Spatial Market Inefficiency in Housing Market:

A Spatial Quantile Regression Approach *

Jiyoung Chae † Anil K. Bera ‡

September 28, 2019

Abstract

This paper empirically tests housing market efficiency in the spatial dimension by using the spatial autoregressive conditional heteroskedastic (ARCH) and spatial quantile regression models. The tests were conducted in terms of both housing returns and squared returns (volatility). The sale price data used is from Cook County residential MLS for the years 2010-2016. The main findings are: housing returns are not spatially correlated but squared returns are spatially correlated, and the spatial dependence of squared returns seems to be stronger for higher squared return quantiles.

Keywords: Housing market, Market efficiency, Spatial dependence, Spatial volatility clustering, Spatial quantile regression.

JEL Classification: G14, R23, R31, C21

*We are grateful to Osman Dogan, Geoffrey J. D. Hewings, Daniel P. McMillen and Suleyman Taspinar for helpful discussions and many pertinent suggestions on an earlier version of the paper; though the remaining shortcomings are solely ours. We thank to Illinois REALTORS for financial support to the Regional Economics Applications Laboratory that provided research funding for Jiyoung Chae.

†Department of Economics, University of Illinois at Urbana-Champaign, Urbana, IL, USA, email: jchae3@illinois.edu.

‡Department of Economics, University of Illinois at Urbana-Champaign, Urbana, IL, USA, email: abera@illinois.edu.

1 Introduction

The efficient market hypothesis (EMH), popularly known as the random walk theory, implies that future prices cannot be predicted by analyzing the past price movements because all historical information is fully incorporated in the current prices (Fama, 1970). There has been a great deal of empirical research on the EMH, but overall, the empirical evidence still remains inconclusive. Also, much of the research is focused on the efficiency of the stock market (Sewell, 2011). The issue of housing market efficiency became especially important after it was found that excess volatility of housing prices in OECD countries has been increasing over time (Bracke, 2013; Claessens et al., 2011) and consequently housing markets are becoming less stable than in the past. Nevertheless, there is a general consensus that the housing market is inefficient and thus predictable with respect to some information. It has been frequently argued that the housing market inefficiency results from its specific features, such as, high transaction costs, infrequent transactions, and a high degree of heterogeneity. Thus information does not diffuse promptly in the market at a sufficient depth and consequently, the prices do not adjust accordingly. While the test methods and the market studied may differ, previous empirical studies generally find that housing returns (or price changes) exhibit positive serial correlation, implying predictability of returns and, consequently housing market inefficiency (Gau, 1984; Linneman, 1986; Guntermann and Smith, 1987; Rayburn et al., 1987; Case and Shiller, 1989; Gyourko and Voith, 1992; Kuo, 1996; Gu, 2002; Schindler, 2013).

However, despite the substantial body of research investigating the market efficiency in the time series context, very little has been written on *spatial* market efficiency even though an important and distinguishing characteristic of real estate is its spatial dimension (Tirtiroglu, 1992). If markets are spatially efficient, then the returns from neighborhood locations should not be informative in predicting the returns of other locations (Meen, 2012). On the other hand, if there is any indication of inefficiency in the housing markets, housing returns in one area could contain useful and valuable information for predicting the returns of neighboring houses. The literature on U.S. housing market suggests that the latter is the case. Tirtiroglu (1992) and Clapp and Tirtiroglu (1994) find evidence of a spatial diffusion process within the Connecticut metropolitan housing markets, whereby price changes of houses in a given sub-market affect changes in contiguous markets. Pollakowski and Ray (1997) also examine the spatial interactions in housing prices across the U.S. census divisions and primary metropolitan statistical areas (MSA). Using a vector autoregressive (VAR) approach, they demonstrate that the market is inefficient and that contiguous regions transmit more influence than non-contiguous regions. Gupta and Miller (2012) consider the issue of diffusion in eight Southern California MSAs and report substantial evidence of temporal causality between them. Other studies have focused on analyzing both return and volatility spillovers in the U.S. housing market, for instance, see Miao et al. (2011) and Zhu et al. (2013). Nevertheless, research on spatial relations in the U.S. regional housing market has been relatively sparse compared to that in the U.K.

The main objective of this paper is to test the spatial market efficiency in the U.S. housing market by investigating the behavior of spatial dependence in housing returns, which is analogous to testing for dependence in the time series context. In particular, we focus on nonlinear behavior of the spatial dependence in the returns using their squared terms since, in empirical finance, even though there is lack of serial correlation in the level of returns, the autocorrelation in the squared returns is found to be significant. In the time series literature, this empirical fact led to the autoregressive conditional heteroskedastic (ARCH) model, proposed by Engle (1982) and its various generalizations; for a survey of these models, see, Bera and Higgins (1993).

Note that squared returns are often used to proxy the actual volatility. Therefore high serial correlations of squared returns indicates that volatility is serially correlated and therefore predictable. This strong serial correlation in squared returns leads to the stylized fact of volatility clustering. The implication of such volatility clustering is that volatility shocks today will affect the expectation of volatility in a future period. The absence of linear autocorrelation but the presence of dependence in squared returns in the time series context naturally suggests that the similar possibility in the spatial context needs to be investigated. In fact, Meen (1999) argue that the spatial dependence in house price variations may arise due to four possible factors. The first is migration, which means that if house prices in one area are high relative to the other regions, then households might migrate to those regions, leading to equalization in the house prices. The second is equity transfer, which is closely related to migration and suggests that households in regions with higher house prices would have greater buying power, leading to higher prices in the other regions if these households would want to move. The third is spatial arbitrage whereby if new information becomes available in one area, this information is transmitted first to contiguous areas, thus allowing investors to acquire properties in lower priced regions, where higher anticipated return on housing investment exist. Finally, the spatial patterns in determinants of prices can simply induce spatial dependence in house prices, even if there is no spatial link. For example, two regions with similar economic conditions can affect one another's prices.

Accommodating this expectation, we investigate the possible spatial dependence in housing returns and suggest that if there is any form of spatial dependence in the returns, regardless of whether it is linear or nonlinear, the market is spatially inefficient and thus returns from the neighboring houses or neighboring areas can be used to predict the returns and/or volatility at given locations. This argument is carried further, suggesting that the market inefficiency would vary at different points of the distribution, i.e. at different quantiles. Our expectation is based upon various findings in the theoretical as well as the empirical finance literature. For example, Veronesi (1999) proposes a rational expectations equilibrium model and finds that prices overreact to bad news in good times and underreact to good news in bad times. Baur et al. (2012) argue that the state dependence can be captured by the different quantiles of the conditional return distribution: upper (lower) quantiles are associated with good (bad) states. Hence, they use quantile autoregressive model, constructed under the framework of quantile regression (Koenker and Bassett, 1978), to study the predictability of stock returns and find that lower quantiles exhibit positive dependence on past returns, while upper quantiles are marked by negative dependence. Several other recent studies have also examined the relation between lagged returns and current returns using quantile regression approaches (Chiang et al., 2010; Ma and Pohlman, 2008; Tsai, 2012). This motivates us to employ the *spatial* quantile regression (SQR) to explore the spatial dependence of housing returns and volatility across different quantiles.

In short, this study contributes to studies of spatial dependence in housing returns and volatility within the framework of housing market efficiency by employing the spatial ARCH and SQR models. Most of the studies in this branch has overwhelmingly focused on evaluating the spatial effects on the first two conditional moments of return distributions, while ignoring other parts of the distributions. There have been several attempts on the empirical analysis of SQR in the housing literature; however, most researchers focus on examining the determinants of housing prices or land prices (Zietz et al., 2008; Kostov, 2009; Liao and Wang, 2012; McMillen, 2012).

For the empirical justification of our formulation, we focus on the Cook County's housing market, the largest and most diverse market in the Chicago metropolitan area, and use the residential property sales price data for the period 2010-2016. Our study reveals a number of new insights into the spatial market efficiency of the housing market. Specifically we find i) while housing returns are not correlated over space, squared returns, which represent volatility, exhibit significant spatial dependence, i.e., spatial market inefficiency and therefore the neighborhood housing returns contain information for spatial prediction, ii) the degree of inefficiency varies over quantiles; the spatial dependence is conspicuously distinct from the lower quantiles to the higher quantiles with a gradually increasing trend, iii) income is negatively associated with squared returns, implying that higher household income leads to lower volatility of housing returns, and represents a gradual decrease with higher quantiles.

The rest of this paper is organized as follows. Section 2 gives an overview of the EMH and related literature. The fair game model and random walk model are defined and discussed to provide the background on testing the spatial market efficiency. Section 3 provides the spatial market efficiency testing framework and Monte Carlo study of the finite sample properties of the proposed test is presented in Section 4. The data are described in Section 5, and the empirical results are reported in Section 6. Finally, in Section 7, we conclude the paper with suggestions for future research.

2 Theoretical Background

2.1 The Efficient Market Hypothesis (EMH)

Fama (1970) defines an efficient market as one in which prices always "fully reflect" all available information. We will focus on two models of market efficiency, the Fair Game Model and the Random Walk Model. It is also necessary to specify what subset of available information is to be "fully reflected" in prices. The classic taxonomy of information sets, due to Roberts (1967), and used by Fama (1970) consists of the following:

- *Weak form* efficiency: The information set includes only historical prices of returns.
- *Semi-strong form* efficiency: The information set includes all publicly available information.
- *Strong form* efficiency: The information set includes all publicly and privately available information.

In this study, we will be concerned only with the weak form efficiency since the objective of the paper is to investigate the *spatial* dependence of housing returns on the neighboring values, which is analogous to examining the dependence of stock returns from the time series of historical returns.

2.1.1 The Fair Game Model

The fair game model is based on the assumption that the conditions of market equilibrium can be stated in terms of expected returns that are formed on the basis of the information set at time t, denoted by Φ_t. Formally,

$$E(P_{i,t+1}|\Phi_t) = [1 + E(r_{i,t+1}|\Phi_t)]P_{i,t}, \tag{1}$$

where $P_{i,t}$ is the price of security i at time t and $E(r_{i,t+1}|\Phi_t)$ is the next period expected return conditional on Φ_t, reflecting the full utilization of the available information. The major empirical implication expressed in (1) is that there is no trading system that achieves excess periodic returns above expectations. Let $z_{i,t+1}$ denote the excess return on security i, i.e.,

$$z_{i,t+1} = r_{i,t+1} - E(r_{i,t+1}), \tag{2}$$

then the sequence of excess returns is a fair game with respect to the information set Φ_t if and only if

$$E(z_{i,t+1}|\Phi_t) = 0, \tag{3}$$

which implies that on average the excess return is zero. Thus, an investor may experience large gains or losses relative to the equilibrium expected return $E(r_{i,t+1})$ in specific periods, but these average out to zero over time.

2.1.2 The Random Walk Model

While the fair game theory assumes that all information is incorporated in expectations, the random walk model is a somewhat extreme variant of the EMH. Here the successive price changes are independent and identically distributed. Letting $f(\cdot)$ denote the density function, the random walk model can be expressed as

$$f(r_{i,t+1}|\Phi_t) = f(r_{i,t+1}), \tag{4}$$

which states that the conditional and marginal probability distributions of future returns are the same and the function $f(\cdot)$ is invariant over time. It also implies that the serial covariance of returns is zero at all leads and lags, and the expected return is equal to the unconditional mean of the distribution $f(\cdot)$ at all times. If $r_{i,t}$ denotes the log return, i.e., $r_{i,t} = p_{i,t} - p_{i,t-1}$, where $p_{i,t} \equiv \log P_{i,t}$, then the log price process, $p_{i,t}$, can be written as

$$p_{i,t} = \mu + p_{i,t-1} + \varepsilon_{i,t}, \tag{5}$$

where $\varepsilon_{i,t}$ is an iid zero-mean random variable and μ denotes the drift parameter. Equation (5) is the most common form of the random walk for stock prices used in the finance literature. Using (5), it is easy to see that under the random walk framework, the condition (3) is satisfied, i.e.,

$$E(z_{i,t+1}|\Phi_t) = \mu - \mu = 0. \tag{6}$$

However as mentioned earlier, the random walk model is more restrictive than the fair game model since it requires successive returns to be independently identically distributed. Fama (1970) thus concludes that empirical tests of the random walk model are more powerful in support of the EMH than tests of the fair game model.

2.2 A Random Walk Test

The simplest model that can be used to test for the random walk in equation (5) is given by, after suppressing the subscript i,

$$r_t = \beta_0 + \beta_1 r_{t-1} + \varepsilon_t, \tag{7}$$

where r_t is the log return, β_0 and β_1 are parameters to be estimated, and $\varepsilon_t \sim iid(0, \sigma_\varepsilon^2)$. If the price follows a random walk, then $\beta_1 = 0$ and thus

$$p_t = \beta_0 + p_{t-1} + \varepsilon_t, \tag{8}$$

the random walk with drift parameter β_0.

Note that equation (7) has constant parameters and the error terms are assumed to follow the usual classical assumptions. With financial markets, the assumption of constant variance may be inappropriate as empirical evidence frequently finds that returns exhibit time-varying conditional variance (volatility). Studies by Park and Bera (1987), Bera, Bubnys, et al. (1988), Bollerslev et al. (1988), Pagan and Schwert (1990), and Nelson (1991) have identified the existence of a significant autoregressive structure in the conditional variance in financial data, particularly in stock returns. Those findings had been extensively used in the literature as evidence against the hypothesis of the standard conditional homoskedastic random walk model. Hence accepting the null hypothesis $H_0 : \beta_1 = 0$ in (7) does not necessarily imply that the market is inefficient.

2.3 Time-Varying Volatility Specification

The ARCH models were first introduced by Engle (1982). An ARCH(q) model for the return series $\{r_t\}$ is expressed as

$$r_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim iid(0, 1), \tag{9}$$

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{q} \alpha_j r_{t-j}^2. \tag{10}$$

Denoting $\Phi_t = \sigma(r_{t-1}, r_{t-2}, ...)$ as the sigma field generated by past information at time t, we have

$$\mathrm{Var}(r_t|\Phi_t) = E(r_t^2|\Phi_t) = E(\sigma_t^2 \epsilon_t^2|\Phi_t) = \sigma_t^2 E(\epsilon_t^2|r_{t-1}, r_{t-2}, ...) = \sigma_t^2, \tag{11}$$

which by definition is function of history. Therefore, although r_t is serially uncorrelated and thus cannot be predicted using its history, which is the evidence of the EMH, the conditional variance of r_t can be predicted.

Now we provide a statistical justification for the ARCH-type models. Consider the following random coefficient version of equation (7) by assuming $\beta_1 = \bar{\beta}_1 + \eta_t$ with $\eta_t \sim iid(0, \sigma_\eta^2)$ and being independent of ε_t. Then

$$\begin{aligned} r_t &= \beta_0 + \beta_1 r_{t-1} + \varepsilon_t \\ &= \beta_0 + (\bar{\beta}_1 + \eta_t) r_{t-1} + \varepsilon_t \\ &= \beta_0 + \bar{\beta}_1 r_{t-1} + \eta_t r_{t-1} + \varepsilon_t. \end{aligned} \tag{12}$$

The first two conditional moments of r_t, conditional on the past information Φ_t, are

$$E(r_t|\Phi_t) = \bar{\beta}_1 r_{t-1}, \tag{13}$$

$$\mathrm{Var}(r_t|\Phi_t) = \sigma_\varepsilon^2 + \sigma_\eta^2 r_{t-1}^2, \tag{14}$$

assuming $\varepsilon_t \sim iid(0, \sigma_\varepsilon^2)$.

Denoting $\sigma_\varepsilon^2 = \alpha_0$ and $\sigma_\eta^2 = \alpha_1$, the equation (14) becomes

$$\mathrm{Var}(r_t|\Phi_t) = \alpha_0 + \alpha_1 r_{t-1}^2, \tag{15}$$

which has the same form as (10) with $q = 1$, i.e., ARCH(1) process. If we further assume normality of η_t and ε_t, then

$$r_t|\Phi_t \sim N(0, \sigma_t^2), \tag{16}$$

which is generally used to estimate the ARCH parameters by maximum likelihood method. The above simple framework illustrates that a *nonlinear* ARCH model could be a manifestation of the time variation of the parameter(s) of a linear model. In spatial data, heterogeneity over space is an ubiquitous phenomenon, thus in spatial context, we can expect space varying nature of the underlying spatial models such as the spatial autoregressive (SAR) model or spatial error model (SEM), leading to spatial ARCH (SARCH)-type models.

3 Spatial Market Efficiency

3.1 Model Specification

From the previous section it follows that a test of spatial market efficiency can be formulated as a test based on the excess return at location i conditional on its neighbors. With a similar notation, let z_i denote the excess return at location i and $\Phi_{j\neq i}$ denote location i's j neighbors' information set, then the spatial counterpart of (3) is

$$E(z_i|\Phi_{j\neq i}) = 0. \tag{17}$$

Then a natural extension of the simple AR model given in (7) to the spatial context is the following SAR model.

$$r_i = \rho \sum_{j\neq i}^{n} w_{ij} r_j + \varepsilon_i, \tag{18}$$

where w_{ij} is the ij-th element of the spatial weight matrix W. In (18), we have put $\beta_0 = 0$ since without loss of generality the population mean of the excess return r_i can be assumed to be zero. The spatial weight matrix specifies the structure of neighborhood relationships, and different forms of W can be employed, for example, one based on contiguity, inverse distance, k-nearest neighbors, or some other scheme. In our application, the simplest and one of the most commonly used contiguity matrix, queen contiguity, is used.[1] Here, w_{ij} is equal to one when observations i and j share a common border or vertex, and a zero value, otherwise. By construction, the diagonal elements of the matrix, w_{ii} are set to zero. Also, for ease of interpretation, the matrix is defined in the row-standardized form, in which the row elements sum to one. The distinguishing feature of the SAR process rests on the specification of the expected returns at location i as a function of the returns of its neighbors as specified by the W matrix, and thus the spatial market efficiency can be tested by testing the null hypothesis $H_0 : \rho = 0$, in (18).

3.2 Spatial ARCH (SARCH) Specification

We should note that accepting the null hypothesis that $\beta_1 = 0$ in (7) does not mean that the market is efficient as the conditional variance can depend on the past information. For spatial data we can expect the presence of conditional heteroskedasticity depending on neighborhood observations. For a simple illustration of this

[1]Distance-based weight matrices (e.g., fixed distance band, inverse distance), have been also considered in empirical analysis for the robustness check in Section 6. The major conclusion still holds.

phenomenon, using our Cook County data which will be described in detail in Section 5, in Figure 1, we plot the quantile maps for returns and squared returns at census tract level. The left figure shades census tracts according to five bins of the returns, with darker shading corresponding to higher returns. Similarly, the right figure shades census tracts according to five bins of the squared returns, with darker shading corresponding to higher squared returns, representing higher volatility. The comparison of the two figures provides evidence that the level of returns appear to be random and not spatially correlated, while squared returns are spatially correlated especially on the south region of Cook County, and the southeast and west regions of the City of Chicago.

To address this concern for the presence of SARCH, we consider the following random coefficient version of SAR model (18) by assuming $\rho = \bar{\rho} + \eta_i$, with $\eta_i \sim iid(0, \sigma_\eta^2)$, and distributed independently of ε_i,

$$\begin{aligned} r_i &= \rho \sum_{j \neq i}^{n} w_{ij} r_j + \varepsilon_i, \qquad \varepsilon_i \sim iid(0, \sigma_\varepsilon^2) \\ &= (\bar{\rho} + \eta_i) \sum_{j \neq i}^{n} w_{ij} r_j + \varepsilon_i \\ &= \bar{\rho} \sum_{j \neq i}^{n} w_{ij} r_j + \eta_i \sum_{j \neq i}^{n} w_{ij} r_j + \varepsilon_i. \end{aligned} \tag{19}$$

Then the first two moments of r_i, conditional on the neighborhood information $\Phi_{j \neq i}$, are given by

$$E(r_i | \Phi_{j \neq i}) = \bar{\rho} \sum_{j \neq i}^{n} w_{ij} r_j, \tag{20}$$

$$\mathrm{Var}(r_i | \Phi_{j \neq i}) = \sigma_\varepsilon^2 + \sigma_\eta^2 \left(\sum_{j \neq i}^{n} w_{ij} r_j \right)^2 = \sigma_\varepsilon^2 + \sigma_\eta^2 N_i^2, \tag{21}$$

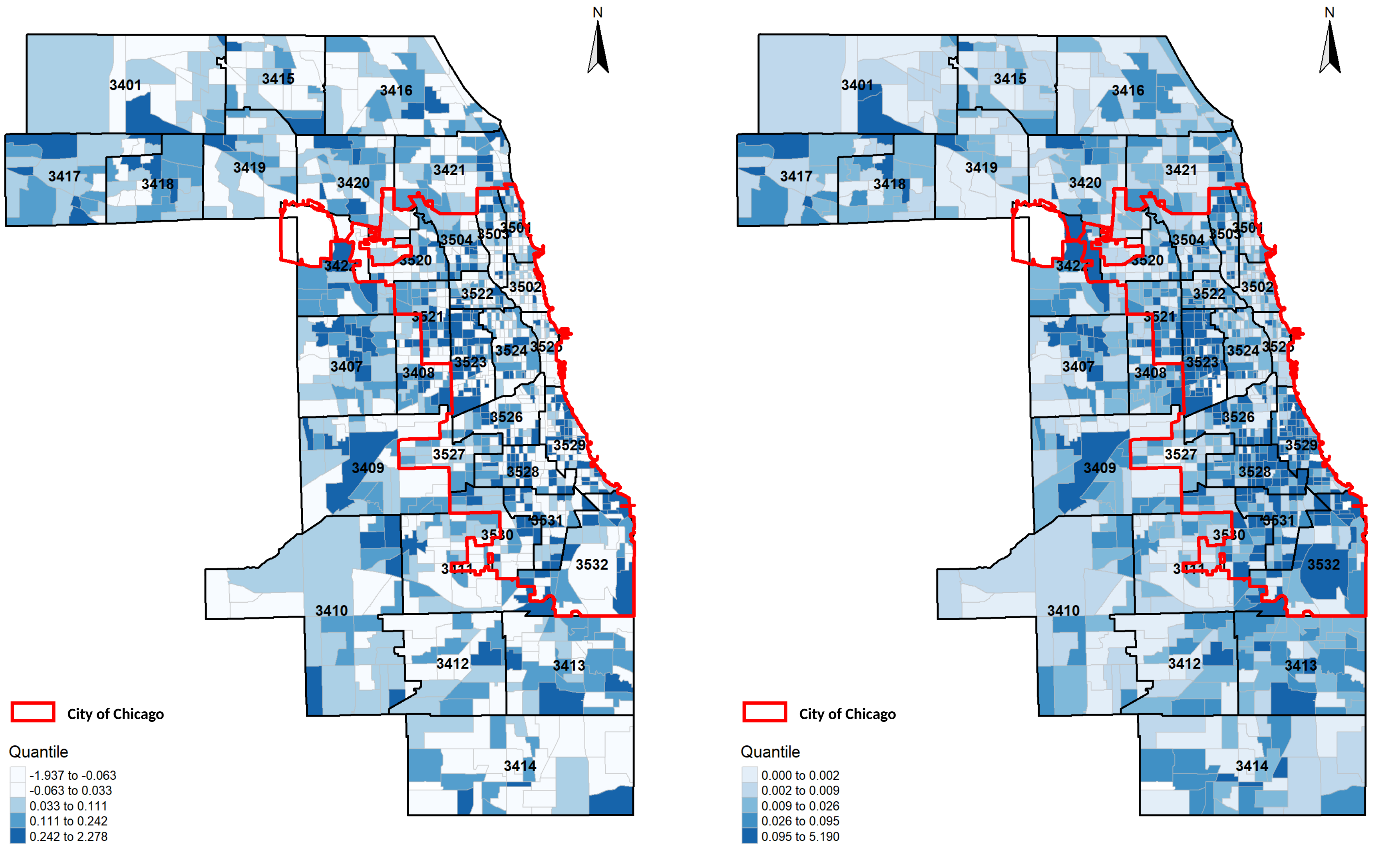
where $N_i = \sum_{j \neq i}^{n} w_{ij} r_j$ can be viewed as the average neighborhood value for the location i, $i = 1, ..., n$. Bera and Simlai (2005) formulated the SARCH specification (21) starting from an SAR model and using the information matrix (IM) test principle of White (1982).

Denoting $\sigma_\varepsilon^2 = \alpha_0$ and $\sigma_\eta^2 = \alpha_1$, (21) becomes

$$\mathrm{Var}(r_i | \Phi_{j \neq i}) = \alpha_0 + \alpha_1 N_i^2, \tag{22}$$

which can be taken as spatial counterpart of (15). In our application, an approximate version of (22) will be used, namely,

$$\mathrm{Var}(r_i | \Phi_{j \neq i}) = E(r_i^2 | \Phi_{j \neq i}) = \alpha_0 + \alpha_1 \sum_{j \neq i}^{n} w_{ij} r_j^2. \tag{23}$$



(a) Returns

(b) Squared returns

FIGURE 1 Quantile maps for returns and squared returns

Thus, while the conditional variance in the time series context is a function of the squares of past observations, in the spatial context it is described by the squares of neighboring observations. The model (23) can be interpreted as a SAR model for *squared* returns, and the parameters α_0 and α_1 can be estimated by using r_i^2 and a spatially lagged variable $\sum_{j\neq i}^{n} w_{ij} r_j^2$, and conducting a SAR-type regression.

3.3 Spatial Quantile Specification

While the above approaches provide parsimonious solutions to investigating the spatial market efficiency by analyzing the spatial dependence patterns in returns and squared returns via the regressions (18) and (23), they do not provide detailed information about the distribution.

Quantile regression (QR) as introduced by Koenker and Bassett (1978) has become a popular robust alternative to extract distributional information. Just as a classical linear regression estimates the overall conditional mean of the dependent variable, QR provides a way to estimate the conditional mean at different quantiles. Therefore, it is well suited to analyzing spatial data in particular situations where the distribution of the dependent variable is highly skewed at certain locations.

QR was incorporated into the SAR model by Kostov (2009) called spatial quantile regression (SQR) model. This model allows the spatial lag parameters ρ and α_1 in equations (18) and (23) to be dependent on each quantile τ, $0 < \tau < 1$, thus allowing for a different degree of spatial dependence at different points of the response distribution. For example, strong spatial dependence may only exist at certain values of τ. At quantile τ, we will use $y(\tau)$ to represent the dependent variable, and $\lambda(\tau)$ to represent the spatial lag parameters ρ and α_1 in (18) and (23), respectively. Incorporating potential independent variables X related to housing returns, the spatial QR (SQR) model can be written compactly in a matrix form as:

$$y(\tau) = \lambda(\tau)Wy + X\beta(\tau) + u, \tag{24}$$

where $y(\tau)$ is the dependent variable (returns or squared returns) in the τ^{th} quantile, W the $n \times n$ spatial weight matrix, X the $n \times k$ matrix of independent variables, $\beta(\tau)$ quantile specific parameters of X and u the vector of error terms.

Clearly the spatially lagged variable Wy present on the right-hand side of (24) is an endogenous variable, thus the conventional QR approach will lead to inconsistent estimator. Two methods have been suggested for consistent estimation by taking account of the endogeneity in any QR set-up: (i) the two-stage quantile regression (2SQR) suggested by Kim and Muller (2004); (ii) the instrumental variable quantile regression (IVQR) proposed by Chernozhukov and Hansen (2006). IVQR was then further extended to include a spatial lag effect (Kostov, 2009; Su and Yang, 2011) and the spatially lagged independent variables are considered as instruments. Zietz et al. (2008), Liao and Wang (2012), McMillen (2012), Zhang and Leonard (2014), and Zhang (2016) apply 2SQR approach to their studies of house prices. Kostov (2009) compares 2SQR and IVQR and considers an application to agricultural land prices.

2SQR resembles the two-stage least squares (2SLS) estimator of Kelejian and Prucha (1998) and IVQR is asymptotically equivalent to the generalized method of moments (GMM). The IVQR method can deal with weak instrument and may have better finite sample properties. On the other hand, the 2SQR procedure has computational efficiency as it requires only two consecutive quantile regressions for each quantile of interest whereas the IVQR method employed in Su and Yang (2011) and Kostov (2009) conducts a search over a range of values for the spatial dependence parameter and thus requires a separate quantile regression to be estimated for each value of the range. From a practical implementation point of view, we employ the 2SQR method for estimation which involves the following two steps. In the first step, for a given τ, we run a QR of Wy on the exogenous variables WX and X, and the predicted $\widehat{Wy}$ is obtained. Then in the second step, we estimate the model (24) by conducting another QR for the same value of τ, by regressing y on $\widehat{Wy}$ and X, and obtain $\hat{\lambda}(\tau)$ and $\hat{\beta}(\tau)$. This procedure is repeated for other quantiles $\tau \in (0, 1)$.

4 Monte Carlo Simulation

In this section, we present small simulation results to investigate the finite sample performance of 2SQR of the proposed model (23) in the quantile setting. In essence, we follow the simulation design of Su and Yang (2011) with a modification on the dependent variable. The specific DGP employed in the simulations has the following form:

$$y_i^2 = \lambda(v_i) \sum_{i=1}^{n} w_{ij} y_j^2 + \beta(v_i)' x_i, \tag{25}$$

where $x_i = (1, x_i^0)$, $x_i^0 \sim iidN(0, 1)$, and

$$\lambda(v_i) = 0.5 + 0.1 F^{-1}(v_i) \tag{26}$$

$$\beta(v_i) = (2.0, 1.0)' + (0.5, 0.5)' F^{-1}(v_i), \tag{27}$$

with $v_i \sim iidU(0, 1)$ and $F(\cdot)$ is chosen to be standard normal. Under these specifications, the values for $\lambda(\tau)$ and $\beta(\tau)' = \{\beta_1(\tau), \beta_2(\tau)\}$ under different τ are summarized in Table 1. The spatial weight matrix is generated according to queen contiguity criteria on regular $m \times m$ grids, leading to a sample size of $n = m^2$. This matrix is row-normalized so that each row sums to unity. The sample sizes used are 100, 400, and 900. Each set of simulation results is based on 1,000 Monte Carlo replications.

TABLE 1 True quantile parameters used in simulations

τ	$\lambda(\tau)$	$\beta_1(\tau)$	$\beta_2(\tau)$
0.25	0.4326	1.6628	0.6628
0.50	0.5000	2.0000	1.0000
0.75	0.5674	2.3372	1.3372

Table 2 presents the Monte Carlo bias together with the standard deviation (StDev) and the root mean squared errors (RMSE) for the 2SQR estimators (Panel A) of the spatial parameter $\lambda(\tau)$ and the slope parameter $\beta_2(\tau)$ at the 0.25th, 0.50th, and the 0.75th quantile. For comparison purpose, we also report the results of IVQR estimators in Panel B.

TABLE 2 Empirical bias, standard deviation (StDev), and RMSE for estimators of $\lambda(\tau)$ and $\beta_2(\tau)$

		$\lambda(\tau)$			$\beta_2(\tau)$		
τ	n	Bias	StDev	RMSE	Bias	StDev	RMSE
Panel A: 2SQR							
0.25	100	0.0297	0.3199	0.3213	0.0732	0.1133	0.1349
	400	0.0319	0.1026	0.1075	0.0738	0.0562	0.0928
	900	0.0377	0.0638	0.0741	0.0712	0.0351	0.0794
0.50	100	−0.0130	0.1947	0.1951	−0.0076	0.1094	0.1096
	400	−0.0077	0.0830	0.0834	0.0012	0.0545	0.0545
	900	−0.0047	0.0560	0.0562	−0.0001	0.0337	0.0337
0.75	100	−0.0607	0.2050	0.2138	−0.0869	0.1140	0.1434
	400	−0.0537	0.0931	0.1075	−0.0744	0.0581	0.0944
	900	−0.0466	0.0596	0.0757	−0.0747	0.0385	0.0840
Panel B: IVQR							
0.25	100	−0.0109	0.2274	0.2277	0.0668	0.1020	0.1219
	400	0.0049	0.0780	0.0782	0.0582	0.0472	0.0749
	900	0.0100	0.0514	0.0524	0.0540	0.0356	0.0646
0.50	100	−0.0133	0.1813	0.1818	−0.0087	0.0970	0.0974
	400	−0.0057	0.0662	0.0665	−0.0001	0.0416	0.0416
	900	−0.0004	0.0394	0.0394	−0.0018	0.0265	0.0265
0.75	100	−0.0227	0.2047	0.2059	−0.0790	0.1065	0.1326
	400	−0.0152	0.0802	0.0816	−0.0591	0.0528	0.0793
	900	−0.0129	0.0503	0.0519	−0.0571	0.0379	0.0685

The results indicate that both estimators behave quite well in general, although a slight edge may go to the IVQR estimator. Also, both standard deviation and RMSE decline across the estimators as the sample size increases, and the magnitude of decrease is generally consistent with the $\sqrt{n}$-asymptotics. As noted in Section 3, the 2SQR is by any means computationally simpler than the IVQR for spatial lag type of models. Given that it has similar properties to the IVQR estimator and the additional computational costs associated with the IVQR estimator may not bring in sufficient advantages to justify their usage over the 2SQR estimator, the 2SQR method is used in empirical analysis.

5 Data

This study was carried out in the Cook County, the largest and most diverse county in the Chicago metropolitan area, consisting of more than 60 percent of the housing units in the area. Several factors were considered in selecting appropriate spatial units for the study to represent neighborhoods. These included homogeneity in terms of socioeconomic status within the spatial unit; large enough population size to minimize the problem with small numbers; data availability; acceptability by urban planners and policymakers; and stability of the boundaries over time for future analyses. Based on these criteria, census tracts were chosen as the spatial units for the analysis, mainly due to their homogeneity with respect to socioeconomic and demographic characteristics.[2] Furthermore, most census data are reported at this level of geography, which allowed us to obtain neighborhood characteristics for our independent variables, and the boundaries of the census tracts follow permanent and easily recognizable physical features.

There are 1,318 census tracts in Cook County, which are the reasonable number of spatial units - neither too big nor too small. For ease of interpretation, we also use Public Use Microdata Areas (PUMA) as reference of local housing submarkets.[3] There are 33 PUMAs in Cook County, 19 of which comprise city communities; the remaining 14 cover the suburban communities. Figure 2 shows a map of Cook County with the 1,318 census tracts and the 33 different PUMAs. The census tracts are outlined with light blue lines, and the PUMAs are outlined in solid black, and their number are labeled on the map using black text. To take a broader view of city versus suburbs, we outline the City of Chicago in a solid red line.

This study employs the residential property sales prices from multiple listing services (MLS) in Cook County for the period 2010-2016, which were accessed from the Illinois REALTORS®. All sales price data were geocoded first and aggregated by the census tract level, and then used to compute the annual median

[2] Census tracts are small, relatively stable spatial units with population ranging between 2,500 and 8,000 with an average of approximately 4,000.

[3] PUMAs are geographic areas defined for statistical use by the U.S. Census Bureau and constructed from census tracts.

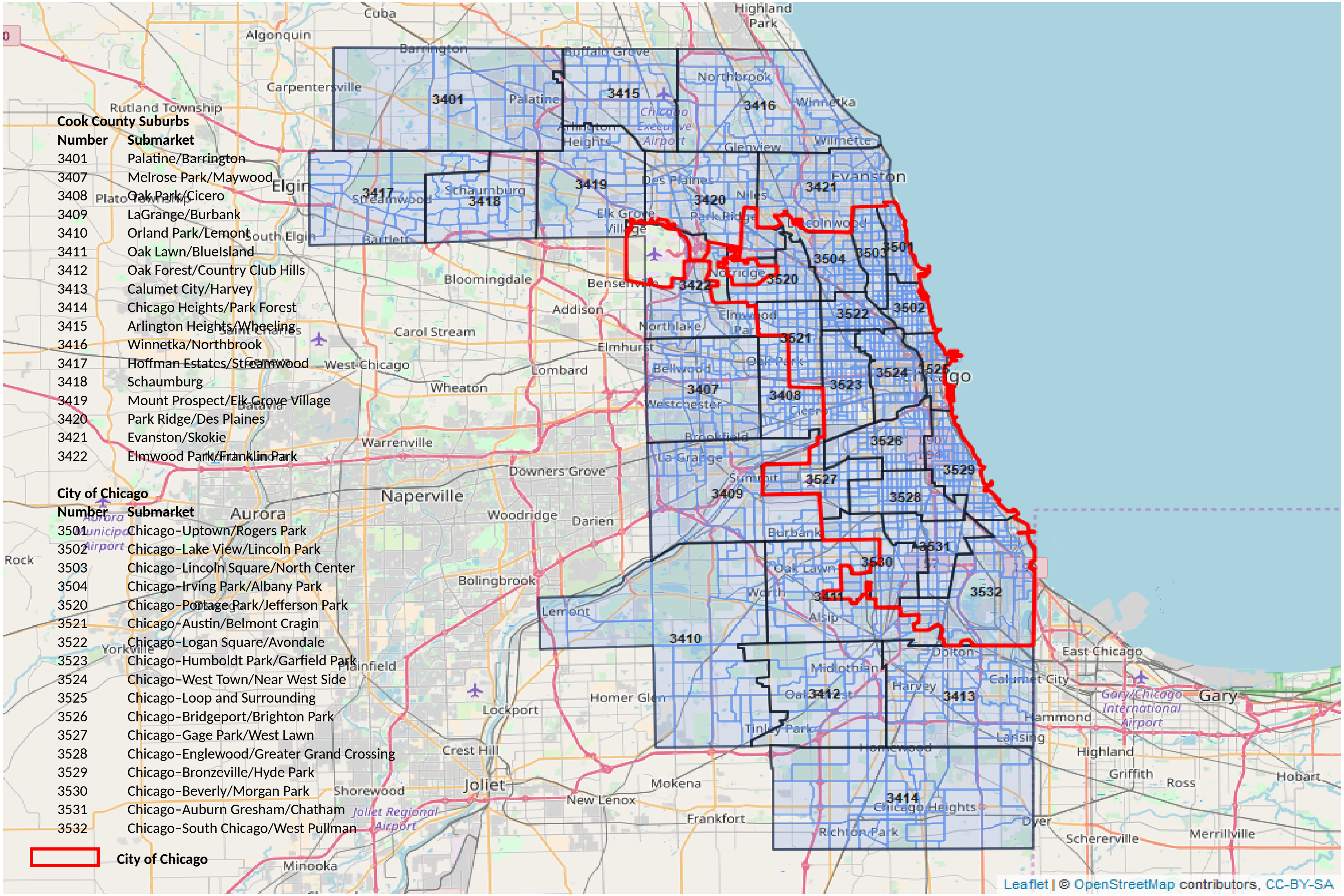


FIGURE 2 Map of 1,318 census tracts for 33 PUMAs in Cook County

FIGURE 2 *Data Source:* U.S. Census Bureau, Institute for Housing Studies, OpenStreetMap.
Note: The listed numbers refer to the codes of the PUMAs upon which the listed housing submarkets are based.

prices for each tract.[4] There has been considerable debate over the merits of using median house price versus repeat sales price, which measures sales prices for the identical property at different points in time, provided that property characteristics do not change between the two sales. This study uses the median price method because the repeat sales method is restricted to properties that have been sold more than once, which may lead to too small estimation samples that are not able to reflect the pure price change of the entire housing market. As the median price includes single sales that can make up a large proportion of total sales, it can be considered more representative of the housing market. Thus the median prices are used to calculate annual housing returns that are obtained from the first difference of log annual median house prices, i.e., $r_{i,t} = \log(P_{i,t}) - \log(P_{i,t-1})$, where at census tract i, $P_{i,t}$ and $P_{i,t-1}$ are, respectively, the median prices at time t and $t-1$, and $r_{i,t}$ is the corresponding log annual return.[5] In order to control for possible factors related to the housing returns, typical hedonic pricing variables are also considered; these variables represent structural, locational, and neighborhood attributes of housing. Two structural variables are considered and constructed as averages: number of bathrooms and floor area. For the locational variable, distance from each census tract to the central business district (CBD) is calculated. For the neighborhood attributes, the American Community Survey (ACS) 5-year data from the U.S. Census Bureau for the period 2010-2016 are used to obtain the socioeconomic variables for each tract, such as unemployment rate, percentage of African Americans, percentage of population with a bachelor's degree, and median income. Table 3 reports summary statistics of the returns and squared returns along with the independent variables, for each year during the study period from 2010 to 2016.

TABLE 3 Descriptive statistics

Year/sample size	2010 ($n = 1246$)		2011 ($n = 1279$)		2012 ($n = 1279$)		2013 ($n = 1284$)		2014 ($n = 1279$)		2015 ($n = 1292$)		2016 ($n = 1279$)	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
Returns	−0.06	0.37	−0.13	0.38	−0.04	0.39	0.13	0.35	0.12	0.30	0.10	0.34	0.10	0.32
Squared returns	0.14	0.44	0.16	0.45	0.15	0.84	0.14	0.78	0.10	0.27	0.12	0.38	0.11	0.30
Bathroom	1.74	0.37	1.76	0.38	1.75	0.36	1.75	0.36	1.78	0.37	1.79	0.36	1.81	0.37
Floor area	6.46	0.50	6.91	0.37	7.01	0.31	7.05	0.30	7.08	0.30	7.10	0.30	7.12	0.28
CBD distance	10.11	7.12	10.02	7.10	10.01	7.10	9.96	7.12	10.00	7.11	9.97	7.10	10.02	7.11
Unemployment	11.13	7.85	12.03	7.85	13.02	8.36	13.80	8.81	13.34	8.89	12.42	8.71	11.13	8.46
African	28.83	37.63	28.82	37.36	28.97	37.34	28.58	37.06	28.14	36.60	28.61	36.82	28.19	36.38
Bachelor	19.58	12.33	19.83	12.58	20.01	12.70	20.07	12.79	20.49	12.92	20.56	12.93	20.97	12.83
Income	10.85	0.47	10.85	0.47	10.84	0.48	10.83	0.49	10.84	0.49	10.84	0.51	10.88	0.51

Note: This table describes mean and standard deviation (StDev) for each year. Floor area (in squared footage), CBD distance (in miles), and median income are expressed in logarithmic term, while Unemployment, African, and Bachelor are expressed in percentage term. To address likely data entry errors, observations with unrealistic or missing values are eliminated, leading to unequal sample sizes across years.

[4] ArcGIS software was used to conduct this task.

[5] No additional components that are often included in computing the total return such as interest and dividends in the financial market data are considered in this analysis as they are not applicable in our context.

Given that we have access to the seven years of the data, a panel study would appear to be the most appropriate and advantageous approach, but we estimate the model as cross-sections as quantile regression estimation for spatial panel models has not been studied in existing literature, with the exception of Dai et al. (2019) who recently investigate the IVQR estimation of general spatial autoregressive panel data model with fixed effects.

6 Empirical Analysis

We first estimate the housing return model (18) including the potential independent variables listed in Table 3 by the method of two-stage least squares (2SLS) with robust standard errors, separately for each of the seven cross-sections. The 2SLS estimation is implemented following Kelejian and Prucha (1998) by using spatially lagged independent variables as instruments for the spatially lagged dependent variable. The results are reported in Table 4.

TABLE 4 Robust 2SLS estimation results based on queen contiguity (dependent variable: returns)

	2010	2011	2012	2013	2014	2015	2016
Wy	0.755***	0.108	0.053	0.182	0.494**	−0.215	0.023
	(0.228)	(0.201)	(0.295)	(0.230)	(0.243)	(0.237)	(0.402)
Constant	−0.414	−0.747	−0.457	1.785***	0.301	0.759**	0.475
	(0.598)	(0.482)	(0.437)	(0.594)	(0.381)	(0.373)	(0.383)
Bathroom	0.094**	0.191***	0.116***	0.147***	0.040	0.112***	0.084**
	(0.040)	(0.038)	(0.044)	(0.046)	(0.037)	(0.040)	(0.039)
Floor area	0.057**	0.060	0.019	−0.076	−0.016	0.039	0.039
	(0.024)	(0.039)	(0.047)	(0.055)	(0.042)	(0.047)	(0.050)
CBD distance	0.000	−0.003**	−0.003**	0.002**	−0.002*	−0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Unemployment	0.001	0.002	0.002	−0.004	−0.004**	−0.002	0.001
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)
Black	0.001	0.000	0.000	0.000	0.000	0.000	0.000
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Bachelor	0.000	−0.001	0.000	0.000	−0.003***	−0.003**	−0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Income	−0.013	−0.009	0.008	−0.126***	−0.006	−0.095***	−0.073**
	(0.055)	(0.047)	(0.040)	(0.046)	(0.029)	(0.036)	(0.034)
n	1246	1279	1279	1284	1279	1292	1279
Pseudo R^2	0.055	0.059	0.028	0.037	0.032	0.033	0.047

Note: Standard errors in parentheses. Pseudo R^2 is computed as the squared correlation between observed and predicted values of the dependent variable. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

As mentioned in Section 3, a queen contiguity spatial weight matrix is used since many census tracts form a relatively uniform rectilinear tessellation across the region. In order to determine the significance of the model, in fact pseudo-significance, a pseudo R^2 is used since the traditional R^2 is not a good measure of fit for the spatial lag models. This pseudo R^2 is computed as the square of the correlation between observed and predicted values of the dependent variable (Anselin, 1995).

It can be seen from Table 4 that the estimated coefficients of spatial lag term (Wy) are not consistent across years in terms of the sign and magnitude and not statistically significant in most years, with two exceptions found in 2010 where the effects of the financial crisis led to significant negative returns for most of tracts and 2014 where housing prices started recovering. There are a few significant independent variables, but they do not play an important role in explaining the housing returns, with pseudo R^2 all below 0.1. As the 2SLS estimates do not provide useful information about the spatial dependence, the discussion now proceeds to the results from the 2SQR model to verify whether there is statistical variation in the coefficients of the spatial lag term across different quantiles.

TABLE 5 2SQR estimation results based on queen contiguity (dependent variable: returns)

	Quantiles (τ)								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Wy	−0.376	−0.742	−0.153	−0.045	−0.460	−0.288	−0.074	−0.078	0.233
	(0.823)	(0.625)	(0.535)	(0.476)	(0.389)	(0.351)	(0.289)	(0.332)	(0.294)
Constant	−0.663	−0.409	−0.249	0.042	0.493*	0.916***	1.222***	1.865***	2.393***
	(0.820)	(0.447)	(0.288)	(0.289)	(0.291)	(0.292)	(0.382)	(0.579)	(0.790)
Bathroom	0.061	0.059*	0.055*	0.068**	0.080***	0.095***	0.124***	0.176***	0.267***
	(0.063)	(0.036)	(0.033)	(0.031)	(0.027)	(0.024)	(0.033)	(0.049)	(0.064)
Floor area	−0.031	−0.034	−0.007	−0.015	−0.008	0.016	0.028	0.058	0.050
	(0.076)	(0.037)	(0.033)	(0.034)	(0.035)	(0.033)	(0.036)	(0.049)	(0.071)
CBD distance	0.006***	0.002*	0.001	0.000	0.000	−0.001	−0.001	−0.002*	−0.004**
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Unemployment	−0.004	−0.004	−0.003	−0.002	−0.001	−0.001	0.000	0.001	0.000
	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.005)
African	−0.003**	−0.001*	0.000	0.000	0.000	0.001	0.001**	0.001**	0.001
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Bachelor	−0.003	−0.004**	−0.002*	−0.002	−0.004***	−0.003***	−0.002**	−0.002	−0.003*
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Income	0.061	0.056	0.025	0.004	−0.034	−0.089***	−0.131***	−0.212***	−0.258***
	(0.070)	(0.041)	(0.028)	(0.027)	(0.028)	(0.028)	(0.035)	(0.047)	(0.072)
n	1,292	1,292	1,292	1,292	1,292	1,292	1,292	1,292	1,292
Pseudo R^2	0.006	0.005	0.000	0.037	0.005	0.022	0.033	0.031	0.042

Note: Standard errors in parentheses. The standard errors for quantile regression are obtained through 500 bootstrap replications. Pseudo R^2 is computed as the squared correlation between observed and predicted values of the dependent variable. Results are based on 2015 data. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

The results are presented in Table 5 in which the numbers in parentheses are the bootstrapping standard errors, that are obtained through 500 bootstrap replications. Due to space limitations, we report only on results from the year 2015. Additional results, consistent with those reported, are available from the authors. According to the estimation results, defining the relationship between the typical hedonic pricing variables and housing returns is not improved by using 2SQR, as indicated by low pseudo R^2 for all quantiles. For some variables the 2SQR results reveal that their association with the housing returns is not stable across quantiles. For other variables that are not statistically significant in the ML and 2SLS estimations, the 2SQR confirms them to remain insignificant over different quantiles. Finally, the estimates of the spatial lag parameter are not statistically significant and substantially vary in magnitude across quantiles, indicating that housing returns are not correlated over space. These findings support the conjecture that the housing market is spatially efficient as long as we consider only the returns.

However, as noted earlier, the absence of spatial linear dependence in returns does not preclude the presence of nonlinear relationship. In Section 3, Panel (b) of Figure 1 clearly identified the spatial clustering for the squared returns. To investigate and quantify the possible presence of nonlinear dependence, we now turn to the model (23) and repeat our exercises using the squared returns, and the results are provided in Table 6. First, and most notably, the spatial coefficients are now highly significant for all years, yielding a considerable improvement in the model fit. This result corroborates the visual conclusions from Panel (b) of Figure 1. Hence, we can conclude that although the returns are spatially uncorrelated, there is significant spatial dependence in the squared returns, indicating that highly volatile locations tend to be surrounded by similar volatile neighborhoods. The presence of spatial dependence in squared returns provides evidence that the housing market is spatially inefficient.

In contrast to the highly significant spatially lagged squared returns, the other independent variables are not significant, with the exception of the floor area and income variables.[6] In particular, the income variable is highly negatively correlated with the squared returns, indicating that higher household income leads to lower volatility of housing returns. This finding is largely in line with those of Hartman-Glaser and Mann (2017), who measure income and housing return volatility at the zip code level within the largest metropolitan statistical areas (MSAs) in the U.S. and conclude that lower income households face higher volatility of housing returns. They argue that this is because collateral constraints are tighter for lower-income areas, causing higher housing return volatility. More specifically, the collateral constraint makes households unable to fully smooth consumption, and thus their marginal rate of substitution (MRS) between housing and other consumption fluctuates with income

[6]It is noteworthy that the coefficients in Table 6 do not directly reflect the marginal effects of the corresponding independent variables on the dependent variable (LeSage and Pace, 2010), we thus need to report the direct, indirect, and total effects of the independent variables. However, the focus of this study is on the spatial dependence parameter and not on the marginal effects of the independent variables.

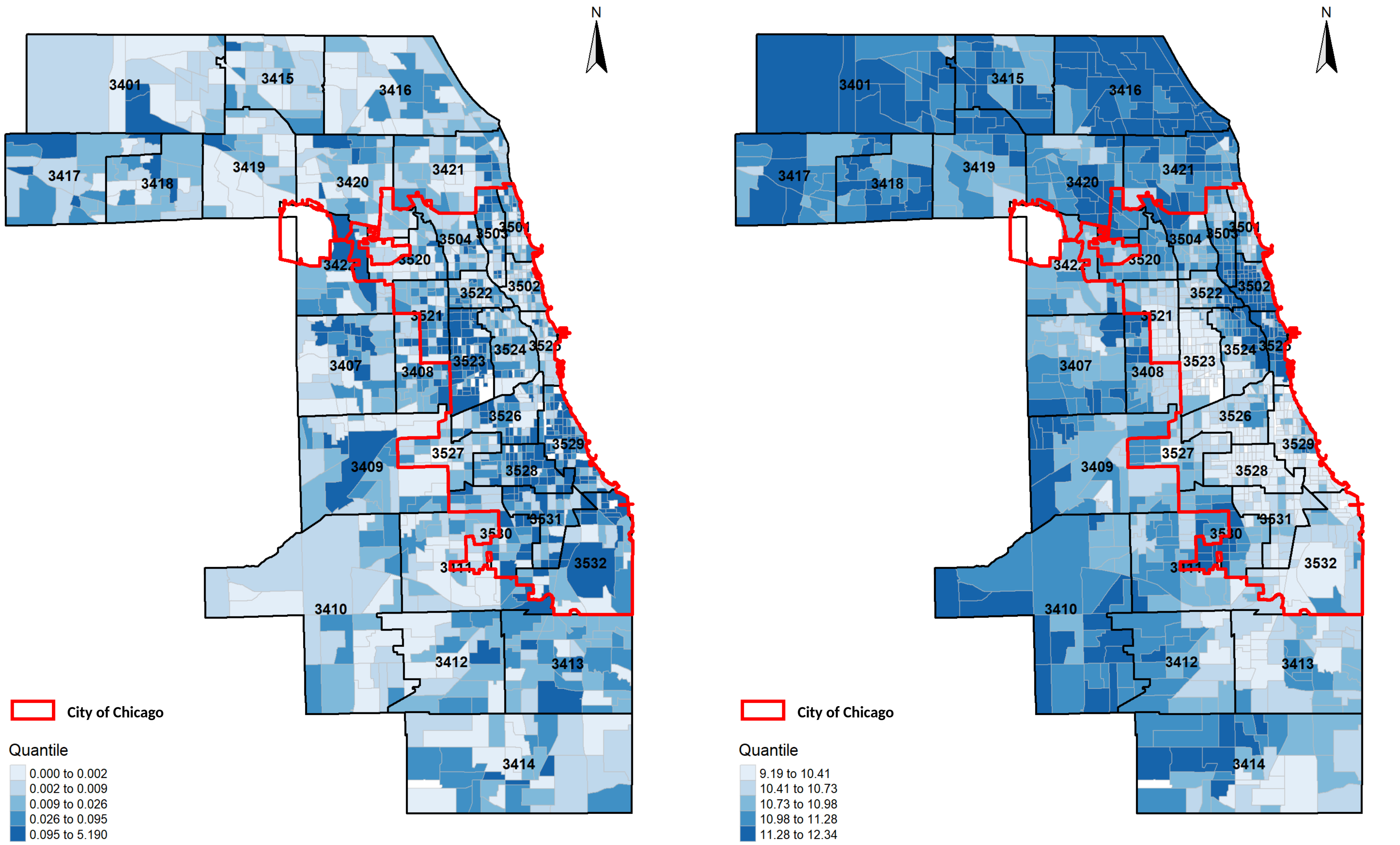
TABLE 6 Robust 2SLS estimation results based on queen contiguity (dependent variable: squared returns)

	2010	2011	2012	2013	2014	2015	2016
Wy	0.587***	0.400**	0.498**	0.485***	0.416***	0.579***	0.441**
	(0.181)	(0.156)	(0.198)	(0.132)	(0.138)	(0.150)	(0.180)
Constant	1.310**	1.666***	0.179	1.781***	0.630**	0.419	0.297
	(0.544)	(0.636)	(0.450)	(0.523)	(0.285)	(0.319)	(0.302)
Bathroom	0.023	0.066*	−0.030	0.029	0.025	0.009	−0.014
	(0.040)	(0.037)	(0.046)	(0.046)	(0.028)	(0.029)	(0.028)
Floor area	0.019	−0.034	0.097**	−0.050	0.006	0.079**	0.120***
	(0.020)	(0.045)	(0.049)	(0.052)	(0.028)	(0.040)	(0.040)
CBD distance	0.000	0.000	−0.001	0.001	−0.001	−0.001	−0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Unemployment	−0.002	0.001	0.002	−0.001	−0.001	0.000	0.001
	(0.003)	(0.003)	(0.004)	(0.002)	(0.001)	(0.002)	(0.002)
Black	0.000	0.001	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Bachelor	0.000	0.001	0.001	0.001	−0.001	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Income	−0.131**	−0.140**	−0.073	−0.134***	−0.058**	−0.087***	−0.099***
	(0.053)	(0.055)	(0.049)	(0.044)	(0.024)	(0.032)	(0.030)
n	1246	1279	1279	1284	1279	1292	1279
Pseudo R^2	0.108	0.059	0.010	0.012	0.086	0.157	0.231

Note: Standard errors in parentheses. Pseudo R^2 is computed as the squared correlation between observed and predicted values of the dependent variable. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

shocks. This in turn leads to endogenous volatility in the housing returns provided that the housing supply is not perfectly elastic. Furthermore, this volatility is greater especially for lower income households because even relatively small income shocks can have significant effects on their financial stability.

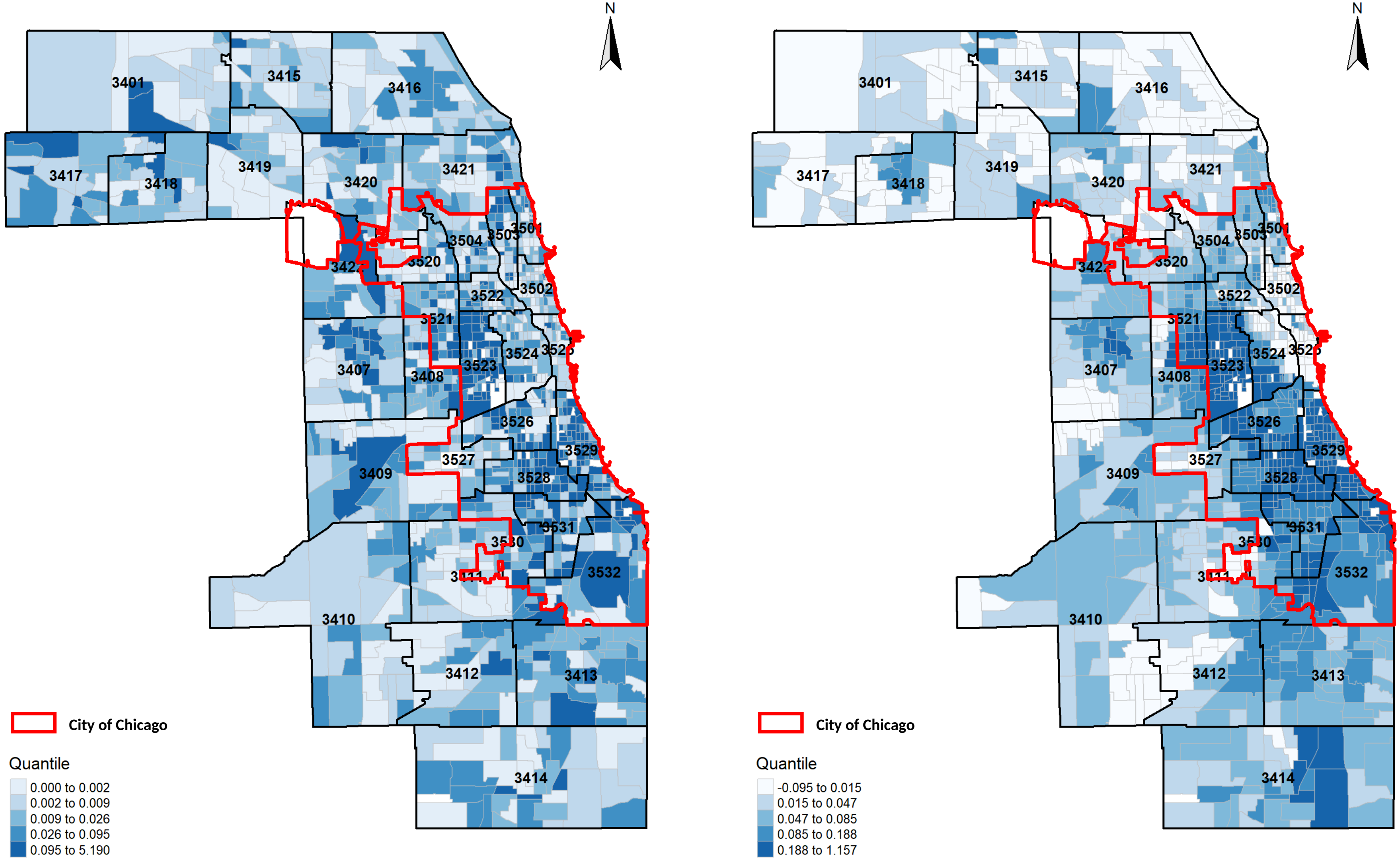
Figure 3 demonstrates these findings visually with quantile maps of the squared returns and income. Note that darker shading in Panel (a) of Figure 3 corresponds to higher volatility and brighter shading of Panel (b) of Figure 3 corresponds to lower income. By comparing the Panels (a) and (b), we observe that high-volatility areas are low-income areas and low-volatility areas are high-income areas. More specifically, most highly volatile areas on the west and south parts of Chicago (PUMAs 3523, 3528, and 3529, as in Figure 2 , namely Humboldt Park, Garfield Park, Englewood, Greater Grand Crossing, Bronzeville, and Hyde Park, are on the lowest end of the income spectrum. These areas are also known as predominately minority (African American and Hispanic) neighborhoods and considered one of the most segregated metro areas in the U.S. By contrast on the lower end of the income spectrum, in places like north side neighborhoods of the city (PUMAs 3502 and 3525) - Lake View, Lincoln Park, and Loop are much less volatile.



(a) Squared returns

(b) Income (in log)

FIGURE 3 Quantile maps for squared returns and income (in log)



(a) Squared returns

(b) Predicted (within sample) squared returns

FIGURE 4 Quantile maps for squared returns and predicted squared returns

Using the 2SLS estimation results, the predicted values of the square returns are calculated and are displayed in Figure 4, Panel (b). For easy comparison, we reproduce quantile map of squared returns (Figure 1, Panel (b)) in Figure 4, Panel (a). Comparing Panels (a) and (b) in Figure 4, we note that the clustering patterns of the squared returns are well captured by the predictions.

Now we estimate 2SQR to portray the different effects of the spatially lagged dependent variable and independent variables at various points of the conditional distribution of squared returns. As presented for the return model, here we only report results for 2015 (results for other years also available from the authors). The results of the 2SQR with bootstrapped standard errors are presented in Table 7.

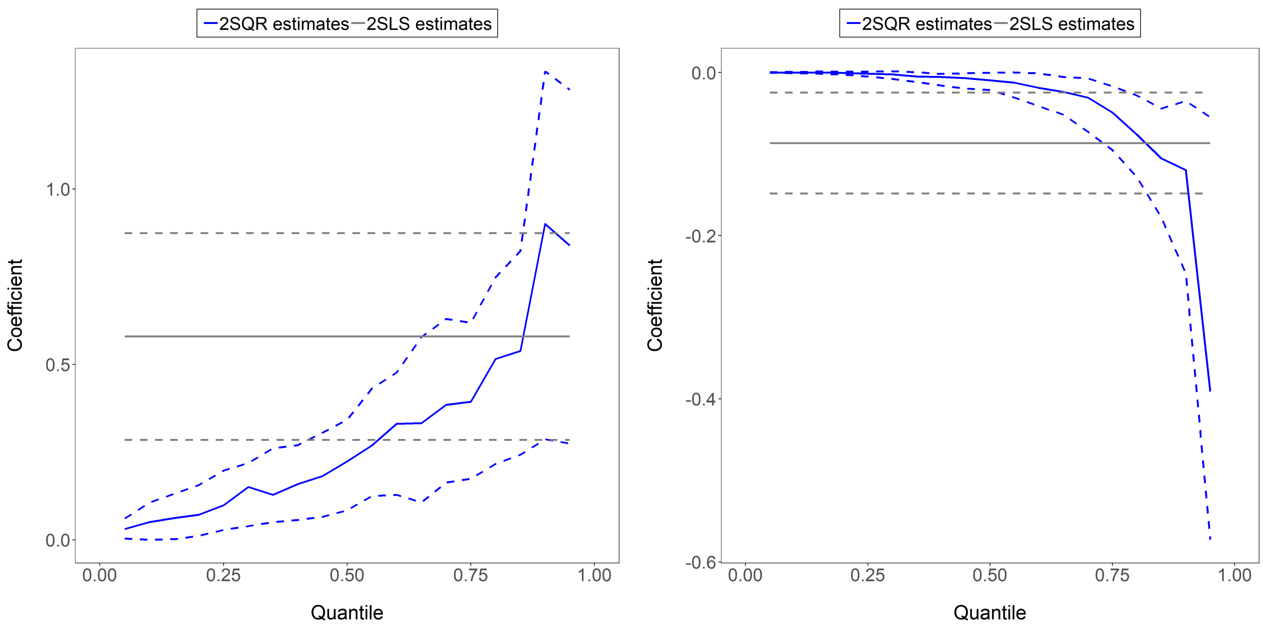
TABLE 7 2SQR estimation results based on queen contiguity (dependent variable: squared returns)

	Quantiles (τ)								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Wy	0.050**	0.072**	0.151***	0.159***	0.223***	0.331***	0.385***	0.515***	0.901***
	(0.025)	(0.035)	(0.049)	(0.052)	(0.067)	(0.087)	(0.107)	(0.141)	(0.263)
Constant	−0.002	0.004	0.019	0.056	0.086	0.150*	0.226	0.583**	0.879
	(0.004)	(0.011)	(0.023)	(0.036)	(0.055)	(0.087)	(0.157)	(0.294)	(0.546)
Bathroom	0.000	0.000	0.002	0.006**	0.010**	0.014**	0.015	0.054**	0.060
	(0.000)	(0.001)	(0.002)	(0.003)	(0.004)	(0.007)	(0.011)	(0.023)	(0.038)
Floor area	0.000	0.000	0.000	0.000	0.001	0.006	0.012	0.021	0.050
	(0.000)	(0.001)	(0.002)	(0.003)	(0.005)	(0.007)	(0.011)	(0.020)	(0.043)
CBD distance	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	−0.002
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Unemployment	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.002
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)	(0.005)
African	0.000	0.000	0.000	0.000*	0.000*	0.000	0.000	0.001	0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.002)
Bachelor	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001**	0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Income	0.000	0.000	−0.002	−0.006	−0.009*	−0.019**	−0.031**	−0.076***	−0.119**
	(0.000)	(0.001)	(0.002)	(0.004)	(0.005)	(0.009)	(0.015)	(0.028)	(0.052)
n	1,292	1,292	1,292	1,292	1,292	1,292	1,292	1,292	1,292
Pseudo R^2	0.130	0.135	0.135	0.144	0.146	0.146	0.149	0.152	0.154

Note: Standard errors in parentheses. The standard errors for quantile regression are obtained through 500 bootstrap replications. Pseudo R^2 is computed as the squared correlation between observed and predicted values of the dependent variable. Results are based on 2015 data. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

According to the results, the estimated coefficients for the spatial lag term are all significantly positive at all quantiles, which is in compliance with the 2SLS estimations presented in Table 6. We also find that the coefficients at upper quantiles are systematically larger than those at the lower quantiles, which indicates that the influence of spatially lagged squared returns increases with quantiles. For the income variable, the general pattern seems to be that at upper quantiles the influence of income on the squared returns is significantly negative with a downward trend as the quantile increases. On the other hand, the coefficients of other independent variables are mostly insignificant with irregular patterns across different quantiles. The floor area which was found to be significant in the 2SLS estimations (column (6) in Table 6) becomes insignificant at all quantiles.

In order to better see the tendency of the coefficients change of our interest across quantiles, two subgraphs are provided to support the illustration of coefficient estimates in Figure 5 according to the estimation results of 2SQR in Table 7, separately for the spatial lag term and income variable. So as to clearly show the change tendency of coefficient, each subgraph exhibits 2SQR coefficient estimates and their 95% confidence intervals at 0.05–0.95 quantiles with step size 0.05. For comparison purpose, the 2SLS coefficient estimates with robust standard errors (from Table 6) and the bounds of their corresponding 95% confidence intervals are also included, which are obviously invariant across quantiles at 0.579 and -0.087 respectively.

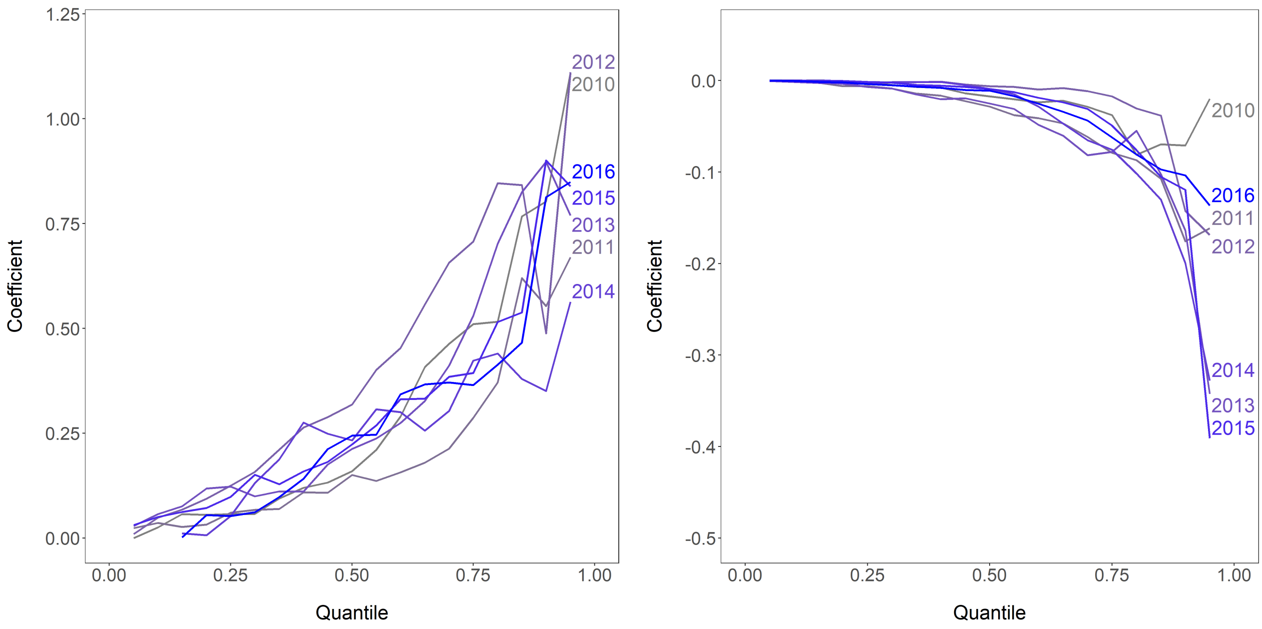


(a) Spatial lag term

(b) Income

FIGURE 5 2SLS and 2SQR coefficient estimates by quantile

As noted in Table 7, Panel (a) in Figure 5 reveals that there is substantial heterogeneity of spatial dependence across quantiles; in particular, there is a clear upward trend of the quantile effects of the spatial dependence. This finding suggests that the housing returns of nearby areas more positively influence those of high-volatility areas than low-volatility areas. On the other hand, Panel (b) in Figure 5 shows that the coefficients on income are negative, with a clear downward trend, suggesting that the decrease of household income leads to more volatile housing returns particularly in areas with higher housing returns. Looking at trends of both coefficients for other years in Figure 6, despite some variation across years, the overall patterns look similar, with the spatial lag term showing a rising trend and the income showing a declining trend across quantiles. Hence, we can conclude that the patterns observed in 2015 can be generalized to other time periods.



(a) Spatial lag term

(b) Income

FIGURE 6 2SQR coefficient estimates by quantile

Such increasing or decreasing spatial patterns across conditional quantiles warrants more research. These patterns might be common in other housing markets, or might be very specific to Cook County. We should note that the Chicago region has a long history of segregation, both racially and socioeconomically. White residents are predominately found in neighborhoods on the North side and surrounding downtown; Black residents in South and West side neighborhoods, and Latin residents in Northwest and Southwest neighborhoods. Even after many nondiscriminatory policies and practices have been applied, the racial segregation has been intensifying partially because of the prevalence of discrimination throughout the housing industry. A recent study by the Institute for Race Research and Public Policy at the University of Illinois at Chicago, “A Tale of Three Cities: the State of Racial justice in Chicago Report,” shows that, due to persistent segregation in Chicago, there are significant economic inequalities among Black, White and Latin residents, even when controlling for educational levels.[7]

We also explore whether the results are changed when ruling out insignificant independence variables and only keeping the spatial lag term and income variable. The estimation results of the parsimonious model are

[7] Henricks et al. (2018)

presented in Table 8 and Table 9 for 2SLS and 2SQR respectively. From Table 8, we note that although the absolute magnitudes of the coefficients of the variables differ, the signs and significance of the coefficients are in line with the previous ones obtained from the full model presented in Table 6, which confirms our main findings are robust with model specification. Further comparison shows that even with a small number of variables, the model performance and fit is not changed significantly. 2SQR results between the full (Table 7) and parsimonious model (Table 9) also suggest similar patterns of the estimated coefficients across quantiles. As can be seen in Figure 7, we find a clear upward trend in the coefficients for the spatial lag term and downward trend in those for the income variable, and as before, these patterns are generalizable to other years with varying magnitude (Figure 8).

TABLE 8 Robust 2SLS estimation results based on queen contiguity (dependent variable: squared returns)

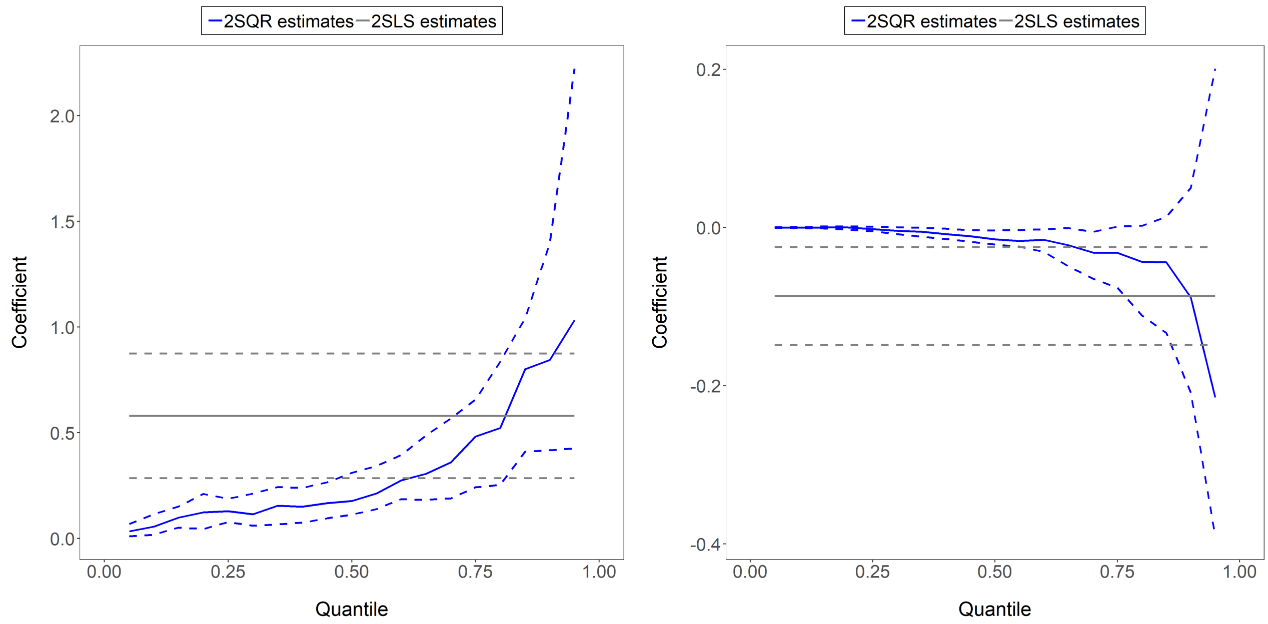
	2010	2011	2012	2013	2014	2015	2016
Wy	0.660***	0.583***	0.676**	0.470**	0.698***	0.750***	0.781***
	(0.222)	(0.214)	(0.279)	(0.190)	(0.148)	(0.190)	(0.175)
Constant	1.662***	1.374**	1.257*	1.171***	0.634***	0.863*	0.664*
	(0.612)	(0.587)	(0.655)	(0.389)	(0.237)	(0.501)	(0.357)
Income	−0.148***	−0.121**	−0.111*	−0.103***	−0.056***	−0.077*	−0.059*
	(0.054)	(0.052)	(0.058)	(0.035)	(0.021)	(0.044)	(0.031)
n	1246	1279	1279	1284	1279	1292	1279
Pseudo R^2	0.111	0.041	0.009	0.015	0.059	0.143	0.219

Note: Standard errors in parentheses. Pseudo R^2 is computed as the squared correlation between observed and predicted values of the dependent variable. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

TABLE 9 2SQR estimation results based on queen contiguity (dependent variable: squared returns)

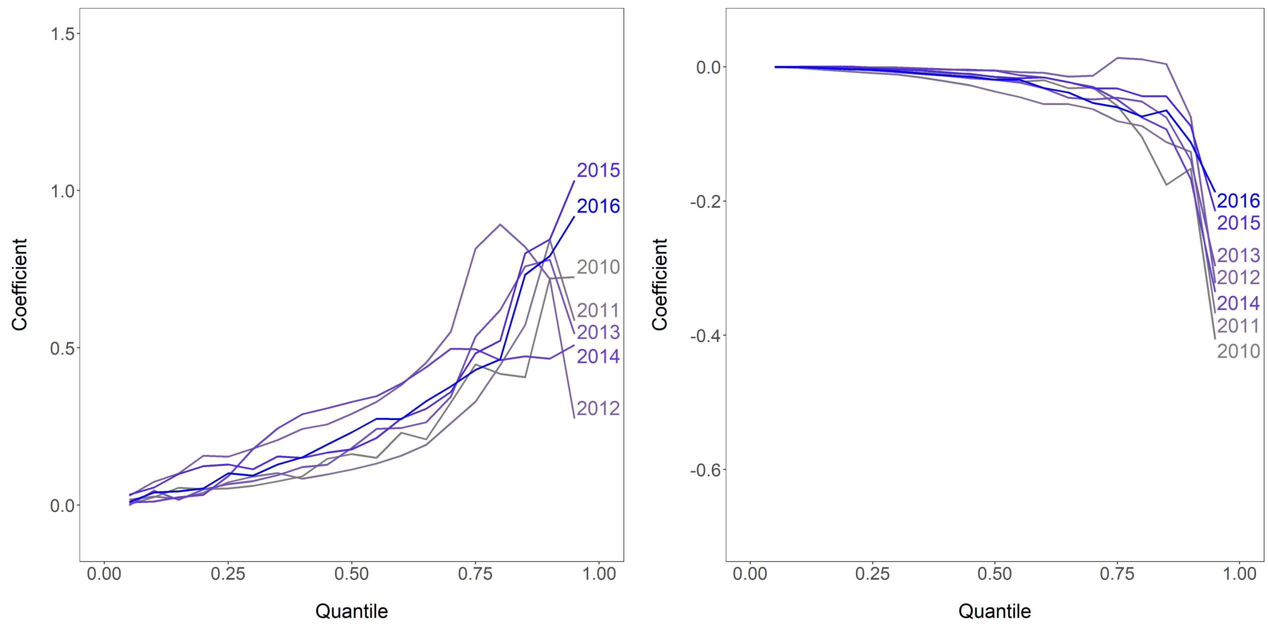
	Quantiles (τ)								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Wy	0.056**	0.124***	0.114***	0.151***	0.177***	0.275***	0.360***	0.523***	0.845***
	(0.027)	(0.038)	(0.039)	(0.040)	(0.048)	(0.054)	(0.095)	(0.145)	(0.242)
Constant	0.001	0.000	0.046**	0.096***	0.173***	0.183**	0.373**	0.508	1.024
	(0.004)	(0.013)	(0.022)	(0.037)	(0.054)	(0.083)	(0.181)	(0.318)	(0.749)
Income	0.000	0.000	−0.004**	−0.008**	−0.015***	−0.015**	−0.032**	−0.043	−0.088
	(0.000)	(0.001)	(0.002)	(0.003)	(0.005)	(0.007)	(0.016)	(0.027)	(0.064)
n	1,292	1,292	1,292	1,292	1,292	1,292	1,292	1,292	1,292
Pseudo R^2	0.128	0.128	0.135	0.138	0.141	0.138	0.142	0.141	0.143

Note: Standard errors in parentheses. The standard errors for quantile regression are obtained through 500 bootstrap replications. Pseudo R^2 is computed as the squared correlation between observed and predicted values of the dependent variable. Results are based on 2015 data. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$



(a) Spatial lag term (b) Income

FIGURE 7 2SLS and 2SQR coefficient estimates by quantile



(a) Spatial lag term (b) Income

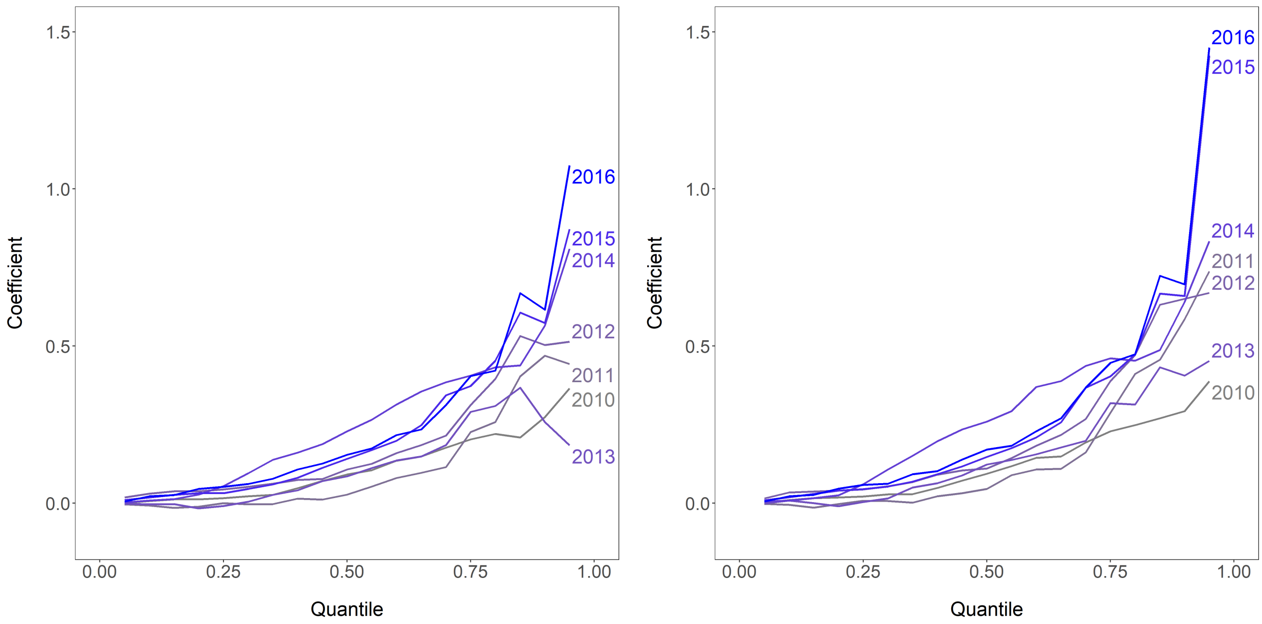
FIGURE 8 2SQR coefficient estimates by quantile

To ensure our results are robust to the choice of spatial weight matrix, we employ two alternative distance-based spatial weight matrices using a fixed distance band and inverse distance. The first one is a binary weight matrix based on a prespecified buffer zone. All census tracts within a prespecified distance receive a one in the weight matrix, with all others assigned zero. The distance is determined from the minimum distance such that every census tract has at least one neighbor. In our case, this distance was set to 3 miles between the centroids. The same prespecified distance is used to specify a inverse distance matrix. The inverse distance method allows neighbors located closer to each other to have higher weights than neighbors located far away. These weight matrices are also row-normalized. 2SLS results presented in Table 10 show that alternative spatial weight matrices give broadly the same estimates and significance levels. It is noteworthy that the two spatial weight matrices are sparser, i.e., most of its elements are assigned zero, thus spillover effects become somewhat less significant. 2SQR results obtained with both alternative specifications also reflect similar patterns in the coefficient estimates of the spatial lag term (Figure 9) and income (Figure 10) across quantiles to those obtained when using the queen contiguity.

TABLE 10 Robust 2SLS estimation results based on alternative weight matrices (dependent variable: squared returns)

	2010	2011	2012	2013	2014	2015	2016
Panel A: Fixed distance band							
Spatial lag term	0.426***	0.511***	0.627***	0.564***	0.659***	0.598***	0.660***
	(0.160)	(0.177)	(0.189)	(0.149)	(0.133)	(0.150)	(0.137)
Constant	2.022***	1.839***	1.189**	0.849**	0.775***	1.274***	1.031***
	(0.471)	(0.510)	(0.590)	(0.377)	(0.281)	(0.391)	(0.302)
Income	−0.180***	−0.163***	−0.106**	−0.075**	−0.069***	−0.114***	−0.092***
	(0.042)	(0.045)	(0.052)	(0.034)	(0.025)	(0.034)	(0.027)
n	1246	1279	1279	1284	1279	1292	1279
Pseudo R^2	0.109	0.088	0.028	0.028	0.134	0.109	0.149
Panel B: Inverse distance							
Spatial lag term	0.430**	0.574***	0.736***	0.573***	0.701***	0.643***	0.711***
	(0.173)	(0.207)	(0.226)	(0.168)	(0.135)	(0.161)	(0.156)
Constant	2.051***	1.451***	0.701	0.802*	0.709**	1.094**	0.787**
	(0.547)	(0.554)	(0.653)	(0.467)	(0.282)	(0.429)	(0.334)
Income	−0.182***	−0.129***	−0.063	−0.071*	−0.063**	−0.097***	−0.070**
	(0.049)	(0.049)	(0.058)	(0.042)	(0.025)	(0.038)	(0.029)
n	1246	1279	1279	1284	1279	1292	1279
Pseudo R^2	0.121	0.094	0.025	0.031	0.132	0.127	0.176

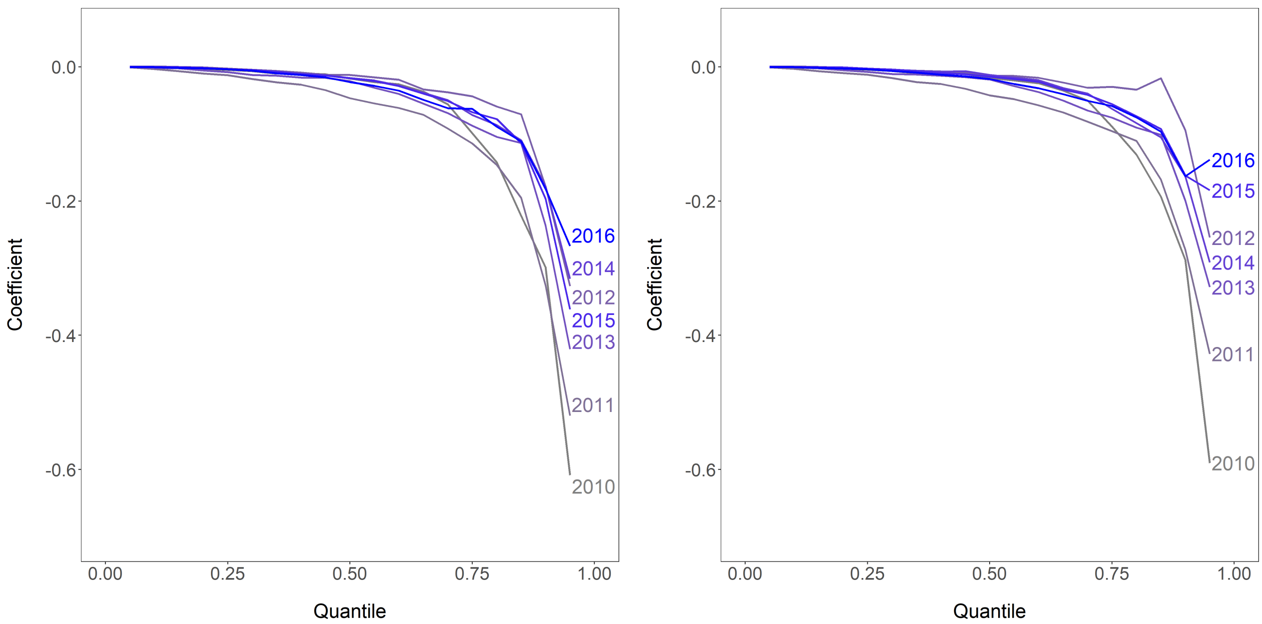
Note: Standard errors in parentheses. Pseudo R^2 is computed as the squared correlation between observed and predicted values of the dependent variable. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$



(a) Fixed distance band

(b) Inverse distance

FIGURE 9 2SQR coefficient estimates for spatial lag term by quantile



(a) Fixed distance band

(b) Inverse distance

FIGURE 10 2SQR coefficient estimates for income by quantile

7 Conclusion

In this paper, we examine the efficiency of housing markets in the spatial dimension, based on the extension of a temporal model for the EMH. The spatial dimension of the housing market efficiency is investigated through spatial ARCH and SQR models and empirical results obtained from those of estimated models are explained not only with general reasons but also with characteristics specific to the regional U.S. housing market. Matching with theoretical expectations, the empirical results suggest that while there is no significant spatial dependence in the level of returns, squared returns (volatility) exhibit significant spatial dependence. The presence of nonlinear spatial dependence in housing returns indicates that housing markets are spatially inefficient, thus the housing returns could contain useful information for predicting the returns of neighboring locations. Another finding is that the income variable is negatively correlated with squared returns, indicating that higher household income leads to lower volatility of housing returns.

Applying the spatial quantile regression technique, we find a clear upward trend of the quantile effects of the spatial dependence; this finding implies a varying degree of market inefficiency and that the housing returns of nearby areas more positively influence those of high-volatility areas than low-volatility areas. It is also shown that the coefficients on income are negative with a clear downward trend, suggesting that the decrease of household income generates housing returns that are more volatile in areas with higher housing returns. The findings of this paper should help housing policy makers make appropriate decisions to create a more stable and sustainable housing market by mitigating housing market volatility, particularly in disadvantaged areas.

Future explorations could explore several facets of the analysis that have received less attention. First, it would be necessary to build a more theoretical framework to reinforce the idea of spatial market efficiency. Second, as mentioned earlier, the robustness of identified spatial patterns along the distribution across different housing markets needs to be examined to determine whether those patterns can be generalized beyond the study area. Also, since this research is currently based on the cross-sectional analysis by focusing on the spatial interactions between neighborhoods, it would be interesting to examine what could be changed when using panel data analysis. The plan also includes applying different types of aggregation of the data as it is well recognized in the empirical literature that the degree of spatial dependence decreases with aggregation of spatial units. Finally it would also be interesting to see how well the model can predict the neighboring returns outside the study area.

References

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical analysis* 27, 93–115.

Baur, D. G., T. Dimpfl, and R. C. Jung (2012). Stock return autocorrelations revisited: A quantile regression approach. *Journal of Empirical Finance* 19, 254–265.

Bera, A. K., E. L. Bubnys, and H. Park (1988). Conditional and unconditional heteroscedasticity in the market model. *Financial Review* 23, 201–214.

Bera, A. K. and M. L. Higgins (1993). ARCH models: properties, estimation and testing. *Journal of economic surveys* 7, 305–366.

Bera, A. K. and P. Simlai (2005). Testing spatial autoregressive model and a formulation of spatial ARCH (SARCH) model with applications. Presented at the Econometric Society World Congress, London.

Bollerslev, T., R. F. Engle, and J. M. Wooldridge (1988). A capital asset pricing model with time-varying covariances. *Journal of political Economy* 96, 116–131.

Bracke, P. (2013). How long do housing cycles last? A duration analysis for 19 OECD countries. *Journal of Housing Economics* 22, 213–230.

Case, K. E. and R. J. Shiller (1989). The efficiency of the market for single-family homes.

Chernozhukov, V. and C. Hansen (2006). Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics* 132, 491–525.

Chiang, T. C., J. Li, and L. Tan (2010). Empirical investigation of herding behavior in Chinese stock markets: Evidence from quantile regression analysis. *Global Finance Journal* 21, 111–124.

Claessens, S., M. A. Kose, and M. E. Terrones (2011). Financial cycles: what? how? when? *International Seminar on Macroeconomics*. Volume 7, 303–344.

Clapp, J. M. and D. Tirtiroglu (1994). Positive feedback trading and diffusion of asset price changes: evidence from housing transactions. *Journal of Economic Behavior & Organization* 24, 337–355.

Dai, X., Z. Yan, M. Tian, and M. Tang (2019). Quantile regression for general spatial panel data models with fixed effects. *Journal of Applied Statistics*, 1–16.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1007.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance* 25, 383–417.

Gau, G. W. (1984). Weak form tests of the efficiency of real estate investment markets. *Financial Review* 19, 301–320.

Gu, A. (2002). The predictability of house prices. *Journal of Real Estate Research* 24, 213–234.

Guntermann, K. L. and R. L. Smith (1987). Efficiency of the market for residential real estate. *Land Economics* 63, 34–45.

Gupta, R. and S. M. Miller (2012). The time-series properties of house prices: a case study of the southern California market. *Journal of Real Estate Finance and Economics* 44, 339–361.

Gyourko, J. and R. Voith (1992). Local market and national components in house price appreciation. *Journal of Urban Economics* 32, 52–69.

Hartman-Glaser, B. and W. Mann (2017). *Collateral constraints, wealth effects, and volatility: Evidence from real estate markets*. Working paper.

Henricks, K., A. E. Lewis, I. Arenas, and D. G. Lewis (2018). A tale of three cities: The state of racial justice in Chicago report.

Kelejian, H. H. and I. R. Prucha (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics* 17, 99–121.

Kim, T.-H. and C. Muller (2004). Two-stage quantile regression when the first stage is based on quantile regression. *Econometrics Journal* 7, 218–231.

Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica* 46, 33–50.

Kostov, P. (2009). A spatial quantile regression hedonic model of agricultural land prices. *Spatial Economic Analysis* 4, 53–72.

Kuo, C.-L. (1996). Serial correlation and seasonality in the real estate market. *Journal of Real Estate Finance and Economics* 12, 139–162.

LeSage, J. P. and R. K. Pace (2010). *Spatial econometric models*. Springer, 355–376.

Liao, W.-C. and X. Wang (2012). Hedonic house prices and spatial quantile regression. *Journal of Housing Economics* 21, 16–27.

Linneman, P. (1986). An empirical test of the efficiency of the housing market. *Journal of Urban Economics* 20, 140–154.

Ma, L. and L. Pohlman (2008). Return forecasts and optimal portfolio construction: a quantile regression approach. *European Journal of Finance* 14, 409–425.

McMillen, D. P. (2012). *Quantile regression for spatial data*. Springer Science & Business Media.

Meen, G. (1999). Regional house prices and the ripple effect: a new interpretation. *Housing studies* 14, 733–753.

Meen, G. (2012). *Modelling spatial housing markets: Theory, analysis and policy*. Volume 2. Springer Science & Business Media.

Miao, H., S. Ramchander, and M. W. Simpson (2011). Return and volatility transmission in US housing markets. *Real Estate Economics* 39, 701–741.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* 59, 347–370.

Pagan, A. R. and G. W. Schwert (1990). Alternative models for conditional stock volatility. *Journal of econometrics* 45, 267–290.

Park, H. Y. and A. K. Bera (1987). Interest-rate volatility, basis risk and heteroscedasticity in hedging mortgages. *Real Estate Economics* 15, 79–97.

Pollakowski, H. O. and T. S. Ray (1997). Housing price diffusion patterns at different aggregation levels: an examination of housing market efficiency. *Journal of Housing Research* 8, 107.

Rayburn, W., M. Devaney, and R. Evans (1987). A test of weak-form efficiency in residential real estate returns. *Real Estate Economics* 15, 220–233.

Roberts, H. V. (1967). Statistical versus clinical prediction of the stock market. unpublished.

Schindler, F. (2013). Predictability and persistence of the price movements of the S&P/Case-Shiller house price indices. *Journal of Real Estate Finance and Economics* 46, 44–90.

Sewell, M. (2011). *Characterization of financial time series*. Research note. UCL Department of Computer Science.

Su, L. and Z. Yang (2011). Instrumental variable quantile estimation of spatial autoregressive models.

Tirtiroglu, D. (1992). Efficiency in housing markets: Temporal and spatial dimensions. *Journal of Housing Economics* 2, 276–292.

Tsai, I.-C. (2012). The relationship between stock price index and exchange rate in Asian markets: A quantile regression approach. *Journal of International Financial Markets, Institutions and Money* 22, 609–621.

Veronesi, P. (1999). Stock market overreactions to bad news in good times: a rational expectations equilibrium model. *Review of Financial Studies* 12, 975–1007.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* 50, 1–25.

Zhang, L. (2016). Flood hazards impact on neighborhood house prices: A spatial quantile regression analysis. *Regional Science and Urban Economics* 60, 12–19.

Zhang, L. and T. Leonard (2014). Neighborhood impact of foreclosure: A quantile regression approach. *Regional Science and Urban Economics* 48, 133–143.

Zhu, B., R. Füss, and N. B. Rottke (2013). Spatial linkages in returns and volatilities among US regional housing markets. *Real Estate Economics* 41, 29–64.

Zietz, J., E. N. Zietz, and G. S. Sirmans (2008). Determinants of house prices: a quantile regression approach. *Journal of Real Estate Finance and Economics* 37, 317–333.