 describes the estimation method. Section 4 describes the construction of the indexes. The results are presented and discussed in Section 5. Section 6 concludes.

2. Hedonic regressions with spatial correlation

2.1. First-order autoregressive spatial model

We specify hedonic regressions at the zip code level and, consequently, treat each zip code as a submarket. A submarket is generally defined as a collection of houses that buyers and sellers view as substitutes (Rothenberg et al., 1991). Empirically, submarkets are distinguished by differences in the hedonic coefficients. Researchers have used a variety of approaches to identify submarkets. In Case et al. (2004, pp. 178–181), for example, Case employs an endogenous approach in which homogenous within-county districts for Fairfax County are identified by a cluster algorithm. He finds that the resulting districts are largely contiguous groups of census tracts. Other approaches include Bourassa et al. (1999, 2003) and Goodman and Thibodeau (1998, 2003). Goodman and Thibodeu construct submarkets by estimating hierarchical models based on school zones in Dallas. Frequently, researchers simply impose submarket boundaries exogenously using municipal boundaries, zip codes groups, school districts or census tracts. This approach is computationally easier to implement. Moreover, Goodman and Thibodeau (2003, p. 200) report that their hierarchical-model approach did not statistically dominate either zip code or census tract defined submarkets. In view of this and the size of our data set (which is considerably larger than those analyzed by Case et al. (2004) and Goodman and Thibodeau (1998, 2003)), we will adopt the convention of using zip codes as the proxy for submarkets.

In what follows, $i(t)$ denotes the ith house sold at time t, and $n(t)$ is the number of sold houses observed at time t. At each time t, a different set of houses is generally sold and, therefore, $i(t)$ and $i(s)$ typically correspond to different houses when $t \neq s$. The data are a pseudo-panel in this sense. We consider a sample period beginning at time S and ending at time T consisting of $n(S,T) = \sum_{t=S}^{T} n(t)$ observations. For each zip code, we assume a first-order spatial autoregressive (SAR) model:

$$\ln p_{i(t)} = x_{i(t)}\beta + \lambda\omega_{i,n(S,T)} \ln P_{n(S,T)} + \sum_{t=S}^{T} \delta_t D_{i(t)} + \varepsilon_{i(t)},$$
$$i(t) = 1(t), \ldots, n(t); \quad t = S, \ldots, T, \tag{1}$$

where $p_{i(t)}$ denotes the observed sales price measured in dollars, β, λ and δ_t are unknown parameters to be estimated, $x_{i(t)}$ is a vector of observed attributes of houses specified below, $P_{n(S,T)}$ is the vector of all $n(S,T)$ prices in the sample observed from time S to T, $\omega_{i,n(S,T)}$ denotes a vector of spatial weights specified below, $D_{i(t)}$ is a dummy that equals one if the transaction occurred in period t and zero otherwise, and $\varepsilon_{i(t)}$ is an i.i.d. error term with mean zero and variance σ^2.

As Anselin and Lozano-Gracia (2008, p. 10) emphasize, a SAR model can be interpreted as specifying the house price as a function of the observed and unobserved attributes of neighboring properties as well as its own attributes through a decay operator defined by the spatial parameter and weight matrix. Eq. (1) can be written in matrix form as follows:

$$\ln P_{n(S,T)} = X_{n(S,T)}\beta + \lambda W_{n(S,T)} \ln P_{n(S,T)} + D_{n(S,T)}\delta + \varepsilon_{n(S,T)}, \tag{2}$$

where, for example, the jth row in $X_{n(S,T)}$ consists of the observed attributes of the jth house. Assume that $S_{n(S,T)}(\lambda) \equiv (I_{n(S,T)} - \lambda W_{n(S,T)})^{-1}$ exists. By Eq. (2):

$$\ln P_{n(S,T)} = S_{n(S,T)}(\lambda)X_{n(S,T)}\beta + S_{n(S,T)}(\lambda)D_{n(S,T)}\delta + S_{n(S,T)}(\lambda)\varepsilon_{n(S,T)}. \tag{3}$$

The reduced form (3) reveals that each individual price also depends on the observed attributes and error terms of the other sampled houses in the neighborhood. The error terms reflect omitted property-level attributes as well as omitted neighborhood amenities. Consequently, spatial correlation in the prices reflects both.[1]

2.2. Data

The extensive FNC data set permits estimation and construction of the indexes at the zip code level. Sample data consists of over one million sales of detached single-family houses in 471 zip codes in the Los Angeles Metropolitan Area (Los Angeles County and Orange County) and the San Diego Metropolitan Area (San Diego County) from February 1999 to June 2008. The data set is a blend of public record and appraisal data that includes sale prices, month and year of each transaction, and detailed property

[1] One alternative to the SAR is the SEM (spatial error model) that assumes $y = X\beta + \varepsilon$ where $\varepsilon = \lambda W\varepsilon + u$. The SEM can be viewed as a version of the SAR model since it implies $y = x\beta - \lambda Wx\beta + \lambda Wy + u$. An interesting application is provided by Case in Case et al. (2004, pp. 178–181). Case estimated two-stage SEM models by adding nearest-neighbor residuals to the hedonic equation. The models were found to predict better than six competing models on the basis of several measures.

attributes. Table 1 describes the variables used to estimate the models. Tables 2 and 3 report the summary statistics by year for Los Angeles and San Diego metropolitan areas.

2.3. Observed attributes: specification of $x_{i(t)}$

The specification of $x_{i(t)}$ is a standard one that includes the age of the house (AGE) measured as the difference between transaction and construction years, gross living area in thousands of square feet (GLA), lot size in thousands of square feet (LOT), the number of bedrooms (BED) and number of bathrooms (BATH). These attributes reflect the quality of a house at the time of a sale and, except for AGE, are expected to have positive effects on the sale prices. Following previous hedonic studies, the squares of AGE, GLA, LOT, BED and BATH are also included to allow for nonlinearities.

The other variables in $x_{i(t)}$ are the average property tax rate imputed for each house (TAX) and the buyer's loan to value ratio (LTV). TAX (in percentage) is measured as

Table 1
Variable descriptions and summary statistics.

Names	Variable descriptions	All sales		Single sales		Repeat sales	
		Means	SD	Means	SD	Means	SD
LA metro							
PRICE	Sale price (dollar)	464,490	363,789	462,259	367,969	467,074	358,870
AGE	Age of house (years)	41.92	22.56	41.33	21.85	42.60	23.34
GLA	Gross living area (1000 square feet)	1.76	0.82	1.81	0.84	1.71	0.80
BED	Number of bedrooms	3.21	0.88	3.25	0.88	3.17	0.88
BATH	Number of bathrooms	2.22	0.96	2.25	0.97	2.18	0.95
LOT	Lot size (1000 square feet)	8.54	10.72	8.99	11.46	8.02	9.77
TAX	Imputed average property tax rate (%)	1.20	0.13	1.20	0.14	1.19	0.12
LTV	Buyer's loan to value ratio (%)	85.74	15.79	84.54	16.19	87.14	15.20
SD metro							
PRICE	Sale price (dollar)	480,349	342,788	478,571	350,964	482,807	331,145
AGE	Age of house (years)	28.62	20.09	29.29	19.61	27.69	20.70
GLA	Gross living area (1000 square feet)	1.89	0.88	1.91	0.89	1.87	0.87
BED	Number of bedrooms	3.33	0.86	3.34	0.85	3.32	0.87
BATH	Number of bathrooms	2.41	0.93	2.40	0.92	2.43	0.93
LOT	Lot size (1000 square feet)	15.48	25.85	16.46	26.91	14.12	24.24
TAX	Imputed average property tax rate (%)	1.13	0.11	1.12	0.11	1.13	0.10
LTV	Buyer's loan to value ratio (%)	83.59	16.54	82.86	16.73	84.59	16.22

Table 2
Summary statistics of variables by calendar years[a]: Los Angeles metro.

N	Total	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
	872,185	95,583	98,685	98,328	106,928	108,981	105,842	105,122	82,967	50,267	19,482
PRICE											
Mean	464,490	279,937	304,901	324,534	374,815	442,965	543,308	625,981	672,504	737,016	608,709
SD	363,789	238,072	260,209	245,442	272,916	312,706	365,914	384,172	404,570	482,790	456,992
AGE											
Mean	41.92	39.85	40.63	39.94	40.66	41.28	42.60	43.14	44.75	46.19	45.76
SD	22.56	20.66	21.38	22.05	22.28	22.70	22.90	23.31	24.04	23.31	22.34
GLA											
Mean	1.76	1.74	1.75	1.75	1.78	1.79	1.77	1.74	1.74	1.82	1.80
SD	0.82	0.80	0.82	0.80	0.83	0.85	0.85	0.81	0.81	0.86	0.81
BED											
Mean	3.21	3.19	3.19	3.21	3.22	3.23	3.22	3.21	3.20	3.23	3.26
SD	0.88	0.86	0.87	0.87	0.88	0.89	0.89	0.88	0.89	0.91	0.87
BATH											
Mean	2.22	2.21	2.22	2.23	2.25	2.24	2.20	2.19	2.17	2.25	2.26
SD	0.96	0.94	0.97	0.94	0.96	0.97	0.97	0.95	0.95	1.00	0.95
LOT											
Mean	8.54	8.45	8.42	8.34	8.46	8.72	8.76	8.57	8.48	8.74	8.39
SD	10.72	10.30	10.23	9.90	10.00	11.22	11.73	11.29	10.84	11.17	9.48
TAX											
Mean	1.20	1.20	1.20	1.20	1.20	1.19	1.19	1.18	1.19	1.24	1.23
SD	0.13	0.13	0.13	0.13	0.12	0.12	0.11	0.11	0.12	0.17	0.16
LTV											
Mean	85.74	86.51	85.89	85.37	84.68	84.57	85.59	86.92	89.01	84.49	79.41
SD	15.79	14.82	15.38	15.58	15.58	16.01	15.98	15.90	15.52	16.88	16.59

[a] Except for year 1999 and year 2008, a calendar year is referred to January to December of the year. Year 1999 is a partial year from February to December, and year 2008 is a partial year from January to June.

Table 3
Summary statistics of variables by calender years[a]: San Diego metro.

N	Total 193,628	1999 23,492	2000 20,905	2001 16,230	2002 19,431	2003 27,235	2004 27,551	2005 23,070	2006 16,459	2007 13,222	2008 6033
PRICE											
Mean	480,349	280,449	319,343	337,342	392,007	472,935	585,445	648,032	657,915	662,423	514,748
SD	342,788	221,593	253,656	238,846	252,141	283,937	331,251	381,016	380,662	431,477	363,114
AGE											
Mean	28.62	24.17	25.10	26.29	27.70	28.76	29.95	31.25	31.66	32.83	32.99
SD	20.09	19.21	19.52	19.37	19.42	19.86	20.14	20.18	20.58	20.88	20.34
GLA											
Mean	1.89	1.92	1.93	1.86	1.88	1.89	1.88	1.86	1.90	1.95	1.87
SD	0.88	0.86	0.91	0.86	0.87	0.89	0.89	0.88	0.89	0.93	0.83
BED											
Mean	3.33	3.36	3.34	3.29	3.29	3.33	3.33	3.31	3.35	3.36	3.37
SD	0.86	0.86	0.87	0.85	0.86	0.86	0.86	0.85	0.87	0.87	0.85
BATH											
Mean	2.41	2.44	2.44	2.38	2.39	2.41	2.41	2.38	2.43	2.47	2.42
SD	0.93	0.89	0.93	0.90	0.90	0.93	0.93	0.93	0.96	0.98	0.90
LOT											
Mean	15.48	15.52	15.90	15.91	15.92	15.07	16.04	15.18	15.18	15.20	13.14
SD	25.85	25.87	26.35	26.36	26.66	25.47	27.05	25.27	25.00	25.16	21.43
TAX											
Mean	1.13	1.13	1.13	1.12	1.12	1.12	1.12	1.12	1.12	1.15	1.14
SD	0.11	0.12	0.11	0.10	0.10	0.10	0.10	0.09	0.10	0.15	0.15
LTV											
Mean	83.59	84.13	82.85	83.33	83.11	82.71	83.20	84.35	86.71	83.26	81.25
SD	16.54	16.02	16.66	16.28	16.33	16.43	16.20	16.07	17.40	17.75	16.93

[a] Except for year 1999 and year 2008, a calendar year is referred to January to December of the year. Year 1999 is a partial year from February to December, and year 2008 is a partial year from January to June.

the ratio of the actual property tax liability and the assessed value last reported for a house. Therefore, TAX reflects the property tax rate last observed for a house.[2] We are unable to calculate the tax rates specific to the time of a sale because NCD constantly updates information on tax liabilities and assessed values and keeps only the most recent records (which do not necessarily match the most recent sales). Since property tax rates typically do not have frequent substantial changes, our imputed property tax rate should represent adequately the tax burden associated with each house during our analysis period. The influence of property tax burdens on housing prices is an empirical question. The theory of capitalization, first formally developed and tested by Oates (1969), predicts that property tax burdens reduce housing prices. On the other hand, higher property taxes are often reflective of better public services in a neighborhood, which increase the housing prices per the Tiebout model (Tiebout, 1956). Existing studies have struggled to distinguish the two effects and found mixed evidence.[3] In the absence of an exhaustive list of controls for local amenity and public services, the coefficient of TAX likely captures the combined effects of tax capitalization and unobserved public services.

The variable LTV is the percent of a sale price that was financed and, thus, measures buyer liquidity. The inclusion of LTV can be motivated by down-payment effects (see, for example, Stein, 1995): the demand for houses falls when buyer liquidity is lower (reflected by higher LTV) because significant down payments are typically required. Consequently, the expected sign of LTV is negative. Alternatively, as suggested in Reichert (1990), LTV captures credit conditions which has a positive effect on housing prices. It should be noted that LTV may be partly endogenous for two reasons. First, the denominator of LTV is the sale price and, consequently, LTV might be correlated with omitted attributes (reflected in the error term) that directly affect price. Second, the numerator is the amount borrowed which is a decision variable for the buyer that depends on unobserved factors such as expected future income streams that might affect offer prices. We experimented with excluding LTV from the matrix of instrumental variables of the BG2SLS estimator described in Section 3, thereby allowing it to be endogenous. This did not appreciably change the coefficients or the index values and, therefore, the endogeniety does not appear to be critical for the purpose of constructing price indexes.

2.4. Spatial correlation: specification of $\omega_{i,n(S,T)}$

Spatially correlated prices arise because: (a) the prices of other houses reflect the unobserved amenities and economic conditions shared by houses in the neighborhood, and (b) buyers have incentives to substitute away from higher-price houses to lower-priced houses. Shared amenities include schools, police, fire, hospitals and commercial

[2] This rate is inclusive of the state-wide nominal tax rate (which is 1% of assessed value per Proposition 13 of the California Constitution) and any special tax rate for the locality.

[3] See for example, Rosen (1982) and Palmon and Smith (1998). Also, ideally, effective tax rates (tax liability divided by the market value of a house) are the preferred measure as opposed to nominal tax rates for testing either hypothesis. However, our data does not allow us to calculate the effective tax rates because the time of tax assessment do not always match the time of the sale.

districts. They are typically confined to relatively small geographical areas. However, consumers do not generally limit their searches to the few blocks of a particular neighborhood. As Goodman and Thibodeau (2007) note, consumers constrained by income are likely to search similar neighborhoods throughout a large area. This suggests that a broad definition of neighborhood such as Census Tracts is needed to capture spatial correlation. Consistent with this usage, the Census website states: "Census Tracts are designed to be homogeneous with respect to population characteristics, economic status and living conditions."

Two different weighting schemes were used to estimate Eq. (1).

The first scheme averages the log of prices across the Census Tract. Each zip code consists of several Census Tracts. The vector $\omega_{i,n(S,T)}$ is specified as an $n_{(S,T)}$ dimensional row vector with jth element $ctrt_{i(t)j(h)}/\sum_{j(h)\in(S,T)} ctrt_{i(t)j(h)}$ where $ctrt_{i(t)j(h)}$ equals one if house $j(h)$ is in the same Census Tract as house $i(t)$ and equals zero otherwise. Note that $\omega_{i,n(S,t)} \ln P_{n(S,T)}$ is the sample mean of $\ln p_{j(h)}$ for all houses sold in the same Census tract as the ith house and, consequently, is an estimate of the conditional expectation, $E(\ln p_{j(h)} \mid ctrt_{i(t)j(h)} = 1)$. In particular, $\omega_{i,n(S,t)} \ln P_{n(S,T)}$ is a nonparametric (uniform kernel) regression estimate of $E(\ln p_{j(h)} \mid ctrt_{i(t)j(h)} = 1)$.

Hence, under this weighting scheme, Eq. (1) specifies that the price for a house is partly determined by an estimate of the expected price in the Census Tract. Higher quality or aggregate demand for a given level of quality increases the expected price. In turn, the prices of individual houses increase because potential buyers would have to pay a higher average price for substitute houses.

The other weighting scheme is a conventional one based on normalized binary indicators of a fixed distance (Anselin, 1988; Dubin, 1998). Here $\omega_{i,n(S,T)}$ is specified as an $n_{(S,T)}$ dimensional row vector with jth element $dist_{i(t)j(h)}/\sum_{j(h)\in(S,T)} dist_{i(t)j(h)}$ where $dist_{i(t)j(h)}$ equals one if the distance between house $i(t)$ and $j(h)$ is less than one half mile and equals zero otherwise. The variable $\omega_{i,n(S,t)} \ln P_{n(S,T)}$ is the sample average of the log price of houses within a half a mile of the ith house and, consequently, is an estimate of $E(\ln p_{j(h)} \mid dist_{i(t),j(h)} = 1)$. Houses in close proximity are not necessarily substitutes. Despite a lack of substitution, however, spatial correlation may still exist among nearby houses due to shared amenities.

Changes in amenities and economic conditions transmit to sale prices through changes in list prices and reservation prices. Decreases in consumer income and less favorable credit conditions, for example, lower the probability of receiving a bid of a given size. In response, sellers lower their reservation and list prices which leads to lower sale prices (Horowitz, 1992; Knight et al., 1994). Our data are limited to sale prices and do not include list prices or reservation prices. The role of list prices, however, has important implications for modeling the correlation between observed sale prices. Selling a house often requires competing with other listed houses for a substantial period of time that may sell before or after the house. Despite differences in sales dates, however, the sale prices of all the houses reflect, in part, conditions of their common time on the market. It follows that the sale price of a house is potentially correlated with past and future as well as current (which means in the same month for our data) sale prices of other houses.

The timing of the sales may still be relevant due to temporal asymmetry in the effects of the other sale prices. For example, past and current sale prices are fully realized and observable at the time of a transaction, whereas future sales are only relevant to the extent that their list prices convey information. Therefore, spatial correlation may be greater between past and current sales than between future and current sales. Another possibility is that current sale prices exert a stronger effect than past and future prices because they are more closely tied to recent market conditions.

Consequently, the sale price of a house is potentially correlated with the sale prices of houses sold at different dates and that correlation may be different for past, current and future sales. To allow for this, we also estimated the following equation under each weighting scheme:

$$\ln p_{i(t)} = X_{i(t)}\beta + \lambda^{past}\omega_{i,n(S,t-1)} \ln P_{n(S,t-1)} + \lambda^{current}\omega_{i,n(t)} \ln P_{n(t)} + \lambda^{future}\omega_{i,n(t+1,T)} \ln P_{n(t+1,T)} + \sum_{j=S}^{T} \delta_j D_{i(t)} + \varepsilon_{i(t)},$$

where $p_{i(t)}$ denotes the observed sale price, β, λ^{past}, $\lambda^{current}$ and λ^{future} are unknown parameters. For example, the variable $\omega_{i,n(S,t-1)} \ln P_{n(S,t-1)}$ captures the effects of past sale prices.

3. Estimation method

We use the Best Generalized Two-Stage Least Squares Estimator (BG2SLS) proposed by Lee (2003) to estimate $\beta, \lambda, \delta_S, \ldots, \delta_T, \sigma^2$. It is an asymptotically efficient version of the spatial 2SLS estimator originally proposed by Kelejian and Prucha (1998). It is not as efficient as maximum likelihood but relies on weaker assumptions and is computationally much faster. To describe it, let $Z_{n(S,T)} = (X_{n(S,T)}, W_{n(S,T)} \ln P_{n(S,T)})$, $\theta = (\beta', \lambda, \delta')'$ and $H_{n(S,T)} = (X_{n(S,T)}$, *matrix of polynomials*), where the matrix of polynomials is constructed from variables in X. Finally, let $\Pi_{H_n} = H_{n(S,T)}(H'_{n(S,T)}H_{n(S,T)})^{-1}H'_{n(S,T)}$.The BG2SLS is computed in two steps:

Step 1: Compute $\tilde{\theta} = (Z'_{n(S,T)}\Pi_{H_n}Z_{n(S,T)})^{-1}Z'_{n(S,T)}\Pi_{H_n}P_{n(S,T)}$.
Step 2: Let $\tilde{H}_n = [X_{n(S,T)}, W_{n(S,T)}(I_{n(S,T)} - \tilde{\lambda}W_{n(S,T)})^{-1}X_{n(S,T)}\tilde{\beta}]$.

Compute the BG2SLS estimator:

$$\hat{\theta}_{\text{BG2SLS}} = (\tilde{H}'_n Z_{n(S,T)})^{-1}\tilde{H}'_n P_{n(S,T)}$$

To estimate the variance we use:

$$\hat{\sigma}^2 = (P_{n(S,T)} - X_{n(S,T)}\hat{\beta} - \hat{\lambda}W_{n(S,T)}P_{n(S,T)} - D_{n(S,T)}\hat{\delta})' \times (P_{n(S,T)} - X_{n(S,T)}\hat{\beta} - \hat{\lambda}W_{n(S,T)}P_{n(S,T)} - D_{n(S,T)}\hat{\delta})/n(S,T)$$

To address heterogeneity across locations and over time, Eq. (2) is estimated for each zip code using monthly data with a 12-month moving window. The moving window accommodates changes over time in contributory values of characteristics and the spatial correlation parameter. Such breaks are likely to occur but at unknown points of times. Identification of structural breaks at unknown points of time is notoriously difficult, particularly for the present problem of estimating individual models for almost 300 zip codes. Consequently, we do not attempt to identify exact break points. Instead, we adopt a strategy of using a relatively short estimation window to mitigate the impact of any breaks. It is well known that pooling a historical data series that contains breaks results in inconsistent parameter estimates. Despite this, however, as Pesaran and Timmermann (2007) recently emphasize, there can be gains in forecasting performance by including both pre- and post-break data provided that the break is not too large. There are of course trade-offs in selecting the length of the estimation window. Shorter estimation windows decrease the likelihood of large breaks but also decrease the number of observations used to estimate the parameters for each zip code. A 12-month window seemed to be a reasonable compromise.

Of the 471 zip codes covered by the sample, 181 had fewer than 150 observed sales in any 12 month window, which we deemed as the minimum needed to adequately estimate the spatial models. We pooled the transactions in these "small" zip codes into larger, adjacent zip codes, bringing the total number of zip code groups to 290 (242 Los Angeles metro and 48 San Diego metro). To pool the zip codes, the average longitude and latitude coordinates was calculated for each. Each small zip code was then pooled into the larger zip code with the closest average coordinates. The four spatial models were then estimated for each of the 102 moving windows for each of the 290 zip code groups. An alternative to the above scheme is temporal pooling. We did not use temporal pooling because this would have required estimation windows much greater than 12 months in most cases.

4. Index construction

The hedonic approach entails imputing the prices of houses with a specified set of quality characteristics. Once the prices are imputed, a constant-quality index can be constructed. The quality of a house depends on its own characteristics as well as those of other houses in the neighborhood. Houses with desirable characteristics enhance the value of other houses while undesirable houses detract value. For this reason, regulations and covenants commonly restrict the characteristics of all houses in a given neighborhood. Therefore, "holding quality constant" for a particular house also requires holding constant the characteristics of other houses. Spatial models make this dependence explicit and allow one to hold constant the observed quality characteristics of other houses (see Eq. (3)).

Conditional expectations are best predictors in the sense of minimizing the expected mean square error loss over all functions of the conditioning variables. Consequently, a natural choice to impute price is an estimate of conditional expectation of price given $X_{n(S,T)}$ which requires an estimate of $\beta, \lambda, \delta, \sigma^2$. The moving window described in the previous section updates the estimates of $\beta, \lambda, \delta, \sigma^2$ each month as follows: $\beta, \lambda, \delta, \sigma^2$ are estimated for January 2000, for example, using the observations from February 1999 to January 2000; $\beta, \lambda, \delta, \sigma^2$ are then re-estimated for February 2000 using the observations from March 1999 to February 2000 and so on. Consequently, there is a different set of estimates of $\beta, \lambda, \delta, \sigma^2$ for each time period. For a given zip code, let $\beta_t, \lambda_t, \delta_t, \sigma_t^2$ denote the population values at time t. We use January 2000 as the base period for the index. Let X_n denote the matrix of characteristics of all n houses in the 12 month window from February 1999 to January 2000, and let $E(p_{i(t)} \mid X_n, \beta_t, \lambda_t, \delta_t, \sigma_t^2)$ denote the conditional expectation implied by Eq. (1) of the ith house at time t. For each zip code, we calculate a monthly Laspeyres-type index defined as the ratio of the average estimated conditional expectations. The index value for the jth zip code at time t is:

$$\pi_j^{t,Jan00} = \frac{n(Jan00)^{-1}\sum_{i\in Jan00} E(p_{i(t)} \mid X_n, \hat{\beta}_t, \hat{\lambda}_t, \hat{\delta}_t, \hat{\sigma}_t^2)}{n(Jan00)^{-1}\sum_{i\in Jan00} E(p_{i(t)} \mid X_n, \hat{\beta}_{Jan00}, \hat{\lambda}_{Jan00}, \hat{\delta}_{Jan00}, \hat{\sigma}^2_{Jan00})} \times 100, \quad t = Jan00, Feb00, \ldots, June08, \tag{4}$$

where $n(Jan00)$ denotes the number of transactions in January 2000 and $\hat{\beta}_t, \hat{\lambda}_t, \hat{\delta}_t, \hat{\sigma}_t^2$ denote the BG2SLS estimates. Eq. (4) tracks the average January 2000 price of housing over time holding constant the observed quality characteristics of all houses in the preceding 12 months. The separately calculated zip code indexes were combined to form metro aggregate price indexes. The aggregate price indexes are weighted averages of the zip code indexes, where the weights are based on the number of detached housing units in zip code according to the 2000 Census.

To derive the expressions for the conditional expectations in (4), let $Q_n(\lambda) = S_n(\lambda)S_n(\lambda)'$, $q_{ij}(\lambda)$ denote the i,jth element of $Q_n(\lambda)$, and $q_{ij}^{-}(\lambda)$ denote the i,jth element of the inverse $Q_n^{-1}(\lambda)$. Under the assumption $\varepsilon_n \mid X_n \sim N(0, \sigma^2 I_n)$, it follows from Eq. (3) and standard results that P_n has a multivariate log-normal distribution with $E(\ln p_j \mid X_n) = s_{j^*}(\lambda)(X_n\beta + D_n\delta)$ where $s_{j^*}(\lambda)$ denotes the jth row of $S_n(\lambda)$, $V(\ln p_j \mid X_n) = \sigma^2 q_{jj}^{-}(\lambda)$ and $E(p_j \mid X_n) = \exp[E(\ln p_j \mid X_n) + V(\ln p_j \mid X_n)/2]$ where p_j denotes the jth element of P_n.

The above expressions are also useful for interpreting the coefficients in the spatial models. Changes in the characteristics of a house change the expected price of the house directly and indirectly through their impact on the expected prices of other houses. The direct effect is the change in the expected price of the house with the expected prices of the other houses held constant and, consequently, can be viewed as the change that would occur if there was no spatial correlation, $\lambda = 0$. The indirect effect is the additional change in the expected price of the house resulting from spatial correlation, $\lambda \neq 0$. Let x_{hj} denote the hth characteristic for the jth house. It can be shown from the above expressions that:

$$\frac{\partial E(p_j \mid X_n)}{\partial x_{hj}} \frac{1}{E(p_j \mid X_n)} = s_{hj^*}(\lambda)\beta_h,$$

where $s_{hj^*}(\lambda)$ is the hth element of $s_{j^*}(\lambda)$. Therefore, 100 $s_{hj^*}(\lambda)\beta_h$ can be interpreted as the percentage change in $E(p_j \mid X_n)$ associated with a one unit increase in x_{hj}. Using a mean value expansion about $\lambda = 0$ we can write $s_{hj^*}(\lambda)\beta_h$ as the sum of direct and indirect effects: $s_{hj^*}(\lambda)\beta_h = s_{hj^*}(0)\beta_h + \lambda\partial s_{hj^*}(\bar{\lambda})/\partial\lambda$, where $\bar{\lambda}$ lies between 0

Table 4
Summary statistics of regression coefficients for LA metro (January 2000 to June 2008, 242 zip code groups).

Variables	OLS model		Census_1 model		Census_2 model		Distance_1 model		Distance_2 model	
	Means	SD	Means	SD	Means	SD	Means	SD	Means	SD
AGE	−0.0010	0.0074	−0.0015	0.0066	−0.0014	0.0068	−0.0010	0.0074	−0.0012	0.0073
AGE^2	−0.000003	0.0001	0.000004	0.0001	0.000002	0.0001	−0.000003	0.0001	0.000002	0.0001
GLA	0.3733	0.1891	0.3296	0.1758	0.3343	0.1769	0.3728	0.1893	0.3301	0.1803
GLA^2	−0.0312	0.0448	−0.0268	0.0464	−0.0277	0.0469	−0.0314	0.0452	−0.0276	0.0484
BED	0.0704	0.1152	0.0737	0.1106	0.0710	0.1109	0.0701	0.1149	0.0703	0.1159
BED^2	−0.0106	0.0163	−0.0103	0.0157	−0.0101	0.0157	−0.0105	0.0163	−0.0100	0.0162
BATH	0.0047	0.1009	0.0040	0.0945	0.0044	0.0947	0.0045	0.0987	0.0058	0.1005
BATH^2	−0.0001	0.0225	−0.0002	0.0208	−0.0002	0.0208	−0.0002	0.0210	−0.0006	0.0229
LOT	0.0363	0.0371	0.0326	0.0360	0.0331	0.0367	0.0364	0.0373	0.0330	0.0383
LOT^2	−0.0010	0.0021	−0.0010	0.0024	−0.0010	0.0024	−0.0010	0.0022	−0.0010	0.0026
TAX	−0.7760	1.0255	−0.6978	0.9186	−0.7086	0.9395	−0.7728	1.0199	−0.7047	0.9405
LTV	−0.0021	0.0022	−0.0019	0.0021	−0.0019	0.0021	−0.0021	0.0022	−0.0019	0.0022
Lambda			0.2011	0.4358			0.0018	0.0072		
Lambda_past					0.0933	0.3465			0.0459	0.3457
Lambda_current					0.1177	0.1790			0.1164	0.2885
Lambda_future					0.0084	0.0440			0.0479	0.2853

Note: All regressions include time dummies.

Table 5
Summary statistics of regression coefficients for SD metro (January 2000 to June 2008, 48 zip code groups).

Variables	OLS model		Census_1 model		Census_2 model		Distance_1 model		Distance_2 model	
	Means	SD	Means	SD	Means	SD	Means	SD	Means	SD
AGE	−0.0018	0.0056	−0.0017	0.0050	−0.0018	0.0051	−0.0018	0.0056	−0.0017	0.0048
AGE^2	0.00001	0.0001	0.000008	0.0001	0.00001	0.0001	0.00001	0.0001	0.00001	0.0001
GLA	0.4119	0.2139	0.3502	0.1803	0.3587	0.1859	0.4104	0.2129	0.3566	0.1794
GLA^2	−0.0337	0.0459	−0.0263	0.0410	−0.0281	0.0421	−0.0335	0.0458	−0.0289	0.0404
BED	0.0786	0.1370	0.0843	0.1200	0.0845	0.1242	0.0781	0.1372	0.0752	0.1222
BED^2	−0.0123	0.0181	−0.0121	0.0161	−0.0121	0.0169	−0.0123	0.0181	−0.0111	0.0162
BATH	0.0090	0.1406	0.0111	0.1282	0.0121	0.1288	0.0094	0.1409	0.0093	0.1252
BATH^2	0.0013	0.0271	0.0006	0.0250	0.0000	0.0251	0.0012	0.0271	0.0007	0.0243
LOT	0.0113	0.0122	0.0100	0.0099	0.0098	0.0098	0.0113	0.0122	0.0101	0.0105
LOT^2	−0.0002	0.0004	−0.0001	0.0003	−0.0001	0.0003	−0.0002	0.0004	−0.0001	0.0003
TAX	−0.0294	0.6701	−0.0571	0.5956	−0.0909	0.5355	−0.0303	0.6648	−0.0525	0.5783
LTV	−0.0014	0.0015	−0.0012	0.0013	−0.0012	0.0013	−0.0014	0.0015	−0.0012	0.0013
Lambda			0.2315	0.2188			0.0011	0.0058		
Lambda_past					0.1196	0.1853			0.0089	0.0478
Lambda_current					0.1066	0.1683			0.1625	0.1795
Lambda_future					0.0053	0.0298			0.0049	0.0287

Note: All regressions include time dummies.

and λ. Since $s_{jh^*}(0) = 1$, it follows that $100\beta_h$ can be interpreted as the percentage change in $E(p_j \mid X_n)$ per unit increase in x_{hj} holding the expected prices of other houses constant, hence the direct effect.

5. Results

5.1. Coefficient estimates

The four spatial models described above and an OLS benchmark model were estimated for each of the 102 months (from January 2000 to June 2008) for each of the 290 zip code groups. The spatial models are summarized as follows:

(1) **Census_1**: $\omega_{i,n(S,T)} = ctrt_{i(t)j(h)}/\sum_{j(h)\in(S,T)} ctrt_{i(t)j(h)}$, $\lambda^{past} = \lambda^{current} = \lambda^{future}$, where $ctrt_{i(t)j(h)} = 1$ house $j(h)$ in same Census Tract as house $i(t)$, and =0 otherwise.
(2) **Census_2**: $\omega_{i,n(S,T)} = ctrt_{i(t)j(h)}/\sum_{j(h)\in(S,T)} ctrt_{i(t)j(h)}$, λ^{past}, $\lambda^{current}$, λ^{future} not restricted.
(3) **Distance_1**: $\omega_{i,n(S,T)} = dist_{i(t)j(h)}/\sum_{j(h)\in(S,T)} dist_{i(t)j(h)}$, $\lambda^{past} = \lambda^{current} = \lambda^{future}$, where $dist_{i(t)j(h)} = 1$ if distance between house $i(t)$ and $j(h) < 1/2$ mile, and =0 otherwise.
(4) **Distance_2**: $\omega_{i,n(S,T)} = dist_{i(t)j(h)}/\sum_{j(h)\in(S,T)} dist_{i(t)j(h)}$, λ^{past}, $\lambda^{current}$, λ^{future} not restricted.

Each estimation model produced a set of coefficient estimates for each of the 102 estimation windows for each zip code group. The magnitude of coefficients varies across locations and over time, but generally have expected signs when significant. Tables 4 and 5 show the sample means and standard deviations of the coefficient estimates from all zip- and time-specific regressions for the Los Angeles and San Diego metropolitan areas separately.[4] For the Census_1 estimates, Tables 6 and 7 show the percentages of coefficients that are significant at the 5% level for all

[4] Note that we cannot infer the degree of significance from the summary statistics provided in Tables 4 and 5.

Table 6
Percentages of significant coefficients (SD metro Census_1 model).

Variable	Positive	Positive and significant	Negative	Negative and significant
AGE	32.5	6.54	67.5	34.5
AGE^2	60.27	27.39	39.73	10.42
GLA	97.26	86.72	2.74	0.18
GLA^2	20.02	4.08	79.98	42.91
BED	77.98	38.89	22.02	2.39
BED^2	20.32	2.08	79.68	38.56
BATH	52.41	13.81	47.59	10.6
$BATH^2$	55	13.62	45	10.05
LOT	95.61	76.49	4.39	0.1
LOT^2	10.6	0.88	89.4	60.09
TAX	40.87	11.56	59.13	27.84
LTV	13.11	1.33	86.89	51.59
Lambda	90.6	74.82	9.4	1.1

Table 7
Percentages of significant coefficients (LA metro Census_1 model).

Variable	Positive	Positive and significant	Negative	Negative and significant
AGE	39.37	10.84	60.63	29.12
AGE^2	52.67	20.88	47.33	14.45
GLA	97.27	82.98	2.73	0.06
GLA^2	19.44	2.81	80.56	43.18
BED	78.25	32.56	21.75	2.08
BED^2	22.02	2.28	77.98	31.58
BATH	52.11	11.42	47.89	10.39
$BATH^2$	50.46	12.24	49.54	10.75
LOT	94.49	79.17	5.51	1.08
LOT^2	11.06	1.96	88.94	63.66
TAX	11.44	1.92	88.56	58.71
LTV	12.42	1.77	87.58	56.44
Lambda	86.88	66.49	13.12	1.92

estimation windows. The results for the other estimation methods are similar. As Tables 6 and 7 reveal, the most consistently significant variables are GLA, LOT and the spatial parameter.

In Tables 4 and 5 the OLS results are presented in the first column for comparison with the spatial models in the remaining columns. As noted above, coefficients represent portions of percentage change in the expected price of a house per unit increase in a characteristic holding the expected prices of other houses constant. The mean coefficient of AGE is negative in all models and in both metropolitan areas. The coefficients of the square of AGE are positive in most models, but they are typically very small, especially when compared to the linear terms. The mean coefficients of GLA, BED, BATH and LOT are consistently positive. However, the means of the squares of GLA, BED and LOT are negative suggesting diminishing returns. The mean coefficient of the square of BATH is negative for the Los Angeles metropolitan area but positive in the San Diego metropolitan area; the former again suggests diminishing returns to having more bathrooms in a house and the later increasing returns. The mean coefficients of property tax rate are negative for both metropolitan areas which is consistent with the capitalization argument. The mean coefficients of the loan to value ratio are negative across the board, suggesting that lower buyer liquidity (higher LTV) decreases housing prices.

In all four spatial models and both metropolitan areas, the mean coefficients of the spatial parameters (λ's) are positive. This is consistent with our expectation that the price of a house is positively influenced by the sale prices of its close substitutes or neighbors. The Census_2 and Distance_2 models explore the role of timing in the spatial effects. We generally find that the spatial effect of housing prices is the greatest for houses sold during the same month, compared to their past and future counterparts. In particular, only the spatial effect in the same month is statistically significant in most cases. This result is plausible because the current sale prices likely convey the most relevant information on the recent market conditions. The one exception is the Census_2 model for SD where the past spatial parameter is just slightly greater than the current (0.1196 versus 0.1066).

To further illustrate the estimation results, Table 8 provides an example zip code (90623) in the Los Angeles metropolitan area for the estimation window ending May 2007. We find that the coefficients are generally robust across model specifications. For this specific zip code and time frame, housing prices are significantly affected by gross living area, number of bedrooms, lot size, and, in some models, spatial correlation. An increase in gross living area by one thousand square feet directly raises the expected price of a house by 19–21%. This effect does not diminish with increasing GLA, since the coefficients of the square terms are statistically insignificant. On the other hand, the number of bedrooms and lot size were found to affect prices positively only up to a threshold with diminishing returns. An additional bedroom would increase the value of a house if it had fewer than four bedrooms, and the effect becomes negative if a house has four or more bedrooms. This result is not surprising since gross living area is controlled for. Holding gross living area constant, additional bedrooms mean smaller rooms, which will likely become undesirable after some point. Similarly, we find that greater lot size would increase the value of a house only for lots below 12,000 square feet.

We found significant spatial effects in two out of four spatial models: Census_1 and Distance_2. The results for the Census_ 1 model suggest that increasing the average log-price in the Census tract by one would lead to an increase in the log price of the house by about 0.19 (roughly equivalent to an elasticity of 0.19). In the Distance_2 models, we find that only the houses sold in the same month have a significant spatial effect, with an elasticity of about 0.16.

5.2. Housing price indexes

The estimated hedonic models were used to construct monthly Laspeyres-type indexes described in Section 4 for each of the 290 zip code groups in the Los Angeles and San Diego metropolitan areas These indexes then were combined to form metro aggregate price indexes. Not surprisingly, which model is "best" by standards such as Akaike or Schwartz information criteria varies by zip code and estimation window. However, across zip codes and estimation windows for both the Los Angeles and San

Table 8
Coefficient estimates for the window ending May 2007, zip code 90623.

Variables	OLS model		Census_1 model		Census_2 model		Distance_1 model		Distance_2 model	
	Coefficients	*t* values	Coefficients	*t* values	Coefficients	*t* values	Coefficients	*t* values	Coefficients	*t* values
AGE	−0.0019	−0.9900	−0.0019	−0.6700	−0.0019	−0.7100	−0.0018	−0.6200	−0.0022	−0.8500
AGE^2	−0.00001	−0.3200	−0.000001	−0.0200	−0.000003	−0.0700	−0.00001	−0.2100	0.00000	0.0000
GLA	0.2127*	3.2600*	0.1919*	3.6900*	0.2063*	3.8700*	0.2069*	3.7100*	0.2019*	3.9400*
GLA^2	−0.0033	−0.1900	0.0003	0.0200	−0.0038	−0.2800	−0.0018	−0.1200	−0.0048	−0.3700
BED	0.1950*	3.0200*	0.1990*	1.9800*	0.1832*	1.8900*	0.1989*	2.0700*	0.1735*	1.7900*
BED^2	−0.0247*	−2.7000*	−0.0253*	−1.8500*	−0.0231*	−1.7400*	−0.0252*	−1.9300*	−0.0217*	−1.6400*
BATH	0.0349	0.8100	0.0477	1.4500	0.0390	1.2000	0.0368	1.0900	0.0343	1.0800
$BATH^2$	−0.0032	−0.3200	−0.0063	−0.8500	−0.0046	−0.6200	−0.0036	−0.4700	−0.0041	−0.5700
LOT	0.0704*	11.0100*	0.0682*	5.3900*	0.0671*	4.9800*	0.0706*	5.3300*	0.0638*	4.9300*
LOT^2	−0.0030*	−7.2000*	−0.0029*	−3.1800*	−0.0028*	−3.0600*	−0.0030*	−3.1400*	−0.0027*	−3.1000*
TAX	−0.0760	−1.3000	−0.0709	−0.9200	−0.0703	−0.9700	−0.0756	−0.9900	−0.0685	−0.9400
LTV	−0.0003	−0.9800	−0.0002	−0.8200	−0.0003	−1.2500	−0.0003	−1.1100	−0.0003	−1.2400
Lambda			0.1932*	2.9500*			0.0041	1.1200		
Lambda_past					0.0001	0.0800			−0.0012	−1.0000
Lambda_current					0.0896	1.3300			0.1650*	2.5300*
Lambda_future					0.0004	0.2500			0.0015	0.5100
N	402		402		402		402		402	
AIC	21.6971		21.6800		21.6998		21.6906		21.6967	
Schwartz criterion	21.9357		21.9285		21.9682		21.9391		21.9651	

Note: 1. All regressions include time dummies, coefficients of which are not shown.
2. AIC and Schwartz Criterion were calculated based on the fitted values of prices instead of the natural log of prices.
* Significant at 10% level.

Table 9
Estimates of market peaks and depreciations.

	LA metro	% Δ to Feb08	SD metro	% Δ to Feb08	Diff. in month
Overall					
CS	Sep-06	−21.58	Nov-05	−23.97	10
FHFA	Dec-06	−6.85	Mar-06	−11.03	9
Repeat sales	May-06	−23.72	Sep-05	−28.59	8
OLS	Apr-07	−17.15	Jul-05	−22.07	21
Census_1	Apr-07	−17.56	Jul-05	−21.88	21
Census_2	Apr-07	−17.22	Jul-05	−23.27	21
Distance_1	Apr-07	−17.07	Jul-05	−21.98	21
Distance_2	Apr-07	−16.88	Jul-05	−22.05	21
Low tier					
CS	Sep-06	−25.89	Jun-06	−31.52	3
Repeat sales	May-06	−27.14	Sep-05	−33.54	8
OLS	Apr-07	−17.82	Jul-05	−24.73	21
Census_1	Apr-07	−18.43	Jul-05	−25.51	21
Census_2	Apr-07	−17.93	Jul-05	−25.70	21
Distance_1	Apr-07	−17.75	Jul-05	−24.71	21
Distance_2	Apr-07	−17.99	Jul-05	−24.49	21
Mid tier					
CS	Jul-06	−24.13	Nov-05	−26.07	8
Repeat sales	May-06	−24.37	Sep-05	−29.55	8
OLS	Apr-07	−17.41	Jul-05	−22.69	21
Census_1	Apr-07	−17.73	Jun-05	−22.48	22
Census_2	Apr-07	−17.43	Jul-05	−24.23	21
Distance_1	Apr-07	−17.51	Jul-05	−22.70	21
Distance_2	Apr-07	−17.13	Jul-05	−22.79	21
High tier					
CS	Jun-06	−16.54	Jun-06	−16.95	0
Repeat sales	May-06	−20.88	Sep-05	−24.03	8
OLS	Apr-07	−16.61	Jul-05	−19.84	21
Census_1	Apr-07	−16.90	Jul-05	−19.35	21
Census_2	Apr-07	−16.65	Jul-05	−21.20	21
Distance_1	Apr-07	−16.37	Jul-05	−19.82	21
Distance_2	Apr-07	−16.46	Jul-05	−20.64	21

Diego metros, the index values produced by the hedonic methods track each other fairly closely.

The indexes cover the period January 2000 to June 2008. The boom- bust cycle that characterized the period has received extensive coverage in the media. Tables 9 and 10 report the estimated peaks and price changes of this cycle for the hedonic indexes and three repeat-sales indexes: S&P/Case-Shiller™ FHFA, and an index constructed by applying the Case-Shiller methodology to the same data used to construct the hedonic indexes. The latter will be referred to as the RS index. Figs. 1 and 2 illustrate the patterns for Census_1, S&P/Case-Shiller™ and the FHFA. The FHFA uses data that are confined to transactions within the conformable loan limits. Consequently, low-tier and high-tier homes tend to be underrepresented in the sample (Rappaport, 2007, p. 52). The FHFA also includes refinancing transactions. The S&P/Case-Shiller™ uses data from public deed records and, thus, covers a wider price range than FHFA. The S&P/Case-Shiller™ also excludes transactions that are not "arms length" such as transfers between family members.

The base period for all indexes is January 2000. In addition to the estimates for all properties in each metro, estimates are also shown for high-tier, medium-tier and low-properties. The tiers are based on percentiles. To create the low-tier price index, we tracked all houses that were sold in the base period at a price less than the 33rd percentile of all prices in the base period (January 2000). For the medium and high-tiers, prices between the 33rd and 66th percentiles, and prices above the 66th percentiles were used.

5.3. Differences between Los Angeles and San Diego metros

In both metros, all indexes in Tables 9 and 10 show large price increases up until about 2005–2007 period followed by large price decreases as the market collapsed. All indexes also show that the overall housing market peaked and declined in the San Diego metro before the Los Angeles metro by an average of about 16 months, and that Los Angeles experienced greater appreciation than San Diego during the boom, and less depreciation after the market peaked. The average price change from January 2000 to the peak for the seven overall indexes is 180% for Los Angeles and 161% for San Diego, while the average price change from the peak to February 2008 is −17% for Los Angeles and −22% for San Diego.

Table 10
Estimates of market peaks and appreciation.

	LA metro	% Δ since Jan00	SD metro	% Δ since Jan00	Diff. in month
Overall					
CS	Sep-06	173.94	Nov-05	150.34	10
FHFA	Dec-06	175.46	Mar-06	144.99	9
Repeat sales	May-06	194.79	Sep-05	161.25	8
OLS	Apr-07	183.32	Jul-05	149.48	21
Census_1	Apr-07	175.30	Jul-05	150.82	21
Census_2	Apr-07	183.77	Jul-05	156.88	21
Distance_1	Apr-07	176.71	Jul-05	149.79	21
Distance_2	Apr-07	175.41	Jul-05	151.71	21
Low tier					
CS	Sep-06	239.81	Jun-06	196.86	3
Repeat sales	Oct-06	293.10	Sep-05	231.09	13
OLS	Apr-07	212.26	Jul-05	172.66	21
Census_1	Apr-07	201.70	Jul-05	174.47	21
Census_2	Apr-07	211.76	Jul-05	182.94	21
Distance_1	Apr-07	203.91	Jul-05	172.73	21
Distance_2	Apr-07	203.94	Jul-05	175.74	21
Mid tier					
CS	Jul-06	183.44	Nov-05	154.56	8
Repeat sales	May-06	203.99	Oct-05	231.09	7
OLS	Apr-07	189.10	Jul-05	151.20	21
Census_1	Apr-07	181.48	Jul-05	152.39	21
Census_2	Apr-07	189.37	Jul-05	160.41	21
Distance_1	Apr-07	182.40	Jul-05	151.39	21
Distance_2	Apr-07	181.96	Jul-05	154.01	21
High tier					
CS	Jun-06	140.26	Jun-06	124.43	0
Repeat sales	Jun-06	122.53	Sep-05	116.40	9
OLS	Apr-07	165.09	Jul-05	137.55	21
Census_1	Apr-07	157.68	Jul-05	137.46	21
Census_2	Apr-07	165.44	Jul-05	141.71	21
Distance_1	Apr-07	159.35	Jul-05	138.01	21
Distance_2	Apr-07	156.93	Jul-05	140.28	21

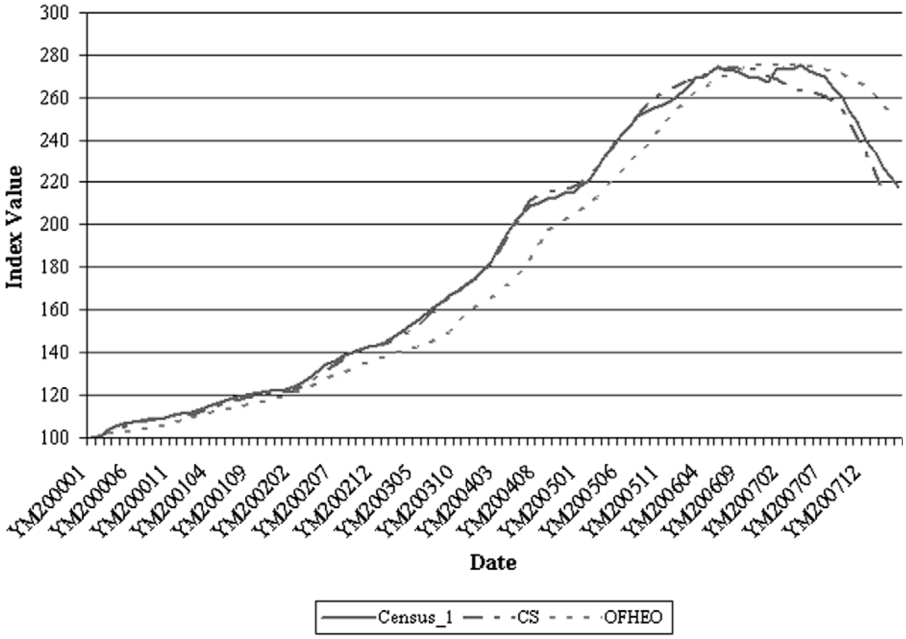

Fig. 1. Overall index value for LA metro.

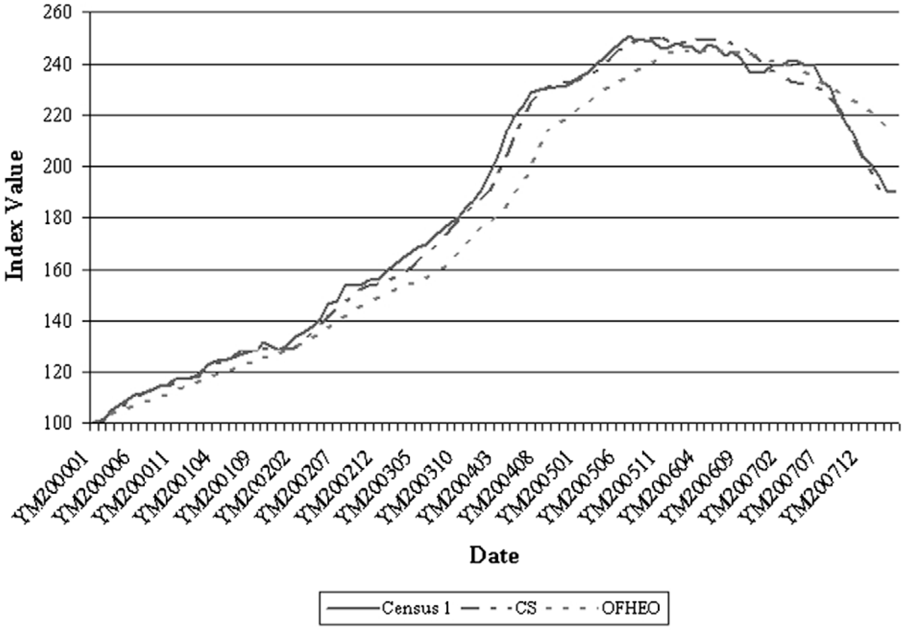

Fig. 2. Overall index value for SD metro.

While a detailed analysis is beyond the scope of this paper, we note some potential factors that might explain the differences between Los Angeles and San Diego. One is that San Diego consistently had larger shares of nonprime loan originations. For every quarter from the first quarter of 2003 up to the third quarter of 2007, the share in San Diego exceeded the share in Los Angeles by an average of almost 2% (JCHS, 2008). Thus, one possibility is that San Diego peaked earlier and had greater post-peak depreciation because the tighter credit standards imposed when the sub prime market collapsed had more impact on prices there. To some extent, this is supported by the price-tier indexes discussed below. We find, for example, that the low-tier markets (which were supported more by subprime loans) experienced greater post-peak depreciation than the high-tier markets.

Other potential factors are vacancy rates and rent controls. In San Diego the 2006 vacancy rate is 7.7% versus 5.2% in Los Angeles according to Census data. In contrast to San Diego, Los Angeles has been under Ellis Act rent controls since 1979 which creates shortages of alternatives to owner occupied housing and, thereby, increases demand. Moreover, a recently passed ordinance in Los Angeles prevents developers from rebuilding apartments just to raise rents (Thornberg, 2007). Shortages in rental housing can spill over into unregulated purchased housing. For example, a shortage of regulated apartments can initially increase the demand and price of substitutes for such unregulated low-tier housing. The price increases in low-tier housing then facilitate trading up through equity effects that, in turn, increase prices in higher tier housing (Meen, 1999; Ho et al., 2007). The rent controls in Los Angeles may have created a persistent shortage of housing that increased the underlying price appreciation, thereby increasing and sustaining that city's pre-peak price boom longer, and moderating the post-peak price decline relative to San Diego.

5.4. Price tiers

There appears to be little difference in the timing of peaks across price tiers for all indexes. For the hedonic indexes, the peaks in the three price-tiers match the dates of the overall peaks in both metros. For the repeat-sales indexes, 9 out of 12 of the price-tier peaks are within 2 months of the overall peaks of the metros and none exceed 8 months.

The mid price-tier indexes display levels of appreciation and depreciation similar to the overall indexes for both Los Angeles and San Diego. The high-tier indexes all show less appreciation than the overall indexes. The difference is relatively small in all cases except for the RS high-tier index

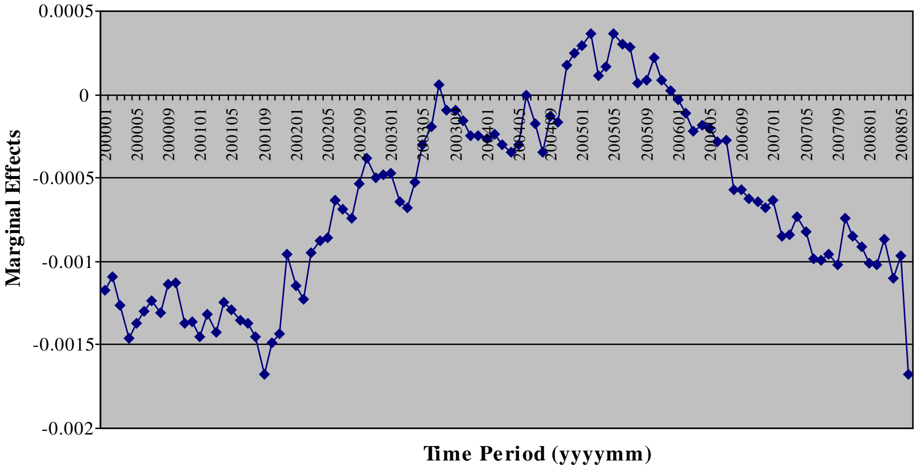

Fig. 3. Average marginal effects of AGE on log housing price in Los Angeles County – measured at AGE = 25.

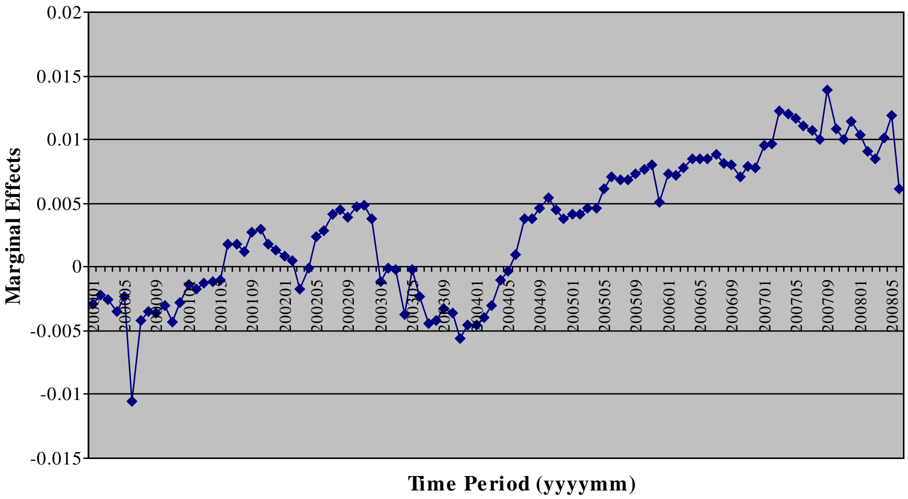

Fig. 4. Average marginal effects of BATH on log housing price in Los Angeles County – measured at BATH = 2.

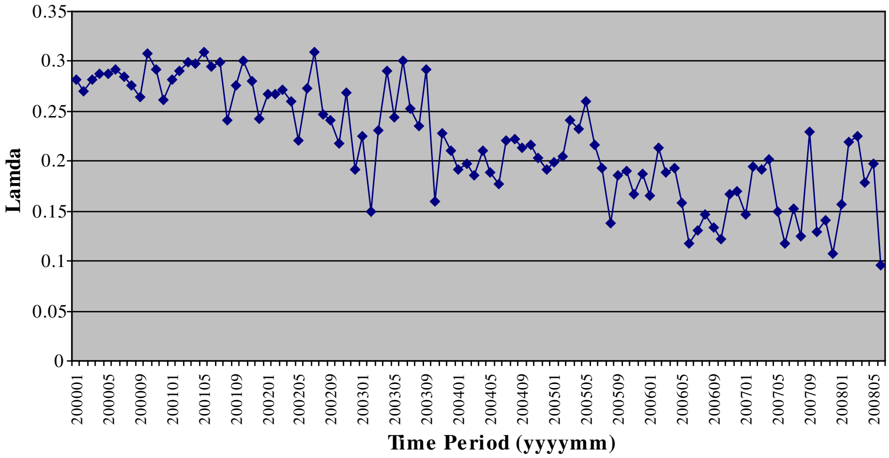

Fig. 5. Average spatial correlation coefficients (Lambda) in Los Angeles County.

which shows significantly less pre-peak appreciation than the overall RS index (72% less in Los Angeles and 45% less in San Diego). In contrast, the low price-tier indexes show greater appreciation pre-peak and greater depreciation post-peak than the other tiers and overall indexes. For example, the average price change from January 2000 to the peak of the seven low tier indexes is 223% for Los Angeles (71% greater than the high tier average) and 187% for San Diego (53% greater than the high tier average), while the average price change from the peak to February 2008 is −20% for Los Angeles (3% less than the high tier average) and −27% for San Diego (7% less than the high tier average).

One explanation for the greater appreciation and depreciation is the greater share of subprime adjustable loans in low-tier housing. For example, more than 40% of such loans in 2006 originated in low-income census tracts compared to 15% for high-income areas (JCHS, 2008). These loans allowed individuals even with poor credit scores to purchase homes and greatly boosted demand and prices. When the rates adjusted and the subprime market collapsed, low-tier housing was hardest hit. Many could not afford the higher payments and were forced into foreclosure which drove down prices. Wealthier home owners, even if they had adjustable rate mortgages often had sufficient income to avoid foreclosure. Low-tier owners also tend to have higher loan to value ratios which magnifies price decreases through "reinforcing effects" (Stein, 1995; Lamont and Stein, 1999): price decreases reduce the equity and, thus, ability of low tier buyers to make down payments which leads to further price decreases.

5.5. Repeat-sales versus hedonic indexes

Although certain results hold across methods, there are notable differences between the results generated by the repeat-sales indexes and the hedonic indexes. Tables 9 and 10 also report the results for three repeat-sales indexes: S&P/Case-Shiller™, FHFA and RS. Since all use the same methodology, differences across the repeat-sales indexes reflect differences in the data sets used. The S&P/Case-Shiller™ and RS indexes display similar patterns that are different than the hedonic indexes, except the RS index displays more pre-peak appreciation and slightly less post-peak depreciation than the S&P/Case-Shiller™. In contrast to S&P/Case-Shiller™, the RS index uses the same data as the hedonic indexes and, consequently, differences between the RS and hedonic indexes reflect purely methodological differences. The overall RS index for Los Angeles estimates that the market peaked at May 2006 which is 11 months earlier than the hedonic indexes which all peaked at April 2007. The overall RS index for San Diego, by contrast, peaks at September 2005 which is 2 months later than the hedonic indexes. For the overall indexes, the time between the San Diego and Los Angeles peaks is 8 months for RS index and 21 months for the hedonic indexes.

Another difference is that from January 2000 to the peak of the housing boom, the RS index shows considerably larger price increases in low-tier housing and smaller price increases in high-tier housing than the hedonic indexes. The RS low-tier index shows a 293% increase for Los Angeles and 231% for San Diego versus the hedonic indexes which average 207% for Los Angeles and 176% for San Diego. In contrast to the low-tier, the RS high-tier index shows 122% increase for Los Angeles and 116% for San Diego versus the hedonic indexes which average 161% for Los Angeles and 139% for San Diego. Post-peak, the RS index shows larger price decreases across all tiers than the hedonic indexes. For example, the RS low-tier index shows a price decrease of 27% for Los Angeles and a decrease of 33% for San Diego versus the hedonic indexes which average decreases of 18% for Los Angeles and 25% for San Diego.

In sum, compared to the hedonic indexes, the RS index shows: (a) earlier peaks in Los Angeles and later peaks in San Diego, (b) more pre-peak appreciation overall and in low- and mid-tier housing in both Los Angeles and San Diego, (c) less pre-peak appreciation in high-tier housing and (d) more post-peak depreciation in both Los Angeles and San Diego across all tiers. What differences between the RS and the hedonic indexes might explain these results?

Both are based on the same data set, but repeat-sales indexes have sources of inefficiency and bias avoided by the hedonic indexes. First, repeat-sales indexes are limited to repeat transactions and, therefore, use considerably fewer transactions. For the sample period of January 2000 to February 2008, the percentages of repeat-sale transactions are just 46% in Los Angeles and 42% in San Diego. The smaller sample sizes of course increase the variances of the regression estimates. In addition, there is selection bias to the extent that the distributions of repeat-sale and single-sale transactions are different. In Los Angeles in 2008, for example, the average price of a repeat-sale house is about $94,000 less than a single-sale house with a standard deviation of about $73,000 less; the average of the annual average price difference across all 9 years is about $30,000 less with an average standard deviation difference of about $31,000 less. Repeat-sale houses in Los Angeles are also somewhat smaller on average (in terms of both living area and lot size) than single-sale houses. Similar differences are found for San Diego. Consistent with these differences, other studies report that frequently sold houses tend to be smaller starter homes (Hwang and Quigley, 2004; Case et al., 1997). Such homes are more likely to have been financed by subprime and adjustable loans, which, as we argued above are likely to appreciate and depreciate more than other houses.

Another difference is that repeat-sale transactions are more likely to be speculative and less likely to be based on fundamentals than single-sale transactions.

In the recent housing bubble, speculators turned over houses on a regular basis to capitalize on rapidly increasing prices. By focusing on frequently traded houses while ignoring single sales, repeat-sales indexes attach undue weight to speculative transactions and thus likely overestimated the appreciation of the housing stock during the boom. As the bubble burst, speculators dumped inventory on the market and unsold properties went into foreclosure. Such sales are likely to cause repeat-sales indexes to overestimate depreciation.

While omitting single-sale houses might have biased repeat-sales indexes upward in most categories of housing, it might have actually caused appreciation to be underestimated in high-tier housing. For example, buyers of high-tier houses typically have higher incomes and, thus, can afford more customized, nonstandard features than buyers of low-tier houses. Such nonstandard features are a source of uncertainty for houses that are not frequently traded and may cause them to be overvalued during booms.

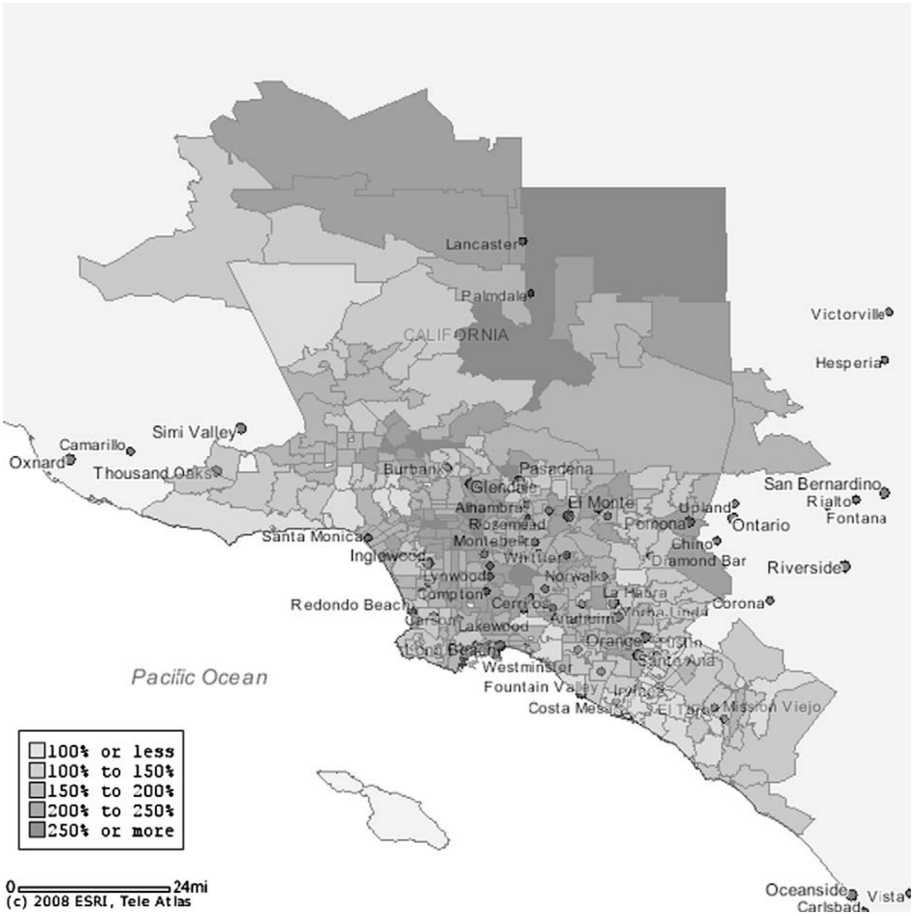

Fig. 6. Appreciation rate from January 2000 to peak for Los Angeles Metropolitan Area (CEN1 index).

Consequently, it may be that high-tier houses that are traded more frequently tend to appreciate less than single-sale high-tier houses because buyers and sellers have better information on the true value. Since single-sale houses are omitted from repeat-sales indexes, overall appreciation in high-tier housing may be underestimated. This effect is reinforced to the extent that single-sale houses are new construction. Newly constructed houses are more precisely aligned with current demands and tastes for amenities than existing houses. Therefore, new houses (which are ignored by repeat-sales indexes) are likely to appreciate more than existing houses particularly at the high end since buyers at that tier have more income to spend on specialized amenities. Consistent with this we find that the estimated effect of AGE is negative in the majority of cases and statistically significant in about half of these (see Tables 6 and 7).

Another source of bias is that repeat-sales indexes assume that the quality of a house does not change between two sale dates, possibly several years apart. Although the gross living area and number of rooms typically do not change, the quality of a house is likely to change due to a mix of improvements and age-related depreciation not offset by expenditures on maintenance (Goetzmann and Spiegel, 1995; Englund et al., 1999; Harding et al., 2007), and changes in neighborhood quality. In contrast to repeat-sales, hedonic indexes can at least partially control for such changes by controlling for changes in the age of the house, and changes in neighborhood quality through the spatial correlation model. During booms the quality of houses between sales is more likely to rise because improvements are more likely to be made when expected returns are greater. Such improvements cause repeat-sales indexes to overestimate price increases under an erroneous assumption of constant quality. During busts, in contrast, the quality of houses between sales is more likely to fall as many sales are foreclosures. As Schuetz, Been and Ellen (2008, p. 309) argue, owners in foreclosure proceedings have less incentive to maintain properties they may lose. Consequently, repeat-sales indexes are likely to overestimate price decreases during busts.

Repeat-sales indexes also assume that the regression coefficients of the quality attributes do not change between the two sale dates. The coefficients represent the contributory values of the quality attributes that are determined by consumer preferences and other market conditions. In contrast to repeat-sales, hedonic indexes can accommodate changes in preferences and market

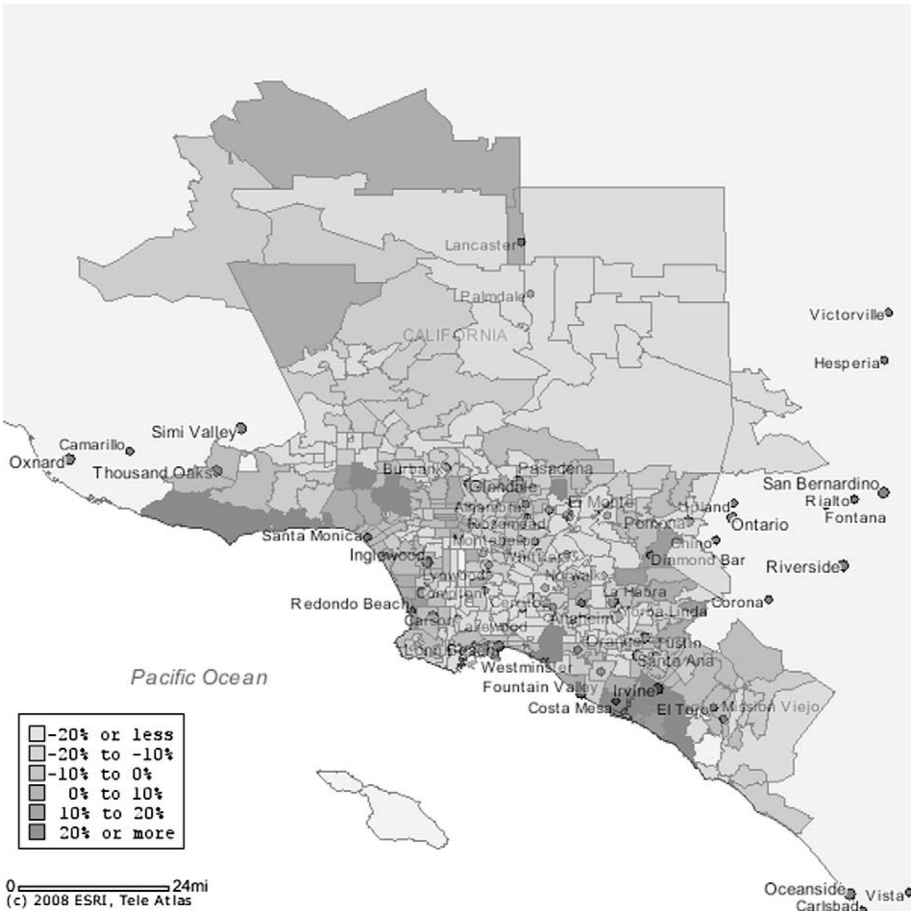

Fig. 7. Appreciation rate from peak to February 2008 for Los Angeles Metropolitan Area (CEN1 index).

Table 11
Correlations between appreciation and foreclosure rates 2007 and 2008.

	Model	OLS	Census_1	Census_2	Distance_1	Distance_2	Average
2007							
LA metro	Pearson correlation coefficient	−0.3337	−0.3943	−0.3319	−0.3362	−0.3030	−0.3398
	P-value	0.0000	0.0000	0.0000	0.0000	0.0000	
SD metro	Pearson correlation coefficient	−0.2681	−0.2691	−0.5870	−0.2812	−0.2167	−0.3244
	P-value	0.0111	0.0108	0.0016	0.0076	0.0402	
2008							
LA metro	Pearson correlation coefficient	−0.4968	−0.4155	−0.4208	−0.4891	−0.4385	−0.4522
	P-value	0.0000	0.0000	0.0000	0.0000	0.0000	
SD metro	Pearson correlation coefficient	−0.3553	−0.3084	−0.3501	−0.3449	−0.2566	−0.3231
	P-value	0.0006	0.0031	0.0862	0.0009	0.0152	

conditions by allowing the coefficients to vary over time. The pattern of estimated coefficients generated by our hedonic models suggests that the repeat-sales assumption of stable coefficients is not in fact justified. Take the Los Angeles County for example. Figs. 3–5 plot the average values of marginal effects of AGE (at AGE = 25), BATH (at BATH = 2) and the spatial correlation coefficient (λ) on the log housing prices in LA for each month during the sample period.[5] On average, a house being a year older than 25 years of age had a negative influence on its price initially, but gradually became slightly positive by early 2005, and then became negative again. The negative effect of AGE diminished in the midst of the boom, which might be explained by unobserved renovations on older homes and a lag between the increasing demand for houses and new constructions in this period. In general, having a third bathroom in a house became increasingly more valuable over time, possibly reflecting a change in preferences. On the other hand, the spatial coefficient declined over time, suggesting a decreasing influence of nearby property values on the value of a particular property. One possibility is that the prices in the housing market were rapidly changing toward the latter part of the sample period, reflecting changing and more uncertain market conditions, so that the recent sales of nearby properties were less relevant.

5.6. Hedonic indexes at the zip code level

The S&P/Case-Shiller™ and FHFA repeat-sales indexes are publicly available only at the metropolitan levels. Another advantage of the extensive FNC National Collateral Database is that hedonic indexes can be constructed for individual zip codes. At the metropolitan level, all indexes show large price increases followed by large price decreases as the market collapsed. However, this pattern did not occur uniformly across zip codes. For example, about 61% of the zip codes in Los Angeles peaked before the metro peak of April 2007 (the earliest being June 2005), while about 30% of the zip codes peaked after the metro peak (the latest being April 2008). Moreover, the Census_1 index indicates for Los Angeles prices rose 175.30% from January 2000 to April 2007 while the price increases for individual zip codes ranged from 123% to 455% during the same period. The Census_1 index also indicates that prices declined 17.56% in Los Angeles from the peak in April 2007 to February 2008 but in the same period about 6% of the zip codes appreciated. The variations in the pre- and post-peak appreciation rates across the Los Angeles metropolitan area are illustrated by the maps in Figs. 6 and 7.

One factor is variation in foreclosure rates across zip codes. Foreclosure activity can accelerate price decreases. Initially, price decreases raise foreclosure rates by reducing homeowner equity. With less equity homeowners are more likely to default on their loans. As foreclosures occur, additional inventory of unsold houses is dumped on the market and prices fall further. Properties sold at foreclosure auctions are likely to be sold at a discount which affects the price of appraisal comparables for neighboring properties (Lin et al., 2009). Other channels of negative spillover effects include neglected and vacant properties that signal an unstable neighborhood (Schuetz et al., 2008). Several recent empirical studies have documented the negative effects of foreclosures on neighboring property values. For a review, see Schuetz et al. (2008).

Hence, price decreases are likely to be larger in areas with higher foreclosure rates. Table 11 reports Pearson correlation coefficients consistent with such patterns. The correlation is between distressed sales and price changes for the zip code groups in Los Angeles and San Diego for 2007 and the first quarter of 2008. The distressed sales rate is defined as the number of distressed sales (which are largely foreclosures) divided by the total number of sales. The data on distressed sales are from FNC's NCD database. The correlations are shown for the five hedonic methods. All correlations are negative and 19 out of 20 are significant at the 5% level, and 14 out of 20 significant at the 1% level.

6. Conclusion

Standard housing price indexes rely on strong constant-quality assumptions, and often give conflicting results. Hedonic indexes can potentially overcome the limitations of the median price and repeat-sales indexes but are seldom constructed because of a lack of data. This is particularly true in the U.S. where extensive public data on housing attributes and prices are not available. Consequently, monitoring developments in the housing market is subject to a great deal of uncertainty. This is unfortunate given that

[5] The marginal effects of AGE and BATH take into account the coefficients for both the linear term and the quadratic term.

housing in the U.S. economy represents the largest source of household wealth. In this paper we constructed hedonic indexes for the Los Angeles and San Diego metropolitan areas using an extensive new data set collected by a mortgage technology company. The data set is considerably more detailed than previously available and covers the boom-bust period of 2000 to June 2008. The indexes were constructed at the zip code level using spatial hedonic regressions with time-varying coefficients that capture correlations within submarkets. Some findings for the hedonic indexes match and therefore lend support to those of S&P/Case-Shiller™ repeat-sales price index. For example, all indexes agree that the San Diego market peaked before Los Angeles. Other findings, however, indicate important differences in the timing of the peaks and the intensity of the cycle over different price-tiers. Such differences could have important implications for the decisions of individual households, real estate investors and government policy makers. In addition, our indexes allow us to examine the housing cycle at the zip code level. We find that the boom-bust cycle did not occur uniformly across zip codes and, consequently, the standard metro-level price indexes that are widely reported in the media are often not a good guide to local conditions.

Acknowledgments

The authors thank Brad Case for numerous suggestions that substantially improved the paper. We are also grateful to the Editor, Henry Pollakowski. Thanks also go to Sankar Bokka, Jennifer Harris and other staff at FNC for providing help with the data. All remaining errors are ours.

References

Anselin, L., Lozano-Gracia, L., 2008. Errors in variables and spatial effects in hedonic house price models of ambient air quality. Empirical Economics 34 (1), 5–34.
Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, Dordrecht.
Baily, M., Muth, R., Nourse, H., 1963. A regression method for real estate price index construction. Journal of American Statistical Association 58, 933–942.
Bourassa, S.C., Hoesli, M., Peng, V.S., 2003. Do housing submarkets really matter? Journal of Housing Economics 12, 12–28.
Bourassa, S.C., Hamelink, F., Hoesli, M., MacGregor, B.D., 1999. Defining housing submarkets. Journal of Housing Economics 8, 160–183.
Case, B., Pollakowski, H.O., Wachter, S.M., 1991. On choosing among house price index methodologies. Real Estate Economics 19 (3), 286–307.
Case, B., Pollakowski, H.O., Wachter, S.M., 1997. Frequency of transaction and house price modeling. Journal of Real Estate Finance and Economics 14, 173–187.
Case, B., Quigley, J.M., 1991. The dynamics of real estate prices. Review of Economics and Statistics 22 (1), 50–58.
Case, K.E., Shiller, R.J., 1987. Prices of single-family homes since 1970: new indexes for four cities. New England Economic Review (September/October), 45–56.
Case, B., Clapp, J., Dubin, R., Rodriguez, M., 2004. Modeling spatial and temporal house price patterns: a comparison of four models. Journal of Real Estate Finance and Economics 29 (2), 167–191.
Case, B., 2006. Housing price indexes. In: Arnott, R.J., McMillan, D.P. (Eds.), A Companion to Urban Economics. Blackwell Publishing, Malden, MA.
Clapham, E., Englund, P., Quigley, J.M., Redfearn, C.L., 2006. Revisting the past and settling the score: index revision for house price derivatives. Real Estate Economics 34, 275–302.
Clapp, J.M., Giaccotto, C., 1999. Revisions in repeat-sales indexes: here today, gone tomorrow? Real Estate Economics 27 (1), 79–104.
Clapp, J.M., 2004. A semiparametric method for estimating local house price indices. Real Estate Economics 32, 127–160.
Dubin, R.A., 1998. Spatial autocorrelation: a primer. Journal of Housing Economics 7, 304–327.
Englund, P., Quigley, J.M., Redfearn, C.L., 1999. The choice of methodology for computing housing price indexes: comparisons of temporal aggregation and sample definition. Journal of Real Estate, Finance and Economics 19 (2), 91–112.
Englund, P., Quigley, J.M., Redfearn, C.L., 1998. Improved price indexes for real estate: measuring the course of Swedish housing prices. Journal of Urban Economics 44, 171–196.
Goodman, A.C., Thibodeau, T.G., 1998. Housing market segmentation. Journal of Housing Economics 7, 121–143.
Goodman, A.C., Thibodeau, T.G., 2003. Housing segmentation and hedonic prediction accuracy. Journal of Housing Economics 12, 181–201.
Goodman, A.C., Thibodeau, T.G., 2007. The spatial proximity of metropolitan area housing submarkets. Real Estate Economics 35 (2), 209–232.
Goetzmann, W.N., Spiegel, M., 1995. Non-temporal components of residential real estate appreciation. Review of Economics and Statistics 77 (1), 199–206.
Griliches, Z., 1961. Hedonic price indexes for automobiles: an econometric analysis of quality change. In: The Price Statistics of the Federal Government. Government Printing Office, Washington, DC.
Harding, J.P., Rosenthal, S.S., Sirmans, C.F., 2007. Depreciation of housing capital, maintenance, and house price inflation: estimates from a repeat sales model. Journal of Urban Economics 61, 193–217.
Hill, R.J., Melser, D., 2007. Hedonic imputation and the price index problem: an application to housing. Economic Inquiry 1, 17.
Horowitz, J.L., 1992. The role of the list price in housing markets: theory and an econometric model. Journal of Applied Econometrics 7, 115–129.
Ho, S.H., Ma, Y., Haurin, D.R., 2007. Domino effects within a housing market: the response of house prices across quality tiers, working paper.
Hwang, M., Quigley, J.M., 2004. Selectivity, quality adjustment and mean reversion in measurement of house values. Journal of Real Estate Finance and Economics 28 (2/3), 161–178.
Joint Center for Housing Studies of Harvard University, 2008. The State of the Nation's Housing 2008. Harvard University.
Kelejian, H.H., Prucha, I.R., 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. Journal of Real Estate Finance and Economics 17, 99–121.
Knight, J.R., Sirmans, C.F., Turnbull, G.K., 1994. List price signaling and buyer behavior in the housing market. Journal of Real Estate, Finance and Economics 9, 177–192.
Lamont, O., Stein, J.C., 1999. Leverage and house price dynamics in U.S. cities. Rand Journal of Economics 30, 498–514.
Lee, L.F., 2003. Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. Econometric Reviews 22 (4), 307–322.
Lin, Z., Rosenblautt, E., Yao, V.W., 2009. Spillover effects of foreclosures on neighborhood property values. Journal of Real Estate Finance and Economics 38, 387–407.
Meen, G., 1999. Regional house prices and the ripple effect: a new interpretation. Housing Studies 14, 733–753.
Meese, R.A., Wallace, N.E., 1997. The construction of residential housing price indices: a comparison of repeat-sales, hedonic-regression, and hybrid approaches. Journal of Real Estate, Finance and Economics 14, 51–73.
Oates, Wallace E., 1969. The effects of property taxes and local public spending on property values: an empirical study of tax capitalization and the Tiebout hypothesis. Journal of Political Economy 77 (6), 957–971.
Palmon, O., Smith, B.A., 1998. New evidence on property tax capitalization. Journal of Political Economy 106 (5), 1099–1111.
Pesaran, M.H., Timmermann, A., 2007. Selection of estimation window in the presence of breaks. Journal of Econometrics 13, 134–161.
Quigley, J.M., 1995. A simple hybrid model for estimating real estate price indexes. Journal of Housing Economics 4, 1–12.
Rappaport, J., 2007. A guide to aggregate house price measures. Federal Reserve Bank of Kansas City Economic Review Second Quarter 2007, 41–71.
Reichert, A.K., 1990. The impact of interest rates, income and employment upon regional housing prices. Journal of Real Estate, Finance and Economics 3, 373–391.

Rosen, K.T., 1982. The impact of Proposition 13 on house prices in Northern California: a test of the interjurisdictional capitalization hypothesis. Journal of Political Economy 90 (1), 191–200.
Rothenberg, J., Galster, G.C., Butler, R.V., Pitkin, J., 1991. The Maze of Urban Housing Markets: Theory, Evidence and Policy. University of Chicago Press, Chicago.
Schuetz, J., Been, V., Ellen, I.G., 2008. Neighborhood effects of concentrated mortgage foreclosures. Journal of Housing Economics 17, 306–319.
Stein, J.C., 1995. Prices and trading volume in the housing market with down-payment effects. The Quarterly Journal of Economics 110 (2), 379–406.
Tiebout, Charles M., 1956. A pure theory of local expenditures. Journal of Political Economy 64 (5), 416–424.
Thornberg, C., 2007. Rent control discourages development of affordable housing. Los Angeles Business Journal. July 2.