



Time-Geographically Weighted Regressions and Residential Property Value Assessment

Jeffrey P. Cohen¹ · Cletus C. Coughlin²  · Jeffrey Zabel³

Published online: 24 July 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

In this study, we develop and apply a new methodology for obtaining accurate and equitable property value assessments. This methodology adds a time dimension to the Geographically Weighted Regressions (GWR) framework, which we call Time-Geographically Weighted Regressions (TGWR). That is, when generating assessed values, we consider sales that are close in time and space to the designated unit. We think this is an important improvement of GWR since this increases the number of comparable sales that can be used to generate assessed values. Furthermore, it is likely that units that sold at an earlier time but are spatially near the designated unit are likely to be closer in value than units that are sold at a similar time but farther away geographically. This is because location is such an important determinant of house value. We apply this new methodology to sales data for residential properties in 50 municipalities in Connecticut for 1994–2013 and 145 municipalities in Massachusetts for 1987–2012. This allows us to compare results over a long time period and across municipalities in two states. We find that TGWR performs better than OLS with fixed effects and leads to less regressive assessed values than OLS. In many cases, TGWR performs better than GWR that ignores the time dimension. In at least one specification, several suburban and rural towns meet the IAAO Coefficient of Dispersion cutoffs for acceptable accuracy.

Keywords Time-geographically weighted regression · Residential property assessment · Coefficient of dispersion · Price-related differential · Percent absolute prediction error

Introduction

Two important considerations in real estate assessment are the accuracy and equity of property assessments. These considerations can be measured in terms of the Coefficient of Dispersion (COD) and the Price-Related Differential (PRD), as described by the

✉ Cletus C. Coughlin
coughlin@stls.frb.org

Extended author information available on the last page of the article

International Association of Assessing Officers (IAAO). “Uniformity” is measured with the COD (IAAO, 2013), while “vertical equity” is measured with the PRD (IAAO, 2003). From a statistical perspective, the percent absolute prediction error (PAPE) is a common measure of predictive accuracy.

There have been many studies in recent decades that consider spatial aspects to more uniformly and equitably estimate assessed values. Among others, these include the nonparametric and/or semi-parametric approaches of Geographically Weighted Regressions (GWR), also commonly referred to as Locally Weighted Regressions (LWR). The LWR methods were developed in Cleveland and Devlin (1988) and later used in urban economics by Meese and Wallace (1991) and McMillen (1996). Among a class of Automated Valuation Methods (AVM), including least squares regressions and fixed effects regression estimations, GWR has been shown to be a favorable alternative in general by Bidanset and Lombard (2014a, b), Borst and McCluskey (2008) and McCluskey et al. (2013), among others. Another set of semi-parametric approaches, known as local polynomial regressions as in Cohen et al. (2017), is intended to address the separate valuation of land and improvements; however, its computational intensity has deterred widespread use in practice.

But little attention has focused on incorporating the time dimension into GWR or LWR models in the context of AVM, such that when assessing properties, the importance of comparables will depend on exactly how long ago they sold. Bringing this time element into the GWR model for the purposes of estimating assessed values and examining the effects on uniformity and vertical equity is a critical contribution of our research. The goal of our paper is to provide the foundations of Time GWR as a method and then demonstrate its potential value for assessors.

Another advantage of our study is that we apply this new methodology to two extensive data sets. We have sales data for all arms-length transactions of residential properties in a sample of 50 municipalities in Connecticut for 1994–2013. We also have transactions of single-family homes for 145 towns in the Greater Boston Area (GBA) for 1987–2012.

An outline of the remainder of the paper is as follows. First, we provide a brief literature review of GWR and then present additional background on the GWR methodology. Then, we explain how we propose to modify the GWR framework to incorporate time variation, which we call Time-Geographically Weighted Regressions (TGWR). We also describe the approach to forecast the “estimated sales price” that is to be used in computing the COD, PRD and PAPE. Next, we describe the data we use for property sales in municipalities in Connecticut (CT) and Massachusetts (MA), followed by the analysis of these data using our TGWR approach. As a part of this analysis, we compare the estimates of COD, PRD and PAPE from using OLS/FE, GWR and TGWR to demonstrate which approach is more appropriate for use in property assessment. Finally, we conclude with a summary of the main findings and some suggestions for future research.

Methodology of GWR and TGWR in AVM

GWR’s popularization in urban economics can be traced back to Meese and Wallace (1991), and software has been developed to estimate GWR models in Stata (see

<https://www.staff.ncl.ac.uk/m.s.pearce/stbgwr.htm>) and more recently there is software for LWR in R (see <https://sites.google.com/site/danielpmcmillen/r-packages>). The basic idea behind GWR (and LWR) is that it allows for a general functional form between the dependent variable in a regression and the explanatory variables. Such a nonparametric approach has been explored in other contexts, commonly known as kernel regression, but the case of GWR utilizes a specific form of the kernel weights based on geographic distances between a given target observation and the other observations in the sample. While McMillen and Redfearn (2010) note that the choice of the functional form of the kernel tends to have little impact on the GWR estimates, a more complicated consideration is the bandwidth parameter. There are various approaches to bandwidth selection, including the Silverman (1986) “Rule of Thumb” bandwidth, and other approaches such as cross-validation. McMillen and Redfearn (2010) and Cohen et al. (2017) provide GWR estimation results for two separate bandwidth choices. We use one bandwidth for the main analysis and then provide some robustness checks using other bandwidths.

First, consider a linear relationship between X (a matrix of observations on property characteristics and neighborhood amenities) and Y (a vector of observations on property sales price), so that:

$$\begin{aligned} Y &= X\beta + u, \\ u &= f + \varepsilon, \\ \varepsilon &\sim N(0, \Omega). \end{aligned} \quad (1)$$

Typically, Y is specified as the log of sales price. It is straightforward to estimate the vector of parameters β by OLS (where f denotes fixed effects), invoking the classical assumptions on Ω .

But if one believes that a linear relationship between property values and characteristics may be too restrictive, one could specify the following general (nonparametric) relationship:

$$Y = h(X) + \varepsilon, \quad (2)$$

where h is some unknown function. We can estimate the marginal effects of X on Y at any point in the sales dataset, by using GWR.

More specifically, GWR can be thought of as a form of weighted least squares (McMillen and Redfearn 2010). To estimate Eq. (2) using GWR we run a weighted least squares regression of $w_{ij}Y_i$ against $w_{ij}X_i$, where $X_i = [x_{1i}, x_{2i}, \dots, x_{mi}]$, m is the number of explanatory variables in the estimation, $w_{ij} = [k_{ij}(\bullet)]^{1/2}$, and $k_{ij}(\bullet)$ is a set of kernel weights. A common choice of kernel weights is the Gaussian form, where $k_{ij} = (\exp(-(d_{ij}/b)^2))$ and d_{ij} is the geographical distance between observations i and j , and b is the bandwidth parameter.

The GWR approach can generate separate estimates of β_i for each target point, i . GWR allows for properties that are further away from a subject property to be given less weight in the estimation, opposed to an arithmetic mean as in the OLS/FE case.

One critical assumption of GWR, however, is that all properties are sold at the same time, or, at the very least, that the time of sale is not a crucial factor to consider in the kernel weights. One might address this by including time fixed effects in Eq. (1). However, as an alternative approach, we propose a methodology similar to that

suggested by Cohen et al. (2014), where the kernel weights function incorporates the date of sale of nearby properties. In such a case, the kernel weights might take the following form:

$$k_{ijt} = \exp\left(-\left(\frac{d_{ij}}{b}\right)^2\right) \cdot \exp\left(-\left(\frac{\tau_{ij}}{b}\right)^2\right) \quad (3)$$

In this version of the kernel weights, τ_{ij} represents the difference in the time between the sale of property i and property j (perhaps in terms of the number of days). More formally, if we order the dates of sale on a line so that the earliest property in our sample had a sale date, $t_1 = 1$, and suppose the second property in our sample had a sales date 5 days later so that $t_2 = 6$, etc., then we define:

$$\begin{aligned} \tau_{ij} &= t_i - t_j, \text{ if } t_i \geq t_j, \\ &= 0 \text{ otherwise.} \end{aligned} \quad (4)$$

In other words, when determining the kernel for each target point, we only consider sales that occurred prior to the sale date of the target point, i . We define the TGWR estimator as a generalization of GWR where the kernel weights are given as k_{ijt} in Eq. (3).

We think the TGWR is an important improvement over GWR since this increases the number of comparable sales that can be used to generate assessed values. Furthermore, it is likely that units that sold at an earlier time but are spatially near the designated unit are likely to be closer in value than units that are sold at a similar time but farther away geographically. This is because location is such an important determinant of house value.

For the purposes of AVM, there are a number of recent studies that have demonstrated the superiority of GWR over OLS, including Lockwood and Rossini (2011) and Borst and McCluskey (2008). The only known studies that use GWR to consider equity and uniformity of assessments are Bidanset and Lombard (2014a, b). However, neither of these two studies incorporate time differentials of sales into the GWR framework. Given the importance in considering recent sales in the appraisal process, we propose the following approach to AVM in a space-time context.

But before introducing our approach, we first discuss some common assessment practices in Connecticut and Massachusetts. In performing revaluations, our informal conversations with several assessors indicate they often estimate revaluations for property i by comparing the ratio of sales to previously assessed values of other nearby properties, j , multiplied by the most recent assessed value of property i . One would expect the importance of nearby properties (j) to diminish with distance from property i , although it is not evident how assessors choose the relevant distance. The date of sale of nearby properties should be an important factor as well, however this might be determined somewhat arbitrarily.

In addition to a property's characteristics, there may be other factors that are important determinants of a property's value, such as school quality, crime, the presence of jobs, and the population density of the town, even for nearby properties that straddle boundaries of different districts. In a regression context, these factors can

be denoted as X , where X is a matrix consisting of m vectors of variables, where m is the total number of characteristics and other factors.

One potential set of approaches to estimate these effects that addresses these issues is as follows. First, for all properties, i , that recently sold (with sales price, Y_i), run a fixed effects regression as in (1) and estimate β . The fixed effects proxy for neighborhood characteristics.

Second, consider the nonparametric GWR and TGWR models. We can estimate the marginal effects of X on Y at any point in the sales dataset, by using GWR and TGWR. This will give us separate estimates of β_i for each i . Then, for these same properties that sold recently, use the estimates of β_i to predict the values of the same properties using eq. (2), separately for GWR and TGWR.

Note that when applying fixed effects, GWR, and TGWR, we exclude future sales from the sample since this is what realistically would be carried out for assessment purposes. When using fixed effects for units that sold in year t , we include sales in year t and $t-1$. When using GWR, we include sales in the previous two years (based on the month of sale). When using TGWR, we only use previous sales that are weighted by the kernel (eq. 3) and limited by the bandwidth.

We follow the approach of Bidanset and Lombard (2014a, b) and calculate the COD and PRD using OLS/FE, GWR, and TGWR. The COD and PRD are measurement standards from the International Association of Assessing Officers (2003). The COD measures, on average, how far each property's ratio of estimated to actual sales price differs from the mean ratio of estimated to actual sales price. It is expressed as a percentage of the median. Thus, a smaller COD indicates a more uniform assessment. Values less than 15% (but not less than 5%) are viewed as acceptable.

Meanwhile, the PRD is a measure of whether high-value properties are assessed at the same ratio to market value as low-value properties. The numerator of the PRD is the ratio of the simple mean of the estimated to actual sales price, while the denominator is the weighted mean of the sum of the estimated prices relative to the sum of the actual sales prices. Values less than 0.98 indicate that high-value properties tend to be "overappraised" relative to low-value properties. Values exceeding 1.03 indicate that high-value properties tend to "underappraised" relative to low-value ones. The goal for PRD is to be in the "neutral" range, i.e., close to 1.0.

The formulas for the COD and PRD are as follows:

$$\text{COD} = \frac{100}{N} \frac{\sum_{i=1}^N \left| \frac{\text{Estimated Price}_i}{\text{Sales price}_i} - \text{Median} \left(\frac{\text{Estimated Price}_i}{\text{Sales price}_i} \right) \right|}{\text{Median} \left(\frac{\text{Estimated Price}_i}{\text{Sales price}_i} \right)} \quad (5)$$

$$\text{PRD} = \frac{\text{Mean} \left(\frac{\text{Estimated Price}_i}{\text{Sales price}_i} \right)}{\frac{\sum \text{Estimated Price}_i}{\sum \text{Sales price}_i}} \quad (6)$$

We calculate the COD and PRD separately for each of the towns in Connecticut and Massachusetts that are in our dataset. Finally, we also calculate the PAPE as a measure of prediction accuracy.¹

Note that we run the regressions (Eq. 1) using the log of sales price as the dependent variable. To convert the predicted values back to levels we use the transformation.

$$\hat{Y} = \hat{\alpha}_0 \exp(\ln \hat{Y}) \quad (7)$$

where

$$\hat{\alpha}_0 = \overline{\exp(\hat{u}_i)} \quad (8)$$

and \hat{u}_i is the residual from the corresponding regression (Wooldridge 2016).

Data

We use sales data for all arms-length transactions of residential properties in a sample of 50 municipalities in Connecticut for 1994–2013. These data come from the Warren Group. An attractive feature of these data is that they may enable us to determine how the OLS/FE and TGWR approaches perform differently for properties in different towns within the state.

To filter out non-arms-length sales, we omit properties with sales price under \$50,000. We eliminate outliers on the high end by dropping a small number of observations with sales price over \$7,000,000. In our sample, as shown in Table 1, the average sales price was approximately \$387,500, with an average lot size of 0.68 acres, living area of 1880 square feet, approximately 7 rooms, 3.19 bathrooms, and an average age at sale date of 47.62 years. The oldest house was 118 years, and there were some houses that were 0 years old (i.e., new construction). Unfortunately we do not have information broken down about the number of half baths; instead, the total baths variable includes the sum of full plus half baths, so that a property with a value of 4 total baths may have 4 full baths or 3 full plus 2 half baths, for example. We have the sale date and latitude/longitude for each property. There are 258,473 properties in this CT dataset.

The 50 municipalities included in the Connecticut sample and the 145 towns included in the Massachusetts sample can be seen in Fig. 1. For the Connecticut sample, there are several urban cities (e.g., Hartford and New Haven), and many suburban locations (such as West Hartford, Greenwich, etc.). The towns in the Massachusetts sample are all located in the Greater Boston Area. We also use transactions of single-family homes in the Greater Boston Area (GBA) for 1987–2012. The data are from the Warren Group for 1987–1994 and CoreLogic for 1995–2012 and cover towns in Bristol, Essex, Middlesex, Norfolk, Plymouth, and Suffolk Counties. These data include the month and year of sale and the exact location (latitude and longitude). Observations with missing sales date or location information are dropped.

¹ The MAE is the mean of the absolute difference of the estimated and true house values divided by the true house value.

Table 1 House-level summary statistics

Variable	Mean	Std Dev	Minimum	Maximum
Connecticut				
Sales Price (\$1,000)	387.50	492.97	50.10	6850.00
Lot Size (acres)	0.68	0.84	0.03	10.00
Living Area (1000 sqft)	1.88	1.06	0.21	92.93
Rooms	6.93	1.81	4.00	87.00
Bedrooms	3.19	1.01	0.00	14.00
Bathrooms	1.98	0.93	1.00	11.50
Age	47.62	25.84	0.00	113.00
Massachusetts				
Sales Price (\$1,000)	318.38	213.23	17.50	1925.00
Lot Size (acres)	0.61	0.77	0.10	10.00
Living Area (1000 sqft)	1.96	0.85	0.50	8.00
Rooms	7	1.61	3.00	23.00
Bedrooms	3.31	0.82	1.00	10.00
Bathrooms	1.71	0.74	1.00	10.00
Half Baths	0.61	0.55	0.00	8.00
Age	44.56	35.06	0.00	200.00

Data for the sample of 50 municipalities in the state of Connecticut - 258,473 arms-length transactions from 1994 to 2013

Data for the 145 towns in the Greater Boston Area - 609,123 arms-length transactions from 1987 to 2012

Sales that were not standard market transactions such as foreclosures, bankruptcies, land court sales, and intra-family sales are excluded. Further, for each year, observations with the bottom and top 1% sales prices are excluded to further guard against non-

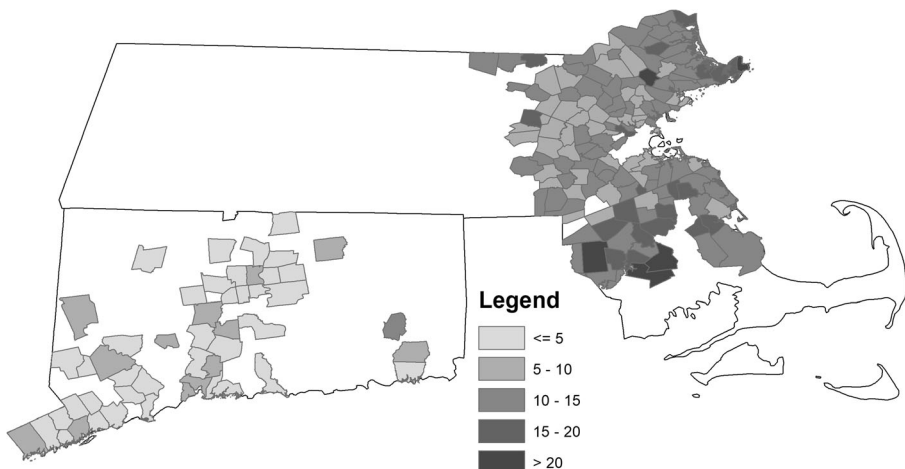


Fig. 1 COD values – OLS

arms-length sales and transcription errors. The data include typical house characteristics: age, living space, lot size, the number of bathrooms, bedrooms, and total rooms. The sample is limited to units with at least one bedroom and bathroom, 3 total rooms and 500 square feet of living space and no more than 10 bedrooms and 10 bathrooms, 25 total rooms, and 8000 square feet of living space, or 10 acres.

The second transaction is excluded for properties that sold twice within 6 months (similar to Case/Shiller) and for properties with two sales in the same calendar year with the same transaction price (likely duplicate records). Properties for which consecutive transactions occurred in the same year or in consecutive years and where the transaction price changed (in absolute value) by more than 100% are excluded. Similarly, properties where consecutive transactions were in year t and $t + j$ and where the transaction price changed (in absolute value) by more than $j00\%$ were excluded for $j = 2, \dots, 12$.

Thirty-two towns in Massachusetts with less than 100 total observations are dropped and 36 census tracts with less than 10 observations are excluded leaving a total of 145 towns and 833 census tracts for a total of 609,123 observations. Descriptive statistics for the data are provided in Table 1.

Results

First, we present OLS results for the entire samples from Connecticut and Massachusetts, including year and town fixed effects in Table 2 to see how the results compare across the two states.^{2,3} The dependent variable is the log of sales price. We include independent variables for log of living area square feet, log of lot size, and categorical variables for the number of total rooms number of bedrooms, number of bathrooms and house age.⁴ In our sample of Connecticut towns, there are 258,473 observations between 1994 and 2013 and the R^2 is 0.844. In our sample of MA towns, there are 609,123 observations between 1987 and 2012 and the R^2 is 0.700.

Generally, the results for Connecticut and Massachusetts are similar although there are some interesting differences that we highlight below. In both regressions, the living square feet, lotsize, and bathroom dummies coefficients are all positive and highly statistically significant and the number of bedrooms has little impact on price. But in Massachusetts, the age of the house is negatively related to price whereas houses age 10 to 30 are worth the most in Connecticut. Finally, house prices increase with the number of rooms in Massachusetts but only matter (positively) for the very largest houses in Connecticut.

Given that our goal is to compare different methods for property value assessment, we estimate hedonic regressions using fixed effects, GWR, and TGWR separately for each town in Connecticut and Massachusetts. Because assessors can only use current and past sales to make their assessments, we limit all samples to sales in the current and prior years of sale. For fixed effects, we estimate a separate regression for each year

² We also have OLS results on a town-by town basis, including year fixed effects. Due to the large volume of output, these OLS results for each of the municipalities are available from the authors upon request.

³ Cohen and Zabel (forthcoming) tried various specifications for the OLS version using the MA dataset, so here we settled on the preferred one from that study.

⁴ The Connecticut data for bathrooms has one variable that is the sum of bathrooms and half bathrooms (counted as 0.5 a bathroom) whereas the Massachusetts data includes separate variables for full bathrooms and half bathrooms.

Table 2 Town fixed effects results-dependent variable: logprice

	CT	MA
Log(living area sqft)	0.458*** (0.032)	0.323*** (0.010)
Log(lotsize)	0.0921*** (0.011)	0.063*** (0.005)
2 bathrooms ⁺	0.0436*** (0.007)	0.051*** (0.003)
3 bathrooms ⁺	0.122*** (0.009)	0.117*** (0.008)
4 or more bathrooms	0.282*** (0.018)	0.070*** (0.012)
Age >=10 and < 30	0.166*** (0.012)	-0.033*** (0.006)
Age >=30 and <=50	0.0588*** (0.008)	-0.127*** (0.007)
Age >=50	-0.0248 (0.013)	-0.177*** (0.011)
2 bedrooms	0.0357** (0.030)	0.010 (0.013)
3 bedrooms	0.0500*** (0.030)	0.019 (0.014)
4 bedrooms	0.0356** (0.028)	0.032** (0.014)
5 or more bedrooms	0.0566*** (0.026)	0.006 (0.014)
5 total rooms	-0.0167* (0.009)	0.024*** (0.005)
6 total rooms	-0.0210 (0.014)	0.026*** (0.007)
7 total rooms	-0.00341 (0.018)	0.034*** (0.008)
8 total rooms	0.0107 (0.022)	0.064*** (0.008)
9 total rooms	0.0293 (0.025)	0.106*** (0.009)
10 to 14 total rooms	0.0577** (0.026)	0.130*** (0.011)
15 or more rooms	0.196*** (0.050)	0.107*** (0.023)
1 half bathroom	- -	0.073*** (0.002)

Table 2 (continued)

	CT	MA
2 or more half bathrooms	–	0.112***
	–	(0.007)
R ²	0.844	0.700
N	258,473	609,123

Year dummies included

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, Clustered standard errors in parentheses

+ For the CT results, the data from Warren Group are not broken up into half baths and full baths; rather, 2 bathrooms include 2 or 2.5 bathrooms and 3 bathrooms includes 3 or 3.5 bathrooms

using the current and previous year's sales. For GWR, we estimate a regression for each sale using sales in the same town in current and previous year. For TGWR, the time dimension of the kernel is limited to sales in the current and previous years.

We use a bandwidth of 200 when applying GWR and TGWR. We offer robustness checks using different bandwidths below. We provide summary statistics for the results using these three estimators in Tables 3 and 4. The means of the point estimates for the different explanatory variables are similar. Not surprisingly, the standard deviations for the GWR and TGWR estimates are larger than for the fixed effects estimates since GWR and TGWR provide different point estimates for each sale.

We provide three statistics for evaluating the effectiveness of the three approaches to property value assessment: PAPE, COD, and PRD.

Table 5 provides the summary statistics for the COD and PRD statistics for OLS, GWR, and TGWR. First, we discuss the COD, which is a measure of accuracy. With the OLS estimates, all 50 towns in Connecticut have COD below the suggested upper bound of 15.0, while all but 24 in Massachusetts are in the acceptable range. This can be seen in Fig. 1. The situation improves in Massachusetts for GWR and TGWR, shown in Figs. 2 and 3, as all towns are below the cutoff for GWR and all towns but one are below the cutoff for TGWR. Further, in both states many towns that were close to the cutoff with OLS are now further below the acceptable cutoff.

Next, we discuss the PRD estimates, which is a measure of equity or progressivity/regressivity. As seen in Fig. 4, the PRD for OLS in Fig. 4 is above the cutoff of 1.03 for all 145 municipalities in MA, implying regressivity throughout the state. In Connecticut, there are several towns with PRD above the cutoff for OLS (see Fig. 4). While in Fig. 5, we see that the situation is mixed in Connecticut when re-estimating with GWR; in Massachusetts PRD improves somewhat with GWR compared with OLS. There is a dramatic improvement with TGWR, as nearly all towns in both states below the PRD cutoff of 1.03, implying neutrality (see Fig. 6).

Finally, we consider the PAPE results; summary statistics for the full set of estimates are given in Table 6. We see that OLS performs significantly worse than GWR and TGWR, whereas TGWR has a lower mean and standard deviation than GWR. The real difference is in the upper tail of the distribution where TGWR

Table 3 Summary statistics for coefficient estimates for town-Level OLS, GWR, and TGWR: CT data, $b = 200$

Variable	Town-level OLS		GWR		TGWR	
	Mean	SD	Mean	SD	Mean	SD
Log(living area sqft)	0.460846	0.147036	0.494622	0.214319	0.493798	0.194698
Log(lotsize)	0.087939	0.05912	0.098415	0.086663	0.097414	0.078184
Age >=10 and < 30	0.158827	0.066766	0.138405	0.162553	0.136647	0.126485
Age >=30 and <=50	0.056267	0.052834	0.043718	0.122411	0.042734	0.087073
Age >=50	-0.03772	0.076631	-0.03105	0.205574	-0.03186	0.09519
2 bathrooms	0.034511	0.026955	0.030772	0.202676	0.030297	0.138418
3 bathrooms	0.107408	0.046833	0.111545	0.227011	0.111214	0.129238
4 or more bathrooms	0.198859	0.100188	0.190862	0.272103	0.195438	0.242935
2 bedrooms	-0.00067	0.068846	0.827136	2.461339	0.199719	1.268229
3 bedrooms	0.000652	0.066883	0.830924	2.465192	0.199368	1.254315
4 bedrooms	-0.0081	0.072914	0.819265	2.468122	0.190706	1.258414
5 or more bedrooms	-0.03853	0.074284	0.772312	2.432396	0.160399	1.264407
5 total rooms	-0.01865	0.077376	0.099089	1.074278	-0.0104	0.343364
6 total rooms	-0.01302	0.08898	0.099435	1.087259	-0.01002	0.371523
7 total rooms	0.002402	0.103602	0.113502	1.107922	0.002799	0.359325
8 total rooms	0.013262	0.114643	0.122437	1.115686	0.011115	0.379564
9 total rooms	0.031008	0.124483	0.133033	1.119346	0.022335	0.391953
10–14 total rooms	0.022581	0.130131	0.124008	1.109783	0.015205	0.410715
15 or more total rooms	-0.0099	0.266177	0.054643	0.829567	-0.06913	0.545862
Constant	7.692293	0.959022	6.850662	2.860154	7.611603	1.842803

Town-Level regressions include year dummies

outperforms GWR. For example, at the town-level in MA, the 95th percentile is 50% PAPE for TGWR, 55% PAPE for GWR, and 88% PAPE for OLS. For CT, the corresponding percentages in the 95th percentile for TGWR, GWR, and OLS are 52%, 57%, and 140%.

Next, we consider the town-level statistics for PAPE since assessors will only be carrying out assessments within their own town. Tables 7 and 8 provide the summary statistics for the town-level distributions of the PAPE. We find similar results as in Table 6; the TGWR outperforms both OLS and GWR. The advantage of TGWR over GWR is most evident at the extremes of the distributions. For an assessor, this could prove quite valuable since avoiding large errors is important.

The PAPE metric provides a clear signal that TGWR is a preferred approach for estimating sales prices. This is underscored by the fact that the extreme values are the properties that in general have fewer sales comparisons for assessors to use in their valuations, so having a methodology such as TGWR that is demonstrated to perform well can be particularly valuable in estimating the values of these high-end properties.

Table 4 Summary statistics for coefficient estimates for town-level OLS, GWR, and TGWR: MA data, $b = 200$

Variable	Town-level OLS		GWR		TGWR	
	Mean	SD	Mean	SD	Mean	SD
Log(living area sqft)	0.315	0.099	0.309	0.177	0.302	0.157
Log(lotsize)	0.070	0.052	0.078	0.093	0.079	0.079
Age >=10 and < 30	-0.012	0.077	-0.008	0.452	0.019	0.226
Age >=30 and <=50	-0.102	0.094	-0.092	0.506	0.108	0.239
Age >=50	-0.153	0.120	-0.142	0.521	0.158	0.250
2 bathrooms	0.057	0.030	0.055	0.152	0.052	0.080
3 bathrooms	0.097	0.065	0.093	0.247	0.085	0.153
4 or more bathrooms	0.054	0.130	0.043	0.402	0.027	0.407
1 half bathroom	0.069	0.052	0.124	0.779	0.156	0.962
2 or more half bathrooms	0.099	0.079	0.145	0.812	0.175	0.984
2 bedrooms	0.039	0.127	1.621	3.710	0.805	2.684
3 bedrooms	0.045	0.124	1.628	3.708	0.815	2.684
4 bedrooms	0.059	0.132	1.637	3.711	0.826	2.689
5 or more bedrooms	0.028	0.137	1.557	3.666	0.786	2.683
5 total rooms	0.025	0.070	0.354	1.818	0.178	1.254
6 total rooms	0.041	0.076	0.393	1.879	0.196	1.257
7 total rooms	0.055	0.084	0.410	1.887	0.211	1.263
8 total rooms	0.077	0.089	0.433	1.890	0.235	1.270
9 total rooms	0.105	0.099	0.455	1.884	0.260	1.271
10–14 total rooms	0.109	0.114	0.441	1.856	0.252	1.254
15 or more total rooms	0.055	0.270	0.075	0.791	0.074	0.946
Constant	9.723	0.675	8.114	4.240	9.162	3.233

Town-Level regressions include year dummies

Robustness Checks

As robustness checks, we consider several separate bandwidths.⁵ For MA, we estimate TGWR using bandwidths of 200 and 500 (we do not consider other comparisons given the huge size of this dataset). The comparison of the PAPE calculated using the TGWR results for the 200 and 500 bandwidths are given in Tables 9 and 10. We can see that TGWR performs better at the lower bandwidth. The results are similar for CT, in the sense that the tails continue to perform better with TGWR.

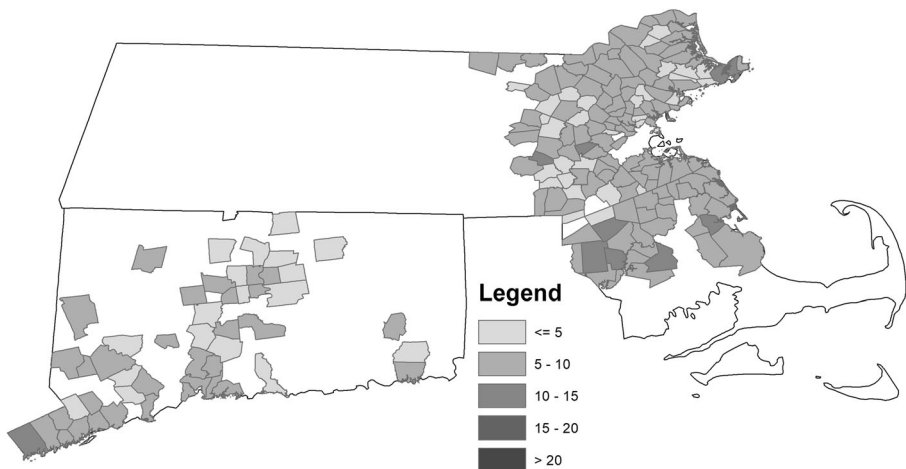
For CT we use bandwidths of 200, 500, and 1000. We examine the PAPE and find that the GWR density is virtually unchanged but the TGWR density is slightly worse at the lower

⁵ We recognize that there are formal procedures, such as using cross validation, to choose the optimal bandwidth. For example, see Fotheringham et al. (2015). In implementing Time GWR at the city level, the selection of the optimal bandwidth, which is likely to differ across cities, is an important step.

Table 5 Summary statistics for COD and PRD $b = 200$

Variable	Mean	SD	Min	Max
CT data				
OLS				
COD	3.44	3.41	-7.43	12.36
PRD	1.04	0.05	0.94	1.14
GWR				
COD	5.75	1.76	2.84	12.77
PRD	1.00	0.001	1.00	1.01
GWR with time				
COD	6.29	1.97	3.03	13.42
PRD	1.02	0.03	0.99	1.09
MA data				
OLS				
COD	11.89	3.47	5.60	28.02
PRD	1.11	0.03	1.06	1.27
GWR				
COD	6.53	2.13	1.18	14.98
PRD	1.05	0.02	1.02	1.15
GWR with time				
COD	6.06	2.17	0.04	15.05
PRD	1.02	0.01	1.00	1.07

percent errors but slightly improved for the larger percent errors. These differences are accentuated slightly further when increasing the bandwidth out to 1000. In other words, across these bandwidths, TGWR does a good job in mitigating large forecast errors that are present with GWR and OLS. As is the case for MA, the TGWR estimator performs better at

**Fig. 2** COD values – GWR (200)

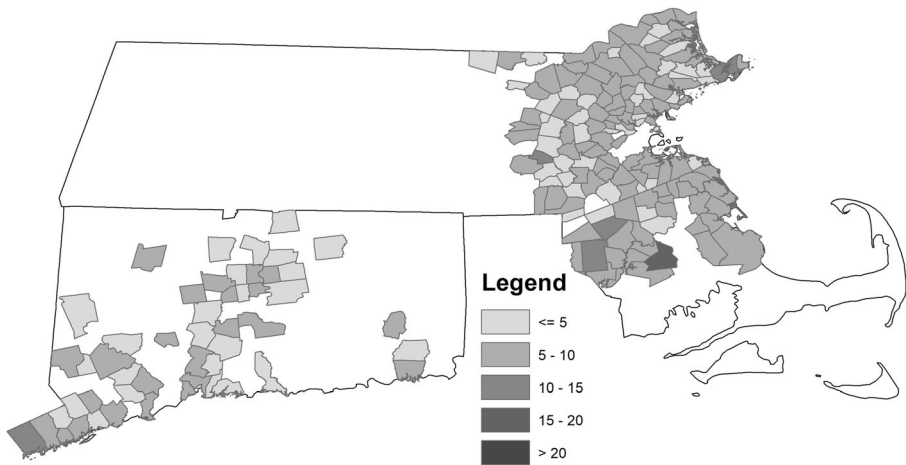


Fig. 3 COD values – TGWR (200)

the lower bandwidths (i.e., the PAPE is lower for $b = 200$ at the extreme values when compared with the extreme values of $b = 500$ and $b = 1000$).

For the $b = 500$ bandwidth, the mean of all of the coefficient estimates across the more than 250,000 weighted least squares regressions in CT and 600,000 observations in MA are similar to the magnitudes of the OLS estimates. But there is substantial variation in these estimates, which is apparent from viewing the standard deviations of the coefficients and their minimum and maximum values. For instance, all regressors have some observations for which the coefficients are negative and some for which the coefficients are positive. In many cases, the standard deviations are large relative to the mean of the coefficients.

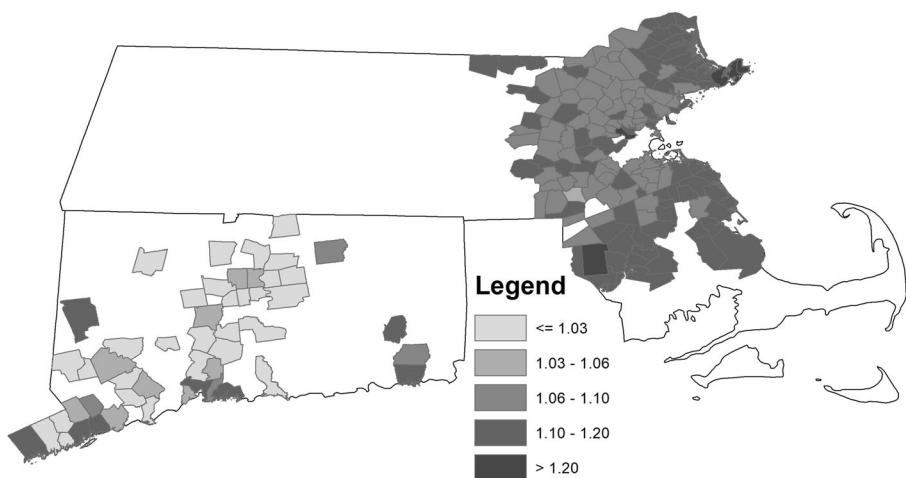


Fig. 4 PRD values – OLS

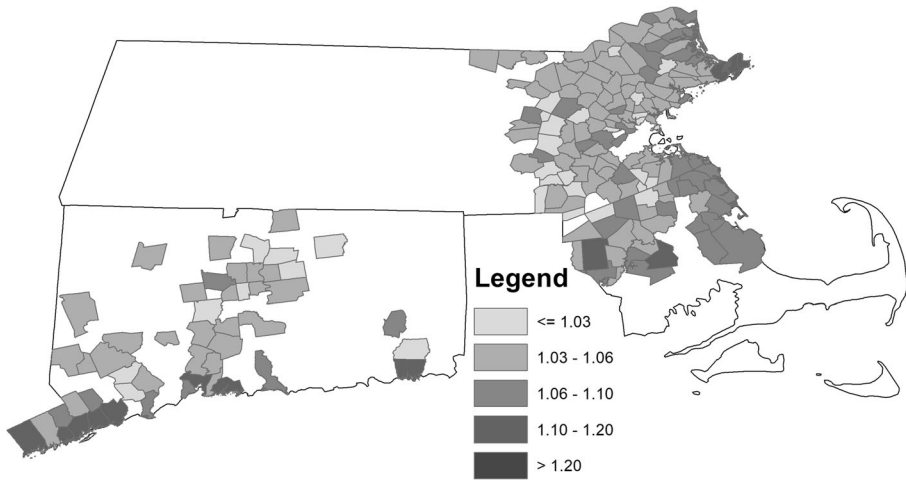


Fig. 5 PRD values – GWR (200)

One might hypothesize that including earlier observations with equal weight (i.e., OLS) will bias the estimates, especially in a sample like ours that spans 1994–2013 for CT and 1987–2012 for MA. Given these OLS, TGWR and GWR estimates, our objective is to use these to calculate predicted values of the sales prices that can be used to calculate the PAPE, COD and PRD, at the town level, separately for the OLS, TGWR and GWR models. We also calculate several separate sets of PAPE, COD and PRD for the GWR estimates, for several different bandwidth estimates. The GWR estimates are very insensitive to a broad range of bandwidths (including one as low as

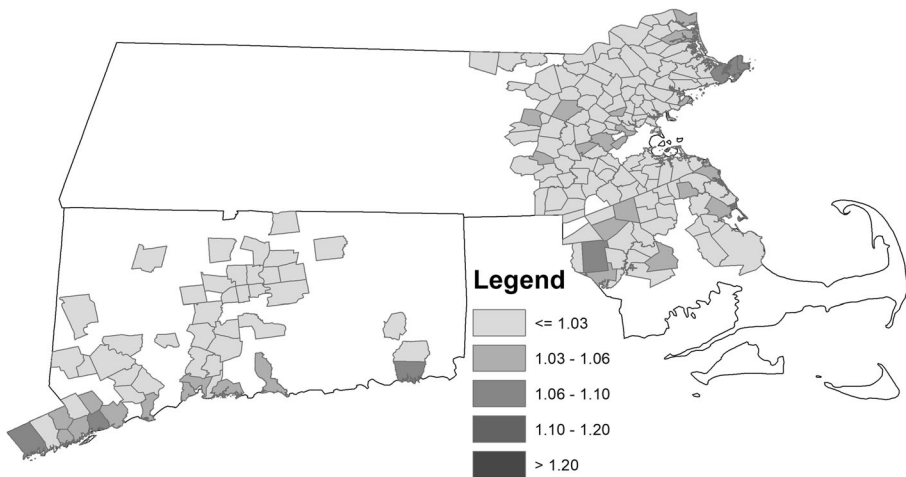


Fig. 6 PRD values – TGWR (200)

Table 6 Summary statistics for percent absolute prediction error: $b = 200$

Statistics	OLS	GWR	TGWR
CT data			
Mean	53.873	20.254	18.928
Standard Dev	70.916	40.119	35.247
Minimum	0.000	0.000	0.000
1st percentile	0.606	0.196	0.190
5th percentile	3.038	1.087	1.065
10th percentile	6.137	2.188	2.187
25th percentile	15.807	5.687	5.644
50th percentile	36.398	12.371	12.187
75th percentile	76.475	22.952	21.956
90th percentile	101.298	41.643	37.900
95th percentile	139.169	61.791	55.831
99th percentile	278.554	137.603	122.055
Maximum	5327.323	4830.420	145.135
MA data			
Mean	24.879	19.401	18.374
Standard Dev	50.336	37.108	33.183
Minimum	0.000	0.000	0.000
1st percentile	0.262	0.210	0.175
5th percentile	1.295	1.048	0.993
10th percentile	2.586	2.091	2.076
25th percentile	6.56	5.352	5.407
50th percentile	13.865	11.562	11.605
75th percentile	24.697	20.722	20.372
90th percentile	46.174	34.851	32.835
95th percentile	79.996	55.104	49.376
99th percentile	228.187	173.892	156.674
Maximum	2506.656	1642.226	1384.39

0.5 and as high as 1000, with several in between). For TGWR, there is slightly more variation in COD, PRD, and PAPE when changing the bandwidth, but not much. Given this stability, along with the large sample sizes for the CT and MA data (250,000+ observations across 50 towns, and 600,000+ observations over 145 towns), and given that we find little variation in the estimates for GWR and TGWR across bandwidths, we present the results for GWR and TGWR for a bandwidth of 200 in the results tables in this paper.

To compare estimates across the two states, we present the COD and PRD at the town-level in Figs. 1, 2, 3, 4, 5, and 6. The PAPE estimates for $b = 200$ are presented in Tables 6, 7, and 8. All calculations are done at the town level. Regressions include the same variables as those using the full sample.

Table 7 Summary statistics for town-level distribution of percent absolute prediction error: CT Data, $b = 200$

Percentile	Mean	Median	SD	Min	Max
OLS					
5th	3.463	3.269	1.169	1.741	7.301
10th	6.995	6.748	2.342	3.76	14.803
25th	17.882	17.152	5.82	9.317	36.965
50th	39.463	38.153	12.29	20.355	79.75
75th	76.226	75	17.484	39.842	127.416
90th	112.139	100	21.102	100	190.177
95th	139.511	131.541	37.637	100	260.808
GWR					
5th	1.059	1.009	0.231	0.674	1.755
10th	2.197	2.012	0.488	1.582	3.62
25th	5.695	5.361	1.227	3.939	9.353
50th	12.45	11.513	2.811	9.201	20.382
75th	22.756	20.65	5.813	16.217	37.9
90th	38.984	35.591	10.039	26.225	62.22
95th	57.32	55.418	12.797	35.597	88.484
TGWR					
5th	1.059	1.009	0.231	0.674	1.755
10th	2.18	2.08	0.451	1.343	3.521
25th	5.645	5.369	1.099	4.019	9.175
50th	12.143	11.316	2.436	8.69	19.505
75th	21.702	19.853	4.771	15.5	34.712
90th	35.755	33.426	8.471	24.466	55.992
95th	51.78	50.187	10.891	30.915	75.803

Conclusion

We develop the TGWR estimator for hedonic pricing models as a new procedure for calculating assessed values. We compare this new methodology to two previously used approaches; GWR and OLS. As a means of comparison, we use the Coefficient of Dispersion (COD) and the Price-Related Differential (PRD); statistics that are commonly employed by the IAAO to examine accuracy and regressivity/progressivity. As a measure of predictive accuracy, we also compare these three methodologies using the percent absolute prediction error (PAPE).

Based on the PAPE metric, TGWR performs better than GWR and OLS. This discrepancy in performance is particularly strong at the extreme percentiles. This is an especially important finding for assessors, because the extreme percentiles (e.g. 95th) tend to have the highest value properties that are often the most difficult to value because of the lack of comparable sales. Also, with the PRD and COD measures, there is variation across towns in terms of the progressivity/regressivity/neutrality. In many

Table 8 Summary statistics for town-Level distribution of percent absolute prediction error: MA Data, $b = 200$

Percentile	Mean	Median	SD	Min	Max
OLS					
5th	1.350	1.331	0.232	0.909	2.422
10th	2.697	2.633	0.425	1.913	4.670
25th	6.840	6.691	1.026	4.717	10.428
50th	14.418	14.160	2.175	10.400	22.600
75th	25.673	24.738	4.362	18.333	44.144
90th	49.757	46.791	14.194	30.892	131.752
95th	87.704	79.427	27.65	45.966	188.901
GWR					
5th	0.996	0.997	0.089	0.774	1.246
10th	2.102	2.104	0.146	1.637	2.622
25th	5.349	5.353	0.222	4.543	6.213
50th	11.556	11.570	0.324	10.505	12.827
75th	20.718	20.686	0.487	18.479	22.21
90th	34.912	34.835	1.283	31.623	41.251
95th	55.410	54.92	4.079	46.633	75.407
TGWR					
5th	0.996	0.997	0.089	0.774	1.246
10th	2.080	2.074	0.159	1.778	2.990
25th	5.415	5.413	0.235	4.771	6.260
50th	11.604	11.574	0.336	10.749	12.602
75th	20.404	20.356	0.509	18.802	21.675
90th	33.021	32.803	1.367	28.858	37.958
95th	50.360	49.337	4.576	41.678	68.912

locations we see PRD estimates that indicate more municipalities are in the neutral range (i.e., less than 1.03) with TGWR than with GWR or OLS.

An important issue that is worthy of further attention is the bandwidth selection, since the COD and PRD results in some of the individual towns are sensitive to bandwidth selected. Ideally, an assessor would use cross-validation (CV) separately on every sale in their municipality to determine a separate optimal bandwidth for the local regression for each property. But, given approximately 900,000 observations in our study, this would be extremely time consuming to implement; as it is, for each bandwidth it takes more than 10 days to solve GWR (and separately, TGWR) for the 900,000+ observations. We recommend that individual assessors use CV on their own municipalities which would be much more feasible to implement in a TGWR or GWR setting. Also, while one possibility to reduce computation time might be to use a sub-sample of the sales in a municipality and interpolate sales prices for other properties, this results in the loss of some valuable information for assessors. Therefore, it is important for individual assessors to use all available data on sales when performing this type of analysis.

Table 9 summary statistics for percent absolute prediction error: CT Data, TGWR

Statistics	Bandwidth		
	200	500	1000
Mean	18.928	20.522	22.518
Standard Dev	35.247	39.247	40.186
Minimum	0	0	0
1st percentile	0.19	0.248	0.289
5th percentile	1.065	1.27	1.49
10th percentile	2.187	2.553	2.987
25th percentile	5.644	6.455	7.574
50th percentile	12.187	13.516	15.853
75th percentile	21.956	23.618	26.695
90th percentile	37.9	39.872	42.114
95th percentile	55.831	59.05	61.046
99th percentile	122.055	133.425	139.354
Maximum	4145.135	5116.346	5575.472

Also, TGWR is a very time-intensive set of computations, which is magnified when considering many different bandwidth options. Therefore, assessors should plan a lot of time to develop Time-GWR estimates since it can take many hours to run for larger

Table 10 Summary statistics for percent absolute prediction error: MA Data, TGWR

Statistics	Bandwidth	
	200	500
Mean	18.291	20.824
Standard Dev	32.995	35.477
Minimum	0.000	0.000
1st percentile	0.175	0.268
5th percentile	0.989	1.34
10th percentile	2.065	2.716
25th percentile	5.386	6.883
50th percentile	11.566	14.235
75th percentile	20.315	23.704
90th percentile	32.729	35.617
95th percentile	49.104	51.699
99th percentile	155.713	166.152
Maximum	1384.39	1533.289

Note for Table 10: Some estimates here differ slightly from those in Table 6, because sample sizes differ due to being conditional on observations where PAPE is observed for all methods. In Table 10 it is TGWR for 200 and 500 bandwidths. In the MA results in Table 6, it is OLS, GWR and TGWR for 200 bandwidth

cities for each bandwidth. If one finds, for instance, that the smaller bandwidths perform better, this implies that when the tails of the distribution shrink and the distribution of the kernel weights become more focused around the target point, one obtains more accurate assessments. But the tradeoff is that we need a sufficient number of observations to make the local regressions work, so an assessor would not want to have too few observations that are given positive weight. Therefore, a detailed bandwidth consideration, such as cross-validation, may be important for assessors in refining this approach. However, this adds additional layers of computational difficulty that need to be considered in future work.

Acknowledgements The views expressed are those of the authors and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors. The authors thank Andrew Spewak for excellent research assistance and Jason Barr, Diana Gutierrez Posada, and others at the D.C. Real Estate Valuation Symposium in October 2018 and at the North American Meetings of the Regional Science Association International in November 2018 for constructive comments.

References

- Bidanset, P. E., & Lombard, J. R. (2014a). The effect of kernel and bandwidth specification in geographically weighted regression models on the accuracy and uniformity of mass real estate appraisal. *Journal of Property Tax Assessment & Administration*, 11(3), 5–14.
- Bidanset, P. E., & Lombard, J. R. (2014b). Evaluating spatial model accuracy in mass real estate appraisal: A comparison of geographically weighted regression and the spatial lag model. *Cityscape*, 16(3), 169–182.
- Borst, R. A., & McCluskey, W. J. (2008). Using geographically weighted regression to detect housing submarkets: Modeling large-scale spatial variations in value. *Journal of Property Tax Assessment & Administration*, 5(1), 21–21.
- Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 83(403), 596–610.
- Cohen, J. P., Osleeb, J. P., & Yang, K. (2014). Semi-parametric regression models and economies of scale in the presence of an endogenous variable. *Regional Science and Urban Economics*, 49, 252–261.
- Cohen, J. P., Coughlin, C. C., & Clapp, J. M. (2017). Local polynomial regressions versus OLS for generating location value estimates. *The Journal of Real Estate Finance and Economics*, 54(3), 365–385.
- Fotheringham, A. S., Crespo, R., & Yao, J. (2015). Geographical and temporal weighted regression (GTWR). *Geographical Analysis*, 47, 431–452.
- International Association of Assessing Officers (IAAO) (2003). Standard on Automated Valuation Models (AVMs). Chicago: International association of assessing officers.
- International Association of Assessing Officers (IAAO). (2013). *Standard on Ratio Studies*. Kansas City: International Association of Assessing Officers.
- Lockwood, T., & Rossini, P. (2011). Efficacy in modelling location within the mass appraisal process. *Pacific Rim Property Research Journal*, 17(3), 418–442.
- McCluskey, W. J., McCord, M., Davis, P. T., Haran, M., & McIlhatton, D. (2013). Prediction accuracy in mass appraisal: A comparison of modern approaches. *Journal of Property Research*, 30(4), 239–265.
- McMillen, D. P. (1996). One hundred fifty years of land values in Chicago: A nonparametric approach. *Journal of Urban Economics*, 40(1), 100–124.
- McMillen, D. P., & Redfeam, C. L. (2010). Estimation and hypothesis testing for nonparametric hedonic house price functions. *Journal of Regional Science*, 50(3), 712–733.
- Meese, R., & Wallace, N. (1991). Nonparametric estimation of dynamic hedonic price models and the construction of residential housing price indices. *Real Estate Economics*, 19(3), 308–332.
- Silverman, B. W. (1986). Density Estimation for Statistical Analysis.
- Wooldridge, J. M. (2016). *Introductory econometrics* (6th ed.). Boston: Cengage Learning.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Jeffrey P. Cohen¹ · **Cletus C. Coughlin**² · **Jeffrey Zabel**³

Jeffrey P. Cohen
Jeffrey.Cohen@business.uconn.edu

Jeffrey Zabel
Jeff.Zabel@tufts.edu

¹ University of Connecticut, Storrs, CT, USA

² Federal Reserve Bank of St. Louis, St. Louis, MO, USA

³ Tufts University, Medford, MA, USA