# LM Tests for Heterogeneous Spatial Correlations with Application in Housing Market

Yang Yang

September 14, 2021

**Abstract**

We consider a spatial autoregressive model with heterogeneous coefficients associated with different type of agents, and develop two different Lagrange multiplier tests, for existence of spatial correlation and the heterogeneity of spatial correlation. With proper assumptions, the two test statistics both have nice asymptotic distributions. For relatively small sample, Monte Carlo simulations show good performance when using asymptotic critical values. Especially, when only part of the individuals are affected by their neighbors, our test for existence of spatial correlation is much more sensitive than the Moran I test. In our empirical application, we investigate the short-run relationship between city size and housing price in northeastern US. From 2006 to 2014, both direct city size effect and neighborhood spillovers to large city areas are dynamic which are potentially driven by credit cycle and heterogeneity among regional income distribution.

## 1 Introduction

In spatial econometrics and statistics, to test the existence of spatial correlation is one of the core problem. The most popular and widely used test is Moran I test which is given by Moran (1950) and Cliff and Ord (1973). In Kelejian and Prucha (2001), the asymptotic property of the Moran I statistic has been investigated in a spatial autoregressive setting. Under some particular regularity conditions for the spatial correlations, the Moran I test statistic is asymptotically Normal.

However, traditional spatial econometrics model, such like spatial autoregressive model, does not consider the heterogeneity across individuals and the induced heterogeneity among their interactions with neighbors. Consequently, as a benchmark test, Moran I may not suitable for this scenario. In recent econometrics literature, there are some researchers begin to introduce heterogeneous coefficient into spatial models, e.g. LaSage and Chih (2016), LaSage, Vance and Chih (2017). Not only theoretical interesting, but also empirically, heterogeneous social network and spillover effect is also an import question in many areas. For example, Matvos and Ostrovsky (2010) try to investigate the heterogeneity and peer effects in mutual fund proxy voting, Yakusheva, Kapinos and Eisenberg (2014) and Patacchini, Rainone and Zenou (2017) try to investigate heterogeneous peer effect in students' behaviors and education. To incorporate with the growing interest in empirical research, it is necessary to develop suitable tests as a replacement of Moran I, as well as distinguishing with the traditional homogeneous situation. For spatial econometrics, a pre-estimation test is even more important since the computational burden is increasing fast as the sample size increases. Even for implement the traditional SAR model by maximum likelihood approach, as we need to optimize a nonlinear function with inverse of $n \times n$ matrix, in large sample applications such as individual level social network with more than 10,000 people, the optimization may take several hours or even days to get the estimator. The heterogeneous case makes it even worse. Thus, pre-estimation test is a good way to determine which model you should use and save valuable time on waiting for results.

In this paper, we develop two Lagrange multiplier tests based on a heterogeneous coefficient spatial autoregressive model formation. One is constructed by least square estimator without the spatial autoregression term, to test the existence of spatial correlation. The other one is constructed by QMLE of SAR model, to test whether the heterogeneity exits in the spatial autocorrelation term. The asymptotic Normality of the test statistics is also proved with Monte Carlo simulation study for finite sample performance. We also provide an empirical example about housing market is given. By investing the housing market in north eastern US from 2006 to 2014, we found a time-varying correlation between city size and housing price, as well as the externality from neighbor areas received. The credit cycle have different impacts to people with different income level, which is a source of the size heterogeneity on housing markets due to the uneven geographic distribution of income. By comparing with post-estimation tests, our tests work well as a pre-estimation benchmark for heterogeneous spatial correlations.

In the following part of this paper, Section 2 is a brief introduction to heterogeneous coefficient spatial autoregressive model. Section 3 is the LM test for existence of spatial correlation. Section 4 is the LM test for heterogeneity among the spatial correlation. Section 5 shows Monte Carlo simulation results of test performance for finite samples, besides some comparisons with Moran I test. Finally, an application is given in Section 6.

# 2 Heterogeneous Coefficient Spatial Autoregressive Model

## 2.1 Data Generating Process

Suppose $n$ individual spatial unites in an economy are located in a region $D_n \subset \mathbb{R}^d$, where the cardinality of $D_n$ is $|D_n| = n$. For convenience, we name these $n$ units as $1, 2, \cdots, n$. The distance between individuals $i$ and $j$ is denotes by $d_{ij}$. For regularity, we need the following assumption:

**Assumption 1:**
*$d_{ij} \geq 1$ for any $i \neq j$.*

**Assumption 2:**

*There are $K$ sub-regions $\left\{D_n^k\right\}_{k=1}^K$ inside $D_n$, which satisfy $\cup_{k=1}^k D_n^k = D_n$ and $D_n^i \cap D_n^j = 0$ for $i \neq j$. $K$ is a constant which does not depend on $n$.*

Here $\left\{D_n^k\right\}_{k=1}^K$ may not be consecutive, which means the individuals inside each $D_n^m$ may have no correlation between each other. For example, in urban economics setting, we may divide counties into metropolitan and micropolitan areas. Another common example in international trade might be developing countries and developed countries. Thus, the sub-regions are divided by different categories of individuals.

The data generating process is the following:

$$y_i = \sum_{k=1}^K \lambda_k h_{i,k}\left(\sum_{j=1}^n w_{ij,n} y_j\right) + x_i^{'}\beta + u_i \tag{1}$$

where $h_{i.k} = \begin{cases} 1 & i \in D_n^k \\ 0 & i \notin D_n^k \end{cases}$ and $u_i \overset{iid}{\sim} \left(0, \sigma^2\right)$. $\lambda_k$ captures the spill-over effect from neighbors for individual $i \in D_n^k$, and $\beta$ captures effects from external regressors $x_i$. Since $h_{i,k}$ is a dummy variable, $h_{i,k}\left(\sum_{j=1}^n w_{ij,n} y_j\right)$ captures the cross-effect of dummy and neighbor regions.

Denote $W_n = \left(w_{ij,n}\right)_{n\times n}$, $y_n = \left(y_{1,n}, \cdots, y_{n,n}\right)^{'}$, $H_{n,k} = diag\left(d_{1,k}, \cdots, d_{n,k}\right)$, $X_n = \left(x_1^{'}, \cdots, x_n^{'}\right)^{'}$, $u_n = \left(u_1, \cdots, u_n\right)^{'}$, then our model can be written as the following matrix form:

$$y_n = \sum_{k=1}^K \lambda_k H_{n,k} W_n y_n + X_n\beta + u_n \tag{2}$$

Easy to see, when $\lambda_k$'s are equal, then it reduces to a mixed SAR model since $\sum_{k=1}^K H_{n,k} W_n = W_n$.

## 2.2 Economic Foundation

Similar to the SAR model, when introducing categorical heterogeneity, it can still be motivated by game theory. It can be regarded as a model on the Nash equilibrium of a static complete information game of different types of players processing with linear-quadratic utilities. Suppose there are $n$individuals, and they choose their actions to maximize their utilities and there are $K$ types of individuals. Let the action for individual $i$with type $k$ be $y_{ni}$, and its cost equals to $\frac{y_{in}^2}{2}$. Suppose individual $i$'s benefit from his action is promotional to his action, and it depends on his characteristics, his type and other individuals' action: $y_{ni}\left(\lambda_k \sum_{j=1}^n w_{ij,n} y_{nj} + x_{ni}\beta + u_{ni}\right)$, which can be substitute or complementary depending on the sign of $\lambda_k$. Then his utility is

$$u_i\left(y_{ni}\right) = y_{ni}\left(\lambda_k \sum_{j=1}^n w_{ij,n} y_{nj} + x_{ni}\beta + v_{ni}\right) - \frac{y_{ni}^2}{2} \tag{3}$$

where $Y_{-i,n} = \left(y_{ni}, \cdots, y_{n,i-1}, y_{n,i+1}, \cdots, y_{n,n}\right)^{'}$, $x_{ni}$ and $v_{ni}$ are known to all individuals. An individual $i$maximizes utility with respect to $y_{ni}$ :

$$\max_{y_{ni}} U_i\left(y_{ni}|Y_{-i,n}\right)$$

where $U_i\left(y_{ni}|Y_{-i,n}\right) = u\left(y_{ni}\right)$. The optimal action for $i$ will be characterized by the optimization of his utility with $y_{ni}$. When the system of equations has a solution, the solution is a Nash equilibrium of this game.

We can also motivate this model by a social interaction setting, where one may have a private and social components in utility:

$$u_i\left(y_{ni}\right) = y_{ni}\left(x_{ni}\beta + v_{ni}\right) - \frac{1}{2}\left(y_{ni} - \lambda_k \sum_{j=1}^{n} w_{ij,n} y_{nj}\right)^2 \tag{4}$$

where the first component represents private utility associated with an action $y_{ni}$ and the second component captures a conformity effect with friends. Unlike the SAR situation in Brock and Durlauf (2001), the conformity effect does depend on the type of $i$ itself. For example, when considering risk spread through mortgages, rich people and poor people will have different financial behaviors with their friends and families due to their budget constraints. Another example, when considering Covid-19 spread, well-educated people and less-educated people may have different responds on surrounding peoples' protection behaviors, e.g. wearing masks and washing hands, so the situation of rich and poor community may have totally distinct infection rate, mortality rate and economic impact. Gender may also be an important source of heterogeneity. Both (3) and (4) type of utilities would give the heterogenous coefficient spatial autoregressive model (HSAR).

## 2.3 Likelihood Function and First Order Conditions

Rewrite (2) into as $\left(I_n - \sum_{k=1}^{K} \lambda_k H_{n,k} W_n\right) y_n = X_n\beta + u_n$, when $\left(I_n - \sum_{k=1}^{K} \lambda_k H_{n,k} W_n\right)^{-1}$ exists, we can transform the model into the following reduced form:

$$y_n = \left(I_n - \sum_{k=1}^{K} \lambda_k H_{n,k} W_n\right)^{-1} \left(X_n\beta + u_n\right)$$

Then, similar to traditional SAR model, we can write down the log-likelihood function of $y_n$ when $u_n \sim N\left(0, \sigma^2 I_n\right)$:

$$\begin{aligned}\ln L_n\left(\Lambda', \beta, \sigma^2\right) = & -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 + \ln\left|S_n(\mathit{\Lambda})\right| \\ & -\frac{1}{2\sigma^2}\left(S_n(\Lambda)y_n - X_n\beta\right)'\left(S_n(\Lambda)y_n - X_n\beta\right)\end{aligned}$$

where $\Lambda = \left(\lambda_1, \cdots, \lambda_k\right)'$ and $S_n(\Lambda) = I_n - \sum_{k=1}^{K} \lambda_k H_{n,k} W_n$.

As $\left(H_{n,k}W_n\right)_{ij} = \begin{cases} w_{ij} & i = k \\ 0 & i \neq k \end{cases}$, we have

$$\left\| \sum_{k=1}^{K} \lambda_k H_{n,k} W_n \right\|_\infty \leq \max_k \left|\lambda_k\right| \left\|W_n\right\|_\infty$$

Thus, a possible parameter space of $\lambda$ would be $\max_k |\lambda_k| < \frac{1}{\|W_n\|_\infty}$, when $\|W_n\|_\infty < \infty$ is the row-sum norm for $W_n$.

The first order conditions are

$$\begin{aligned}\frac{\partial \ln L_n(\theta)}{\partial \lambda_k} &= -\frac{1}{\sigma^2}\left(S_n(\Lambda)y_n - X_n\beta\right)' \frac{\partial S_n(\Lambda)}{\partial \lambda_k} y_n + tr\left(S_n^{-1}(\Lambda)\frac{\partial S_n(\Lambda)}{\partial \lambda_k}\right) \\ &= \frac{1}{\sigma^2}\left[S_n(\Lambda)y_n - X_n\beta\right]' H_{n,k}W_n y_n - tr\left[S_n^{-1}(\Lambda)H_{n,k}W_n\right]\end{aligned}$$

$$\frac{\partial \ln L_n(\theta)}{\partial \beta'} = \frac{1}{\sigma^2}X_n'\left(S_n(\Lambda)y_n - X_n\beta\right)$$

$$\frac{\partial \ln L_n(\theta)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\left(\sigma^2\right)^2}\left(S_n(\Lambda)y_n - X_n\beta\right)'\left(S_n(\Lambda)y_n - X_n\beta\right)$$

where $\theta = \left(\Lambda', \beta, \sigma^2\right)'$. Based on the first order conditions, we can construct different tests for different purposes.

# 3 Test for Existence of Spatial Correlation

## 3.1 Construct the LM statistic

To construct LM tests and discuss the asymptotic distribution, we need to put some basic regularity assumptions first.

**Assumption 3:**
*For $\forall k = 1, \cdots, K$, we have $\lim_{n\to\infty} \frac{|D_n^k|}{n} = c_k$ where $c_k$ is a non-zero positive constant and $\sum_{k=1}^K c_k = 1$, i.e. there exist a stationary distribution of types as n$\to \infty$ and the probability of each type would not shrink to zero.*

**Assumption 4:**
*$\{u_{i,n}\}_{i\in D_n}$ are i.i.d with mean zero and variance $\sigma^2$. Its moment $E\left(|u|^{4+\gamma}\right)$ for some $\gamma > 0$ exits.*

**Assumption 5:**
*$w_{ij,n}$ are at most of order $h_n^{-1}$, denoted by $O(1/h_n)$ uniformly in all $i$ and $j$, i.e. for some real constant $c$, there exists a finite integer $N$, such that for all $n \geq N$, $|h_n w_{ij,n}| < c$ for all $i$ and $j$. The rate sequence $\{h_n\}$ can be bounded or divergent. As a normalization, $w_{ii,n} = 0$.*

**Assumption 6:**
*The ratio $h_n/n \to 0$ as $n \to \infty$.*

**Assumption 7:**

*The matrix $S_n$ is nonsingular.*

**Assumption 8:**
*The sequences of matrices $\{W_n\}$ and $\left\{S_n^{-1}(\Lambda_0)\right\}$ are uniformly bounded in both row and column sums.*

**Assumption 9:**
*The elements of $X_n$ are uniformly bounded constants for all $n$. The $\lim_{n\to\infty} \frac{1}{n} X_n^{'} X_n$ exists and is nonsingular.*

**Assumption 10:**
*$\left\{S_n^{-1}(\Lambda)\right\}$ are uniformly bounded in either row or column sums, uniformly in $\lambda$ in a compact parameter space $\Theta$. The true $\lambda_0$ is in the interior of $\Theta$.*

Assumption 3 is a regularity assumption for the heterogeneity structure. If there is no stationary limiting distribution of different types, our discussion for asymptotic situation will be meaningless which will be showed in detail later. Also, we require Assumption 4-10 are similar to Assumption 1-7 in Lee (2004) which established the asymptotic theory for MLE and QMLE for SAR model, despite $\lambda$ is a vector instead of a scalar. So, the form of $S_n(\lambda)$ also changed a little bit as defined in Section 2.3. The parameter space $\mathit{\Lambda}$ can be the one we discussed before. Detailed interpretation of these assumptions is in Section 2 of Lee (2004). Additional regularity conditions will be showed subsequently when needed.

With considering heterogeneity, traditional Moran I test may not work well since it assume homogeneous correlation among different types of regions. To test whether there exist spatial correlation, we need to test whether all the types of regions are affected by neighbors. Thus, we need to test $H_0 : \lambda_k = 0 \; for \; \forall k = 1, \cdots, K$. The alternative becomes $H_1 : \exists k, \; \lambda_k \neq 0$. Under $H_0$, the HSAR model becomes a linear regression model, thus least square estimator and MLE are consistent and $\sqrt{n}-$convergence. The first order derivatives become

$$\begin{aligned}\frac{\partial \ln L_n\left(0,\beta,\sigma^2\right)}{\partial \lambda_k} &= \frac{1}{\sigma^2}\left(y_n - X_n\beta\right)^{'} H_{n,k} W_n y_n - tr\left(H_k W_n\right) \\ &= \frac{1}{\sigma^2}\left(y_n - X_n\beta\right)^{'} H_{n,k} W_n y_n\end{aligned}$$

$$\frac{\partial \ln L_n\left(0,\beta,\sigma^2\right)}{\partial \beta^{'}} = \frac{1}{\sigma^2} X_n^{'}\left(y_n - X_n\beta\right)$$

$$\frac{\partial \ln L_n\left(0,\beta,\sigma^2\right)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\left(\sigma^2\right)^2}\left(y_n - X_n\beta\right)'\left(y_n - X_n\beta\right)$$

Let $\hat{\theta} = \left(0, \hat{\beta}^{'}, \hat{\sigma}^2\right)^{'}$ be the MLE of linear regression model, we should have

$$\frac{\partial \ln L_n\left(\hat{\theta}\right)}{\partial \lambda_k} = \frac{1}{\hat{\sigma}^2}\left(y_n - X_n\hat{\beta}\right)^{'} H_{n,k}W_n y_n = \frac{1}{\hat{\sigma}^2}\hat{u}^{'} H_{n,k}W_n y_n$$

$$\frac{\partial \ln L_n\left(\hat{\theta}\right)}{\partial \beta'} = \frac{1}{\hat{\sigma}^2}X_n^{'}\left(y_n - X_n\hat{\beta}\right) = \frac{1}{\hat{\sigma}^2}X_n^{'}\hat{u}_n$$

$$\begin{aligned}\frac{\partial \ln L_n\left(\hat{\theta}\right)}{\partial \sigma^2} &= -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2\left(\hat{\sigma}^2\right)^2}\left(y_n - X_n\hat{\beta}\right)^{'}\left(y_n - X_n\hat{\beta}\right) \\ &= -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2\left(\hat{\sigma}^2\right)^2}\hat{u}_n^{'}\hat{u}_n\end{aligned}$$

Since $\frac{\partial \ln L_n(\hat{\theta})}{\partial \beta} = 0$ and $\frac{\partial \ln L_n(\hat{\theta})}{\partial \sigma^2} = 0$, we have $\sigma^2 = \frac{1}{n}\hat{u}^{'}\hat{u}$ and $X_n^{'}\hat{u} = 0$. Then to test $H_0$, we need to focus on the following score functions:

$$g_{k,n}\left(\hat{\theta}\right) \equiv \frac{\partial \ln L_n\left(\hat{\theta}\right)}{\partial \lambda_k} = \frac{1}{\hat{\sigma}^2}\hat{u}_n^{'}H_{n,k}W_n y_n$$

Next, we will use these score functions to construct an LM test statistic. From the FOCs derived in Section 2.3, we can easily get the second order conditions:

$$\frac{\partial^2 \ln L_n(\theta)}{\partial \lambda_k^2} = -\frac{1}{\sigma^2}\left(H_{n,k}W_n y_n\right)^{'} H_{n,k}W_n y_n - tr\left[\left(S_n(\lambda)^{-1} H_{n,k}W_n\right)^2\right]$$

$$\begin{aligned}\frac{\partial^2 \ln L_n(\theta)}{\partial \lambda_h \partial \lambda_k} &= -\frac{1}{\sigma^2}\left(H_{n,l}W_n y_n\right)^{'} H_{n,k}W_n y_n \\ &\quad - tr\left[S_n(\lambda)^{-1} H_{n,l}W_n S_n(\lambda)^{-1} H_{n,k}W_n\right] \\ &= -tr\left[S_n(\lambda)^{-1} H_{n,l}W_n S_n(\lambda)^{-1} H_{n,k}W_n\right]\end{aligned}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial \beta' \partial \lambda_k} = -\frac{1}{\sigma^2}X_n^{'}H_{n,k}W_n y_n$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial \sigma^2 \partial \lambda_k} = -\frac{1}{\left(\sigma^2\right)^2}\left(S_n(\lambda)y_n - X_n\beta\right)^{'} H_{n,k}W_n y_n$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial \beta' \partial \beta} = -\frac{1}{\sigma^2}X_n^{'}X_n$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial \sigma^2 \partial \beta} = -\frac{1}{\left(\sigma^2\right)^2}X_n^{'}\left(S_n(\lambda)y_n - X_n\beta\right)$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial \left(\sigma^2\right)^2} = \frac{n}{2\left(\sigma^2\right)^2} - \frac{1}{\left(\sigma^2\right)^3}\left(S_n(\lambda)y_n - X_n\beta\right)'\left(S_n(\lambda)y_n - X_n\beta\right)$$

since $H_l H_k = 0$ for $\forall l \neq k$ and $\frac{\partial S_n^{-1}(\Lambda)}{\partial \lambda_k} = -S_n^{-1}(\Lambda)\frac{\partial S_n(\Lambda)}{\partial \lambda_k}S_n^{-1}(\Lambda)$. Under $H_0$, from the first order condition, we also have $\frac{\partial^2 \ln L_n(\hat{\theta})}{\partial\sigma^2\partial\beta} = -\frac{1}{(\hat{\sigma}^2)^2} X_n' \left(y_n - X_n\hat{\beta}\right) = 0$. As $H_k$'s are idempotent, the remained non-zero terms in Hessian matrices are

$$\frac{\partial^2 \ln L_n\left(\hat{\theta}\right)}{\partial\lambda_k^2} = -\frac{1}{\hat{\sigma}^2} y_n' W_n' H_{n,k} W_n y_n - tr\left[\left(H_{n,k}W_n\right)^2\right]$$

$$\frac{\partial^2 \ln L_n\left(\hat{\theta}\right)}{\partial\lambda_h\partial\lambda_k} = -tr\left(H_{n,l}W_nH_{n,k}W_n\right)$$

$$\frac{\partial^2 \ln L_n\left(\hat{\theta}\right)}{\partial\beta'\partial\lambda_k} = -\frac{1}{\hat{\sigma}^2} X_n' H_{n,k} W_n y_n$$

$$\frac{\partial^2 \ln L_n\left(\hat{\theta}\right)}{\partial\sigma^2\partial\lambda_k} = -\frac{1}{\left(\sigma^2\right)^2} \hat{u}_n' H_{n,k} W_n y_n$$

$$\frac{\partial^2 \ln L_n\left(\hat{\theta}\right)}{\partial\beta'\partial\beta} = -\frac{1}{\hat{\sigma}^2} X_n' X_n$$

$$\frac{\partial^2 \ln L_n\left(\hat{\theta}\right)}{\partial\left(\hat{\sigma}^2\right)^2} = \frac{n}{2\left(\hat{\sigma}^2\right)^2} - \frac{1}{\left(\hat{\sigma}^2\right)^3}\hat{u}_n'\hat{u}_n$$

By likelihood equation $E_\theta\left(\frac{\partial^2 \ln L_n(\theta)}{\partial\theta\partial\theta'}\right) + E_\theta\left(\frac{\partial \ln L_n(\theta)}{\partial\theta}\frac{\partial \ln L_n(\theta)}{\partial\theta'}\right) = 0$, as MLE of linear regression model is $\sqrt{n}-$ convergence, we can construct the following LM statistic to test $H_0$:

$$LM_1 = -g_n\left(\hat{\theta}\right)' E\left(\frac{\partial^2 \ln L_n\left(\hat{\theta}\right)}{\partial\theta\partial\theta'}\right)^{-1} g_n\left(\hat{\theta}\right) \tag{5}$$

where $g_n\left(\hat{\theta}\right) = \frac{\partial \ln L_n(\hat{\theta})}{\partial\theta} = \left(g_{1,n}\left(\hat{\theta}\right), \cdots, g_{K,n}\left(\hat{\theta}\right), 0, \cdots, 0\right)'$ with FOCs for parameters other than $\lambda_k$ are zeros. In the next section, we will derive the asymptotic distribution of $LM_1$.

## 3.2 Asymptotic Distribution of $LM_1$

To derive the asymptotic distribution of $LM_1$, we need to derive the asymptotic distribution of the score function $g_n\left(\hat{\theta}\right)$. Let $a = \left(a_1, \cdots, a_K\right)'$ be an arbitrary vector of real numbers, then we have

$$f_n\left(a, \hat{\theta}\right) = \sum_{k=1}^{K} a_k g_n\left(\hat{\theta}\right) = \sum_{k=1}^{K} a_k \frac{1}{\hat{\sigma}^2}\hat{u}_n' H_{n,k} W_n y_n = \frac{1}{\hat{\sigma}^2}\hat{u}_n' H_{a,n} W_n y_n$$

where $H_{a,n} = \sum_{k=1}^{K} a_k H_{n,k}$ which is a diagonal matrix with diagonal elements $h_{a,n,ii} = a_k$ for $i \in D_n^k$. Next, we will prove the asymptotic Normality of $\frac{1}{\sqrt{n}} f_n$ which implies jointly asymptotic Normality of $\frac{1}{\sqrt{n}} g_n\left(\hat{\theta}\right)$ despite the trivial cases when $a = 0$.

Since $\hat{u}_n = \left[I - X_n \left(X_n^{'} X_n\right)^{-1} X_n\right] u_n \equiv M_n u_n$, $y_n = X_n \beta_0 + u_n$, we have

$$\begin{aligned}
\hat{u}_n^{'} H_{a,n} W_n y_n &= u_n^{'} M_n H_{a,n} W_n \left(X_n \beta_0 + u_n\right) \\
&= u_n^{'} M_n H_{a,n} W_n X_n \beta_0 + u_n^{'} M_n H_{a,n} W_n u_n \\
&= u_n^{'} H_{a,n} W_n X_n \beta_0 + u_n^{'} X_n \left(X_n^{'} X_n\right)^{-1} X_n^{'} H_{a,n} W_n X_n \beta_0 \\
&+ u_n^{'} H_{a,n} W_n u_n + u_n^{'} X_n \left(X_n^{'} X_n\right) X_n^{'} H_{a,n} W_n u_n
\end{aligned} \tag{6}$$

For each $a$, denote $\widetilde{W}_n(a) = H_{a,n} W_n$, then each element $\widetilde{w}_{a,n,ij} = a_k w_{ij,n}$ for for $i \in D_n^k$. Thus, For a given real vector $a$, $\widetilde{w}_{a,n,ij}$ are also $O(1/h_n)$ uniformly for all $i$ *and* $j$ follows Assumption 5 since $a$is a finite vector. $\left\{\widetilde{W}_n(a)\right\}$ is also uniformly bounded in row and column sums follows Assumption 8.

To continue our discussion, we need the following additional assumption:

**Assumption 11:**
*For each $a$, $\lim_{n\to\infty} \frac{1}{n} \left(H_{a,n} W_n X_n \beta_0\right)^{'} H_{a,n} W_n X_n \beta_0$ exists.*

By Assumption 9 and 11, we have

$$\begin{aligned}
&\frac{1}{\sqrt{n}} u_n^{'} X_n \left(X_n^{'} X_n\right)^{-1} X_n^{'} H_{a,n} W_n X_n \beta_0 \\
=& \frac{u_n^{'} X_n}{n} \left(\frac{X_n^{'} X_n}{n}\right)^{-1} \frac{1}{\sqrt{n}} X_n^{'} H_{a,n} W_n X_n \beta_0 \\
=& o_p(1)
\end{aligned}$$

and

$$\begin{aligned}
&\frac{1}{\sqrt{n}} u_n^{'} X_n \left(X_n^{'} X_n\right)^{-1} X_n^{'} H_{a,n} W_n u_n \\
=& \frac{u_n^{'} X_n}{n} \left(\frac{X_n^{'} X_n}{n}\right)^{-1} X_n H_{a,n} W_n \frac{1}{\sqrt{n}} u_n \\
=& o_p(1)
\end{aligned}$$

Thus from equation (6), we have

$$
\begin{aligned}
\frac{1}{\sqrt{n}} f_n\left(a, \hat{\theta}\right) &= \frac{1}{\hat{\sigma}^2 \sqrt{n}} \left[ \left( \widetilde{W}_n(a) X_n \beta_0 \right)^{'} u_n + u_n^{'} \widetilde{W}_n(a) X_n u_n \right] + o_p(1) \\
&\equiv \frac{1}{\hat{\sigma}^2 \sqrt{n}} \left( A_n^{'} u_n + u_n^{'} B_n u_n \right) + o_p(1) \qquad (7)
\end{aligned}
$$

where $A_n(a) = \widetilde{W}_n(a) X_n \beta_0$ and $B_n(a) = \frac{1}{2}\left[\widetilde{W}_n(a) + \widetilde{W}_n^{'}(a)\right]$. Since $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$, by Slutsky's theorem, the remaining work is to discuss the limiting distribution of the following linear-quadratic form:

$$
Q_n\left(a, \hat{\theta}\right) = A_n^{'} u_n + u_n^{'} B_n u_n
$$

By Assumption 5, there might be two cases: $\{h_n\}$ is bounded or $\lim_{n\to\infty} h_n = \infty$. When $\{h_n\}$ is bounded, by Assumption 8 and Assumption 9, $A_n$ and $B_n$ are also uniformly bounded in column and row sum. Thus, by Assumption 4, we can apply central limit theorem for linear-quadratic forms in Kelejian and Prucha (2001) on $Q_n\left(a, \hat{\theta}\right)$, and get

$$
\frac{Q_n\left(a, \hat{\theta}\right) - E\left[Q_n\left(a, \hat{\theta}\right)\right]}{\sigma_{Q_n}} \xrightarrow{d} N(0,1)
$$

where $\sigma_{Q_n}$ is the variance of $Q_n\left(a, \hat{\theta}\right)$. This result will imply asymptotic Normality of $\frac{1}{\sqrt{n}} f_n\left(a, \hat{\theta}\right)$.

When $\lim_{n\to\infty} h_n = \infty$, $\frac{1}{\sqrt{n}} A_n^{'} u_n$ will dominate $\frac{1}{\sqrt{n}} u_n^{'} B_n u_n$ by Assumption 11. As $\widetilde{w}_{a,n,ij}$ are also $O(1/h_n)$ as we proved before, we have

$$
var\left(\frac{1}{\sqrt{n}} u_n^{'} B_n u_n\right) = o\left(\frac{1}{h_n}\right) \Rightarrow \frac{1}{\sqrt{n}} u_n^{'} B_n u_n = o_p(1)
$$

Simultaneously, we have $\frac{1}{\sqrt{n}} A_n^{'} u_n = O_p(1)$. Thus, by applying Lyapunov CLT's central limit theorem on $\frac{1}{\sqrt{n}} A_n^{'} u_n$ , we can also get the asymptotic Normality of $\frac{1}{\sqrt{n}} f_n\left(a, \hat{\theta}\right)$. As $a$is an arbitrary real vector, we can get the jointly asymptotic Normality of $\frac{1}{\sqrt{n}} g_n\left(\hat{\theta}\right)$.

Since we have $K$ constraints, we should have $K$ degree of freedom. However, without Assumption 3, we may have degenerate issue in some cases. Let's focus on the situation when $\lim_{n\to\infty} h_n = \infty$. Easy to see, there are only $\left|D_n^k\right|$ non-zero terms in vector $H_{a,n} W_n X_n \beta_0$ and each term is $O(n/h_n)$ by Assumption 5 and 9. WLOG, reorder the observations and let $H_{n,k} = \begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix}$. Then, we have

$$
\frac{1}{\sqrt{n}} u_n^{'} H_{a,n} W_n X_n \beta_0 = \sqrt{\frac{\left|D_n^k\right|}{n}} \frac{1}{\sqrt{\left|D_n^k\right|}} \sum_{i=1}^{\left|D_n^k\right|} \left(H_{a,n} W_n X_n \beta_0\right)_{i,n} u_{i,n}
$$

If $\lim_{n\to\infty} \frac{|D_{n,k}|}{n} = 0$, then we will have $\frac{1}{\sqrt{n}} u_n^{'} H_{a,n} W_n X_n \beta_0 \xrightarrow{p} 0$, its no longer $O_p(1)$. Also, if $\lim_{n\to\infty} \frac{|D_{n,k}|}{n} = 0$ does not exist, then the limiting distribution of $\frac{1}{\sqrt{n}} u_n^{'} H_{a,n} W_n X_n \beta_0$ also does not exist. Then, as we stated before, when $\lim_{n\to\infty} h_n = \infty$, this term will dominate the asymptotic distribution of $\frac{1}{\sqrt{n}}g_{k,n}\left(\hat{\theta}\right)$, we will either have degenerate issue or diverge issue. Thus, Assumption 3 is necessary to regulate our model and make sure each score function have proper asymptotic distribution. With all above assumptions hold, we can get that

$$LM_1 = -g_n\left(\hat{\theta}\right)' E\left(\frac{\partial^2 \ln L_n\left(\hat{\theta}\right)}{\partial\theta\partial\theta'}\right)^{-1} g_n\left(\hat{\theta}\right) \xrightarrow{d} \chi^2(K)$$

# 4 Test for Heterogeneity among Spatial Correlation

## 4.1 Construct the LM statistic

Besides the existence of the spatial heterogeneity, we are also interested in whether the heterogeneity exist among the spatial correlation. Thus, we want test $H_0 : \rho_1 = \cdots = \rho_K$, and the alternative is $H_1 : \rho_i \neq \rho_j, \exists i \neq j$. If $H_0$ is true, then the HSAR is reduced to a SAR model. Lee (2004) had proved the consistency and asymptotic Normality of QMLE of SAR. Here we just consider the regular mixed regression case discussed Section 3 in Lee (2004) without including $W_nX_n$ term into the regression. To make sure the QMLE of SAR exist, we need the following assumption:

**Assumption 12:**
*The* $\lim_{n\to\infty}\frac{1}{n}(X_n, G_nX_n\beta_0)'(X_n, G_nX_n\beta_0)$ *exists and is non-singular.*

where $G_n = W_nS_n^{-1}(\Lambda_0)$ and $\Lambda_0 = \left(\underbrace{\lambda_0, \cdots, \lambda_0}_{K}\right)'$. $\lambda_0$ is the true parameter of the SAR model. This assumption makes sure $G_nX_n\beta_0$ and $X_n$ are not asymptotically multi-collinear, and it is a sufficient condition for global identification for the SAR model. Let $\bar{\theta} = \left(\bar{\Lambda}', \bar{\beta}', \bar{\sigma}^2\right)'$ be the constrained estimator of HSAR under $H_0$, where $\left(\bar{\lambda}, \bar{\beta}, \bar{\sigma}^2\right)$ is the QMLE for SAR and $\bar{\Lambda} = \left(\underbrace{\bar{\lambda}, \cdots, \bar{\lambda}}_{K}\right)'$. Under Assumption 1-10 and 12, QMLE of SAR is consistent and $\sqrt{n}-$ convergence.

Then, we have

$$\begin{aligned}\frac{\partial \ln L_n\left(\bar{\theta}\right)}{\partial\lambda_k} &= \frac{1}{\bar{\sigma}^2}\left[S_n\left(\bar{\Lambda}\right)y_n - X_n\bar{\beta}\right]' H_{n,k}W_ny_n - tr\left[S_n^{-1}\left(\bar{\Lambda}\right)H_{n,k}W_n\right] \\ &= \frac{1}{\bar{\sigma}^2}\bar{u}_n' H_{n,k}W_ny_n - tr\left[\left(I_n - \bar{\lambda}W_n\right)^{-1}H_{n,k}W_n\right]\end{aligned}$$

$$\frac{\partial \ln L_n\left(\bar{\theta}\right)}{\partial\beta'} = \frac{1}{\bar{\sigma}^2}X_n'\bar{u}_n$$

$$\frac{\partial \ln L_n\left(\bar{\theta}\right)}{\partial \sigma^2} = -\frac{n}{2\bar{\sigma}^2} + \frac{1}{2\left(\bar{\sigma}^2\right)^2}\bar{u}_n^{'}\bar{u}_n$$

Clearly, $\frac{\partial \ln L_n\left(\bar{\theta}\right)}{\partial \beta} = 0$ and $\frac{\partial \ln L_n\left(\bar{\theta}\right)}{\partial \sigma^2}$ under $H_0$, then we need to focus on the following score functions:

$$h_{k,n}\left(\bar{\theta}\right) = \frac{\partial \ln L_n\left(\bar{\theta}\right)}{\partial \lambda_k} = \frac{1}{\bar{\sigma}^2}\bar{u}_n^{'}H_{n,k}W_n y_n - tr\left[\left(I_n - \bar{\lambda}W_n\right)^{-1}H_{n,k}W_n\right]$$

Also, for the second order derivatives, we have the following non-zero terms:

$$\frac{\partial^2 \ln L_n\left(\bar{\theta}\right)}{\partial \lambda_k^2} = -\frac{1}{\bar{\sigma}^2}\left(H_{n,k}W_n y_n\right)^{'}H_{n,k}W_n y_n - tr\left[\left(\left(I_n - \bar{\lambda}W_n\right)^{-1}H_{n,k}W_n\right)^2\right]$$

$$\frac{\partial^2 \ln L_n\left(\bar{\theta}\right)}{\partial \lambda_h \partial \lambda_k} = -tr\left[\left(I_n - \bar{\lambda}W_n\right)^{-1}H_{n,l}W_n\left(I_n - \bar{\lambda}W_n\right)^{-1}H_{n,k}W_n\right]$$

$$\frac{\partial^2 \ln L_n\left(\theta\right)}{\partial \beta' \partial \lambda_k} = -\frac{1}{\bar{\sigma}^2}X_n^{'}H_{n,k}W_n y_n$$

$$\frac{\partial^2 \ln L_n\left(\theta\right)}{\partial \sigma^2 \partial \lambda_k} = -\frac{1}{\left(\bar{\sigma}^2\right)^2}\bar{u}_n^{'}H_{n,k}W_n y_n$$

$$\frac{\partial^2 \ln L_n\left(\theta\right)}{\partial \beta' \partial \beta} = -\frac{1}{\bar{\sigma}^2}X_n^{'}X_n$$

$$\frac{\partial^2 \ln L_n\left(\theta\right)}{\partial \left(\sigma^2\right)^2} = \frac{n}{2\left(\bar{\sigma}^2\right)^2} - \frac{1}{\left(\bar{\sigma}^2\right)^3}\bar{u}_n^{'}\bar{u}_n$$

By likelihood equation $E_\theta\left(\frac{\partial^2 \ln L_n(\theta)}{\partial \theta \partial \theta'}\right) + E_\theta\left(\frac{\partial \ln L_n(\theta)}{\partial \theta}\frac{\partial \ln L_n(\theta)}{\partial \theta'}\right) = 0$, as QMLE of SAR model is $\sqrt{n}-$ convergence, we can construct the following LM statistic to test $H_0$:

$$LM_2 = -h_n\left(\bar{\theta}\right)^{'}E\left(\frac{\partial^2 \ln L_n\left(\bar{\theta}\right)}{\partial \theta \partial \theta'}\right)^{-1}h_n\left(\bar{\theta}\right) \tag{8}$$

where $h_n\left(\bar{\theta}\right) = \frac{\partial \ln L_n\left(\bar{\theta}\right)}{\partial \theta} = \left(h_{1,n}\left(\bar{\theta}\right), \cdots, h_{K,n}\left(\bar{\theta}\right), 0, \cdots, 0\right)^{'}$ with FOCs for parameters other than $\lambda_k$ are zeros. In the next section, we will derive the asymptotic distribution of $LM_2$.

## 4.2 Asymptotic Distribution of $LM_2$

Similar to Section 3.2, we consider to derive the asymptotic distribution of linear combination of the score functions:

$$
\begin{aligned}
&\xi_n\left(a,\bar{\theta}\right)\\
=&a^{'}h_n\left(\bar{\theta}\right)\\
=&-\frac{1}{\bar{\sigma}^2}\bar{u}_n^{'}\left(\sum_{k=1}^{K}a_kH_{n,k}\right)W_ny_ntr\left[\left(I_n-\bar{\lambda}W_n\right)^{-1}\left(\sum_{k=1}^{K}a_kH_{n,k}\right)W_n\right]\\
=&\frac{1}{\bar{\sigma}^2}\bar{u}_n^{'}H_{a,n}W_ny_n-tr\left[\left(I_n-\bar{\lambda}W_n\right)^{-1}H_{a,n}W_n\right]
\end{aligned}
$$

Since $y_n=S_n\left(\lambda_0\right)^{-1}\left(X_n\beta_0+u_n\right)$ and $\bar{u}_n=S_n\left(\bar{\lambda}\right)y_n-X_n\bar{\beta}$, the following term in $\xi_n$ can be decomposed as

$$
\begin{aligned}
&\bar{u}_n^{'}H_{a,n}W_ny_n\\
=&\left[S_n\left(\bar{\lambda}\right)y_n-X_n\bar{\beta}\right]^{'}H_{a,n}W_nS_n\left(\lambda_0\right)^{-1}\left(X_n\beta_0+u_n\right)\\
=&y_n^{'}S_n^{'}\left(\bar{\lambda}\right)H_{a,n}W_nS_n\left(\lambda_0\right)^{-1}X_n\beta_0+y_n^{'}S_n^{'}\left(\bar{\lambda}\right)H_{a,n}W_nS_n\left(\lambda_0\right)^{-1}u_n\\
&-\bar{\beta}^{'}X_n^{'}H_{a,n}W_nS_n\left(\lambda_0\right)^{-1}X_n\beta_0-\bar{\beta}^{'}X_n^{'}H_{a,n}W_nS_n\left(\lambda_0\right)^{-1}u_n\\
=&\left[S_n\left(\lambda_0\right)^{-1}\left(X_n\beta_0+u_n\right)\right]^{'}S_n^{'}\left(\bar{\lambda}\right)H_{a,n}W_nS_n\left(\lambda_0\right)^{-1}X_n\beta_0\\
&+\left[S_n\left(\lambda_0\right)^{-1}\left(X_n\beta_0+u_n\right)\right]^{'}H_{a,n}W_nS_n\left(\lambda_0\right)^{-1}u_n\\
&-\bar{\beta}^{'}X_n^{'}H_{a,n}W_nS_n\left(\lambda_0\right)^{-1}X_n\beta_0-\bar{\beta}^{'}X_n^{'}H_{a,n}W_nS_n\left(\lambda_0\right)^{-1}u_n\\
=&\beta_0^{'}X_n^{'}S_n^{'}\left(\lambda_0\right)^{-1}S_n^{'}\left(\bar{\lambda}\right)H_{a,n}W_nS_n\left(\lambda_0\right)^{-1}X_n\beta_0\\
&+u_n^{'}S_n^{'}\left(\lambda_0\right)^{-1}S_n^{'}\left(\bar{\lambda}\right)H_{a,n}W_nS_n\left(\lambda_0\right)^{-1}X_n\beta_0\\
&+\beta_0^{'}X_n^{'}S_n^{'}\left(\lambda_0\right)^{-1}H_{a,n}W_nS_n\left(\lambda_0\right)^{-1}u_n+u_n^{'}S_n^{'}\left(\lambda_0\right)^{-1}H_{a,n}W_nS_n\left(\lambda_0\right)^{-1}u_n\\
&-\bar{\beta}^{'}X_n^{'}H_{a,n}W_nS_n\left(\lambda_0\right)^{-1}X_n\beta_0-\bar{\beta}^{'}X_n^{'}H_{a,n}W_nS_n\left(\lambda_0\right)^{-1}u_n
\end{aligned}
\tag{9}
$$

By continuous mapping theorem and Slutsky's theorem, since $\bar{\theta}\xrightarrow{p}\theta_0$, we have

$$
\begin{aligned}
&\beta_0^{'}X_n^{'}S_n^{'}\left(\lambda_0\right)^{-1}S_n^{'}\left(\bar{\lambda}\right)H_{a,n}W_nS_n\left(\lambda_0\right)^{-1}X_n\beta_0\\
&-\beta_0^{'}X_n^{'}H_{a,n}W_nS_n\left(\lambda_0\right)^{-1}X_n\beta_0\xrightarrow{p}0
\end{aligned}
$$

and

$$
\begin{aligned}
&\bar{\beta}^{'}X_n^{'}H_{a,n}W_nS_n\left(\lambda_0\right)^{-1}X_n\beta_0\\
&-\beta_0^{'}X_n^{'}H_{a,n}W_nS_n\left(\lambda_0\right)^{-1}X_n\beta_0\xrightarrow{p}0
\end{aligned}
$$

Then, we have

$$
\beta_0^{'}X_n^{'}S_n^{'}\left(\lambda_0\right)^{-1}S_n^{'}\left(\bar{\lambda}\right)H_{a,n}W_nS_n\left(\lambda_0\right)^{-1}X_n\beta_0-\bar{\beta}^{'}X_n^{'}H_{a,n}W_nS_n\left(\lambda_0\right)^{-1}X_n\beta_0=o_p\left(1\right)
$$

Denote the following matrices:

$$\begin{aligned} C_n\left(a,\bar{\theta}\right) &= \beta_0^{'}X_n^{'}S_n^{'}\left(\lambda_0\right)^{-1}W_n^{'}H_{a,n}S_n\left(\bar{\lambda}\right)S_n\left(\lambda_0\right)^{-1} \\ &+ \beta_0^{'}X_n^{'}S_n^{'}\left(\lambda_0\right)^{-1}H_{a,n}W_nS_n\left(\lambda_0\right)^{-1} \\ &- \bar{\beta}^{'}X_n^{'}H_{a,n}W_nS_n\left(\lambda_0\right)^{-1} \end{aligned}$$

and

$$\begin{aligned} D_n\left(a\right) &= \frac{1}{2}S_n^{'}\left(\lambda_0\right)^{-1}H_{a,n}W_nS_n\left(\lambda_0\right)^{-1} \\ &+ \frac{1}{2}S_n\left(\lambda_0\right)^{-1}W_n^{'}H_{a,n}S_n^{'}\left(\lambda_0\right)^{-1} \end{aligned}$$

Then, from (9), we can get the following result:

$$\begin{aligned} \frac{1}{\sqrt{n}}\bar{u}_n^{'}H_{a,n}W_ny_n &= \frac{1}{\sqrt{n}}\left[C_n\left(a,\bar{\theta}\right)u_n + u_n^{'}D_n\left(a\right)u_n\right] + o_p\left(1\right) \\ &\equiv \frac{1}{\sqrt{n}}\widetilde{Q}_n\left(a,\bar{\theta}\right) + o_p\left(1\right) \end{aligned}$$

Similar to Assumption 11, now we need the following assumption as an regulation:

**Assumption 11':**
*For any* $a$, $\lim_{n\to\infty}\frac{1}{n}D_n^{'}\left(a\right)D_n\left(a\right)$ *exists.*

Then, similar to the discussion in Section 3.2, for the two different situations $\{h_n\}$ is bounded or $\lim_{n\to\infty}h_n = \infty$, due to Assumption 4, 5, 8 and 9, we have the following result by different types of CLTs:

$$\frac{\widetilde{Q}_n\left(a,\bar{\theta}\right) - E\left[\widetilde{Q}_n\left(a,\bar{\theta}\right)\right]}{\sigma_{\widetilde{Q}_n}} \xrightarrow{d} N\left(0,1\right)$$

where $\sigma_{\widetilde{Q}_n}$ is the variance of $\widetilde{Q}_n\left(a,\bar{\theta}\right)$.

Next, lets move to the term $tr\left[\left(I_n - \bar{\lambda}W_n\right)^{-1}H_{a,n}W_n\right]$. As $tr\left(ABC\right) = tr\left(BCA\right) = tr\left(CAB\right)$ for any $n\times n$ matrix $A, B$ and $C$, we have

$$\begin{aligned} \left|tr\left[\left(I_n - \bar{\lambda}W_n\right)^{-1}H_{a,n}W_n\right]\right| &= \left|tr\left[H_{a,n}W_n\left(I_n - \bar{\lambda}W_n\right)^{-1}\right]\right| \\ &\le \max_{k=1,\cdots K}\left|a_k\right|\left|tr\left(W_n\left(I_n - \bar{\lambda}W_n\right)^{-1}\right)\right| \end{aligned}$$

Since $W_n$ and $\left(I_n - \bar{\lambda}W_n\right)^{-1}$ are uniformly bounded in both column and row sum norm, $W_n\left(I_n - \bar{\lambda}W_n\right)^{-1}$ is also uniformly bounded in both column and row sum norm, thus $tr\left(W_n\left(I_n - \bar{\lambda}W_n\right)^{-1}\right) = O_p\left(1\right)$. Then,

$$\frac{1}{\sqrt{n}}\xi_n\left(a,\bar{\theta}\right)=\frac{1}{\hat{\sigma}^2}\frac{1}{\sqrt{n}}\widetilde{Q}_n\left(a,\bar{\theta}\right)+o_p\left(1\right)$$

is asymptotically Normal. Thus, $\frac{1}{\sqrt{n}}h_n\left(\bar{\theta}\right)$ is jointly asymptotic Normal. With Assumption 3, by similar argument as in Section 2.3, each term in $\frac{1}{\sqrt{n}}h_n\left(\bar{\theta}\right)$ is not degenerate.

Since we have $(K-1)$ equation constraints in $H_0:\rho_1=\cdots=\rho_K$, the degree of freedom of $LM_2$ should be $(K-1)$. In fact, we have

$$\begin{aligned}\sum_{k=1}^{K}h_{k,n}\left(\bar{\theta}\right)&=\sum_{k=1}^{K}\left\{\frac{1}{\bar{\sigma}^2}\bar{u}_n^{'}H_{n,k}W_ny_n-tr\left[\left(I_n-\bar{\lambda}W_n\right)^{-1}H_{n,k}W_n\right]\right\}\\&=\frac{1}{\bar{\sigma}^2}\bar{u}_n^{'}W_ny_n-tr\left[\left(I_n-\bar{\lambda}W_n\right)^{-1}W_n\right]\\&=0\end{aligned}$$

since it equals to the first order condition of MLE of SAR model. Thus, we have

$$LM_2=-h_n\left(\bar{\theta}\right)^{'}E\left(\frac{\partial^2\ln L_n\left(\bar{\theta}\right)}{\partial\theta\partial\theta'}\right)^{-1}h_n\left(\bar{\theta}\right)\xrightarrow{d}\chi^2\left(K-1\right)$$

# 5 Monte Carlo Simulations

## 5.1 Basic Settings

In this section, we try to do Monte Carlo simulations to see whether the test we derived in Section 3 and 4 work well. And then, we will show the performance of the unconventional LM statistics. For each test statistic, we will evaluate its performance by three indicators: 1. test size under $H_0$; 2. test power under $H_1$ ; 3. simulated critical values. All these three indicators will be showed for $\alpha=10\%$, $\alpha=5\%$ and $\alpha=1\%$.

To simulate the SAR and spatial log-ARCH like model, we need to first simulate the spatial correlations among regions which satisfy our Assumption 1, 2, 5 and 6. Here we construct the row-stochastic nearest neighbor spatial weight matrix $W_n=(w_{ij,n})$ using LeSage's econometrics toolbox. The procedure is:

1. Generate two random vectors of coordinates as the geographic location for each observation;
2. Find $l$ nearest neighbors for each observation according to their spatial distances and denote the corresponding $w_{ij,n}=1$, otherwise $w_{ij,n}=0$;
3. Row-normalize $W_n$.

We will consider two different situations when $l=5$ and $l=10$.

In our simulation exercises, there are three different DGP we need to consider:

**DGP 1**: Linear Regression Model

$$y_n=X_n\beta+u_n,u_n\sim\left(0,\sigma^2I_n\right)$$

**DGP 2**: Spatial Autoregressive Model

$$y_n = \lambda W_n + X_n\beta + u_n, , u_n \sim \left(0, \sigma^2 I_n\right)$$

**DGP 3**: Heterogeneous Coefficient Spatial Autoregressive Model

$$y_n = \sum_{k=1}^{K} \lambda_k H_{n,k} W_n y_n + X_n\beta + u_n, , u_n \sim \left(0, \sigma^2 I_n\right)$$

From DGP 1 to 4, for external regressor $x_{i,n}$, we consider a two regressors case: $x_{i,n} = (x_{1,i,n}, x_{2,i,n})'$ where $x_{1,i,n}$ is the intercept and $x_{2,i,n} \overset{iid}{\sim} N(0,1)$. Also, we consider the following two parameter setting for $\beta$ and $\sigma^2$ : $\left(\beta', \sigma^2\right) = [(1,1), 4]$ and $\left(\beta', \sigma^2\right) = [(2,-5), 1]$ in DGP 1 and DGP 2. For the SAR coefficient in DGP 2, we use $\rho = 0.5$ and $\rho = -0.4$. Each round of simulation with different number of regions and parameter settings will be replicated for 1,000 times. In the next two sessions, we will show the performance of $LM_1$ and $LM_2$ separately.

## 5.2 Performance of $\mathbf{LM_1}$

By using DGP 1 and DGP 3, we can simulate the test size, power and critical values of $LM_1$ in Section 3 with finite samples. For the heterogeneity structure, we consider two-category case and the ratio between the regions of these two categories are fixed at $4:1$. To simulate the power, we consider the the following two different alternatives: $(\lambda_1, \lambda_2) = (0.5, -0.2)$ and $(\lambda_1, \lambda_2) = (0, 0.4)$. The first alternative is the situation when different types of regions or agents have distinct response to neighbors, or are affected by neighbor's spill-over effects differently. The second alternative is the situation when only part of the regions or agents are affected by neighbors, while the others may affect others but are not affected by others. Also, besides Normal distributed residuals, we also consider re-centered Gamma distribution and uniform distribution, to see whether $LM_1$ also works for non-Normal case. The simulation results are showed in Table 1-3.

Table 1 shows the test size of $LM_1$ when using $\chi^2(2)$ critical value at 5% significance level. For all the three types of residuals, when sample size is small, it tends to over reject the true model but

Table 1: Test Size of $LM_1$ ($\chi^2_{0.95}(2) = 5.9915$)

| n | neighbors | residuals | $\left(\beta^{'}, \sigma^2\right) = [(1,1), 4]$ | $\left(\beta^{'}, \sigma^2\right) = [(2,-5), 1]$ |
|---|---|---|---|---|
| 100 | $l = 5$ | $N\left(0, \sigma^2\right)$ | 0.068 | 0.071 |
| | | $\sigma\left[\Gamma(2.25, 2) - 4.5\right]$ | 0.072 | 0.073 |
| | | $\sigma U\left[-\sqrt{3}, \sqrt{3}\right]$ | 0.067 | 0.061 |
| | $l = 10$ | $N\left(0, \sigma^2\right)$ | 0.078 | 0.077 |
| | | $\sigma\left[\Gamma(2.25, 2) - 4.5\right]$ | 0.057 | 0.071 |
| | | $\sigma U\left[-\sqrt{3}, \sqrt{3}\right]$ | 0.073 | 0.066 |
| 200 | $l = 5$ | $N\left(0, \sigma^2\right)$ | 0.055 | 0.067 |
| | | $\sigma\left[\Gamma(2.25, 2) - 4.5\right]$ | 0.074 | 0.054 |
| | | $\sigma U\left[-\sqrt{3}, \sqrt{3}\right]$ | 0.057 | 0.058 |
| | $l = 10$ | $N\left(0, \sigma^2\right)$ | 0.058 | 0.064 |
| | | $\sigma\left[\Gamma(2.25, 2) - 4.5\right]$ | 0.054 | 0.059 |
| | | $\sigma U\left[-\sqrt{3}, \sqrt{3}\right]$ | 0.067 | 0.054 |
| 400 | $l = 5$ | $N\left(0, \sigma^2\right)$ | 0.048 | 0.049 |
| | | $\sigma\left[\Gamma(2.25, 2) - 4.5\right]$ | 0.050 | 0.048 |
| | | $\sigma U\left[-\sqrt{3}, \sqrt{3}\right]$ | 0.056 | 0.048 |
| | $l = 10$ | $N\left(0, \sigma^2\right)$ | 0.052 | 0.054 |
| | | $\sigma\left[\Gamma(2.25, 2) - 4.5\right]$ | 0.053 | 0.062 |
| | | $\sigma U\left[-\sqrt{3}, \sqrt{3}\right]$ | 0.051 | 0.057 |

Table 2: Test Power of $LM_1$ ($\chi^2_{0.95}(2) = 5.9915$)

| n | neighbors | residuals | $\left(\lambda_1, \lambda_2, \beta', \sigma^2\right) = [0.5, -0.2, (1,1), 4]$ | $\left(\lambda_1, \lambda_2, \beta', \sigma^2\right) = [0, 0.4, (2,-5), 1]$ |
|---|---|---|---|---|
| 100 | $l = 5$ | $N\left(0, \sigma^2\right)$ | 0.912 | 0.933 |
| | | $\sigma\left[\Gamma(2.25, 2) - 4.5\right]$ | 0.730 | 0.437 |
| | | $\sigma U\left[-\sqrt{3}, \sqrt{3}\right]$ | 0.931 | 0.996 |
| | $l = 10$ | $N\left(0, \sigma^2\right)$ | 0.877 | 0.763 |
| | | $\sigma\left[\Gamma(2.25, 2) - 4.5\right]$ | 0.526 | 0.395 |
| | | $\sigma U\left[-\sqrt{3}, \sqrt{3}\right]$ | 0.830 | 0.997 |
| 200 | $l = 5$ | $N\left(0, \sigma^2\right)$ | 0.998 | 1 |
| | | $\sigma\left[\Gamma(2.25, 2) - 4.5\right]$ | 0.955 | 0.776 |
| | | $\sigma U\left[-\sqrt{3}, \sqrt{3}\right]$ | 0.995 | 1 |
| | $l = 10$ | $N\left(0, \sigma^2\right)$ | 0.951 | 1 |
| | | $\sigma\left[\Gamma(2.25, 2) - 4.5\right]$ | 1 | 0.547 |
| | | $\sigma U\left[-\sqrt{3}, \sqrt{3}\right]$ | 0.964 | 1 |
| 400 | $l = 5$ | $N\left(0, \sigma^2\right)$ | 1 | 1 |
| | | $\sigma\left[\Gamma(2.25, 2) - 4.5\right]$ | 1 | 0.985 |
| | | $\sigma U\left[-\sqrt{3}, \sqrt{3}\right]$ | 1 | 1 |
| | $l = 10$ | $N\left(0, \sigma^2\right)$ | 1 | 1 |
| | | $\sigma\left[\Gamma(2.25, 2) - 4.5\right]$ | 0.964 | 0.726 |
| | | $\sigma U\left[-\sqrt{3}, \sqrt{3}\right]$ | 0.999 | 1 |

Table 3: Simulated Critical Values of $LM_1$ with Normal Residuals

| n | neighbors | $\left(\beta', \sigma^2\right) = [(1,1), 4]$ | | | $\left(\beta', \sigma^2\right) = [(2,-5), 1]$ | | |
|---|---|---|---|---|---|---|---|
| | | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 |
| 100 | $l = 5$ | 4.8871 | 6.5476 | 11.4063 | 5.3041 | 6.8218 | 11.1602 |
| | $l = 10$ | 5.4134 | 6.9741 | 10.4855 | 5.2785 | 6.9584 | 11.4887 |
| 200 | $l = 5$ | 4.6411 | 5.6946 | 9.2097 | 5.1939 | 6.5236 | 9.7176 |
| | $l = 10$ | 4.8791 | 6.3121 | 8.7272 | 4.9353 | 6.4395 | 11.3451 |
| 400 | $l = 5$ | 4.701 | 5.8109 | 9.3248 | 4.4700 | 5.9705 | 8.6374 |
| | $l = 10$ | 4.6101 | 6.1810 | 10.2791 | 4.7262 | 6.1095 | 9.3792 |

not seriously. As the sample size goes larger, the test size will decrease and be around the theoretical value 5%. From Table 3, we can see that the simulated critical values under Normal case is getting close to the asymptotic critical values of $\chi^2(2)$, which are 4.6052, 5.9915 and 9.2103 for 90%, 95% and 99% critical level. When residuals are Gamma or Uniform distributed, the simulated critical values perform similar to the Normal case which are not showed here.

Table 3 shows the test power of $LM_1$ for two different alternatives. The critical value used here is also $\chi^2(2)$ critical value at 5% significance level. When residual is Normal, it is a powerful test. Even with 100 observations, it rejects over 70% of alternatives. As sample size gets larger, the test power goes to 1. When residuals are not Normal, the test power for small sample highly depends on the residual distribution. For 5 neighbor cases, when residual is uniformly distributed, the test power is as large as the Normal case. However, when residual is re-centered Gamma distribution, the test power shrinks to around only 40% in some cases. When sample size is large, the test power will increase and converge to 1, just like the Normal case.

Traditionally, Moran I test developed in Moran, Cliff and Ord (1973) is widely used to test whether there exist spatial correlation, it does not work well in our heterogeneity setting, especially for some situations. In Table 4, we show the test power of Moran I test of our Monte Carlo simulation.

From Table 4, we can see Moran I test works very bad for the parameter setting $\left(\lambda_1, \lambda_2, \beta', \sigma^2\right) = [0, 0.4, (2, -5), 1]$. In this setting, 20% of the regions are affected by their neighbors, and the remaining 80% of regions may still impact their neighbors but not affected by neighbors. Although in this case, spatial correlation exists, the Moran I test can not reject the null hypothesis well. Even with 400 observations, the simulated test power is only around 15%-25% which is too low. As a comparison, in Table 2, we can see our $LM_1$ has a much better performance on rejecting the alternative models in this situation. Even with 100 observations, the test power for Normal and uniform cases are higher than 90%, and it is about 40% for re-centered Gamma case which is low but still way better than Moran I which is only less than 10%. For the other parameter setting, Moran I works well but the test power are still less than our $LM_1$. Thus, empirically, if there exist categorical heterogeneity among regions, we suggest you to use our $LM_1$ test instead of Moran I, otherwise you may have a very large chance to falsely accept the null hypothesis and ignore the spatial correlation among regions. Then, your estimators for external regressors will also be biased by using linear regressions.

Table 4: Test Power of Moran I (5% Significant Level)

| n | neighbors | residuals | $\left(\lambda_1, \lambda_2, \beta^{'}, \sigma^2\right) = [0.5, -0.2, (1, 1), 4]$ | $\left(\lambda_1, \lambda_2, \beta^{'}, \sigma^2\right) = [0, 0.4, (2, -5), 1]$ |
|---|---|---|---|---|
| 100 | $l = 5$ | $N\left(0, \sigma^2\right)$ | 0.724 | 0.055 |
| | | $\sigma\left[\Gamma(2.25, 2) - 4.5\right]$ | 0.701 | 0.091 |
| | | $\sigma U\left[-\sqrt{3}, \sqrt{3}\right]$ | 0.684 | 0.103 |
| | $l = 10$ | $N\left(0, \sigma^2\right)$ | 0.442 | 0.085 |
| | | $\sigma\left[\Gamma(2.25, 2) - 4.5\right]$ | 0.433 | 0.083 |
| | | $\sigma U\left[-\sqrt{3}, \sqrt{3}\right]$ | 0.502 | 0.059 |
| 200 | $l = 5$ | $N\left(0, \sigma^2\right)$ | 0.906 | 0.177 |
| | | $\sigma\left[\Gamma(2.25, 2) - 4.5\right]$ | 0.943 | 0.148 |
| | | $\sigma U\left[-\sqrt{3}, \sqrt{3}\right]$ | 0.926 | 0.213 |
| | $l = 10$ | $N\left(0, \sigma^2\right)$ | 0.972 | 0.132 |
| | | $\sigma\left[\Gamma(2.25, 2) - 4.5\right]$ | 0.708 | 0.102 |
| | | $\sigma U\left[-\sqrt{3}, \sqrt{3}\right]$ | 0.713 | 0.111 |
| 400 | $l = 5$ | $N\left(0, \sigma^2\right)$ | 0.998 | 0.176 |
| | | $\sigma\left[\Gamma(2.25, 2) - 4.5\right]$ | 0.998 | 0.189 |
| | | $\sigma U\left[-\sqrt{3}, \sqrt{3}\right]$ | 0.997 | 0.237 |
| | $l = 10$ | $N\left(0, \sigma^2\right)$ | 0.960 | 0.193 |
| | | $\sigma\left[\Gamma(2.25, 2) - 4.5\right]$ | 0.947 | 0.146 |
| | | $\sigma U\left[-\sqrt{3}, \sqrt{3}\right]$ | 0.920 | 0.143 |

Table 5: Test Size of $LM_2$ ($\chi^2_{0.95}(2) = 5.9915$)

| n | neighbors | residuals | $\left(\lambda, \beta', \sigma^2\right) = [0.5, (1,1), 4]$ | $\left(\lambda, \beta', \sigma^2\right) = [-0.4, (2,-5), 1]$ |
|---|---|---|---|---|
| 100 | $l=5$ | $N\left(0,\sigma^2\right)$ | 0.078 | 0.077 |
| | | $\sigma\left[\Gamma(2.25,2) - 4.5\right]$ | 0.112 | 0.054 |
| | | $\sigma U\left[-\sqrt{3}, \sqrt{3}\right]$ | 0.067 | 0.080 |
| | $l=10$ | $N\left(0,\sigma^2\right)$ | 0.097 | 0.069 |
| | | $\sigma\left[\Gamma(2.25,2) - 4.5\right]$ | 0.093 | 0.058 |
| | | $\sigma U\left[-\sqrt{3}, \sqrt{3}\right]$ | 0.070 | 0.071 |
| 200 | $l=5$ | $N\left(0,\sigma^2\right)$ | 0.062 | 0.065 |
| | | $\sigma\left[\Gamma(2.25,2) - 4.5\right]$ | 0.095 | 0.067 |
| | | $\sigma U\left[-\sqrt{3}, \sqrt{3}\right]$ | 0.060 | 0.057 |
| | $l=10$ | $N\left(0,\sigma^2\right)$ | 0.067 | 0.058 |
| | | $\sigma\left[\Gamma(2.25,2) - 4.5\right]$ | 0.085 | 0.057 |
| | | $\sigma U\left[-\sqrt{3}, \sqrt{3}\right]$ | 0.045 | 0.059 |
| 400 | $l=5$ | $N\left(0,\sigma^2\right)$ | 0.054 | 0.047 |
| | | $\sigma\left[\Gamma(2.25,2) - 4.5\right]$ | 0.082 | 0.050 |
| | | $\sigma U\left[-\sqrt{3}, \sqrt{3}\right]$ | 0.056 | 0.049 |
| | $l=10$ | $N\left(0,\sigma^2\right)$ | 0.059 | 0.054 |
| | | $\sigma\left[\Gamma(2.25,2) - 4.5\right]$ | 0.077 | 0.049 |
| | | $\sigma U\left[-\sqrt{3}, \sqrt{3}\right]$ | 0.048 | 0.048 |

## 5.3 Performance of $LM_2$

By using DGP 2 and DGP 3, we can simulate the test size, power and critical values of $LM_2$ in Section 4 with finite samples. For the heterogeneity structure, we consider three-category case and the ratio between the regions of different categories are fixed at 3 : 5 : 2. To simulate the power, we consider the the following two different alternatives: $(\lambda_1, \lambda_2, \lambda_3) = (0.5, -0.2, 0.7)$ and $(\lambda_1, \lambda_2) = (0, 0.4, 0.1)$. Similar to Section 5.2, besides Normal distributed residuals, we also consider re-centered Gamma distribution and uniform distribution, to see whether $LM_1$ also works for non-Normal case. The simulation results are showed in Table 5-7.

Table 6: Test Power of $LM_2$ ($\chi^2_{0.95}(2) = 5.9915$)

| n | neighbors | residuals | $\left(\lambda_1, \lambda_2, \lambda_3, \beta', \sigma^2\right) = [0.5, -0.2, 0.7, (1,1), 4]$ | $\left(\lambda_1, \lambda_2, \lambda_3, \beta', \sigma^2\right) = [0, 0.4, 0.1, (2,-5), 1]$ |
|---|---|---|---|---|
| 100 | $l=5$ | $N\left(0,\sigma^2\right)$ | 0.790 | 0.921 |
| | | $\sigma\left[\Gamma(2.25,2) - 4.5\right]$ | 0.496 | 0.218 |
| | | $\sigma U\left[-\sqrt{3}, \sqrt{3}\right]$ | 0.827 | 0.862 |
| | $l=10$ | $N\left(0,\sigma^2\right)$ | 0.834 | 0.673 |
| | | $\sigma\left[\Gamma(2.25,2) - 4.5\right]$ | 0.315 | 0.186 |
| | | $\sigma U\left[-\sqrt{3}, \sqrt{3}\right]$ | 0.603 | 0.882 |
| 200 | $l=5$ | $N\left(0,\sigma^2\right)$ | 0.971 | 0.989 |
| | | $\sigma\left[\Gamma(2.25,2) - 4.5\right]$ | 0.775 | 0.453 |
| | | $\sigma U\left[-\sqrt{3}, \sqrt{3}\right]$ | 0.982 | 0.997 |
| | $l=10$ | $N\left(0,\sigma^2\right)$ | 0.946 | 0.931 |
| | | $\sigma\left[\Gamma(2.25,2) - 4.5\right]$ | 0.531 | 0.343 |
| | | $\sigma U\left[-\sqrt{3}, \sqrt{3}\right]$ | 0.967 | 0.984 |
| 400 | $l=5$ | $N\left(0,\sigma^2\right)$ | 1 | 1 |
| | | $\sigma\left[\Gamma(2.25,2) - 4.5\right]$ | 0.963 | 0.715 |
| | | $\sigma U\left[-\sqrt{3}, \sqrt{3}\right]$ | 1 | 1 |
| | $l=10$ | $N\left(0,\sigma^2\right)$ | 0.999 | 0.998 |
| | | $\sigma\left[\Gamma(2.25,2) - 4.5\right]$ | 0.811 | 0.580 |
| | | $\sigma U\left[-\sqrt{3}, \sqrt{3}\right]$ | 1 | 1 |

Table 7: Simulated Critical Values of $LM_2$ with Normal Residuals

| n | neighbors | $\left(\lambda, \beta', \sigma^2\right) = [0.5, (1,1), 4]$ | | | $\left(\lambda, \beta', \sigma^2\right) = [-0.4, (2,-5), 1]$ | | |
|---|---|---|---|---|---|---|---|
| | | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 |
| 100 | $l=5$ | 5.4320 | 7.2831 | 10.8773 | 5.3287 | 6.7644 | 10.0293 |
| | $l=10$ | 5.8685 | 7.9904 | 11.6299 | 5.2176 | 7.1535 | 11.6175 |
| 200 | $l=5$ | 4.9169 | 6.2664 | 10.6303 | 4.9273 | 6.4021 | 10.2525 |
| | $l=10$ | 5.0219 | 6.5657 | 10.2726 | 5.0179 | 6.3613 | 9.5738 |
| 400 | $l=5$ | 4.7358 | 6.1294 | 8.8924 | 4.8382 | 6.1149 | 9.4095 |
| | $l=10$ | 4.7364 | 6.3350 | 9.7782 | 4.6703 | 6.1025 | 9.4557 |

From Table 5-7, we can see the performance of $LM_2$ is similar to $LM_1$. With small sample, it tend to over reject $H_0$ when it is true. As sample size goes larger, the size and simulated critical value are getting closer to asymptotic values. For the test power, it also increases along with sample size. Another thing we should notice is that the performance when the residuals are re-centered Gamma is significantly bad then the other two cases. A potential reason is that $\Gamma(2.25, 2)$ is not a symmetric distribution. Although QMLE for SAR model works well for distributions other than Normal asymptotically, higher order information such like skewness and kurtosis may create bias in small sample situation. In general, with more than 200 observations, it works well enough.

# 6 Application: City Size and Housing Market

## 6.1 Motivation and Data Description

City size is often discussed in the urban economics literatures. Many previous research had provided empirical evidences that city size is a very important spatial demographical heterogeneity which has large impact on regional economic development and labor market. Segal (1976) started to focus on the impact of city size on productivity, and identified significant higher output per capita in large cities due to returns to scale after controlling capital, labor and some other factors. Moomaw (1981) extended results in Segal (1976) and found that productivity advantage of large cities are much larger for the non-manufacturing sector than the manufacturing sector. Recent years, researchers also focused on some other aspects of economic issues related to city size. Chritstoffersen and Sarkissian (2009) observed performance improvements of the same equity fund manager at the same fund in financial centers but not elsewhere, which is an evidence of knowledge spillovers in large cities. Baum-Snow and Pavan (2012) focused on the significant wage premium in large cities showed in 2010 Census and tried to explain it by a labor search model. Baum-Snow and Pavan (2013) identified a strong positive monotonic relationship between wage inequality and city size from late 1970s to early 2000s. Since local housing demand and supply are closely correlated with local economic development, city size may have a large impact on housing market. Moreover, since the industry and income distribution performs very different inside a large and small cities, the spatial correlation of housing market inside each sub-regions of a large city may also by very different. Financial risk correlated with the housing market and the derivative market generated by housing mortgage may also closely affected by the city size, and have different spatial correlation structures in large and small cities.

Due to our interest in large cities, we focus on Northeastern United States which has the largest

Table 8: Summary Statistics of Annual Change of HPI (%) in Northeastern US

| | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 |
|---|---|---|---|---|---|---|---|---|---|
| Mean | 7.09 | 2.07 | -1.38 | -4.45 | -3.08 | -2.28 | -1.72 | 0.21 | 1.36 |
| Minimum | -4.15 | -4.25 | -9.5 | -19.83 | -11.91 | -9.53 | -8.34 | -6.99 | -6.08 |
| Maximum | 32.27 | 12.75 | 10.39 | 6.70 | 6.54 | 7.15 | 6.50 | 8.08 | 11.88 |
| s.t.d | 4.05 | 3.00 | 3.33 | 3.97 | 2.81 | 2.65 | 2.39 | 2.02 | 2.86 |

megalopolis in the world. Additionally, By Census Bureau's definition, there are 9 states belongs to this geographical concept: Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island and Vermont. Combined the definitions from Association of American Geographers and Geographical Society of America, we also include Delaware, Maryland and Washington, D.C. to our sample. So, totally we have Since the Metropolitan Statistical Areas (MSA) and Combined Statistical Areas (CSA) are defined by United States Office of Management and Budget (OMB) on a county basis, we also use county level data as our observations. From Federal Housing Finance Agency (FHFA), we can access the annual house price indexes at county level[1]. The annual percentage change of HPI is a good approximation of annual average housing return in a county. Although the HPI does not perfectly match the return of housing market since it counts total price, not price per $sq.ft$, as county is a relatively large area which contains different types of houses, tradings in one year should be a good mixture of different types. Since this data set is still under development, too old historical data and most recent data are not available for some counties. Due to this issue, the sample period used in this paper is limited from 2006 to 2014. As this period contains the tail of the housing boom during early 2000s, and the global financial crisis, it is a good time window for us to investigate how the city size plays a roll in spatial risk spreading in housing market. Additionally, due to lack of trading data, 5 counties including Sullivan County, PA (FIPS: 42113), Cameron County, PA(FIPS: 42023), Forest County, PA (FIPS: 42053), Juniata County, PA (FIPS: 42067), and Hamilton County, NY (FIPS:36061), are not included in the database. Population size of these 5 counties are very small, with less than 30,000 in Juniata county, PA and less than 10,000 in the other four. We can expect that the house trading be inactive and not have significant impact on neighbor counties. Thus, totally we have 240 counties in our sample area. Summary statistics of the annual percentage change of HPI is showed in Table 8. Also, we reported the summary statistics of annual real GDP growth in Table 9, which are publicly accessible from Bureau of Economic Analysis (BEA) [2]. In general, the housing price and GDP growth had similar trends in the sample years, but the housing market recovered much slower during the financial crisis.

[1]See https://www.fhfa.gov
[2]See https://www.bea.gov

Table 9: Summary Statistics of Annual Real GDP Growth (%) in Northeastern US

| | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 |
|---|---|---|---|---|---|---|---|---|---|
| Mean | 3.44 | 0.47 | 0.86 | -1.26 | 3.13 | 0.50 | 0.39 | 0.90 | 1.53 |
| Minimum | -48.9 | -22.2 | -10.9 | -19.1 | -12.9 | -10.4 | -15.5 | -8.7 | -6 |
| Maximum | 19 | 14.1 | 30.7 | 30 | 29.3 | 20.4 | 30.8 | 24.4 | 37.2 |
| s.t.d | 4.97 | 3.92 | 3.83 | 4.30 | 4.06 | 3.45 | 3.59 | 3.47 | 4.61 |

To distinguish large cities and their neighbors, we use population size in 2010 census as the indicator and both consider MSA and CSAs. The 240 counties in our sample belong to 53 MSAs defined by OMB. In this paper, we consider the largest 10 MSAs with more than 1 million populations as large city group. Additionally, we also consider their encompassing CSA which contains neighborhood regions, including some smaller MSAs and Micropolitan Statistical Areas (µSA) (smaller town areas with smaller population size). By definition from OMB, CSAs represent multiple metropolitan or micropolitan areas that have an employment interchange of at least 15%, the counties inside the CSA around a large MSA can be viewed together with the MSA as a integrated labor and commodity market. Thus, in this paper, we consider a broader definition of a city, and identified large city regions as the counties belongs to the encompassing CSAs listed in Table 10[3]. In total, there are 107 counties identified as large city regions.

From Table 11, although large cities and other areas had similar economic performance during financial crisis, the situation of housing market was a totally different story. By grouping the cities, in Table 11, we show the group average of their annual HPI change (%) and real GDP growth (%). Referring to US business cycle identified by NBER[4], the economic recession started at December 2007 and ended at June 2009. However, it took two more years for housing market to recover until 2013. More importantly, different regions had distinct performances. The housing market in the large city areas have both relatively better performances during expansion and bad performances during recession on average. In the next session, we will try to investigate the effect of city size on housing market in detail.

[3]For the full list of MSA in the US and their population sizes in 2010 Census, see https://en.wikipedia.org/wiki/List_of_metropolitan_statistical_areas

[4]See https://www.nber.org/research/data/us-business-cycle-expansions-and-contractions

Table 10: MSAs in Northeastern US with >1 Million Population Size in 2010 Census

| Rank | MSA | 2010 Census Population | Encompassing CSA |
|---|---|---|---|
| 1 | New York City-Newark-Jersey City | 18,897,109 | New York-Newark |
| 2 | Philadelphia-Camden-Wilmington | 5,965,343 | Philadelphia-Reading-Camden |
| 3 | Washington-Arlington-Alexandria | 5,649,540 | Washington-Baltimore-Arlington |
| 4 | Boston-Cambridge-Newton | 4,552,402 | Boston-Worcester-Providence |
| 5 | Baltimore-Columbia-Towson | 2,710,489 | Washington-Baltimore-Arlington |
| 6 | Pittsburgh, PA MSA | 2,356,285 | Pittsburgh-New Castle-Weirton |
| 7 | Providence-Warwick | 1,600,852 | Boston-Worcester-Providence |
| 8 | Hartford-East Hartford-Middletown | 1,212,381 | Hartford-East Hartford |
| 9 | Buffalo-Niagara Falls | 1,135,509 | Buffalo-Cheektowaga-Cattaraugus |
| 10 | Rochester | 1,079,671 | Rochester-Batavia-Seneca Falls |

Table 11: Group Average of △HPI (%) and △GDP(%) in Northeastern US

| | | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 |
|---|---|---|---|---|---|---|---|---|---|---|
| Large | △HPI (%) | 7.10 | 0.57 | -3.25 | -6.24 | -3.53 | -3.17 | -2.08 | 0.55 | 2.73 |
| | △GDP(%) | 3.31 | 0.81 | 0.64 | -1.49 | 2.68 | 0.32 | 0.78 | 0.44 | 0.82 |
| Other | △HPI (%) | 7.09 | 3.28 | 0.12 | -3.01 | -2.72 | -1.56 | -1.42 | -0.07 | 0.27 |
| | △GDP(%) | 3.55 | 0.21 | 1.03 | -1.07 | 3.49 | 0.64 | 0.09 | 1.27 | 2.09 |

## 6.2 Empirical Strategy and Results

Due to our data set and purpose, although we have observations for several years, in this paper, we will use cross-sectional regression for each year instead of panel models. On one hand, due to our interest on large cities, we would like to identify the large city fixed effect. However, as the large cities are time invariant in our sample periods, fixed effect can not identified in a panel model. On the other hand, during the urban formation, housing price will influence the dynamic of industries thus have an impact on the spatial distribution of population. In the long-run, the reverse effect of housing price to city size is hard to be captured and controlled in a reduced form regression model framework. Thus, to separately consider the effect of city size to housing price, using yearly time window is a good choice. In such a short time period, the reverse effect of housing price to city size would be negligible. More importantly, spatial panel model pre-assume persistent and time-invariant spatial correlations, however for housing market, there might exist some time variant risk spillovers across regions which highly correlated with the economic

To investigate the effect of city size to housing price, especially to the spatial correlation among the housing prices, we would like to compare different specifications. The following models are considered:

**Model 1**: Linear Regression

$$y_{i,t} = \beta_0 + \beta_1 \Delta realGDP_{i,t} + \beta_2 Large_i + State_i + \varepsilon_i$$

**Model 2**: Spatial Autoregressive Model

$$y_{i,t} = \beta_0 + \rho \sum_{j=1}^{n} w_{ij} y_{j,t} + \beta_1 Large_i + \beta_2 \Delta realGDP_{i,t} + \beta_3 \sum_{j=1}^{n} w_{ij} \Delta realGDP_{j,t} + State_i + \varepsilon_i$$

**Model 3**: Heterogenous Spatial Autoregressive Model

$$y_{i,t} = \beta_0 + \rho_L Large_i \sum_{j=1}^{n} w_{ij} y_{j,t} + \rho_S (1 - Large_i) \sum_{j=1}^{n} w_{ij} y_{j,t} + \beta_1 Large_i$$
$$+ \beta_2 \Delta realGDP_{i,t} + \beta_3 \sum_{j=1}^{n} w_{ij} \Delta realGDP_{j,t} + State_i + \varepsilon_i$$

for $t = 2006, \cdots, 2014$ and $i$ belongs to our sample counties. In the regressions above, $y_{i,t}$ is the annual change of HPI (%) of county $i$in year $t$. $\Delta realGDP_{i,t}$ is the annual real GDP growth rate (%) of county $i$in year $t$. $Large_i$ is the large city dummy, which equals 1 if the county $i$ belongs to the CSAs listed in Table 17. To control the policies on housing markets and other unobserved factors correlated to local economy, geography and demography, state level fixed effect, which is denoted as $State_i$, is also included in the regressions as the dummy variables for states. Since most of the states in Northeastern are small and with small variations among counties on climate and pollution, and the local policies regulating housing market are mostly state level, the state dummy would be a good control for regional specific short-run factors which may affect local housing demand. The real GDP growth captures the economics condition of each county. In Model 2 and Model 3, we include neighborhood effect of local economic performances. To estimate the direct neighborhood effect through the price channel, control the neighborhood effect through local economies channel is necessary. Annual real GDP growth rate is a good approximation for local economic performance throughout a year.

The neighborhood among counties are defined by whether they are adjacent to each other or not, i.e. sharing land borders. We assume the spatial correlations for a county with all its neighbors are homogeneous, and the total effect of neighbor are normalized to 1. Thus, we have

$$w_{ij} = \begin{cases} 1/n_i & if\ j\ is\ adjacent\ to\ i \\ 0 & else \end{cases}$$

where $n_i$ is total number of county $i$. Thus, $\sum_{j=1}^{n} w_{ij} y_{j,t}$ and $\sum_{j=1}^{n} w_{ij} \Delta realGDP_{j,t}$ are the average of annual HPI change (%) and GDP growth rate (%) of county $i$'s neighbors in year $t$, which is similar to linear-in-means setting in peer effect literatures. The only difference is that there is no such a group concept in our setting. Thus even if the neighborhoods are symmetric, neighbors may have different neighbors despite other group members. In some regional economic literatures, neighborhood relationship are defined by distance between two regions. However, since county is a relatively large area instead of a single point, how to define the geographical distance between counties is quit difficult and more or less subjective. Without additional information, adjacency should be the most objective and natural way to capture the geographical spatial correlations. Thus, now we can see clearly, Model 1 totally ignore the spatial correlations among the counties, and assume the city size only have a direct effect on local housing market. Model 2 and Model 3 added spatial correlations among counties by two different channels: housing market itself as well as economic growth. The difference between Model 1 and Model 2, is that whether the city size

Table 12: Test Results of Moran's $I$, $LM1$ and $LM2$

| | | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 |
|---|---|---|---|---|---|---|---|---|---|---|
| Moran | Statistic | 8.39 | 2.06 | 6.09 | 7.99 | 5.74 | 7.86 | 10.07 | 1.37 | 2.08 |
| | p-value | .00 | .04 | .00 | .00 | .00 | .00 | .00 | .17 | .04 |
| $LM1$ | Statistic | 87.83 | 9.56 | 41.75 | 65.92 | 29.13 | 69.44 | 119.91 | 2.60 | 6.70 |
| | p-value | .00 | .01 | .00 | .00 | .00 | .00 | .00 | .27 | .04 |
| $LM2$ | Statistic | 8.07 | 3.73 | 2.05 | .62 | .10 | 2.61 | 1.14 | 1.33 | 4.72 |
| | p-value | .00 | .05 | .15 | .43 | .75 | .11 | .29 | .25 | .03 |

has an effect on the spatial correlations of housing market, i.e. whether the interaction of housing market among neighbor counties are heterogeneous driven by their size. Comparison between these three models can help us understand better about the city size effect in housing market through different channels.

Before estimation, some statistical tests for spatial correlations can be done. To test whether spatial correlation exists, i.e. differentiate between Model 1 and Model 2/3, we can implement Moran's I test and the $LM1$ test which we give in Section 3. To test whether the spatial correlation is heterogeneous which caused by city size, i.e. differentiate between Model 2 and Model 3, we can implement $LM2$ test which we give in Section 4. The test results are showed in Table 12, with reporting the value of test statistics and the p-value using asymptotic critical values (For Moran's $I$, we report the standardized test statistic value instead of the original Moran's $I$). For Moran's $I$, $H_0$ is $\rho = 0$ in Model 2; for $LM_1$ and $LM_2$, their $H_0$'s are $\rho_H = \rho_L = 0$ and $\rho_H = \rho_L$.

For the existence of the spatial correlation, p-value of Moran's $I$ and $LM1$ both suggest the existence of spatial correlation despite in 2013. For most sample years, the test statistics rejects the null hypothesis at 5% significant level. By contrast, results of $LM2$ are pretty different. In 2006 and 2014, the $LM2$ rejects the null hypothesis at 5% significant level, and rejects the null hypothesis at 10% significant level in 2007. However, in other sample years, the test statistic is insignificant. The results indicate a very interesting phenomenon: the spatial correlation among housing prices has some degrees of heterogeneity by city size, which is correlated with the general performance of housing market. The significant years, 2006, 2007 and 2014, are housing boom years, when both the housing prices in large cities and small cities were rising up. When the housing prices went down, the difference of spatial correlation seems to disappear. As $\sum_{j=1}^{n} w_{ij} \mathit{\Delta realGDP}_{j,t}$ had been included when constructing $LM2$, it indicates that the heterogeneity is not driven by neighborhood effect of economic performance.

Table 13: Results of Model 1 (Linear Regression)

| | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 |
|---|---|---|---|---|---|---|---|---|---|
| $\beta_0$ | 8.28*** | 2.19*** | .73 | -3.34*** | -2.55*** | -1.81*** | -1.44** | -.77 | -1.23** |
| | (.92) | (.64) | (.64) | (.74) | (.61) | (.58) | (.56) | (.51) | (.65) |
| $\beta_1$ | -.90* | -2.27*** | -1.83*** | -1.50*** | .35 | -.56* | .16 | .70** | 2.67*** |
| | (.54) | (.38) | (.38) | (.43) | (.35) | (.034) | (.33) | (.31) | (.39) |
| $\beta_2$ | .11** | -.01 | -.02 | .06 | .12*** | .20*** | .00 | .05 | -0.01 |
| | (.05) | (.04) | (.04) | (.04) | (.04) | (.05) | (.034) | (.03) | (0.04) |
| $R^2$ | .32 | .39 | .52 | .55 | .42 | .37 | .26 | .16 | .31 |

The next step is to estimate Model 1, 2 and 3 for each year to further investigate the size effect. For Model 1 and Model 2, the maximum likelihood estimators are used in constructing $LM1$ and $LM2$. The consistency and asymptotic Normality of QMLE/MLE for SAR model had been proved in Lee (2004). To estimate the Model 3, recall equation (2), we can view it as a special case of a higher order spatial autoregressive model with two different types of spatial correlations. In our setting, we can view $H_1 W_n$ and $H_2 W_n$ as two different spatial correlation matrices. Gupta and Robinson (2018) had investigated the asymptotic properties of pseudo-maximum likelihood estimator for higher order SAR model. However, in their simulation exercise, the finite sample bias is relatively large when the true residual is not Normally distributed with small samples. Besides, the post estimation $t$-statistic does not have good performance especially when sample size is small even for Normal case. Here, we implement the MLE procedure to estimate our Model 3, but since we have a relatively small sample with only 240 observations, the post-estimation $t-$tests may have over-reject or lack of power issue which is not as trustable as the $LM$ tests we implemented before.

Estimation results are reported in Table 13, 14 and 15, with all the parameters and their standard deviations except for state level fixed effects. The standard deviation reported in Table 14 and 15 are derived from asymptotic variance of the maximum likelihood estimators. We also reported $R^2$ for each specification. Although by likelihood approach, the sum square of residual is not minimized, $R^2$ can still be a good measure of fitness. Pseudo-$R^2$ or other likelihood based benchmarks are not adopted in this paper, although MLE for SAR and HSAR are consistent, the true residual distribution may not be Normal. Thus, the maximized likelihood value may not be the true likelihood of our sample. In this sense, traditional $R^2$ would be a better benchmark since it is a distribution-free measure. We also report the post-estimation $t-$test results for $H_0 : \rho_H = \rho_L$, as a comparison and performance benchmark for our pre-estimation $LM2$ test.

Table 14: Results of Model 2 (SAR)

| | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 |
|---|---|---|---|---|---|---|---|---|---|
| $\rho$ | .51*** | .20** | .41*** | .46*** | .32*** | .46*** | .59*** | .09 | .10 |
| | (.07) | (.09) | (.07) | (.07) | (.08)) | (.07) | (.06) | (.10) | (.09) |
| $\beta_0$ | 4.00*** | 1.71*** | .58 | -1.61*** | -2.19*** | -1.3** | -.53 | -.72 | -1.22 |
| | (.94) | (.64) | (.57) | (.68) | (.65) | (.53) | (.47) | (.50) | (.64) |
| $\beta_1$ | -.87* | -2.02*** | -1.23*** | -.82** | .61* | -.26 | .41 | .67** | 2.58*** |
| | (.46) | (.38) | (.35) | (.38) | (.32) | (.30) | (.27) | (.30) | (.39) |
| $\beta_2$ | .09** | -.01 | -.02 | .02 | .07** | .15*** | .01 | .05 | -.02 |
| | (.04) | (.04) | (.04) | (.04) | (.03) | (.04) | (.03) | (.04) | (.04) |
| $\beta_3$ | .07 | .05 | .12 | .13 | .19*** | .17* | .00 | .03 | .06 |
| | (.09) | (.08) | (.09) | (.08) | (.07) | (.09) | (.07) | (.07) | (.07) |
| $R^2$ | .34 | .40 | .54 | .58 | .46 | .42 | .29 | .16 | .31 |

By comparing results showed on Table 13, 14 and 15, we can clearly see that Model 3 specification has significantly higher $R^2$ than Model 1 and 2 in most of the sample years. By considering the heterogeneous neighborhood effect corresponding to city size, Model 3 can explain the variance of annual HPI change in our sample regions much better except for 2013. Due to the limitation of the yearly data, there might be some structural change or seasonal issue in the middle of the year which can not be captured by annual level economic performance and state level fixed effect. It might be a potential reason why all of three specifications do not work well for 2013. Except 2013, our yearly HSAR regression performs well in goodness of fitting. By comparing results showed on Table 12 and Table 15, the pre-estimation test $LM_2$ for heterogeneous spatial correlation works pretty well. The asymptotic p-value of post-estimation $t-$statistic are almost identical to the asymptotic p-value of $LM2$ in all our sample years. Also, results of $\rho_L$ and $\rho_S$ in Table 15 also indicate the good performance of $LM1$.

After comparison between models, let us focus on the effect of city size identified by Model 3. On one hand, the results of $\beta_1$ indicates a time-varying difference among the performance of

Table 15: Results of Model 3 (HSAR)

| | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 |
|---|---|---|---|---|---|---|---|---|---|
| $\rho_L$ | .67*** | .36*** | .50*** | .50*** | .35*** | .34*** | .53*** | .21 | .33** |
| | (.08) | (.12) | (.09) | (.08) | (.11) | (.10) | (.10) | (.15) | (.14) |
| $\rho_S$ | .31*** | .05 | .33*** | .41*** | .31*** | .55*** | .66*** | -.02 | -.08 |
| | (.10) | (.12) | (.09) | (.08) | (.10) | (.09) | (.08) | (.13) | (.12) |
| $\beta_0$ | 5.50*** | 2.08*** | .60 | -1.77** | -2.23*** | -1.15** | -.43 | -.80 | -1.3** |
| | (1.09) | (.67) | (.56) | (.69) | (.66) | (.53) | (.47) | (.50) | (.63) |
| $\beta_1$ | -3.35*** | -2.72*** | -1.01*** | -.50 | .73 | -.71* | .22 | .56* | 1.92*** |
| | (.99) | (.53) | (.38) | (.57) | (.50) | (.40) | (.32) | (.32) | (.50) |
| $\beta_2$ | .10*** | -.01 | -.03 | .02 | .07** | .14*** | .01 | .05 | -.02 |
| | (.04) | (.04) | (.04) | (.04) | (.03) | (.04) | (.03) | (.04) | (.04) |
| $\beta_3$ | .11 | .05 | .13 | .13 | .19*** | .16* | .00 | .02 | .06 |
| | (.09) | (.08) | (.09) | (.08) | (.07) | (.09) | (.07) | (.07) | (.07) |
| $R^2$ | .84 | .59 | .61 | .82 | .75 | .67 | .53 | .17 | .44 |

Table 16: Post Estimation $t$-test for $H_0 : \rho_L = \rho_S$

| | | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 |
|---|---|---|---|---|---|---|---|---|---|---|
| $t$-statistic | Statistic | 2.81 | 1.93 | 1.42 | .78 | .30 | -1.60 | -1.05 | 1.14 | 2.13 |
| | p-value | .01 | .05 | .16 | .44 | .76 | .11 | .29 | .25 | .03 |

housing market in large city areas and the others. From 2006 to 2008, the US economy shifted from booming to depression, as well as the housing price, there is a strong and significant negative effect among the large city areas. On average, the housing price in large city areas had more than 1% drop than other areas in North eastern US after controlling neighborhood, economic growth and policy effects. However, as the economy recovered from the depression slowly, the gap became ambiguous and insignificant. Finally, after the housing market recovered to expansion in 2014, the gap flipped to positive and significant, which is nearly 2% more annual growth in large cities. On the other hand, the size heterogeneity among cities also affect the spatial correlations among housing markets. When the housing market was expansion in 2006, 2007 and 2014, large city regions had significantly higher correlations with their neighbors. When neighborhood regions had 1% increase on housing price on average, the positive externality to large city areas are more than 0.3% than to other areas. During the depression, the gap between the externality disappear, although the spatial correlations are still significant expect in 2013. To our surprise, both the economic performance of a county itself and the neighborhood economic performances do not affect the housing market significantly. Local real GDP growth rate only had marginal impact on short-run housing price change in the sample years which was not even stable. Externality through the neighbors' economic growth to local housing price was also weak and insignificant.

## 6.3 Why city size matters?

First, let us focus on the direct city size effect. Since the estimation results showed a dynamic correlation among city size and housing price which is highly correlated with the business cycle, in this section, we tried to investigate the origin of this relationship. Based on empirical results in literatures and some facts showed in 2010 Census, the key factor here would be geographical distribution of mortgage delinquency, which is driven by the different income distributions in large cities and small locations.

During the financial crisis around 2008, housing mortgage delinquency was one of the key risk source. By Mian and Sufi (2009), the housing price appreciation during 2001-2005 and the subsequent mortgage defaults during 2005-2007 are driven by the rapid expansion in the supply of mortgages. Further, in Adelino, Schoar and Severino (2015), by tracking individual data of household income, credit score and loan data, they found out the middle class people with relatively high income and FICO score contributed increasing share of total housing mortgage delinquency from 2003 and 2006. In Mian and Sufi (2015), they pointed out that the contribution to the total dollar rise in household debt was strongest among individuals in the 20th to 60th percentile of the initial credit score distribution from 2000 to 2007, however 73% and 68% of the total amount of delinquent debt in 2007 and 2008 were contributed by the bottom 40% of the credit score groups. After large amount of delinquencies happened, the house corresponding to delinquent mortgage would highly probably be processed foreclosure, and banks would be more cautious on approval for new mortgages, there would be two simultaneous shocks to the housing market: positive to housing supply and negative to housing demand, which would drive the housing price to decrease.

Based on the analysis before, during the delinquency boom from 2005 to 2008, mortgages to low-income class kept contributing the major part as well as the increasing contribution of middle-income class, the income distribution of different areas is an important factor on how large the shock is to the housing market. Not only the absolute income level itself, but also the relative income to housing price matters a lot. On one hand, inside a region, if the income inequality is more serious, there will be more low-income and low middle-income people who need to raise their leverage to buy a house relative to the population size. On the other hand, a region with relatively higher housing price comparing to income will make the low-income and middle-income classes even harder to buy their own house, and force them to borrow more money from banks. The housing market of a region with a higher proportion of relatively "poor" people would be benefit from the increasing of housing loans during the housing booms especially when banks expand sub-prime mortgage , but will definitely received more shocks when the financial crisis happens.

Unfortunately, large city areas are facing both of the issues, highly unequal distribution of absolute income and relatively low income comparing to home price. On one hand, Baum-Snow and Pavan (2013) identified a strong positive monotonic relationship between wage inequality and city size during 1979~2007 in the US. Even after controlling the difference among skills and experiences, the rapid growth in wage inequality in larger locations can explain at least 23% of the nationwide increase in the variance of log hourly wages. On the other hand, based on data from Joint Center for Housing Studies (JCHS) of Harvard University[5], county belongs to large MSAs and surrounding areas have much higher home price to income ratios, especially during the housing boom from 2002 to 2006. As we analyzed before, when large amount of delinquencies happened starting from 2005, the negative shock to housing market in large city regions would be larger. After the financial system recovered and housing market expanded again, due to credit expansion, the inequality would drive the housing price to grow. Our estimation result of $\beta_1$ showed on Table 15 captures this dynamic effect almost perfectly for different sample years which is also on the different stage of the financial crisis. The results here also provide indirect evidence that dynamic of inequality is a major source of financial crisis, which is investigated in Kumhof, Ranciere and Winant (2015), with considering the degree of geographical inequality and relative inequality comparing to the housing price level in different regions.

After investigate the direct effect of size, let us move to the heterogenous spatial correlation corresponding to the city size. Our result shows that the counties belongs to large cities and surrounding areas received significantly larger positive externality from neighborhood regions than small locations when the housing price was rising in 2006, 2007 and 2014. However, when the market performed bad, this difference become insignificant. It seems not intuitive since large cities in general have better job opportunities, education and medical resources and entertainment facilities which

[5]See https://www.jchs.harvard.edu/blog/price-to-income-ratios-are-nearing-historic-highs

are more attractive to people. Also, the economy and labor market in large cities and surrounding areas in general are more integrated, thus have more inter-county transitions of people. These factors are not affected by business cycle. However, if we consider the geographical income inequality, the credit cycle might explain this effect. As we pointed out before, people are relatively poor comparing the higher housing price. For a potential migrator from a neighborhood area, it is easier for them to purchase a house during credit expansion. Also, more importantly, as an investor, he/she will only be motivated to buy a house if the housing price in targeting area is increasing. Thus, although large city areas are more attractive, the real estate assets may not be. Instead, some people may give up to find a job in large cities, and move to small towns or rural areas. Then, externality from this type of migration to housing market of small locations would increase, since housing markets receive smaller degree of shocks and housing prices are more persistent in small locations as we analyzed before. This may partially explain the disappeared gap of neighborhood effect observed in our sample.

As a conclusion, we can see that size of a city have a large impact on housing market, and the impact is dynamic and linked to business cycle and credit cycle. Not only the dynamic of housing price in large cities and small locations are significantly different, their correlations with neighbors are also distinct. Due to existing empirical evidence, the difference can be partially explained by income inequality and housing mortgages. In existing regional economics and housing finance literatures, the difference had not been investigated much yet. Since it might be helpful to policy makers to regulate housing market, it would be beneficial to keep tracking on it.

# References


[1] Abhimanyu Gupta, Peter M. Robinson, "Pseudo maximum likelihood estimation of spatial autoregressive models with increasing dimension", *Journal of Econometrics*, Volume 202, Issue 1, January 2018, 92-107

[2] Andrew Cliff, Keith Ord, "Testing for Spatial Autocorrelation Among Regression Residuals", *Geographical Analysis*, July 1972, Volume 4, Issue 3, 267-284

[3] Andrew Cliff, Keith Ord, "Spatial Autocorrelation", 1973, Pion, London

[4] Atif Mian, Amir Sufi, "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis", *The Quarterly Journal of Economics*, Volume 124, Issue 4, November 2009, 1449-1496

[5] Atif Mian, Amir Sufi, "Household Debt and Defaults from 2000 to 2010: Facts from Credit Bureau Data", NBER working paper 21203, May 2015

[6] David Segal, "Are There Return to Scale in City Size?", *The Review of Economics and Statistics*, Vol. 58, No. 3 (Aug., 1976), 339-350

[7] Eleonora Patacchini, Edoardo Rainone, Yves Zenou, “Heterogeneous peer effect in education”, *Journal of Economic Behavior & Organization*, Volume 134, February 2017, 190-227

[8] Gregor Matvos, Micheal Ostrovsky, “Heterogeneity and peer effects in mutual fund proxy voting”, *Journal of Financial Economics*, Volume 98, Issue 1, October 2010, 90-112

[9] Harry H. Kelejian, Ingmar R. Prucha, “On the asymptotic distribution of the Moran I test statistic with applications”, *Journal of Econometrics*, Volume 104, Issue 2, September 2001, 219-257

[10] Jeffrey P. Cohen, Cletus C. Coughlin, “Spatial hedonic models of airport noise, proximity, and housing prices”, *Journal of Regional Science*, VOL. 48, NO. 5, 2008, 859–878

[11] James P. LeSage, Yao-Yu Chih, “Interpreting heterogeneous coefficient spatial autoregressive panel models”, *Economic Letters*, Volume 142, May 2016, 1-5’

[12] James P. LeSage, Colin Vance, Yao-Yu Chih, “A Bayesian heterogeneous coefficients spatial autoregressive panel data model of retail fuel duopoly pricing”, *Regional Science and Urban Economics*, Volume 62, January 2017, 46-55

[13] Kumhof, Michael, Romain Rancière, and Pablo Winant. 2015. "Inequality, Leverage, and Crises." *American Economic Review*, 105 (3): 1217-45.

[14] Lung-Fei Lee, “Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models”, *Econometrica*, Vol. 72, No. 6 (November, 2004), 1899-1925

[15] Manuel Adelino, Antoinette Schoar, Felipe Severino, “Loan Originations and Defaults in the Mortgage Crisis: The Role of the Middle Class”, *The Review of Financial Studies*, Volume 29, Issue 7, July 2016, 1635-1670

[16] Michele Aquaro, Natalia Bailey and M. Hashem Pesaran, “Estimation and inference for spatial models with heterogeneous coefficients: An application to US house prices”, *Journal of Applied Econometrics*, June 2020, Volume 36, Issue 1, 18-44

[17] Nathaniel Baum-Snow, Ronni Pavan, “Understanding the City Size Wage Gap”, *The Review of Economic Studies*, Volume 79, Issue 1, January 2012, 88-127

[18] Nathaniel Baum-Snow, Ronni Pavan, “Inequality and City Size”, *The Review of Economics and Statistics*, December 2013, 95(5): 1535-1548

[19] Olga Yakusheva, Kandice A. Kapinos, Daniel Eisenberg, “Estimating Heterogeneous and Hierarchical Peer Effects on Body Weight Using Roommate Assignments as a Natural Experiment”, *The Journal of Human Resources*, Winter 2014, Vol. 49, No. 1, 234-261

[20] P. A. P. Moran, "A Test for the Serial Independence of Residuals", *Biometrika,* Vol. 37, No. 1/2 (Jun. 1950), 178-181

[21] Ronald L. Moomaw, "Productivity and City Size: A Critique of the Evidence", *The Quarterly Journal of Economics*, Volume 96, Issue 4, November 1981, 675-688

[22] Ryan R. Brady, "The spatial diffusion of regional housing prices across U.S. states", *Regional Science and Urban Economics*, Volume 46, May 2014, 150-166

[23] Susan E.K. Christofferen, Sergei Sarkissian, "City size and fund performance", *Journal of Financial Economics*, Volume 92, Issue 2, May 2009, 252-275

[24] William A. Brock, Steven N. Durlauf, "Interaction-based models", *Handbook of Econometrics*, Volume 5, 2001, Chapter 54, 3297-3380

[25] Yong Tu, Hua Sun, Shi-Ming Yu, "Spatial Autocorrelation and Urban Housing Market Segmentation", *The Journal of Real Estate Finance and Economics*, 34, 385-406(2007)