Spatial Economic Analysis

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rsea20

Covid-19 impact on US housing markets: evidence from spatial regression models

Jim Lee & Yuxia Huang

To cite this article: Jim Lee & Yuxia Huang (2022): Covid-19 impact on US housing markets: evidence from spatial regression models, Spatial Economic Analysis, DOI: 10.1080/17421772.2021.2018028

To link to this article: https://doi.org/10.1080/17421772.2021.2018028

Published online: 10 Jan 2022.

Submit your article to this journal

Article views: 263

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=rsea20

SPATIAL ECONOMIC ANALYSIS
https://doi.org/10.1080/17421772.2021.2018028

Routledge Taylor & Francis Group
Regional Studies Association
Check for updates

Covid-19 impact on US housing markets: evidence from spatial regression models

Jim Lee [a] and Yuxia Huang [b]

ABSTRACT
This paper empirically investigates the conventional wisdom that urban residents have reacted to the Covid-19 pandemic by fleeing city centres for the suburbs. A conventional panel model of US ZIP code-level data provides mixed evidence in support of a shifting housing preference for more space or neighbourhoods farther from the urban core. Regressions accounting for spatial dependence and spatial heterogeneity show strong support of an urban flight within metro areas, but this local phenomenon is uneven across broad regions of the United States. The finding of geographical disparity underscores both the local as well as the regional nature of housing market conditions.

KEYWORDS
housing markets, spatial dependence, spatial heterogeneity, spatial autoregression, geographically weighted regression

JEL C4, R1
HISTORY Received 00 March 2021; in revised form 15 November 2021

INTRODUCTION

Central cities in the United States have gained vitality since the turn of the 21st century. In the downtown neighbourhoods of metro areas, growth in the share of the prime working population aged 25–54, the overall educational attainment and household income level, and home values has all outpaced their suburbs and rural counterparts (Fry, 2020). In the face of the Covid-19 pandemic, however, a growing body of anecdotal evidence from local housing markets across the nation points to a pause and probable reversal of this demographic trend (Campisi, 2020; Speianu, 2020).

After Covid-19 outbreaks took hold across the United States in early 2020, home sales declined precipitously when stay-at-home orders, business lockdowns and social distancing mandates reduced house searching and transaction activities. As housing markets rebounded along with overall economic recovery beginning in May, divergent trends between urban and suburban markets emerged (Speianu, 2020). Within New York City, for instance, home prices in the Lowest East Side of Manhattan dropped more than 50% year-on-year in October, while sales grew in the city's so-called bedroom communities around New York City, such as Kingston in Ulster County and the lower end of Sussex County (Campisi, 2020). Narratives of an urban flight

CONTACT
[a] **(Corresponding author)** jlee@tamucc.edu
Texas A&M University – Corpus Christi, Corpus Christi, TX, USA. CONTACT
[b] Texas A&M University – Corpus Christi, Corpus Christi, TX, USA. lucy.huang@tamucc.edu

© 2022 Regional Studies Association

accord with an observed shift in travel trips from large metro areas' urban cores, such as Chicago downtown, to the suburbs (Grogan, 2020).

Despite abundant media stories about urban dwellers moving out of the metropolitan centres and into the suburbs, Handbury (2020) argues that an urban flight is a myth without 'scientific evidence' or rigorous analysis, and the observed demographic shifts are limited to large cities such as San Francisco and New York. Still, Liu and Su (2020) show evidence of a shifting homebuyer preference toward less densely populated suburbs across the nation during the early stage of the pandemic through June.

The objective of this study is to reconcile the ongoing controversy over the prevalence of a housing market shift within metro areas across the United States. To this end, we adopt a spatial econometric approach that allows inferences on the key factors of interest, namely the distance from a city centre and population density, to be dependent on the spatial or geographical characteristics that are not directly observable in the data-generating processes being modelled (Basile et al., 2014; Geniaux & Martinetti, 2018). One aspect of such characteristics is spatial dependence in the context that one neighbourhood's housing market conditions are similar to its nearby neighbourhoods' (Brady, 2011; Holly et al., 2011; Ioannides & Zabel, 2003). We explore spatial dependence with a spatial autoregressive (SAR) model (LeSage & Pace, 2009). Alternatively, in line with the adage 'there is no such thing as a US housing market', a geographically weighted regression (GWR) model allows for spatial heterogeneity, which reflects varying model relationships across neighbourhoods or submarkets of a broader study area (Fotheringham et al., 2002).

Spatial dependence and spatial heterogeneity are cornerstones of spatial-oriented studies (Geniaux & Martinetti, 2018). SAR models have gained popularity in the applied econometric literature on housing markets (Basile et al., 2014; Bhattacharjee et al., 2012; Holly et al., 2011; Kuethe & Pede, 2011; Pace et al., 2000). However, the conventional SAR framework does not delineate submarkets based on market characteristics that are spatially heterogeneous. Because spatial heterogeneity arises largely from the lack of spatial stability in model relationships, it has received more attention in the theoretical spatial literature than in the econometric modelling literature (Basile et al., 2014). Spatial heterogeneity is typically modelled by locally weighted regression, such as a non-parametric estimator (McMillen, 1996) and GWR (Fotheringham et al., 2002).

GWR modelling is found predominantly in the field of geographical information systems (GIS) (Bidanset & Lambard, 2014; Brunsdon et al., 1999; Fotheringham et al., 2015; Huang et al., 2010; Li et al., 2017; Lu et al., 2017; Wu et al., 2014) due in part to the requirement for spatial data and a calibration process that is computationally intensive. The present study is one of the early attempts that investigate the US housing markets with a recently advanced SAR-GWR panel setting (Yu, 2010), which pools time-series and cross-section data together. This framework allows us to explore changing market conditions across neighbourhoods within metro areas of the United States while allowing for disparity across broad regions.

The rest of the paper proceeds as follows. The next section describes the data and motivation for our empirical analysis. The third section outlines the spatial econometric models that we apply in this study, followed by a discussion of regression results in the fourth section. The fifth section contains a conclusion along with suggestions for extensions in future research.

DATA DESCRIPTION

Housing market data

Our empirical work on the post-Covid-19 housing markets in the United States builds on the recent studies of Liu and Su (2020) and Zhao (2020). The alternative housing market outcome variables are online viewings of a typical home, the number of houses sold (single and multifamily units) and the median sales prices. We obtained monthly ZIP code-level data of median home prices from Zillow.com and the other housing data from Realtor.com. Online views of properties are considered a novel yet useful indicator of house shopping activity and homebuyers' preferences, especially when stay-at-home and lockdown orders in early 2020 prevented in-person showings (Zhao, 2020). Based on the most recent available data from various sources, our dataset consists of 11,317 US ZIP codes between April and December 2020. The sample covers about 70% of all urban and suburban communities in the United States.[1]

Figure 1 provides an overview of changing US housing market conditions in 2020, as measured alternatively by year-on-year per cent changes in online views, sales, and median home prices. For each month between April and December 2020, the cross-sectional averages of ZIP code-level data are computed for three alternative subsamples based on (1) the distance between a ZIP code area and its closest metro centre and (2) the area's population density.

As explained below, distance is measured by driving time. Each plot in the first column of Figure 1 compares the monthly cross-sectional averages of one of the three housing market indicators for ZIP codes within the bottom 5% of the full sample by distance (i.e., the range of distance up to the 5th percentile) against the middle 50% (interquartile) and the top 25% (fourth quartile). In comparison with the median of about 31 min, the 5th percentile represents about a 5-minute driving distance from a metro centre.

According to the first plot, growth in online views for houses within this measure of downtown neighbourhoods closely followed the rest of the nation through April. Since then, the downtown neighbourhoods have trailed other parts of the metro area, especially those in the fourth quartile of distance with an average driving time of about 60 min.

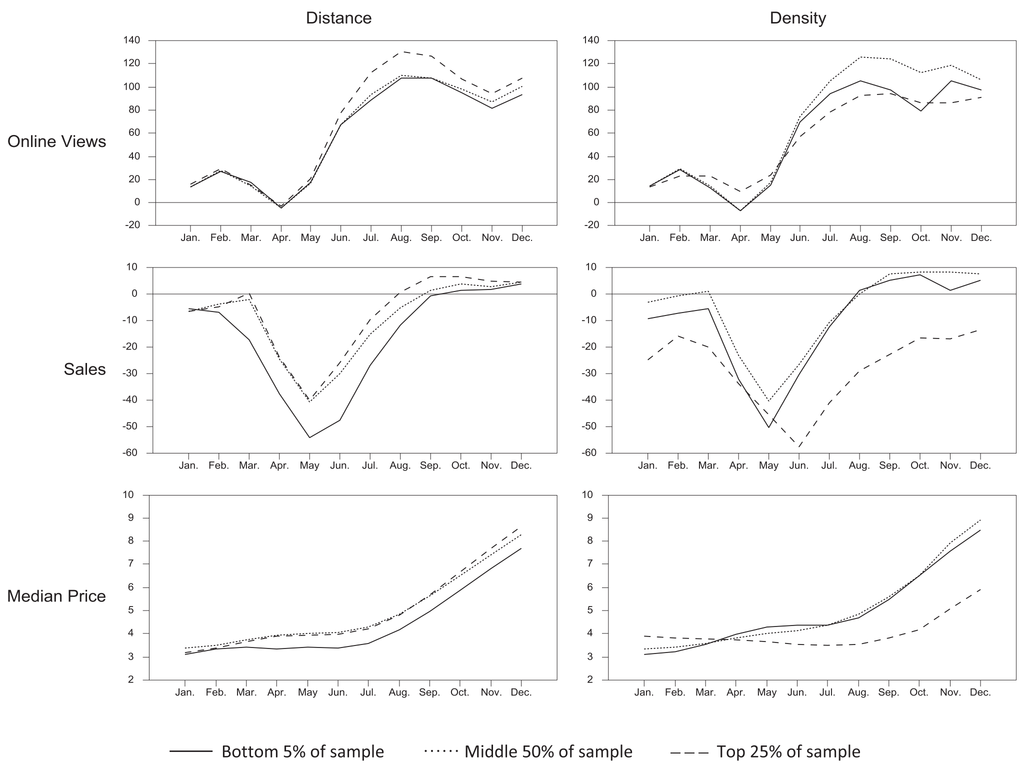

Figure 1. Year-on-year per cent change in US housing market indicators, 2020.

Despite growth in online viewing activity, house sales were below the 2019 levels until September. Precipitous declines in sales between April and May were likely the outcomes of statewide shelter-in-place orders and lockdowns that closed non-essential businesses, including real estate agencies. Home sales fell more near city centres than the suburbs during the first half of 2020, but their gaps diminished by year end.

Housing markets across the nation recovered immediately after states reopened their economies. An acceleration in home prices over the second half of 2020 helped raise the appreciation of the US median home price to 5.2% on average between 2019 and 2010. One driver for the strong overall market recovery amid a historic economic recession is the Federal Reserve's aggressive monetary easing measures, leading the 30-year fixed mortgage rate to slide below 3% by mid-2020 (Zhao, 2020). Median home prices tended to rise faster among ZIP codes farther from city centres.

The plots in the second column of Figure 1 display the corresponding monthly data of subsamples defined by neighbourhoods of different levels of population density. The patterns of shifting housing market conditions are similar to those for subsamples defined by distance. Relative to other parts of the metro areas, the growth of online viewings in the least populated neighbourhoods (lowest 5%) fluctuated less over time, as sales fell disproportionately among the most densely populated neighbourhoods (fourth quartile). By comparison, ZIP codes with population density in the interquartile distribution rebounded the most after April.

Together, the plots in Figure 1 reveal uneven patterns of changing US housing market conditions across different neighbourhoods in the face of the Covid-19 pandemic. During this period, there was an apparent housing demand shift from densely populated downtown neighbourhoods to less dense city outskirts and suburbs. These observations motivate our empirical work to explore the drivers behind the observed changes in local housing market conditions.

Housing demand

Housing demand reflects people's desire to live and work in certain locales. Liu and Su (2020) offer several reasons for a shift of homeowners' preference away from locations close to downtown or densely populated neighbourhoods since the onset of the Covid-19 pandemic. First, the prevalence of work from home has reduced the need for many employees to live close to their workplace. City centres tend to have a greater share of jobs that allow for work from home (Veuger et al., 2020).

The second reason for the shifting housing demand preference toward suburbs stems from access to amenities. Other than employment concentration, cities typically offer more leisure and consumption amenities. Access to restaurants and other amenities, such as movie theatres and shopping malls, became less attractive in the face of business lockdowns, capacity restrictions and social distancing practices during the pandemic. Because of the perks and amenities that downtowns have offered historically, housing demand and thus home prices tend to be higher in those neighbourhoods. As the benefits of living within a city have diminished, residents have instead opted for suburban neighbourhoods that offer relatively more affordable homes.

Another reason for homeowners to avoid dense neighbourhoods is the high risk of contracting the coronavirus in crowded places. Social distancing is more difficult in high-rise apartments and public transit systems in cities. People likely feel safer in more spacious homes and less crowded places in the suburbs.

Given the focus of our study, the key measures of local characteristics for analysing housing demand are population density and the travel distance between the centre of a ZIP code area and the central business district (or downtown) of its closest metro area. Instead of the Euclidean (straight-line) geographical length, our distance measure is in terms of minutes of driving time. Although these two alternative distance measures are highly correlated (with a correlation coefficient > 0.8), we consider driving time a better measure of distance perceived by homebuyers.[2] The travel distance between each ZIP code area and its closest metro downtown is calculated through the street network using ESRI's ArcGIS software.

To evaluate the impact of the pandemic on housing demand, our empirical model includes a measure of the local Covid-19 infection rate. As data on confirmed Covid-19 cases are available at the county level, a ZIP code area's monthly case rate since March 2020 draws on the number of monthly county-level new cases divided by the county population.

Following Liu and Su (2020), our regressions also control for other local characteristics, including the number of jobs and the share of jobs that are compatible with work from home, the number of restaurants per capita and the per capita income level. People tend to prefer to live close to their workplace in order to minimize the commuting time (Glaeser & Gottlieb, 2009). The number of jobs per capita reflects the extent of employment opportunities. As discussed above, the pandemic might have reduced the attractiveness of living in neighbourhoods with relatively more jobs.

Other things being equal, an area with a larger fraction of jobs that can be performed remotely also reduces the desirability of employees to live close to their workplaces, especially during the pandemic. To construct ZIP code-level data for the share of jobs that are compatible with remote work, we followed the methodology developed by Dingel and Neiman (2020) for metro areas. We first evaluated each occupational information network (O*NET) occupation's compatibility with remote work. Next, we matched the compatibility measures with the North American Industry Classification System (NAICS) job distributions of each ZIP code through an industry–occupation crosswalk.

As discussed above, city downtowns typically offer relatively more consumption amenities. Our regressions include the number of restaurants per capita as a proxy. Regressions also include income per capita to control for socioeconomic heterogeneity across neighbourhoods.[3] Moreover, we control for the effect of pre-pandemic housing demand with the inclusion of the local median house price level in 2019.

Some of those explanatory variables may be highly correlated, potentially leading to multicollinearity in regressions. For instance, a suburban community farther from the city centre may also have a lower population density. Nonetheless, the correlation coefficients between explanatory variables described above are all < 0.5 and their variance inflation factors (VIFs) are < 5. These diagnostic statistics mitigate the concern of multicollinearity.

EMPIRICAL MODELS

In our study of N locations and T periods, the panel-data regression model as a comparison benchmark can be expressed as:

$$y_{it} = X_{it}\beta + \varepsilon_{it},\ i = 1,\ \ldots,\ N;\ t = 1,\ \ldots,\ T \tag{1}$$

$$\varepsilon_{it} = \alpha_i + \tau_t + \epsilon_{it}, \tag{2}$$

where y_{it} denotes a housing market outcome; X_{it} is a matrix of K explanatory variables; and β is a $K \times 1$ vector of coefficients. The error term ε_{it} includes unobserved time-invariant individual or spatially fixed effects captured by α_i, common time effects captured by τ_t, and an i.i.d. error ϵ_{it}.

Equation (1) abstracts from any kind of indirect or spatial effects due to interactions or spillovers between nearby locations. In reality, nearby housing markets tend to be interrelated, generating the widely recognized 'neighborhood' effects (Brady, 2011; Holly et al., 2011; Ioannides & Zabel, 2003). Homeowners likely check prices in their neighbourhoods before putting a house on the market. Proximity to the same sources of positive or negative externalities, such as air emissions from an industrial plant, may have similar impacts on nearby communities. On the other hand, public services, such as school quality and street conditions, and local regulations that affect housing supply tend to vary from one community to another (Basile et al., 2014). Empirical models that ignore these spatial attributes are subject to misspecification that may have profound implications for our understanding of changes in housing market outcomes.

To explore unobserved geographical or spatial interactions, we extend the panel-data model by incorporating the SAR framework (LeSage & Pace, 2009). Corresponding to equation (1), the SAR that captures the spatial linkages of the observations of the dependent variable can be expressed as:

$$y_{it} = \rho 1^{\top} \boldsymbol{W} y_t + X_{it}\beta + \varepsilon_{it}, \tag{3}$$

where ρ is a scalar parameter known as the SAR coefficient capturing spatial spillover effects, $1^{\top} = (1, 1, \ldots, 1)$ is an N-vector of ones, $y_t = (y_{1t}, y_{2t}, \ldots, y_{Nt})^{\top}$, and $\boldsymbol{W}$ is an $N \times N$ spatial weight matrix that consists of diagonal elements $w_{ij} = 0$ for $i = j$. Spatial dependence is captured by the spatially lagged dependent variable, $\boldsymbol{W}y_t$. Following a popular approach in the spatial literature (Anselin, 1988), we construct the spatial matrix $\boldsymbol{W}$ by contiguity, so that the relationships between each location i and its neighbour that shares a common border (edge or corner) are weighted equally.[4] Because the spatial lag variable is correlated with the error term, that is, $\mathrm{Cov}(\boldsymbol{W}y_t, \varepsilon_{it}) \neq 0$, least-squares estimates of equation (3) are biased and inconsistent (Basile & Minguez, 2018). For this reason, the panel SAR model is typically estimated using a maximum likelihood (ML) procedure described by Elhorst (2009).

Spatial autocorrelation expands the information set in a regression model to include information from neighbouring locations. Spatial dependence is apparent in housing markets and thus SAR has become increasingly standard in the real estate literature (Basile et al., 2014; Bhattacharjee et al., 2012; Holly et al., 2011; Kuethe & Pede, 2011; Pace et al., 2000; Pace & Gilley, 1997). Still, SAR is a 'global' model incapable of accounting for unobserved spatial heterogeneity. Model relationships may be unevenly distributed across a broad study area such as the United States. Some explanatory variables may affect housing market outcomes in some but not all ZIP code areas.

The allowance for coefficients to vary across different locations has made GWR a popular approach to modelling local housing markets (Huang et al., 2010).[5] GWR essentially extends 'global' regression models to spatial heterogeneity by allowing coefficients to be estimated 'locally'. Following the conventional notation, an extension of equation (3) to GWR can be written as:

$$y_{it} = \rho(u_i, v_i) 1^{\top} \boldsymbol{W} y_t + X_{it}\beta(u_i, v_i) + \varepsilon_{it}, \tag{4}$$

where (u_i, v_i) denotes the respective longitude and latitude coordinates of observation i; and the coefficients $\rho(u_i, v_i)$ and $\beta(u_i, v_i)$ reflect values at that location. In contrast to constant coefficient estimates for all locations in the 'global' model, equation (4) allows the geographically weighted coefficients to vary across different observations in different locations, and therefore it is likely to capture heterogeneous 'local' effects.

As pointed out above, the spatially lagged dependent variable is correlated with the error term. To remove this source of endogeneity, the SAR-GWR model is estimated by a spatial two-stage least squares procedure (Anselin, 1988; Geniaux & Martinetti, 2018; Kelejian & Prucha, 1998), which relies on a set of instruments $H = [\boldsymbol{X}, \boldsymbol{WX}, \boldsymbol{W}^2\boldsymbol{X}, \ldots]$. One major advantage of this procedure is higher efficiency in iterated computation for local regressions, particularly in the case of a large sample size such as ours.

As for SAR, GWR requires the construction of a spatial weight matrix $\Omega(u_i, v_i) = \mathrm{diag}(\omega_{i1}, \omega_{i2}, \ldots, \omega_{iN})$. The diagonal elements in $\Omega(u_i, v_i)$ represent the geographical weights of each observation of the i-th location (Fotheringham et al., 2002). In practice, this diagonal weight matrix typically assumes that data observations closer to location i have a greater effect in estimating a coefficient than observations located farther from that location. Our specific spatial weighting scheme draws on the 'adaptive' bandwidth approach, which determines a different window size for each location i, so that all locations have the same number of nearest neighbours.[6] For each regression point i, the weight of data observations of another location j in estimating each regression point is given by a bi-square kernel function for the diagonal elements of $\Omega(u_i, v_i)$:

$$\omega_{ij} = \begin{cases} \left[1 - \left(\frac{d_{ij}}{h_i}\right)^2\right]^2, & \text{if } d_{ij} < h \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where $d_{ij}^2 = [(u_i - u_j)^2 + (v_i - v_j)^2]$ is the Euclidean distance between location i and location j; and h_i is the bandwidth (i.e., window size) at location i.

To identify the spatial weight matrix $\boldsymbol{W}$, we apply a moment estimator that optimizes a goodness-of-fit statistic for the model estimation with respect to the specified $\boldsymbol{W}(h_\omega)$ (Geniaux & Martinetti, 2018). The search for the optimal value of h_ω is performed for each bi-square kernel. Our final choice corresponds to the pair of kernel/bandwidth that generates the minimum Akaike information criterion (AIC) for the SAR-GWR model. In our study, the optimal bandwidth is 809 ZIP codes for the online views and sales data, and 854 ZIP codes for the median price data. These window sizes correspond to about 13 broad regions across the United States.

Housing market modelling is a popular empirical application of GWR, given the restriction $\rho(u_i, v_i) = 0$ in equation (4) (Basile et al., 2014; Bhattacharjee et al., 2016; Brunsdon et al., 1999; Fotheringham et al., 2015; Huang et al., 2010; Wrenn & Sam, 2014; Wu et al., 2014). Using a Monte Carlo experiment, Geniaux and Martinetti (2018) find the performance of the mixed SAR-GWR model to be more robust than conventional models, such as ordinary least squares (OLS), SAR and GWR. Building on the above spatial regression studies that rely on cross-section data, we extend the SAR-GWR model to a panel data setting as detailed in Basile and Minguez (2018), Lin (2011) and Yu (2010). More specifically, we estimate our panel housing data with fixed effects.[7] The advantages of a panel framework over time series or cross-section regressions include increased efficiency with more degrees of freedom and less multicollinearity (Hsiao, 2014).

EMPIRICAL RESULTS

Global regression estimates

To study the impact of Covid-19 on US housing market conditions in the wake of the pandemic, we apply the alternative regression techniques outlined in the third section above to balanced monthly panels of ZIP code-level data between April and December 2020. The coronavirus began spreading across the United States in March. To allow for the time it took to affect the US housing market, the monthly local case rate (COVID), which represents the extent of contagion, enters the models with a one-month lag. December was the latest month for which housing data were available in this study.

Table 1 shows the estimation results of a total of 101,853 observations (11,317 ZIP codes × 9 months) for each of the three alternative measures of housing market outcomes: the number of online views of a typical home, the units of homes sold and the local median house price.[8] The dependent variables are expressed as the year-on-year per cent change in a ZIP code minus its metro area's year-on-year per cent change. In line with the study of changing preference for density by Ahlfeldt et al. (2015, p. 2154), this difference-in-differences specification is a popular approach in event studies, such as the aftermath of a disaster. Essentially, our regression models compare the relative change of a neighbourhood's housing market in the wake the pandemic.

Table 1. Panel data estimation results.

	Base		SAR		SAR-GWR		
	(1)		(2)		(3)		
					Minimum	Mean	Maximum
(a) *Online views*							
COVID	0.82	(0.30)	−4.06	(1.65)***	−63.86	−3.75	52.10
Distance × COVID	1.31	(10.93)*	0.33	(3.13)*	−17.27	0.06	22.93
Density × COVID	−0.05	(2.18)**	−0.05	(2.16)**	−15.43	−0.41	7.29
Jobs × COVID	−0.38	(4.05)**	−0.13	(1.54)	−18.16	0.40	22.21
Remote Work × COVID	2.15	(4.60)*	1.89	(4.51)*	−85.03	0.12	121.37
Restaurants × COVID	0.05	(5.87)*	0.04	(4.82)*	−10.94	0.15	5.60
Income × COVID	−1.04	(3.67)*	−0.64	(2.54)*	−39.07	1.83	47.65
2019 Price × COVID	−0.05	(2.38)**	−0.20	(1.78)***	−41.94	1.23	32.62
Spatial lag Wy			0.74	(15.19)*	−0.67	0.57	1.03
Spatial LR test			20563	*		26579	*
Adjusted R^2	0.30		0.44			0.52	0.30
Log-likelihood	−507,054		−496,772			−483,483	−507,054
(b) Sales: COVID	−2.25	(1.65)****	−2.30	(2.11)**	−31.67	−9.74	72.97
Distance × COVID	0.42	(2.76)*	0.06	(0.48)	−34.89	0.26	31.34
Density × COVID	−0.22	(3.79)*	−0.11	(2.23)**	−12.00	0.15	16.23
Jobs × COVID	0.04	(0.33)	−0.02	(0.23)	−26.15	−0.20	25.25
Remote Work × COVID	−0.51	(0.85)	0.60	(1.24)	−161.58	1.26	164.71
Restaurants × COVID	0.04	(3.26)*	0.03	(3.76)*	−29.91	−0.08	7.72
Income × COVID	0.48	(1.34)	−0.33	(1.13)	−70.10	1.39	95.00
2019 Price × COVID	−0.38	(2.44)*	−0.09	(0.69)	−42.52	−0.93	28.11
Spatial lag Wy			0.84	(21.06)*	−0.70	0.32	1.07
Spatial LR test			36,119*	*		18,048	*

Adjusted R^2	0.15		0.36			0.51	0.15
Log-likelihood	−528,441		−510,382			−501,358	−528,441
(c) Median Price: COVID	0.58	(4.74)*	0.29	(2.89)*	−18.95	1.05	24.68
Distance × COVID	0.17	(0.32)	0.80	(1.92)**	−1.52	0.05	0.80
Density × COVID	0.67	(1.25)	−0.06	(2.35)**	−0.70	−0.03	0.29
Jobs × COVID	−2.27	(5.41)*	−0.62	(1.78)***	−1.68	0.02	0.74
Remote Work × COVID	−8.11	(3.89)*	−1.38	(0.80)	−4.60	0.14	4.28
Restaurants × COVID	0.40	(2.43)*	0.19	(6.03)*	−0.29	0.01	0.31
Income × COVID	−2.02	(1.79)****	−2.07	(1.98)**	−2.08	−0.09	1.34
2019 Price × COVID	−0.94	(2.18)**	−0.13	(2.29)**	−1.12	−0.03	1.58
Spatial lag Wy			0.87	(20.82)*	−0.45	0.78	1.08
Spatial LR test			33,641*		13,815*		
Adjusted R^2	0.12		0.37			0.42	0.12
Log-likelihood	−225,380		−208,559			−201,852	−225,380

Note: The sample comprises monthly data of 11,317 ZIP code areas between April and December 2020. All specifications control for unobserved individual and time fixed effects. Numbers in parentheses are absolute t-statistics based on standard errors robust to clustering at the metro-area level. *, ** and ***Statistical significance at the 1%, 5% and 10% levels, respectively.

All panel regressions also control for individual- and time-fixed effects. As described in the second section above, the explanatory variables capture various local characteristics of interest. To capture interactions with the pandemic effects, the values of these variables, which are first expressed in natural logarithmic terms, are multiplied by the one-month lagged COVID variable.

For comparison purposes, column (1) of Table 1 displays the regression results without spatial effects. The coefficient estimates for Covid are statistically significant in the housing sales and price regressions. The positive estimate reflects rising local median housing prices compared with the metro area averages during the pandemic period. However, the results do not provide statistical support for any relationship between the relative median housing prices and the distance or density variable that interacts with the Covid-19 case rate.

In the regression for online viewing activity, all explanatory variables except Covid are statistically significant. These estimates suggest that, along with the interaction with the local caseloads, houses in neighbourhoods that are less densely populated or farther from a metro core received relatively more online viewings. Evidence for the changing behaviour among prospective homebuyers is also evident in the regression for home sales but not for home prices.

According to the estimation results in column (1) of Table 1, house viewings and prices tended to increase less among neighbourhoods where jobs were more plentiful. However, controlling for the pandemic impacts, prospective homebuyers browsed more online for houses in neighbourhoods with a greater share of remote-work-compatible jobs and more restaurant choices. This perhaps reflects the continued preference for living in areas with relatively more remote-work-compatible jobs, which are typically high-skilled white-collar occupations, and other perks. The negative coefficient estimates on the pre-pandemic house prices and income levels suggest a shift of housing demand toward more affordable neighbourhoods.

Column (2) of Table 1 lists regression results for the SAR model depicted by equation (2), which augments the benchmark model with a spatially lagged dependent variable. The contiguity-based spatial weight matrix was constructed with a median of six neighbouring ZIP codes. The coefficient estimates for the spatially lagged term range between 0.74 and 0.87, indicating that housing market conditions are similar among nearby neighbourhoods. A notable improvement in the adjusted R^2's supports the presence of spatial interdependence among local housing markets, as captured by the SAR specification. Based on the log-likelihood values of the base and SAR models, the high likelihood ratio (LR) statistics for testing spatial effects implies that the coefficient estimates without accounting for spatial dependence in the dependent variables may be biased associated with omitted variables.[9]

Because the spatial lag coefficient $\rho \neq 0$, the interpretation of the coefficients of explanatory variables is different from a conventional least-squares interpretation. The positive values of the SAR coefficient estimates imply a larger size of the estimate for an explanatory variable's 'total' effects, which consist of a ZIP code area's own or 'direct' effect and the spatial or 'indirect' effect of its neighbours (LeSage & Pace, 2009).

However, the results between our key variables of interest – distance and density – are mixed. For instance, given the spatial autocorrelation coefficient estimate of 0.74 in the regression for online views, the estimate for distance's 'total' effects is 1.27 (0.33/(1 – 0.74)), which is close to the corresponding estimate of 1.31 in the base model. By contrast, the corresponding estimate in the house price regression is remarkably larger when unobserved neighbourhood effects are taken into consideration (6.15 versus 0.17 in the base model). Evidence of spillover, or diffusion effects, in local housing markets is also well documented in the literature (Brady, 2011; Holly et al., 2011; Pace et al., 2000).

Column (3) of Table 1 shows the results of the SAR-GWR model that allows for unobserved spatial heterogeneity in addition to spatial dependence. The overall goodness-of-fit criteria in light of the adjusted R^2's and LR statistics (testing against the SAR specification) suggest probable bias in previous findings due to spatial variations across the sample. Improvement in the predictive accuracy of the SAR and GWR is well documented in the spatial literature (e.g. Bidanset & Lambard, 2014; Fotheringham et al., 2015; Geniaux & Martinetti, 2018).

The SAR-GWR model generates 'local' coefficient estimates for each ZIP code location. Table 1 lists their means along with the minimum and maximum values. Despite the means of those 'local' coefficient estimates that are close to their counterparts in columns (1) and (2), their ranges are remarkably wide. For instance, in the regression for online views, the coefficient estimates of the distance interactive term are 0.06 on average among all ZIP codes, but their values range between –17.27 and 22.93. This underscores the extent of heterogeneity across the local housing markets in the United States.

Local GWR model estimates

One advantage of the GWR model is the results for individual ZIP codes instead of the 'global' results for all observations together. Figures 2–5 show Choropleth maps of 'local' model estimates for individual ZIP codes. Figure 2 contains maps of local R^2's. For online views, R^2's tend to be higher among areas in the Western region along the Pacific coast. For house sales, relatively higher R^2's cluster in the Midwestern region near Chicago. For house prices, R^2's are relatively higher along the East and West Coasts, and across the state of Texas where new home construction and thus overall housing supply growth have been relatively limited. The comparative results across the three maps reflect the fact that online viewings of properties do not necessarily result in home purchases, and home prices are driven not only by demand but also by supply.

Figures 3–5 display the geographical distribution of 'total' local coefficient estimates for some key explanatory variables. Only coefficients that are statistically significant at the 10% level or better are visible. Each ZIP code's coefficient estimates for 'total' effects include estimates of that ZIP code's own (direct) effects and spatial spillover (indirect) effects. Figure 3 shows the local estimates on the Covid variable. For instance, for ZIP code 10001 inside New York City, the 'direct effect' coefficient estimate for online views is –30.55. Given its local estimate of 0.69 on the spatial lag, the corresponding 'total' coefficient estimate becomes –98.55 (–30.55/(1 – 0.69)).

In the online views and sales regressions, the 'total' coefficient estimates by ZIP code are quite evenly split between positive and negative entries. This reflects the corresponding statistically insignificant 'global' estimates in the non-GWR models, even though the average 'local' estimates from the SAR-GWR model are positive. For the pandemic's impact on online views and sales, negative local estimates are clustered largely near the East Coast between the states of New York and South Carolina.

In the home price regression, the 'local' estimate on Covid is positive on average (column (3) in Table 1). This positive mean reflects proportionally more positive than negative 'local' estimates, the latter of which are confined mostly along the East and West Coasts. Areas with positive estimates experienced home price increases despite relatively more local Covid-19 caseloads.

Figure 4 shows local estimated coefficients for the 'total' effects of the distance interactive variable. For all three housing market outcomes, coefficient estimates are higher in communities surrounding Washington DC in the Northeastern United States. In the online views and home price regressions, estimates are higher among some communities on the West Coast. In the Southern region, estimates of distance's effect are relatively high among Gulf of Mexico coastal communities between the states of Louisiana and South Carolina. The vast majority of local estimates are positive, meaning that housing demand and price increases were relatively stronger across neighbourhoods farther from city centres. In the regressions of online views and sales, negative estimates are evident near the city of Atlanta in Georgia and in parts of the New England region between Boston and New York City.

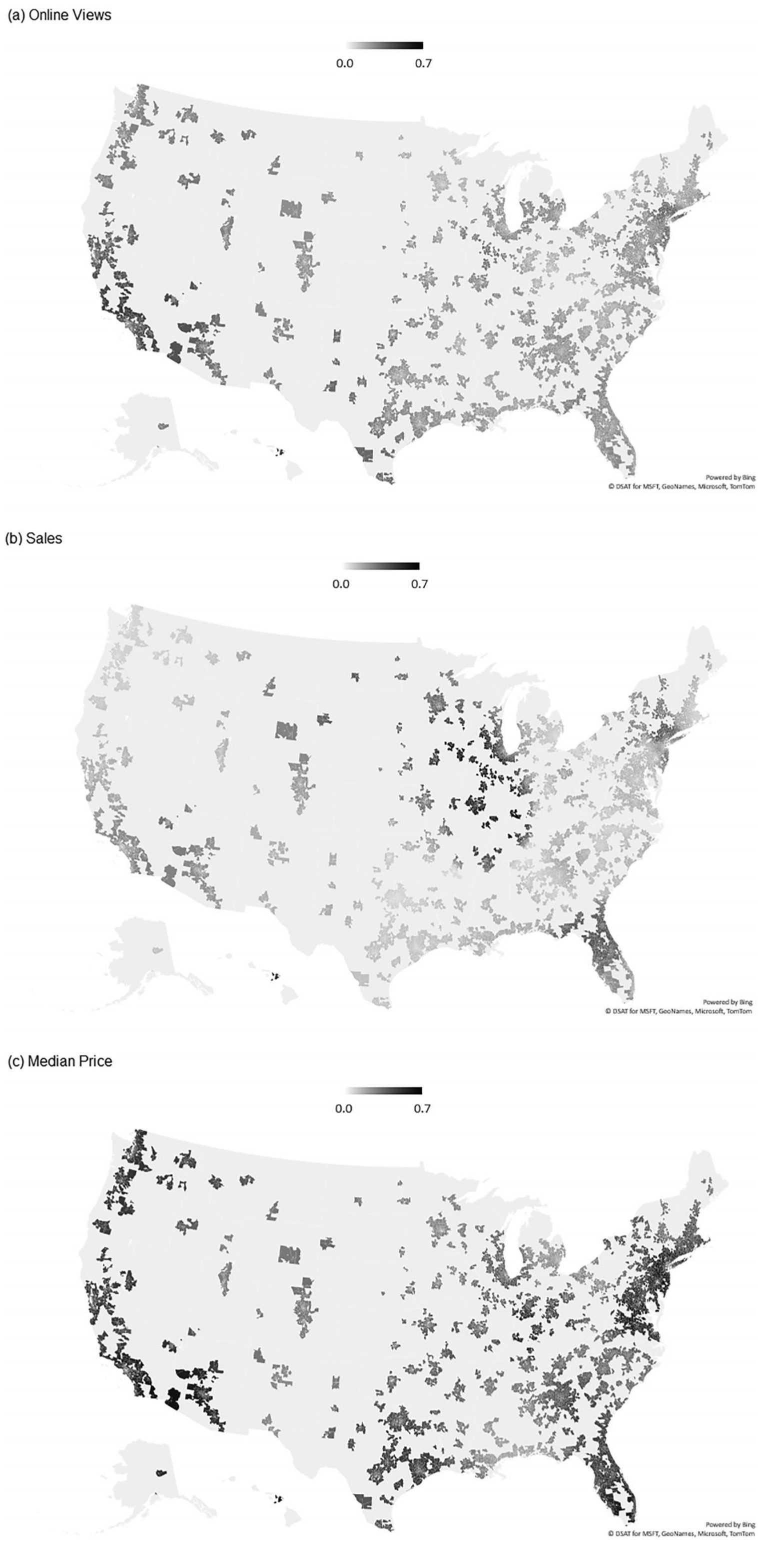

Figure 2. Local R^2's.

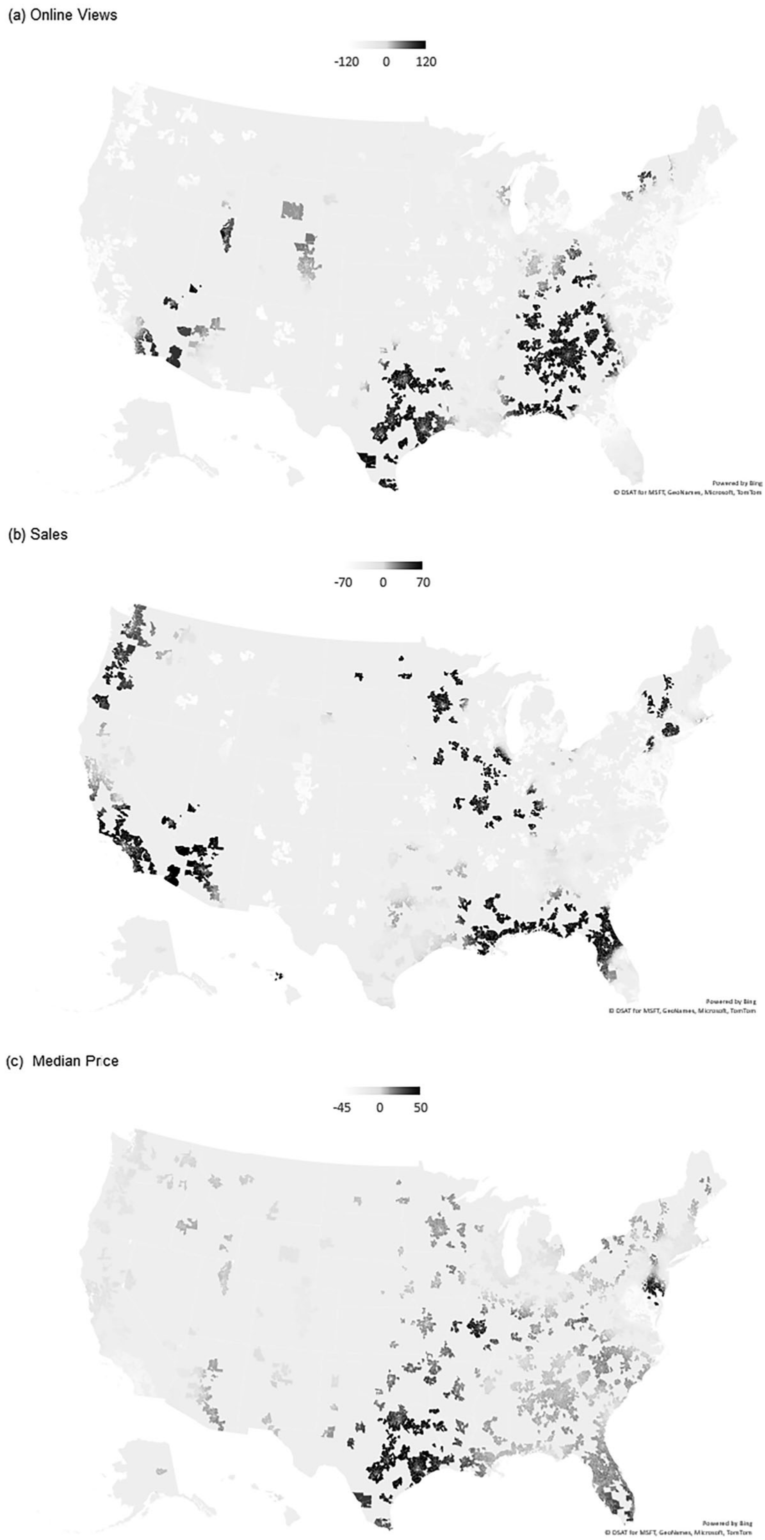

Figure 3. Local 'total' coefficient estimates of Covid.

Figure 4. Local 'total' coefficient estimates of Distance × COVID.

Figure 5. Local 'total' coefficient estimates of Density × COVID.

Figure 5 contains corresponding maps for the density interactive variable. For all three housing market outcomes, the local coefficient estimates are mostly negative as evidence in support of a shifting preference toward less densely populated areas. Their absolute sizes are larger among communities along the Atlantic Coast. The estimated impact of density on home sales also tends to be larger among areas in Southern California on the West Coast.

In the online views and sales regressions, the coefficient estimates are positive among some ZIP codes in New York City's outskirts. For house sales, positive coefficient estimates also appear among some neighbourhoods in the state of Illinois south of Chicago. Despite relatively higher population density than the national average, housing demand and thus house sales in those suburbs, or bedroom communities, away from downtown Manhattan or Chicago were rising.

Figures 2–5 together reveal varying housing market conditions across broad regions of the United States. Despite the observed regional patterns, the focus of this study is disparity *within* a region. As the SAR-GWR 'local' model estimates indicate, some local housing markets were affected more disproportionately than others by the pandemic, but some areas or submarkets even experienced changing market conditions opposite to the nationwide market captured by the 'global' model.

To illustrate spatial variation within a region, Figure 6 provides close-up views of the local estimates for the 'total' effects of the distance interactive variable in the Northeastern region surrounding New York City. Except for Boston, the coefficient estimates are smaller or statistically indifferent from zero at the core of major metro areas, such as New York City, Philadelphia, Baltimore and Washington DC. While estimates tend to be larger for those metro areas' outlying communities, the estimates for home sales and prices are instead negative in some neighbourhoods outside Philadelphia and New York City. These disparate patterns essentially reflect the nonlinear relationships of spatially correlated variables.

CONCLUSIONS

We have empirically investigated the conjecture that urban residents have responded to Covid-19 outbreaks in the United States by fleeing city centres for the suburbs. Regression results of our 'global' models indicate that Liu and Su's (2020) overall finding of an urban flight at the onset of the pandemic prevailed through the end of 2020. Our findings also help explain the scepticism about the prevalence of the changing household behaviour nationwide (Handbury, 2020) by adding a spatial or geographical perspective to the nationwide findings.

According to the traditional regression model applied to ZIP code-level data, online viewings of homes and home sales have grown more rapidly among neighbourhoods that are less densely populated or farther from a metro area's downtown. However, corresponding evidence based on home prices as alternative market outcomes is absent. Ignoring spatial dependence and spatial heterogeneity captured by the SAR-GWR framework, conventional regression models might have generated biased inferences on the impact of Covid-19 on local US housing markets.

Empirical results become more robust across alternative measures of market outcomes when regressions allow for both spatial interactions and intrinsic characteristics that are unevenly distributed over space. The finding of a so-called spatial lag effect among ZIP codes supports the spillover or neighbourhood effects across adjacent local housing markets; the finding of spatial heterogeneity underscores the extent of disparities among subregions of the national housing market.

Global regressions mask valuable information in local units or regions. Housing demand surged disproportionately in the East and West Coasts during the pandemic. Without corresponding increases in home listings from sellers and thus homes sold, home prices appreciated more in those regions than the rest of the nation. Home prices tended to rise even more

Figure 6. Local 'total' coefficient estimates of Distance × COVID for the Northeastern United States.

among suburban neighbourhoods than city downtowns. Still, the regional nature of housing markets within a nation should make us careful not to overgeneralize.

The Covid-19 pandemic has underscored the importance of people's mobility and thus the spatial aspect of our economy. Spatial analysis helps us better understand changing housing market conditions across the United States. Still, the workhorse of our study may benefit from additional extensions in future research. For instance, in local housing markets, spatial interactions are likely to be dynamic in nature in the sense that the housing market condition in one location affects the housing market of its nearby locations in subsequent periods. Even though it is too early to tell whether the pandemic will have a lasting or irreversible impact on the US housing markets, spatial regression that also accounts for temporal variations in model coefficients can shed light on changing market conditions at least in the short run.

From this perspective, a panel model that allows for time-varying coefficients might help us better understand how the pandemic has affected house prices, which seem to be slower to adjust over time than other market outcomes. In the cross-section setting, local effects in both space and time have been captured by geographically and temporally weighted regression (GTWR) as an extension to GWR (Fotheringham et al., 2015; Huang et al., 2010; Wu et al., 2014). It would, therefore, be fruitful to apply GTWR to panel data. Incorporating the temporal perspective would potentially help us understand how changing housing market conditions would evolve through different stages of the pandemic.

As evident in Figures 2–6, evidence on the regional or submarket patterns of housing markets in the United States suggests that some communities share similar socioeconomic characteristics. Clearly, GWR is inadequate for variables that have global effects and are independent from individual locations. This motivates the consideration of clusters in covariate effects as another extension to our framework for spatial autocorrelation and heterogeneity, as suggested by Ma et al. (2020). This extension would potentially help us better identify submarkets within a nation.

NOTES

[1] The sample covers more than 380 US metro areas and matches the US Census Bureau's 2019 Rural–Urban Commuting Area (RUCA) definition of urban and suburban areas that have at least 10% of commuting flows to an urbanized area or metro core. The dataset excludes rural areas for which our online databases have limited data observations.

[2] Regression results are nevertheless robust to either distance measure.

[3] We have also considered other socioeconomic variables, such as the crime rate and the minority population share. Their estimation results are not statistically meaningful and thus are not included in the results reported in this paper.

[4] An alternative class of spatial weight is based on distance instead of contiguity, but a preliminary analysis indicates that the contiguity-based weights fit our data substantially better than do distance-based weights.

[5] GWR assumes that all processes in the model operate at the same spatial scale or bandwidth. However, some processes may operate over a local scale, while other processes operate over a broader, regional scale. Fotheringham et al. (2017) have relaxed this assumption with a multiscale GWR (MGWR), which allows different processes to operate at different spatial scales or bandwidths and thus a spatial weighting matrix for each coefficient of a particular location. Despite substantially more computationally intensive, preliminary analysis of our data with MGWR offers no improvement in the goodness-of-fit statistics (adjusted R^2 and Akaike information criterion – AIC) over GWR.

[6] Alternatively, the 'fixed' bandwidth determines a fixed distance for each location by allowing the number of nearest neighbours to vary. Because of the uneven distribution of our ZIP code-

level data over space, the adaptive bandwidth has the advantage that each regression point i has an identical amount of local data points for local coefficient estimates.

[7] Panel regressions of SAR and SAR-GWR are run with an extension to the MGWRSAR package (Geniaux & Martinetti, 2021) in the statistical software R with the PLM routine, as described by Geniaux and Martinetti (2018) and Yu (2010).

[8] In addition to the three market outcomes, we performed regressions for the number of home listings, but most results are not statistically meaningful and thus are not reported in this paper.

[9] In addition to the LR tests and the t-statistics for testing the null hypothesis H_0: $\rho = 0$ in the SAR models, the popular Moran's I and Lagrange multiplier (LM) tests (Anselin, 1988) on regression residuals provide strong support for the presence of spatial autocorrelation.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

ORCID

Jim Lee http://orcid.org/0000-0002-9830-3012
Yuxia Huang http://orcid.org/0000-0002-8042-8992

REFERENCES

Ahlfeldt, G. M., Redding, S. J., Sturm, D. M., & Wolf, N. (2015). The economics of density: Evidence from the Berlin Wall. *Econometrica*, *83*(6), 2127–2189. https://doi.org/10.3982/ECTA10876

Anselin, L. (1988). *Spatial econometrics: Methods and models*. Kluwer.

Basile, R., & Minguez, R. (2018). Advances in spatial econometrics: Parametric vs. semiparametric spatial autoregressive models. In P. Commendatore, I. Kubin, S. Bougheas, A. Kirman, M. Kopel, & G. I. Bischi (Eds.), *The economy as a complex spatial system* (pp. 81–106). Springer Proceedings in Complexity.

Basile, R., Minguez, R., Montero, J. M., & Mur, J. (2014). Modelling regional economic dynamics: Spatial dependence, spatial heterogeneity and nonlinearities. *Journal of Economic Dynamics and Control*, *48*, 229–245. https://doi.org/10.1016/j.jedc.2014.06.011

Bhattacharjee, A., Castro, E., Maiti, T., & Marques, J. (2016). Endogenous spatial regression and delineation of submarkets: A new framework with application to housing markets. *Journal of Applied Econometrics*, *31*(1), 32–57. https://doi.org/10.1002/jae.2478

Bhattacharjee, A., Castro, E., & Marques, J. (2012). Spatial interactions in hedonic pricing models: The urban housing market of Aveiro, Portugal. *Spatial Economic Analysis*, *7*(1), 133–167. https://doi.org/10.1080/17421772.2011.647058

Bidanset, P. E., & Lambard, J. R. (2014). 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. https://doi.org/10.1080/24694452.2017.1352480

Brady, R. (2011). Measuring the diffusion of housing prices across space and over time. *Journal of Applied Econometrics*, *26*(2), 213–231. https://doi.org/10.1002/jae.1118

Brunsdon, C., Fotheringham, A. S., & Charlton, M. (1999). Some notes on parametric significance tests for geographically weighted regression. *Journal of Regional Science*, *39*(3), 497–524. https://doi.org/10.1111/0022-4146.00146

Campisi, N. (2020, October 23). Manhattan real estate fizzles as prices, demand plummets. Forbes.com. https://www.forbes.com/sites/advisor/2020/10/23/manhattan-real-estate-fizzles-as-prices-demand-plummets/

Dingel, J. I., & Neiman, B. (2020). How many jobs can be done at home? *Journal of Public Economics*, *189*, 104235. https://doi.org/10.1016/j.jpubeco.2020.104235

Elhorst, J. P. (2009). Spatial panel data models. In M. M. Fischer, & A. Getis (Eds.), *Handbook of applied spatial analysis* (pp. 377–407). Springer.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically weighted regression: The analysis of spatially varying relationships*. John Wiley and Sons.

Fotheringham, A. S., Crespo, R., & Yao, J. (2015). Geographical and temporal weighted regression (GTWR). *Geographical Analysis*, *47*(4), 431–452. https://doi.org/10.1111/gean.12071

Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale geographically weighted regression (MGWR). *Annals of the American Association of Geographers*, *107*(6), 1–19. https://doi.org/10.1080/24694452.2017.1352480

Fry, R. (2020, July 29). Prior to COVID-19, urban core counties in the U.S. were gaining vitality on key measures. *Social & Demographic Trends*, Pew Research Center. https://www.pewsocialtrends.org/2020/07/29/prior-to-covid-19-urban-core-counties-in-the-u-s-were-gaining-vitality-on-key-measures/

Geniaux, G., & Martinetti, D. (2018). A new method for dealing simultaneously with spatial autocorrelation and spatial heterogeneity in regression models. *Regional Science and Urban Economics*, *72*, 74–85. https://doi.org/10.1016/j.regsciurbeco.2017.04.001

Geniaux, G., & Martinetti, D. (2021). GWR and MGWR with spatial autocorrelation. https://cran.r-project.org/web/packages/mgwrsar/mgwrsar.pdf

Glaeser, E. L., & Gottlieb, J. D. (2009). The wealth of cities: Agglomeration economies and spatial equilibrium in the United States. *Journal of Economic Literature*, *47*(4), 983–1028. https://doi.org/10.1257/jel.47.4.983

Grogan, T. (2020, October 29). Surprising changes to traffic congestion in urban areas. Streetlightdata.com. https://www.streetlightdata.com/changes-to-traffic-congestion-in-urban-areas/

Handbury, J. (2020, December 1). COVID-19 and the myth of urban flight. *Wharton Business Daily*. https://knowledge.wharton.upenn.edu/article/covid-19-and-the-myth-of-urban-flight/

Holly, S., Pesaran, M. H., & Yamagata, T. (2011). The spatial and temporal diffusion of house prices in the UK. *Journal of Urban Economics*, *69*(1), 2–23. https://doi.org/10.1016/j.jue.2010.08.002

Hsiao, C. (2014). *Analysis of panel data* (3rd ed.). Cambridge University Press.

Huang, B., Wu, B., & Barry, M. (2010). Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, *24*(3), 383–401. https://doi.org/10.1080/13658810802672469

Ioannides, Y. M., & Zabel, J. E. (2003). Neighbourhood effects and housing demand. *Journal of Applied Econometrics*, *18*(5), 563–584. https://doi.org/10.1002/jae.740

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*(1), 99–121. https://doi.org/10.1023/A:1007707430416

Kuethe, T. H., & Pede, V. O. (2011). Regional housing price cycles: A spatio-temporal analysis using US state-level data. *Regional Studies*, *45*(5), 563–574. https://doi.org/10.1080/00343400903497897

LeSage, J., & Pace, K. (2009). *Introduction to spatial econometrics*. SRC Press.

Li, S., Ye, X., Lee, J., Gong, J., & Qin, C. (2017). Spatiotemporal analysis of housing prices in China: A big data perspective. *Applied Spatial Analysis and Policy*, *10*(3), 421–433. https://doi.org/10.1007/s12061-016-9185-3

Lin, Z. (2011). ML estimation of spatial panel data geographically weighted regression model. 2011 *International Conference on Management and Service Science*, Wuhan, 2011, 1–4.

Liu, S., & Su, Y. (2020). The impact of the COVID-19 pandemic on the demand for density: Evidence from the U.S. housing market. Federal Reserve Bank of Dallas, Research Department Working Papers, No. 2024.

Lu, B., Brunsdon, C., Charlton, M., & Harris, P. (2017). Geographically weighted regression with parameter-specific distance metrics. *International Journal of Geographical Information Science*, *31*(5), 982–998. https://doi.org/10.1080/13658816.2016.1263731

Ma, Z., Xue, Y., & Hu, G. (2020). Heterogeneous regression models for clusters of spatial dependent data. *Spatial Economic Analysis*, *15*(4), 459–475. https://doi.org/10.1080/17421772.2020.1784989

McMillen, D. P. (1996). One hundred fifty years of land values in Chicago: A nonparametric approach. *Journal of Urban Economics*, *40*(1), 100–124. https://doi.org/10.1006/juec.1996.0025

Pace, R. K., Barry, R., Gilley, O. W., & Sirmans, C. F. (2000). A method for spatial–temporal forecasting with an application to real estate prices. *International Journal of Forecasting*, *16*(2), 229–246. https://doi.org/10.1016/S0169-2070(99)00047-3

Pace, R. K., & Gilley, O. W. (1997). Using the spatial configuration of the data to improve estimation. *Journal of the Real Estate Finance and Economics*, *14*(3), 333–340. https://doi.org/10.1023/A:1007762613901

Speianu, S. (2020, September 24). Urban vs suburban growth report: The grass is greener in the suburbs. Realtor.com. https://www.realtor.com/research/urban-vs-suburban-growth-report-september-2020

Veuger, S., Brooks, L., & Begley, J. (2020, May 7). What jobs, location, and telework can tell us about the coronavirus in the region. American Enterprise Institute. https://www.aei.org/articles/what-jobs-location-and-telework-can-tell-us-about-the-coronavirus-in-the-region/

Wrenn, D. H., & Sam, A. G. (2014). Geographically and temporally weighted likelihood regression: Exploring the spatiotemporal determinants of land use change. *Regional Science and Urban Economics*, *44*, 60–74. https://doi.org/10.1016/j.regsciurbeco.2013.10.005

Wu, B., Li, R., & Huang, B. (2014). A geographically and temporally weighted autoregressive model with application to housing prices. *International Journal of Geographical Information Science*, *28*(5), 1186–1204. https://doi.org/10.1080/13658816.2013.878463

Yu, D. (2010). Exploring spatiotemporally varying regressed relationships: The geographically weighted panel regression analysis. In E. Guilbert, B. Lees, & Y. Leung (Eds.), *International archives of the photogrammetry, remote sensing and spatial information sciences* (Vol. 38, pp. 134–139). International Society for Photogrammetry and Remote Sensing.

Zhao, Y. (2020). The US housing market during COVID-19: Aggregate and distributional evidence, International Monetary Fund, Working Paper, WP/20/212. Center for Economic Policy Research Press. https://cepr.org/file/9694/download?token=lBYLZegm