Should we Fear the Shadow? House Prices, Shadow Inventory, and the Nascent Housing Recovery [1]

Hua Kiefer∗, Leonard Kiefer†, and Thomas Mayock∗
∗ *Office of the Comptroller of the Currency* , † *Freddie Mac*

Version: June, 2013

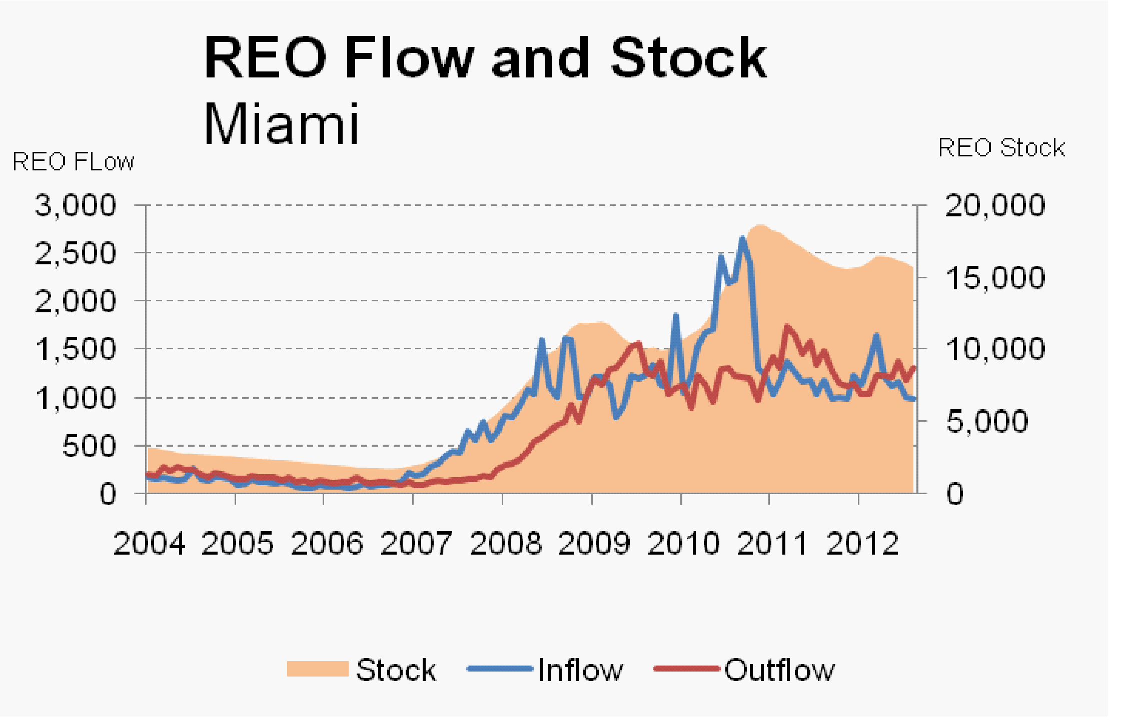
Although a broad-based increase in house prices has been observed over the past year, not everyone is convinced the rise of house prices will persist and lead to a steady recovery of the economy. The main reason for this skepticism is uncertainty about the "shadow inventory": foreclosed homes held by investors or as REOs, which have not yet hit the market but likely will as market prices rise. The volume of shadow inventory itself in local markets is largely unknown, as is its impact on the housing market. This study quantifies the size of the shadow inventory and investigates the spatial impact of the of the out-flow of shadow inventory. The scope of our study is a set of housing markets that vary in both their historic housing price volatility as well as institutional factors - such as foreclosure law statutes - that may influence the relationship between the shadow inventory and house price dynamics. To address the endogeneity that characterizes the spatial interaction of house prices and the out-flow of the shadow inventory, we utilize a simultaneous equation system of spatial autoregressions (SESSAR). The model is estimated using measures of the shadow inventory derived from DataQuick's national transaction history database and county-level house price indices provided by Lender Processing Services. Lastly, because our estimate - as well as all other existing estimates - of the shadow inventory relies upon string matching algorithms to identify entry into and exit out of REO status, we validate the accuracy of our measures of REOs using loss mitigation data from the OCC Mortgage Metrics database.

1. INTRODUCTION

After five years of declines, national house price indexes bottomed out in 2012. In the last half of 2012 and the first quarter of 2013, house prices have risen across most of the country. The rebound in house prices was especially pronounced in those areas that were hardest hit by the housing crash. As many markets posted double digit percentage gains in house prices, fears of a renewed housing bubble have surfaced. Scarred by the previous decade, economists, policymakers, and market watchers are all asking: "*Are the recent house prices gains sustainable?*"

One possible headwind against sustained house price growth is the so-called shadow inventory. Precise definitions of the shadow inventory vary, but it is generally accepted to include homes in the foreclosure process, as well as homes held by banks and financial institutions (REO) prior to liquidation. Researchers and policymakers alike have expressed concern over the rapid growth in the number of bank-owned properties in the wake of the collapse of the housing market. A growing literature (e.g., Harding et al. (2009); Campbell et al. (2011)) suggests that such properties generate significant negative externalities that are capitalized into the values of surrounding homes. Although the existence of foreclosure externalities is widely accepted, the exact mechanism generating these externalities is the subject of debate. The most common explanation for the existence of foreclosure spillovers is that while a property is

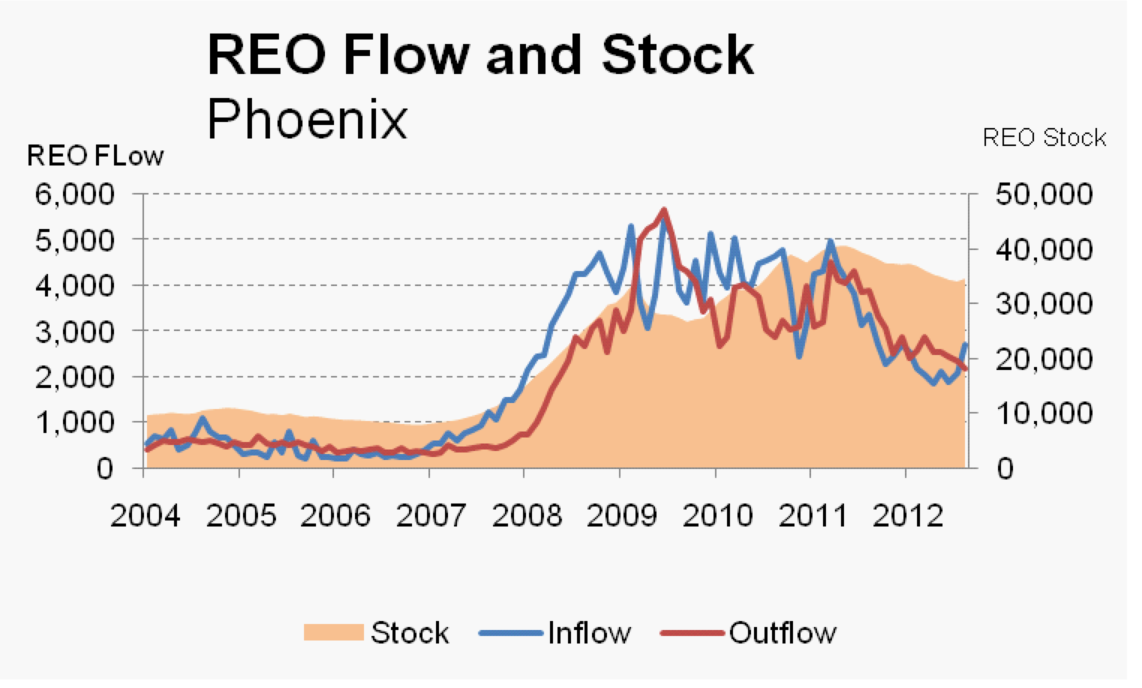
[1] The views expressed in this paper are those of the authors alone and do not reflect those of the Office of the Comptroller of the Currency and Freddie Mac.



in REO status, neighborhood quality declines because the vacant and oftentimes dilapidated REO properties reduce the physical attractiveness of a neighborhood and may attract crime (Ellen et al. (2011)). If the REO spillover effect is solely attributable to such effects and the sale of an REO property results in the property being better maintained, then the release of the shadow inventory should actually improve the price of surrounding properties over time. A number of papers (e.g., Harding et al. (2009); Mayock and Ihlanfeldt (2013)) provide evidence that the spillover effect is quite long lasting, however, with formerly bank-owned properties reducing nearby property values up to two years after exiting from REO status. The persistence of this effect has been hypothesized to be the result of REO sales serving as comparables in the negotiations for non-distressed properties or the REOs being sold to investors and converted to rental properties, which are also known to produce spillovers (Coulson and Li, 2013).

If post-REO spillovers are larger in magnitude than the spillovers associated with bank-owned properties, then the sale of large quantities of REO properties could exert substantial downward pressure on home values, particularly in several of the markets where price growth since 2012 has been strong. For instance, as of September of 2012, Phoenix and Miami had posted year-over-year price gains of 20 percent and 10 percent, respectively. The REO stock in these markets, however, is still very large; as of the 3rd Quarter 2012, over 15,000 properties are REO in Miami and about 30,000 properties are REO in Phoenix. The extent to which price gains in such markets can be maintained depends critically on the relationship between the liquidation of the shadow inventory and price dynamics. Although much recent research has studied the how the entry into foreclosure affects nearby house prices, the existing literature provides little guidance about the impact that the outflow of properties from REO status has on housing values. In this paper, we set out to gain a deeper understanding of this relationship so as to better know what to expect as the substantial shadow inventory moves through the foreclosure pipeline and ultimately hits the market.

Our study requires detailed estimates of both movements in housing values as well as the shadow inventory. Unfortunately, estimates of the shadow inventory are not widely available. Many data vendors do provide estimates of the REO haircut and its variation over time;



although such estimates are useful, they cannot be used to identify the impact of the outflow of REO properties on the values of non-distressed homes in the same area. The first contribution of this paper it thus the construction of shadow inventory measures for a collection of U.S. markets, which we construct using DataQuick's national transaction history database. Because our shadow inventory measure is an estimate that, like all other existing estimates of the REO stock, relies upon string matching algorithms to identify entry into and exit out of REO status, we validate the accuracy of our stock measures using loss mitigation data from the OCC Mortgage Metrics database.

The second contribution of our paper is the estimation of the relationship between REO sales and housing prices. Identifying this relationship is complicated by two distinct sources of endogeneity. First, it is our expectation that the flow out of the REO stock is simultaneously determined with housing prices. As prices rise, banks wishing to mitigate loan losses will face stronger incentives to bring REO properties to market. For reasons stated above, it is also the case that the sale of REO properties may exert a marketwide impact on housing values. The second source of endogeneity in our study is spatial in nature. Several researchers have demonstrated the existence of spatial autocorrelation in housing price models.[2] In and of itself, such spatial autocorrelation does not introduce bias or inconsistency to non-spatial estimators. However, if the spatial autocorrelation is indicative of serious model misspecification such as the joint determination of covariates across space, standard estimators may be rendered inconsistent. To address these two types of endogeneity, we build on the generalized spatial three-stage least squares (GS3SLS) model of Kelejian and Prucha (2004), which was designed to estimate a simultaneous equation system of spatial autoregressions (SESSAR) on cross sectional data. Because we utilize panel data, however, the GS3SLS method of Kelejian and Prucha (2004) cannot be used directly because of the presence of panel-specific heterogeneity; to address this heterogeneity, we utilize the same estimation technique as in Kiefer and Kiefer (2011) and demean our data (Baltagi (2001)) before the implementation of the GS3SLS

[2] See Can (1992), Can and Megbolugbe (1997), Pace and Gilley (1997), and Pace, Barry, and Clapp (1998) as examples.

estimator.[3] [4] This approach allows us to control for possible bias arising from unobserved heterogeneity, spatial correlation, and simultaneity.

The rest of the paper is organized as follows. Section 2 provides a brief review of the literature on the relationship of distressed sales and house prices. Section 3 describes our unique dataset that allows us to measure the REO net flow. Section 4 describes the economic and statistical models used. Section 5 presents the results of our estimation. Section 6 concludes with an overview of the policy implications of our results.

2. LITERATURE REVIEW

The recent collapse of the housing market and the resulting surge in foreclosures stimulated interest in the impact that bank-owned real estate has on surrounding properties. The large and growing literature on this topic suggests a consensus amongst researchers that distressed homes reduce the value of nearby properties and that this spillover effect can persist for years following the liquidation of a property out of REO status. The strong interest in foreclosure issues notwithstanding, no study has yet investigated the relationship between the stock of REO properties – the "shadow inventory" – and housing prices. One of the impediments to such research is likely the dearth of reliable estimates of the REO stock.[5] Ellen et al. (2012) note, for instance, that "despite the policy attention REO properties have received, our understanding of the size and nature of current REO stocks, as well as what becomes of properties after being sold, is extremely limited or anecdotal" (p. 3). In fact, as of this writing, only three studies have attempted to measure the REO stock, and each of these studies has been primarily descriptive in nature.

Immergluck (2009) uses data from 16 mortgage servicers aggregated by Lender Processing Services (LPS) to estimate the size of the REO stock in metropolitan areas throughout the country between August 2006 and August of 2008. Because the LPS data is drawn from a subset and not the universe of servicers, the REO data derived from this database only represents a subset of the actual REO stock. To generate estimates of the full REO stock, the author utilizes a weighting scheme based on a more comprehensive delinquency report to rescale the REO counts in the LPS data. The resulting estimates reveal that the REO stock grew substantially over this period, particularly in market such as Phoenix, Las Vegas, and Miami that exhibited dramatic price growth during the housing boom. Smith and Duda (2009) utilize county recorder data to analyze the REO stock in Chicago between 2005 and 2008; like Immergluck (2009), they find a spike in REO inventory over the sample period, with a large concentration of the REO properties in minority communities.

The most recent study to estimate the REO stock is Ellen et al. (2012) who utilize deed data from Atlanta, Miami, and New York City to document the evolution of the REO stock. Unlike the two previous studies that only covered the early years of the housing downturn, the Ellen et al.'s data frame covers the depths of the crisis, spanning from 2002 through 2011. The authors document a sizable increase in the REO stock in each of the three cities between 2006 and 2008. The stock of REOs then began to decline in each of the three cities throughout

[3] Although, we could recover the estimates of the fixed effects using the results of the GS3SLS estimates, since the fixed effects are nuisance parameters, we do not report their estimated values in the paper.

[4] A small number of papers have employed spatial econometric techniques in a multi-equation framework. Steinners and Fisher(1974) first tried to incorporate spatial interactions of employment and population in their intra-urban population and employment model. Boarnet (1994) extended Carlino and Mills' (1987) model by introducing spatial cross-correlation into the employment and population equation system. Henry, Barkley, and Bao (1997) further extended Boarnet's (1994) model by adding interaction terms between urban growth rates and the spatial cross-regressive lags as additional explanatory variables. Henry, Schmitt, and Piguet (2001) augmented Carlino and Mills' (1987) and Boarnet's models by incorporating spatial auto-regressive lags, and compared these augmented models with Henry et al.'s (1997) model. Gebremariam, Gebremedhin, Schaeffer, and Jackson (2008) extended the conventional two equations system to a five-equation system, allowing the interdependences (spatial autocorrelation, spatial cross-correlation and cross-equational correlation) among employment growth, migration behavior, household income, and local public services.

[5] The literature on foreclosure contagion effects has been reviewed extensively in Frame (2010) and Ihlanfeldt and Mayock (2013) and will not be summarized here.

2009. In New York City, this decline continued throughout the sample period, whereas in Atlanta and Miami, the REO stock grew throughout 2010 before beginning to decline in 2011. The authors note that in each of the markets the decline in the REO inventory after the initial peak was driven by an increase in REO liquidation, whereas the most recent decline in the stock has primarily been the result of a reduced flow into the REO stock.

In sum, although a number of studies have investigated the spillover effects of foreclosed homes, there is no existing evidence on the relationship between housing prices and the speed of exit from the REO stock. The literature on the shadow inventory is limited to just three descriptive studies, and only one of these studies (Ellen et al., 2012) captured the full extent of the downturn in the housing market.

3. DATA

The data that we use to estimate our empirical models is derived from two primary sources: DataQuick's transaction database, which we use to construct our measures of the stock and flow of the shadow inventory; and the county-level house price index from Lender Processing Services (LPS). Moreover, to validate the efficacy of our algorithm in counting REO entries, we cross-reference to a unique dataset known as OCC Mortgage Metrics that contains precise loss mitigation and monthly performance outcomes for about 64% of U.S. mortgages.

3.1. REO Inventory

The DataQuick transaction database provides a history of sales and financing activity on residential housing units throughout the United States. For the purposes of this study, the transaction records are restricted to single-family residences and the time period spanning from January 1^{st} 2006 to August 31^{st} 2012. Our sample covers the peak of the housing bubble, the subsequent dramatic decline in prices, and a period of price stabilization. Our sample thus provides a complete picture of the evolution of the REO inventory through a nationwide boom-bust cycle. Because each property in the DataQuick data is geocoded, we can identify the spatial relationship among properties, a relationship that is needed to estimate our spatial models.

In addition to the characteristics included in most standard real estate transaction databases such as the sales price, transaction date, and property characteristics, the DataQuick data also includes a distressed sale indicator that we utilize to identify a property's entry into and exit from the REO stock. Specifically, the distressed sale indicator identifies transactions that fall into one of the following six categories

1. The buyer is identified as a government entity or bank/lender via a trustee's deed upon sale[6];
2. A trustee's deed upon sale is filed and the buyer is not a government entity nor bank/lender and the deed's type is not a redemption certificate;
3. The property is transferred to a financial institution, or a government agency (including GSEs);
4. A bank/lender transferred the property to the guarantor within a short time period (18 to 24 months) from acquiring the title;
5. A bank/lender or government entity transferred the property to a private party (REO liquidation);
6. A sale between private parties with a sale price at least 5% less than total outstanding loan (inferred short sale).

To construct the REO stock, we use these 6 categories to identify entries into and exits from the REO stock. Entries are defined as sales falling into Class 1 through Class 4, and the date on which the transaction took place serves as the property's date of entry into the REO stock. In the event that multiple entries are identified on the same property within a short amount

[6] A trustee's deed upon sale is a deed of foreclosure.

of time without a subsequent exit, the date of the earliest of these dates is defined as the entry date.[7] Once a property is identified with an REO entry, we then search all sales of the property following entry to identify the exit from the REO stock. If a sale event of class 5 is found, the transaction date of this event is then recorded as the REO exit date. Because of limitations of the string matching algorithm used to identify distressed transactions, however, it can be the case that the true exits from the REO stock are not classified as distressed transactions.[8] To correct for this shortcoming, we also consider non-distressed sales as possible exits from the REO stock. On a given property, the date of exit from the REO stock is then defined as the earliest of (1) the earliest Class 5 transaction following entry into the REO stock and (2) the first non-distressed sale after entry into the REO stock. Finally, we cap the REO duration at 3 years to further correct for missing records in the database.[9] Once the entry and exit dates are determined for each property, it is straightforward to construct measures of the stock of housing currently in REO status as well as the flow into and out of REO status. In our analysis below, the REO liquidation rate is computed by dividing the difference of REO exits and REO entries by the REO stock measure.

3.2. Cross Comparison with CoreLogic Report

To evaluate the quality of our REO measures, it would be beneficial to conduct a cross comparison of the REO numbers in this study with some publicly available REO data. To the best of our knowledge, the only public resource on REO measures is the reports produced by CoreLogic (i.e., "the Market Pulse" or previously called "US housing and Mortgage Trend"). In these reports, CoreLogic presents their estimated numbers of total sales, total REO sales, total short sales, share of distressed sales among all sales, and month supply of currently distressed homes (i.e., the number of months it takes to sell all the distressed homes in stock at the current liquidation speed) for the entire country, each of the 50 states, and the top 25 CBSAs. Because our study is state based, we pull CoreLogic's statistics of our 4 targeted states and report our measures along with CoreLogic's numbers in Figure(1).The sample time period is chosen for both August 2010 and August 2011[10]. There are two distinctions of the definitions between the measures in this study and that from CoreLogic: 1). our measures are related only to single-family residential units (SFRs), while CoreLogic's numbers include SFRs and condominiums; 2). as for the statistics of month supply of currently distressed homes, we define the total number of distressed homes as the current REO inventory, while CoreLogic uses the number of loans that are at least 90 days past due (DPD). In view of the slight difference in the variable definitions, the values of our measure are quite close to those of CoreLogic. It is worth noting that the REO sale counts from CoreLogic are not always greater than our values, in spite of the fact that CoreLogic covers both condominiums and SFRs (e.g., MD in 08/2010 and FL in 08/2011). This suggests that CoreLogic may be conservative in its identification of REO sales. Overall, these two sets of numbers agree with each other in terms of magnitude.

[7] Ideally, Class 1 and Class 2 should capture all the REO entries, and Class 3 and Class 4 would be associated with transactions following entry. Because of errors in public records or DataQuick's distress identification algorithm, however, it is possible for the true entry into the REO stock to go unrecorded. In such cases, a property would be associated with a Class 3 or Class 4 transaction with no subsequent Class 1 or Class 2 transaction. Because Class 3 and Class 4 transactions cannot generally occur without the property first being transferred to a financial institution, we flag such transactions as entries into the REO stock.

[8] For instance, for a sale to be characterized as a Class 5 transaction, the algorithm must recognize the seller as a financial institution. If the formatting of the seller name field was such that DataQuick's algorithm did not identify the seller as a financial institution, the true exit from the stock will not be captured by the distressed transaction field.

[9] If no Class 5 distressed sale or non-distressed sale is found within 3 years after a property enters the REO stock, we code the REO exit date for this property as 3 years since the date of entry.

[10] CoreLogic produced the report at a nonregular basis (usually every 2-4 months). Because their online pdf files are copy/paste proctected starting from the 2012 reports, we only take a snapshot for year 2010 and 2011.

FIG. 1 Close Values of REO Measures from DQ and CoreLogic

		AZ	FL	GA	MD
August, 2010	*Total Sales_ DQ*	8,005	18,023	4,877	4,381
	Total Sales_CL	9,087	27,116	6,676	3,100
	REO Sales_DQ	3,601	6,146	1,434	762
	REO Sales_CL	3,431	7,515	1,660	442
	Short Sales_DQ	1,670	3,411	270	454
	Short Sales_CL	1,382	3,598	388	219
	Distressed Share_DQ	65.8%	53.0%	34.9%	27.8%
	Distressed Share_CL	53.0%	41.0%	30.7%	21.3%
	Month Supply_DQ	12.7	10.6	23.7	8.6
	Month Supply_CL	13.0	20.8	19.5	24.5

		AZ	FL	GA	MD
August, 2011	*Total Sales_ DQ*	10,006	20,634	6,803	4,659
	Total Sales_CL	11,814	33,344	9,689	7,367
	REO Sales_DQ	4,890	5,804	2,514	811
	REO Sales_CL	3,879	5,784	1,849	824
	Short Sales_DQ	2,234	5,000	500	613
	Short Sales_CL	1,784	4,570	833	652
	Distressed Share_DQ	71.2%	52.4%	44.3%	30.6%
	Distressed Share_CL	47.9%	31.1%	27.7%	20.0%
	Month Supply_DQ	10.1	9.6	21.3	8.5
	Month Supply_CL	6.8	15.9	12	10.3

DQ represents our measures using DataQuick transaction database; CL represents measures from COreLogic reports.

3.3. Cross Comparison with OCC Mortgage Metrics Data

The values of REO sales from CoreLogic report line up with our REO exit measure. To validate our REO entry measure, we leverage the loan performance information from a unique date source - the OCC Mortgage Metrics data. This dataset consists of origination and monthly servicing observations of over 34 million mortgages (64% of all US residential mortgages) serviced by large U.S. banks regulated by the OCC. The monthly servicing data has been collected since January 2008. Due to the size of the data and the enormous time required to match addresses between the two databases, our validation of REO entry between the DataQuick and Mortgage Metrics data in this section is restricted to one state in our sample (Florida).

The servicing information collected monthly includes detailed information of mortgage payment, delinquency status, updated LTV, workout resolution, and etc. The key variable for our REO entry identification purpose is the foreclosure status updated monthly in the dataset. The mortgage companies report the foreclosure status of each loan in their monthly report. The status of a loan can be: 1). not in foreclosure (including cured from foreclosure); 2). foreclosure start/pre-sale – any mortgage that has been referred to an attorney to initiate legal foreclosure proceedings but has not yet gone to foreclosure sale; 3). post-sale foreclosure – any loan where the bank has obtained title at foreclosure sale, but the property is not yet actively being marketed (typically this will include loans that are in redemption or being repaired); 4). REO - for any mortgage where the bank has obtained title at foreclosure sale and the property is on the market and available for sale or other instances where the bank has obtained title but the availability for sale is not known. Case 1) is coded for loans not in foreclosure, while case 2) through 4) capture the entire foreclosure procedure. Essentially, our algorithm of counting REO entry based on observing a title transfer (to lender/bank, guarantor, government entity, or GSE), so case 3) and 4) would trigger our REO entry counting.[11] More specifically, the month when the foreclosure status is recorded as case 3) or 4) for the first time is coded as the REO entry time.

Both the DataQuick and Mortgage Metrics data include address information of every residential property. To validate our measure of REO entry, we can try to match the list of REO entries derived separately from these two datasets (DataQuick vs Mortgage Metrics) based on property address and REO entry date information[12]. Because banks and county offices differ in the recording time (sometimes the deed transfer appears in the county office a few months later than when it is recorded in the bank system; other times it is the other way around), we need to establish a matching rule of the allowed time lag between the two REO entry dates identified respectively using the DataQuick and Mortgage Metrics data. We therefore select a month within our observation window when the REO entries are mostly populated (i.e., June 2010 with 9,698 REO entries) for the Mortgage Metrics data, and then match the REO entries in this month with the entire REO list from DataQuick data. After plotting for the distribution of matched cases of REO entries across time, we found most of the matches can be located ± 12 months from the identified REO entry date (about 85% of matches are within an observation window of [June 2009, June 2011]). So we require a maximum of 12 months in the absolute value of the difference between two identified REO entry dates in addition to the address-matching rule. Particularly, we match the properties with an REO entry in a given month from the Mortgage Metrics data to those REO properties identified with an observation window of ± 12 months from the given month in the DataQuick data and report the match rate in Figure (**??**) below. The first column specifies the reference month of REO entries found in Mortgage Metrics data; the second column reports the number REO entries identified in the Mortgage Metrics data for the reference time; the third and fourth column

[11] Case 2) is excluded for the reason that lenders/banks have not acquired the title yet, so there is no record in the county office reflecting the property transfer. Either case 3) or 4) indicates a title transfer to the lender/bank, which triggers our counting for REO entry.

[12] To account for potential multiple REO entries of the same property it is necessary to match with the REO entry date as well as the address.

FIG. 2 REO Entry Validation - FL State: Match MM with DQ

Reference Time yyyymm	# of REO Entry from MM	Matched Cases	Matching Rate	Reference Time yyyymm	# of REO Entry from MM	Matched Cases	Matching Rate
201001	2,944	1,895	64.37%	***201101***	1,575	771	48.95%
201002	2,357	1,582	67.12%	***201102***	1,099	576	52.41%
201003	3,337	2,206	66.11%	***201103***	1,351	720	53.29%
201004	3,401	2,250	66.16%	***201104***	1,313	658	50.11%
201005	3,487	2,314	66.36%	***201105***	1,720	979	56.92%
201006	9,698	5,507	56.78%	***201106***	2,143	1,307	60.99%
201007	4,399	2,815	63.99%	***201107***	2,214	1,362	61.52%
201008	4,693	3,092	65.89%	***201108***	1,935	1,185	61.24%
201009	5,554	3,825	68.87%	***201109***	2,027	1,222	60.29%
201010	3,094	1,897	61.31%	***201110***	4,163	2,949	70.84%
201011	2,349	1,405	59.81%	***201111***	2,804	1,816	64.76%
201012	1,934	1,000	51.71%	***201112***	1,717	1,050	61.15%
Reference Time yyyymm	# of REO Entry from MM	Matched Cases	Matching Rate				
201201	2,545	1,733	68.09%				
201202	2,511	1,670	66.51%				
201203	2,600	1,720	66.15%				
201204	2,310	1,476	63.90%				
201205	2,328	1,558	66.92%				
201206	2,455	1,675	68.23%				
201207	2,319	1,472	63.48%				
201208	2,787	1,864	66.88%				
201209	2,822	1,763	62.47%				
201210	3,230	1,827	56.56%				

present the matched cases in number and in percentage. For example, 2,944 properties are identified entering REO stock in January, 2010. When these properties are matched with the REO entry list in DataQuick for a time period of [January 2009, January 2011], 1,895 cases are returned, which is 64.37% among the total 2,944 cases.

We have reported the matching rate for the time period between January 2010 and October 2012, which is the observation time period of the Mortgage Metrics data. The matching rate is usually around 60%. Ideally, if the mortgage servicers always flag the foreclosure status of their loans on time, and our algorithm for REO entry is perfect, we should see close to 100% matching rate (or close to 85% allowing for the disparity in recording time and with a restriction of the matching time window). The reality is that banks do not always record the change of foreclosure status accurately. For instance, the definition of "foreclosure status" suggest that case 3) must proceed case 4). But we find there are quite some properties flagged with case 4) without any case 3) happened in the past. Also the discrepancy of the recording time between the bank and county office systems suggest a potential delay in documenting the change of status. That being said, we have to admit that our REO measures are not perfect and can be improved.[13] However, the reported matching rate in Figure (??) and close values to the CoreLogic report in Figure (1) suggest that our REO measures are reasonable enough to be used in our regression analysis.

3.4. House Price Index and Other Economic Variables

We use county level house price index from Lender Processing Services (LPS) as another important input of our model. This HPI index series was developed using LPS' proprietary property and loan data (covers about 83% of properties in the US) as well as publicly available county-record data. Among the multiple HPI data tiers at the county level (e.g., by property type/price dependent path/seasonal adjustment), we select the series specifically for single family residence, averaged over all price tier, and nonseasonally-adjusted. Though the monthly refreshed HPI goes back to February 1991, we restrict the observation window to be January 2006 to August 2012 for consistency.

To help control for macroeconomic conditions we include two other variables. From the Bureau of Labor Statistics we get county level employment statistics. Local area job growth is a good measure of the health of the labor market and the broader economy. In all specifications we use the year over year percent change for the employment variable. Using annual differencing helps to account for nonstationary behavior of this variable. From DataQuick we construct a measure of total home sales, short sales, and cash sales in a given month. During the recession both employment and home sales fell sharply, while short sales increased. The inclusion of these variables help to control for the business cycle.

3.5. Summary Statistics

It is well known that judicial and non-judicial foreclosure proceedings can result in widely different foreclosure duration and terms.[14] The interaction between REO liquidation and house price can differ under these two systems. Additionally, the difference in the volitity of house price movement might potentially shape the movement patterns as well. Therefore, based on the different foreclosure law statutes and housing market condition, we select four local markets as our target study areas : Arizona, Florida, Georgia, and Maryland. Among these four states, FL and MD are judicial states with only judicial foreclosures allowed, and AZ and GA are non-judicial states with both judicial and non-judicial foreclosures. The housing

[13] More particularly, with the recent access to the buyer and seller's information in the DataQuick data, we will update our method to add more controls over the distress indicator for better capturing the REO entry and exit.

[14] Definitions from Mortgage Bankers Association: A judicial foreclosure is a court proceeding that begins when the lender files a complaint and records a notice in the public land records announcing a claim on the property to potential buyers, creditors and other interested parties. The complaint asks the court to allow the lender to foreclose its lien and take possession of the property as a remedy for non-payment. The defendant (borrower) is permitted to dispute the facts (such as show that payments were made), offer defenses or present counterclaims by answering the complaint, filing a separate suit, and / or by attending a hearing arranged by the court. If the defendant shows there are differences of material facts, a trial will be held by the court to determine if foreclosure should occur. If the court determines the homeowner did default and that the debt is valid, it will issue a judgment in favor of the servicer for the total amount owed, including costs for the foreclosure process. Next, the court will authorize a sheriff's sale. The sale is an auction of the property open to anyone, and must be held in a public place. The individual with the highest bid is granted the property. After the sale is confirmed by the court, the deed, which transfers ownership, is prepared, recorded and the highest bidder becomes the owner of the property.

The requirements for non-judicial foreclosure are established by state statute; there is no court intervention. When the default occurs, the homeowner is mailed a default letter and in many states a Notice of Default is recorded, at or about the same time. The homeowner may cure the debt during a prescribed period; if not, a Notice of Sale is mailed to the homeowner, posted in public places, recorded at the county's recorder's office, and published in area newspapers / legal publications. When the legally required notice period (determined by each state) has expired, a public auction is held and the highest bidder becomes the owner of the property, subject to recordation of the deed. Prior to the sale, if the borrower disagrees with the facts of the case, he or she can try to file a lawsuit to enjoin the trustee's sale.

FIG. 3 Summary Statistics of Target States: AZ, FL, GA, and MD

	Arizona							Florida						
	2006	2007	2008	2009	2010	2011	2012	2006	2007	2008	2009	2010	2011	2012
HPA *% change of HPI*	0.58	-0.31	-1.24	-0.31	-0.33	0.16	1.44	0.47	-0.54	-0.88	-0.12	-0.19	0.29	0.97
REO Outflow *log of level*	-0.95	-0.77	-0.17	0.40	0.45	0.51	0.54	-0.79	-0.74	-0.11	0.36	0.44	0.48	0.36
REO Inflow *log of level*	-1.43	-0.86	0.06	0.53	0.67	0.65	0.37	-1.11	-0.45	0.08	0.34	0.55	0.19	0.40
Employment *% annual change*	3.67	2.40	1.52	-3.60	-1.44	-1.91	-0.63	3.56	1.75	-1.82	-4.34	-0.91	1.17	0.59
Sales *log of level*	0.50	0.40	0.09	-0.03	-0.24	-0.38	-0.34	0.32	0.06	-0.22	-0.19	-0.12	0.08	0.08
Cash Purchase *log of level*	0.43	0.23	-0.08	-0.03	-0.18	-0.18	-0.18	0.13	-0.20	-0.29	-0.11	0.04	0.21	0.23
Short Sales *log of level*	-0.87	-0.69	-0.30	0.22	0.40	0.56	0.68	-0.95	-0.77	-0.22	0.25	0.43	0.58	0.67

	Georgia							Maryland						
	2006	2007	2008	2009	2010	2011	2012	2006	2007	2008	2009	2010	2011	2012
HPA *% change of HPI*	0.52	0.08	-0.64	0.06	-0.17	-0.66	0.82	0.46	-0.05	-0.57	-0.15	-0.30	-0.22	0.84
REO Outflow *log of level*	-0.98	-0.56	-0.26	-0.18	0.11	0.77	1.10	-0.90	-0.60	-0.06	0.33	0.64	0.55	0.05
REO Inflow *log of level*	-0.77	-0.58	-0.38	-0.41	0.39	0.79	0.95	-0.99	-0.51	0.08	0.41	0.81	0.25	-0.05
Employment *% annual change*	2.53	1.70	-0.70	-5.75	-0.22	0.85	1.60	2.02	-0.27	0.05	-3.37	0.03	0.77	0.78
Sales *log of level*	-0.50	-0.27	-0.48	-0.60	0.33	0.53	0.99	0.55	0.29	-0.08	-0.22	-0.13	-0.17	-0.24
Cash Purchase *log of level*	-0.30	-0.22	-0.41	-0.57	0.30	0.45	0.75	0.34	0.04	-0.18	-0.17	0.04	0.07	-0.14
Short Sales *log of level*	-0.40	-0.22	-0.19	0.02	0.16	0.26	0.36	-0.84	-0.69	-0.31	0.14	0.38	0.60	0.72

markets in AZ, Fl, and GA are considered as volatile with big bubbles; while the housing market in MD is relatively steady during the peak and downturn.

The summary statistics of variables used for our model fitting are presented for each of our study markets in the Figure (**??**). All the values of the variables are in logarithmic or percentage.[15] For instance, REO outflow/inflow, sales, short sales, and cash purchases are the natural logarithm of the original numbers; HPA is the log difference of HPI; employment is the annual % change of payroll employment.

4. ECONOMIC MODELS

This section presents the economic models used for estimation.

4.1. OLS Regression

First, we estimate a simple linear regression. We model the interactions of three variables, Inflow into REO, Outflow from REO, and House prices. Our unit of observation is county by month. For each of our four states we track the inflow/outflow of properties into/out of REO and monthly changes in house prices. We also include as controls county level employment, as percentage changes to help control for local economic conditions. Also, we use the logarithm of total sales, short sales, and cash transactions (i.e., housing transactions paid in full by cash) as additional controls of other properties of the local housing market, such as liquidity, distressed

[15] Some variables receive a value of zero for certain time priods (e.g, no short sales recorded for a specific month). To prevent from taking logarithm of zeros, we replace the those zero values with a small value of 1.

status, and supply. Foremployment, sales, short sales, and cash sales, we take a one month lag to help control for potential endogeneity of these variables, but otherwise we treat them as exogenous for our purposes. For inflow and outflow of REO we take the logarithm of the actual amount to be consistent with other measures (e.g., sales, short sales, and cash sales).

Because REO outflow represents additional sales on the market, we expect that REO outflow will negatively impact local house prices. Insofar as increased REO inflow measures weakening economic conditions REO inflow might be negatively correlated with house prices, but the direct connection should be small. REO outflow out to be positively correlated with REO inflow –greater increases in the REO stock should presage greater outflow in the future. House prices is expected to negatively impact REO inflow, but positively impact outflow.

4.2. Spatial Simultaneous Equations System

Previous OLS regression assumes away potential spatial spillover effects of house price, REO outflow, and REO inflow from neighborhood counties. The feedback effects between house price and REO inflow/outflow is restricted in a time lag format. The spatial simultaneous equation system adopted in this section allows for spatial spillovers (e.g., HPA from nearby location affecting neighbors) and the cross spillover (e.g., interactions between HPA and REO outflow/inflow) taking place simultaneously. Essentially, our simultaneous equations system accomodate an instantaneous response of house price to REO outflow, REO inflow to house price, REO outflow to REO inflow, and etc.; as well as the simultaneous spatial comovement of house price/REO outflow/REO inflow with that of the neighbors.

We first write the spatial autoregressive functions for house price, REO outflow, and REO inflow respectively, and then combine these three equations and write them in a linear form. Now we have a spatial simultaneous equations system with observations available for spatial unit $i = 1, \cdots, N$ over time period $t = 1, \cdots, T$. In each time t, we have an $y_{1,N}(t)$ equation representing the house price function; an $y_{2,N}(t)$ equation representing the REO outflow function; and an $y_{3,N}(t)$ equation representing the REO inflow function as

$$\begin{aligned} y_{1,N}(t) &= \lambda_1 W_N y_{1,N}(t) + (y_{2,N}(t), y_{3,N}(t))\beta_1 + X_{1,N}(t)\gamma_1 + u_{1,N}(t), & (1)\\ y_{2,N}(t) &= \lambda_2 W_N y_{2,N}(t) + (y_{1,N}(t), y_{3,N}(t))\beta_2 + X_{2,N}(t)\gamma_2 + u_{2,N}(t), & (2)\\ y_{3,N}(t) &= \lambda_3 W_N y_{3,N}(t) + (y_{2,N}(t), y_{3,N}(t))\beta_3 + X_{3,N}(t)\gamma_2 + u_{3,N}(t), & (3)\end{aligned}$$

where for each equation $j = 1, 2, 3$, $y_{j,N}(t)$ is the $N \times 1$ vector of cross sectional observations on the dependent variable in the jth equation at time t, $X_{j,N}(t) = (x_{j1,N}(t), x_{j2,N}(t), \cdots, x_{jk_j,N}(t))$ is an $N \times k_j$ matrix of cross sectional observations on k_j exogenous variables at time t with associated parameters in the $k_j \times 1$ vector of $\gamma_j = (\gamma_{j1}, \gamma_{j2}, \cdots, \gamma_{jk_j})'$of the jth equation, $u_{j,N}(t)$ is an $N \times 1$ disturbance vector in the jth equation at time t, W_N is an $N \times N$ weights matrix of known constants,[16] whose ijth entry is w_{ij}, represents the proximity between the observation in location i and that of location j, βs are 2×1 parameter vectors representing the feedback effect across equations, and λs are scalar coefficients reflecting the spatial impact. Equation (1) corresponds to the house price equation, equation (2) corresponds to the REO outflow equation, and Equation (3) corresponds to the RO inflow equation. The simultaneity arises from two sources: the spatial simultaneity is from the inclusion of the spatial autoregressive lags, represented by $W_N y_{1,N}(t)$ in the price equation, $W_N y_{2,N}(t)$ in the REO outflow equation, and $W_N y_{3,N}(t)$ in the REO inflow equation; and the feedback simultaneity is from the specification of multi-equation system, represented by $y_{2,N}(t)$ and $y_{3,N}(t)$ in the price equation, $y_{1,N}(t)$ and $y_{3,N}(t)$ in the REO outflow equation, and $y_{1,N}(t)$ and $y_{2,N}(t)$ in the REO inflow equation.

[16] We are assuming that the system only involves one weights matrix. This assumption is made for ease of presentation, but also seems to be the typical specification in applied work. Our results can be generalized in a straight forward way to the case in which each spatially lagged variable depends upon a weights matrix which is unique to that variable.

To define the elements in the weight matrix, we take a more economically meaningful approach than the conventionally adopted contiguity weight matrix based on physical adjacency criteria. Ideally, w_{ij}, the ijth element of W_N represents the significance of the impact from j relative to i's other neighbors on i. In the context of the interaction pattern across various local housing markets, the individual migration data over geographic borders can potentially serves as a strong indicator of the relative importance of one market to another. For instance, both Navajo County and Maricopa County(Phoenix Metro) border Gila County in AZ. However, after studying the migration data between 2009 and 2010, we find only 84 individuals migrated to Gila from Navajo, while 892 migrated from Maricopa to Gila. The fact that Maricopa is the most populous county in the state (e.g., Maricopa and Pinal comprise the Phoenix Metropolitan Area.) and Navajo is primarily a desert suggests the impact from Maricopa to Gila should be larger than that from Navajo even though they both border Gila (The contiguity matrix tends to assign a same weight from Maricopa and Navajo to Gila.). We therefore employ the migration inflow statistics (i.e., approximated by the nmber of personal exemptions) between 2009 and 2010 from IRS for the specification of our weight matrix W_N. More specifically, if there are c_{ij} individuals migrating from location j to location i, the ijth element of W_N, w_{ij}, is then defined as c_{ij} (before row normalization). The diagonal elements of W_N, w_{ii}s, always receive a value of zero. The reference time of the migration inflow data is chosen as between 2009 and 2010 for the reason that this time period is in the middle of our observation window.

The equation system allows 2 types of simultaneity: feedback simultaneity, and spatial auto-regressive simultaneity, which are represented by $y_{j,N}(t)s$, and $W_N y_{i,N}(t)$ (for $i \neq j$) respectively in each equation. Despite the directly specified spatial autocorrelation terms of $W_N y_{1,N}(t)$, $W_N y_{2,N}(t)$,and $W_N y_{3,N}(t)$ in the system, it is possible that the disturbance terms present some degree of spatial interdependence due to omission of spatially autocorrelated regressors. We will conduct Moran I test based on the asymptotic distribution of the statistic derived by Anselin and Kelejian (2003) before proceed to include spatially correlated disturbances in our specification. If the null hypothesis of no spatial autocorrelation in the error terms is rejected, we assume the disturbances follow a first-order spatial autoregressive process as,

$$\begin{aligned} u_{1,N}(t) &= \rho_1 W_N u_{1,N}(t) + \varepsilon_{1,N}(t), & (4)\\ u_{2,N}(t) &= \rho_2 W_N u_{2,N}(t) + \varepsilon_{2,N}(t), & (5)\\ u_{3,N}(t) &= \rho_3 W_N u_{3,N}(t) + \varepsilon_{3,N}(t), & (6) \end{aligned}$$

where for each equation $j = 1, 2, 3$, $\varepsilon_{j,N}(t)$ denotes an $N \times 1$ vector of innovations and ρ_j is the scalar spatial autoregressive coefficient in the corresponding equation.

Collecting both equations, the system is written as a matrix form,

$$\begin{aligned} Y_N(t) &= Y_N(t)B + X_N(t)C + \overline{Y}_N(t)\Lambda + U_N(t),\\ U_N(t) &= \overline{U}_N(t)R + E_N(t), \end{aligned}$$

with

$$\begin{aligned} \overline{Y}_N(t) &= (\overline{y}_{1,N}(t), \overline{y}_{2,N}(t), \overline{y}_{3,N}(t)), \text{ and } \overline{y}_{j,N}(t) = W_N y_{j,N}(t),\\ \overline{U}_N(t) &= (\overline{U}_{1,N}(t), \overline{U}_{2,N}(t), \overline{U}_{3,N}(t)), \text{ and } \overline{U}_{j,N}(t) = W_N U_{j,N}(t),\\ \forall\, j &= 1, 2, 3, \end{aligned}$$

where $Y_N(t) = (y_{1,N}(t), y_{2,N}(t), y_{3,N}(t))$ with a dimension of $N \times 3$, $X_N(t) = (X_{1,N}(t), X_{2,N}(t), X_{3,N}(t))$ with a dimension of $N \times (k_1 + k_2 + k_3)$, $U_N(t) = (u_{1,N}(t), u_{2,N}(t), u_{3,N}(t))$ with a dimension of $N \times 3$, $E_N(t) = (\varepsilon_{1,N}(t), \varepsilon_{2,N}(t), \varepsilon_{3,N}(t))$ with a dimension of $N \times 3$, and

$$B_{3\times 3} = \begin{bmatrix} 0 & \beta_{21} & \beta_{31}\\ \beta_{12} & 0 & \beta_{32}\\ \beta_{13} & \beta_{23} & 0 \end{bmatrix}, C_{(k_1+k_2+k_3)\times 3} = \begin{bmatrix} \gamma_1 & 0 & 0\\ 0 & \gamma_2 & 0\\ 0 & 0 & \gamma_3 \end{bmatrix},$$

$$\Lambda_{3\times 3}=\begin{bmatrix}\lambda_1 & 0 & 0\\ 0 & \lambda_2 & 0\\ 0 & 0 & \lambda_3\end{bmatrix}, R_{3\times 3}=\begin{bmatrix}\rho_1 & 0 & 0\\ 0 & \rho_2 & 0\\ 0 & 0 & \rho_3\end{bmatrix}.$$

Noting that Λ and R are both diagonal, which implies that the spatial correlations in the dependent variables and disturbances are assumed to be within each equation, but not cross equations. The spatial simultaneity represented by the spatial auto-regressive lag is modeled via $\overline{y}_{j,N}(t)$, $j=1,2,3$.The ith element of $\overline{y}_{j,N}(t)$ is given by

$$\overline{y}_{ij,N}(t)=\sum_{r=1}^{N} w_{ir,N}y_{rj,N}(t).$$

If the weights $w_{ir,N}$ is nonzero, units i and r are said to be neighbors, otherwise, observation i and observation r are independent.

The model is now expressed in a vector form that will more clearly reveal its solution for the endogenous variables. Let

$$\begin{aligned}
y_N(t) &= vec(Y_N(t))=(y_{1,N}(t)', y_{2,N}(t)', y_{3,N}(t)')',\\
x_N(t) &= vec(X_N(t))=(x_{11,N}(t)',\cdots, x_{1k_1,N}(t)', x_{21,N}(t)',\\
&\quad \cdots, x_{2k_2,N}(t)', x_{31,N}(t)',\cdots, x_{3k_3,N}(t)')',\\
u_N(t) &= vec(U_N(t))=(u_{1,N}(t)', u_{2,N}(t)', u_{3,N}(t)')',\\
\overline{y}_N(t) &= vec(\overline{Y}_N(t))=(\overline{y}_{1,N}(t)', \overline{y}_{2,N}(t)', \overline{y}_{3,N}(t)')',\\
\overline{u}_N(t) &= vec(\overline{U}_N(t))=(\overline{u}_{1,N}(t)', \overline{u}_{2,N}(t)', \overline{u}_{3,N}(t)')'\\
\varepsilon_N(t) &= vec(E_N(t))=(\varepsilon_{1,N}(t)', \varepsilon_{2,N}(t)', \varepsilon_{3,N}(t)')'.
\end{aligned}$$

Noting that $\overline{y}_N(t)=(I_3\otimes W_N)y_N(t)$ with $\otimes$ representing the Kronecker product operator. If A_1 and A_2 are conformable matrices, that $vec(A_1A_2)=(A_2'\otimes I)vec(A_1)$, it follows that

$$\begin{aligned}
y_N(t) &= B_N^* y_N(t)+C_N^* x_N(t)+u_N(t), \qquad (7)\\
u_N(t) &= R_N^* u_N(t)+\varepsilon_N(t),
\end{aligned}$$

where $B_N^*=[(B'\otimes I_N)+(\Lambda'\otimes W_N)]$, $C_N^*=(C'\otimes I_N)$, and $R_N^*=(R'\otimes W_N)$. Assuming the row vector of $E_N(t)$, $\{\varepsilon_{1,i}(t),\varepsilon_{2,i}(t),\varepsilon_{3,i}(t): 1\le i\le n\}$, is distributed identically and independently with zero mean and variance covariance of $\Sigma=\begin{bmatrix}\sigma_{11} & \sigma_{12} & \sigma_{13}\\ \sigma_{21} & \sigma_{22} & \sigma_{23}\\ \sigma_{31} & \sigma_{32} & \sigma_{33}\end{bmatrix}$, which has a dimension of 3×3, thus the innovations entering the disturbance process are spatially uncorrelated but correlated across equations, which is analogous to the specification of classical simultaneous equation model. Therefore,

$$\begin{aligned}
E\varepsilon_N(t) &= 0,\\
E\varepsilon_N(t)\varepsilon_N(t)' &= \Sigma\otimes I_N.
\end{aligned}$$

Stacking the observations over T time periods for each equation, it yields

$$\begin{aligned}
y_{1,NT} &= \beta_{12}y_{2,NT}+\beta_{13}y_{3,NT}+\lambda_1 W_{NT}y_{1,NT}+X_{1,NT}\gamma_1+u_{1,NT}, \qquad (8)\\
u_{1,NT} &= \rho_1 W_{NT}u_{1,NT}+\varepsilon_{1,NT},
\end{aligned}$$

$$
\begin{aligned}
y_{2,NT} &= \beta_{21} y_{1,NT} + \beta_{22} y_{3,NT} + \lambda_2 W_{NT} y_{2,NT} + X_{2,NT}\gamma_2 + u_{2,NT}, \\
u_{2,NT} &= \rho_2 W_{NT} u_{2,NT} + \varepsilon_{2,NT},
\end{aligned} \tag{9}
$$

and

$$
\begin{aligned}
y_{3,NT} &= \beta_{31} y_{1,NT} + \beta_{32} y_{2,NT} + \lambda_3 W_{NT} y_{3,NT} + X_{3,NT}\gamma_3 + u_{3,NT}, \\
u_{3,NT} &= \rho_3 W_{NT} u_{3,NT} + \varepsilon_{3,NT},
\end{aligned} \tag{10}
$$

with W_{NT} defined as

$$
W_{NT} = (I_T \otimes W_N),
$$

where for $j = 1, 2, 3$,

$y_{j,NT} = (y_{j,N}(1)', \cdots, y_{j,N}(t)', \cdots y_{j,N}(T)')'$ with a dimension of $NT \times 1$,
$u_{j,NT} = (u_{j,N}(1)', \cdots, u_{1,N}(t)', \cdots u_{j,N}(T)')'$ with a dimension of $NT \times 1$,
$\varepsilon_{j,NT} = (\varepsilon_{j,N}(1)', \cdots, \varepsilon_{1,N}(t)', \cdots \varepsilon_{j,N}(T)')'$ with a dimension of $NT \times 1$,
$X_{j,NT} = (X_{j,N}(1)', \cdots, X_{j,N}(t)', \cdots X_{j,N}(T)')'$ with a dimension of $NT \times k_j$. The observations are ordered with t (i.e., time period index) being the slow running index, and i (i.e., spatial unit index) being the fast running index. For simplicity, equations (8), (9), and (10) are expressed as ($j = 1, 2, 3$)

$$
\begin{aligned}
y^*_{NT} &= \widetilde{B}^*_{NT} y^*_{NT} + \widetilde{C}^*_{NT} x^*_{NT} + u^*_{NT}, \\
u^*_{NT} &= \widetilde{R}^*_{NT} u_{NT} + \varepsilon^*_{NT},
\end{aligned} \tag{11}
$$

in which $y^*_{NT} = (y'_{1,NT}, y'_{2,NT}, y'_{3,NT})'$, $x^*_{NT} = (X'_{1,NT}, X'_{2,NT}, X'_{3,NT})'$,[17] $u^*_{NT} = (u'_{1,NT}, u'_{2,NT}, u'_{3,NT})'$,
$\varepsilon^*_{NT} = (\varepsilon'_{1,NT}, \varepsilon'_{2,NT}, \varepsilon'_{3,NT})'$, $\widetilde{B}^*_{NT} = [(B' \otimes I_{NT}) + (\Lambda' \otimes W_{NT})]$, $\widetilde{C}^*_{NT} = (C' \otimes I_{NT})$, and $\widetilde{R}^*_{NT} = (R' \otimes W_{NT})$.

$$
\begin{aligned}
y_{j,NT} &= Z_{j,NT}\delta_j + u_{j,NT}, \\
u_{j,NT} &= \rho_j W_{NT} u_{j,NT} + \varepsilon_{j,NT},
\end{aligned} \tag{12}
$$

where $Z_{j,NT} = (y_{-j,NT}, \overline{y}_{j,NT}, X_{j,NT})$ with $y_{-j,NT}$, $\overline{y}_{j,NT}$, and $X_{j,NT}$ representing the corresponding matrices of observations on the system endogenous variables, spatially lagged variables, and exogenous variables, for instance, $Z_{1,NT} = (y_{2,NT}, y_{3,NT}, (I_T \otimes W_N)y_{1,NT}, X_{1,NT})$, and $\delta_j = (\beta'_j, \lambda_j, \gamma'_j)$ with λ_j denoting the jjth element of Λ.The innovation in equation (12) follows that

$$
\begin{aligned}
E\varepsilon_{j,NT} &= 0, \\
E\varepsilon_{j,NT}\varepsilon'_{l,NT} &= \sigma_{jl} I_{NT}, \text{ for } j, l = 1, 2, 3,
\end{aligned} \tag{13}
$$

where σ_{jl} is the element of matrix Σ.

To derive the responses of the dependent variable, $y_{1,NT}$ and $y_{2,NT}$, to a one unit of structural shock, we can take the partial derivative of the structural equation system with respect to $\begin{bmatrix} \varepsilon_{1,NT} \\ \varepsilon_{2,NT} \\ \varepsilon_{3,NT} \end{bmatrix}$ as,

$$
\partial \begin{bmatrix} y_{1,NT} \\ y_{2,NT} \\ y_{3,NT} \end{bmatrix} / \partial \begin{bmatrix} \varepsilon_{1,NT} \\ \varepsilon_{2,NT} \\ \varepsilon_{3,NT} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}
$$

[17] The repeated variales in $X_{1,NT}$, $X_{2,NT}$, and $X_{3,NT}$ are included in x^*_{NT} for once.

or

$$= \left(I_{3NT} - \begin{bmatrix} 0_{NT} & \beta_{12} \times I_{NT} & \beta_{13} \times I_{NT} \\ \beta_{21} \times I_{NT} & 0_{NT} & \beta_{23} \times I_{NT} \\ \beta_{31} \times I_{NT} & \beta_{32} \times I_{NT} & 0_{NT} \end{bmatrix} - \begin{bmatrix} \lambda_1 W_{NT} & 0_{NT} & 0_{NT} \\ 0_{NT} & \lambda_2 W_{NT} & 0_{NT} \\ 0_{NT} & 0_{NT} & \lambda_3 W_{NT} \end{bmatrix} \right)^{-1} \tag{14}$$

$$\cdot \left(I_{3NT} - \begin{bmatrix} \rho_1 W_{NT} & 0_{NT} & 0_{NT} \\ 0_{NT} & \rho_2 W_{NT} & 0_{NT} \\ 0_{NT} & 0_{NT} & \rho_3 W_{NT} \end{bmatrix} \right)^{-1}$$

in which 0_{NT} denotes an $NT \times NT$ zero matrix, I_{3NT} denotes a $3NT \times 3NT$ identity matrix. Because variables in $y_{1,NT}$, $y_{2,NT}$,and $y_{3,NT}$ equation system are all measured as natural log terms (e.g., house price appreciation is the log difference of house price indices; REO outflow and inflow are both logarithms), the derivative results are in fact estimated elasticities. Matrices of A_{11}, A_{12},and, A_{13} reflect elasticities of HPI: A_{11} represents the percentage change of HPI in response to a 1 unit of structural shock in the HPA equation, $\frac{\partial y_{1,NT}}{\partial \varepsilon_{1,NT}}$, A_{12} represents the percentage change of HPI in response to a 1 unit of structural shock in the REO outflow equation (i.e., because REO outflow is defined as the logarithm of REO exit number, 1 unit of change in REO outflow means 1% change of REO exit number), $\frac{\partial y_{1,NT}}{\partial \varepsilon_{2,NT}}$, A_{13} represents the percentage change of HPI in response to a 1 unit of structural shock in the REO inflow equation (i.e., because REO inflow is defined as the logarithm of REO entry number, 1 unit of change in REO inflow means 1% change of REO entry number), $\frac{\partial y_{1,NT}}{\partial \varepsilon_{3,NT}}$; matrices of A_{21}, A_{23}, and A_{23} reflect marginal impact of REO outflow: A_{21} represents the percentage change HPI in response to a structural shock in the REO outflow equation (i.e., 1% change of REO outflow number), $\frac{\partial y_{2,NT}}{\partial \varepsilon_{1,NT}}$, A_{22} represents the percentage change of REO outflow number in response to a structural shock in the REO outflow equation, $\frac{\partial y_{2,NT}}{\partial \varepsilon_{2,NT}}$, and A_{23} represents the percentage change of REO outflow number in response to a structural shock in the REO inflow equation, $\frac{\partial y_{2,NT}}{\partial \varepsilon_{3,NT}}$, and etc. For example, the ijth element of $A_{12}(t)$ represents county i's HPI change rate in response to a one unit of structural shock of REO outflow in county j at time t. The $NT \times NT$ elements in the matrices of $A_{kl}(\forall k, l = 1, 2, 3)$, can be computed based on the estimated parameter values of SESSAR model in Figure (**??**). Because serial correlation is excluded in our model specification, $A_{kl}s$ are all block diagonal matrices with T blocks of $N \times N$ square matrices. Moreover, the tth block matrix of A_{kl}, $A_{kl}(t)$, is exactly the same as the sth block matrix of $A_{kl}(S)(\forall s = 1, ..., T)$.

5. SPATIAL J TEST FOR AN ALTERNATIVE SPECIFICATION OF WEIGHT MATRIX

For the purpose of identifying whether alternative specifications of weight matrix add signicanly to the explanatory power of the null model, we adopt the spatial J test introduced by Kelejian (2008) and Kelejian and Piras (2011) with a slight modification to accommodate the simultaneity across equations. The well established J test is designed to test a null model against non-nested alternatives. Kelejian (2008) and Kelejian and Piras (2011) extend the application of J test into a spatial framework. Though their spaital J test procedure allows for testing the null model against multiple alternative models (e.g., alternative specifications of regressor matrix, alternative specifications of weight matrix, alternative specifications of disturbance term, or some combination of these three) at the same time, it is set within a single equation model context. We modify their spatial J test procedure by incorporating the GS3SLS technique for a better suit for the simultaneous equation system in this paper.

The intuition of the spatial J test is to evaluate whether the alternative models help to explain the dependent variable in addition to the null model. So, the test is based on an augmented model with elements derived from both the null and alternative specifications. If the parameters in front of the elements representing the alternative models are identified as jointly significant, the null model should be rejected.

5.1. An Alternative Model H_1 with a Weight Matrix of M_N

Recall that the weight matrix W_N in our null model is based on the migration inflow data from IRS: if county j sends c_{ij} individuals to county i, then the ijth element of W_N is set equal to c_{ij} before row normalization; and the diagonal elements of W_N are set to zeros. For testing purpose, we adopt the conventionally used distance based matrix as an alternative weight matrix specification: the ijth element of W_N, w_{ij}, is defined as the inverse of the distance between two centroids (before row normalization) of county i and county j. We use M_N to indicate the alternative distance based weight matrix. The alternative model with the new specification of the weight matrix, say H_1, is then written as

$$\begin{aligned}
y_{1,N}(t) &= \lambda_1^a M_N y_{1,N}(t) + \beta_{12}^a y_{2,N}(t) + \beta_{13}^a y_{3,N}(t) + X_{1,N}(t)\gamma_1^a + u_{1,N}^a(t),\\
y_{2,N}(t) &= \lambda_2^a M_N y_{2,N}(t) + \beta_{21}^a y_{1,N}(t) + \beta_{22}^a y_{3,N}(t) + X_{2,N}(t)\gamma_2^a + u_{2,N}^a(t),\\
y_{3,N}(t) &= \lambda_3^a M_N y_{3,N}(t) + \beta_{31}^a y_{1,N}(t) + \beta_{32}^a y_{2,N}(t) + X_{3,N}(t)\gamma_3^a + u_{3,N}^a(t),\\
\text{where } u_{1,N}^a(t) &= \rho_1^a M_N u_{1,N}^a(t) + \varepsilon_{1,N}^a(t),\\
u_{2,N}^a(t) &= \rho_2^a M_N u_{2,N}^a(t) + \varepsilon_{2,N}^a(t),\\
\text{and } u_{3,N}^a(t) &= \rho_3^a M_N u_{3,N}^a(t) + \varepsilon_{3,N}^a(t).
\end{aligned}$$

The superscript of a distinguishes the parameters and innovations in the alternative model from the null model. Besides the specification of the weight matrix, the other components of the alternative are the same as the null. After stacking the observations over T time periods, we have

$$\begin{aligned}
y_{1,NT} &= \beta_{12}^a y_{2,N}(t) + \beta_{13}^a y_{3,N}(t) + \lambda_1^a M_{NT} y_{1,NT} + X_{1,NT}\gamma_1^a + u_{1,NT}^a,\\
y_{2,NT} &= \beta_{21}^a y_{1,N}(t) + \beta_{22}^a y_{3,N}(t) + \lambda_2^a M_{NT} y_{2,NT} + X_{2,NT}\gamma_2^a + u_{2,NT}^a,\\
y_{3,NT} &= \beta_{31}^a y_{1,N}(t) + \beta_{32}^a y_{2,N}(t) + \lambda_3^a M_{NT} y_{3,NT} + X_{3,NT}\gamma_3^a + u_{3,NT}^a,\\
\text{where } u_{1,NT}^a &= \rho_1^a M_{NT} u_{1,NT}^a + \varepsilon_{1,NT}^a,\\
u_{2,NT}^a &= \rho_2^a M_{NT} u_{2,NT}^a + \varepsilon_{2,NT}^a,\\
\text{and } u_{3,NT}^a &= \rho_3^a M_{NT} u_{3,NT}^a + \varepsilon_{3,NT}^a,
\end{aligned} \tag{15}$$

where $M_{NT} = (I_T \otimes M_N)$ or similarly

$$\begin{aligned}
y_{1,NT} &= Z_{1,NT}^a \delta_1^a + u_{1,NT}^a, \text{where } u_{1,NT}^a = \rho_1^a M_{NT} u_{1,NT}^a + \varepsilon_{1,NT}^a,\\
y_{2,NT} &= Z_{2,NT}^a \delta_2^a + u_{2,NT}^a, \text{where } u_{2,NT}^a = \rho_2^a M_{NT} u_{2,NT}^a + \varepsilon_{2,NT}^a,\\
y_{3,NT} &= Z_{3,NT}^a \delta_3^a + u_{3,NT}^a, \text{where } u_{3,NT}^a = \rho_3^a M_{NT} u_{3,NT}^a + \varepsilon_{3,NT}^a,
\end{aligned} \tag{16}$$

in which $Z_{1,NT}^a = [y_{2,NT}, y_{3,NT}, M_{NT}y_{1,NT}, X_{1,NT}]$, $Z_{2,NT}^a = [y_{1,NT}, y_{3,NT}, M_{NT}y_{2,NT}, X_{2,NT}]$, $Z_{3,NT}^a = [y_{1,NT}, y_{2,NT}, M_{NT}y_{3,NT}, X_{3,NT}]$, and $\delta_j^a = (\beta_j^{a\prime}, \lambda_j^a, \gamma_j^{a\prime})'$ for $j = 1, 2, 3$. The innovations in equation (16) follows that

$$\begin{aligned}
E\varepsilon_{j,NT}^a &= 0,\\
E\varepsilon_{j,NT}^a \varepsilon_{l,NT}^{a\prime} &= \sigma_{jl}^a I_{NT}, \text{ for } j, l = 1, 2,
\end{aligned}$$

where σ_{jl}^a is the element of $\Sigma^a = \begin{bmatrix} \sigma_{11}^a & \sigma_{12}^a & \sigma_{13}^a \\ \sigma_{21}^a & \sigma_{22}^a & \sigma_{23}^a \\ \sigma_{31}^a & \sigma_{32}^a & \sigma_{33}^a \end{bmatrix}$, which is the variance covariance matrix of $\{\varepsilon_{1,i}^a(t), \varepsilon_{2,i}^a(t), \varepsilon_{3,i}^a(t) : 1 \leq i \leq n\}$.

Before conducting the spatial J test, we take the spatial Cochrane-Orcutt transform of the null model (e.g., by premultiplying $(I_{NT} - \rho W_{NT})$ on both sides of the equation (12)) for an augmented model as,

$$y^*_{j,NT} = Z^*_{j,NT}\delta_j + \varepsilon_{j,NT}, \tag{17}$$

where $y^*_{j,NT} = (I_{NT} - \rho W_{NT})y_{j,NT}$, and $Z^*_{j,NT} = (I_{NT} - \rho W_{NT})Z_{j,NT}$. The predictors of $y^*_{j,NT}$ based on the null and the alternative models are then constructed repectively. The augmented dependent variable, $y^*_{j,NT}$,is then written as a function of both predictors, and the spatial J test is a test against the parameters of the alternative predictor of $y^*_{j,NT}$. Kelejian and Piras (2011) propose 2 predictors based on the alternative models, denoted H_1. We choose the second predictor because of the computational simplicity:[18]

$$\begin{aligned} E[y^*_{j,NT}|H_1] &= (I_{NT} - \rho W_{NT})E[y_{j,NT}|H_1] \\ &= (I_{NT} - \rho W_{NT})Z^a_{j,NT}\delta^a_j \\ &= Z^{*^a}_{j,NT}\delta_j, \end{aligned}$$

where $Z^a_{j,NT}$ and δ^a_j are defined the same as in equation (16), and $Z^{*^a}_{j,NT} = (I_{NT} - \rho W_{NT})Z^a_{j,NT}$. The augmented model can be rewritten as a function of the predictors derived from the null, H_0, and the alternative, H_1,

$$\begin{aligned} y^*_{j,NT} &= E[y^*_{j,NT}|H_0] + \alpha_j E[y^*_{j,NT}|H_1] + \widetilde{\varepsilon}_{j,NT} \\ &= Z^*_{j,NT}\delta_j + \alpha_j Z^{*^a}_{j,NT}\delta^a_j + \widetilde{\varepsilon}_{j,NT}. \end{aligned} \tag{18}$$

The unknown parameter vector δ_j has a dimension of $k_j + 3$, where k_j is the number of exogenous variables in $X_{j,NT}$, and α_j is a scalar parameter, which should be zero if the null model is the true model.

To determine the distribution of the estimated α_j, we need to take into account of the correlation in the disturbance term of the above equation. To be analogous to the specification of classical simultaneous equation model and consistent with the null and alternative models, the innovations entering the disturbance term in equation (18) are specified as spatially uncorrelated but correlated across equations. Thus,

$$\begin{aligned} E\widetilde{\varepsilon}_{j,NT} &= 0, \\ E\widetilde{\varepsilon}_{j,NT}\widetilde{\varepsilon}'_{l,NT} &= \widetilde{\sigma}_{jl}I_{NT}, \text{ for } j,l = 1,2,3, \end{aligned} \tag{19}$$

where $\widetilde{\sigma}_{jl}$ is the element of $\widetilde{\Sigma} = \begin{bmatrix} \widetilde{\sigma}_{11} & \widetilde{\sigma}_{12} & \widetilde{\sigma}_{13} \\ \widetilde{\sigma}_{21} & \widetilde{\sigma}_{22} & \widetilde{\sigma}_{23} \\ \widetilde{\sigma}_{31} & \widetilde{\sigma}_{32} & \widetilde{\sigma}_{33} \end{bmatrix}$, which is the variance covariance matrix of $\{\widetilde{\varepsilon}_{1,i}(t), \widetilde{\varepsilon}_{2,i}(t), \widetilde{\varepsilon}_{3,i}(t) : 1 \le i \le n\}$. Combining equations (18) and (19), we have the augmented equation system ready for the spatial J test,

$$\begin{aligned} y^*_{j,NT} &= Z^*_{j,NT}\delta_j + \alpha_j Z^{*^a}_{j,NT}\delta^a_j + \widetilde{\varepsilon}_{j,NT}, \text{for } j = 1,2,3, \\ \text{where } E\widetilde{\varepsilon}_{j,NT} &= 0, \text{and } E\widetilde{\varepsilon}_{j,NT}\widetilde{\varepsilon}'_{l,NT} = \widetilde{\sigma}_{jl}I_{NT}, \text{ for } j,l = 1,2,3. \end{aligned} \tag{20}$$

5.2. Modified Spatial J Test

The spatial J test focuses on the estimated value of α_j. If the null model is correctly specified and the alternative model does not have explanatory power, the estimated value of α_j should not be significantly different from zero. To derive a consistent and efficient estimate of α_j and correctly specify its asymptotical distribution for our spatial simultaneous equation

[18] We have only one alternative specification to test against. According to Kelejian and Piras (2011), when the number of alternative specifications, G, is equal to one, the J test has the same power under H_1 with either of the predictors.

system, we modify the spatial J test procedure described in Kelejian and Piras (2011) by incorporating the GS3SLS technique. Specifically, we adopt the GS3SLS method 3 times: one for the null model of equation (12), one for the alternative model of equation (15), and one for the augmented model of equation (20). The detailed test procedure is provided as follows.

Step 1: Estimate the parameters, δ_j and ρ_j of the null model, and δ_j^a of the alternative model using the GS3SLS method. Denote the estimated coefficients as $\widehat{\delta} = (\widehat{\delta}_1', \widehat{\delta}_2', \widehat{\delta}_3')'$, $\widehat{\rho} = (\widehat{\rho}_1, \widehat{\rho}_2, \widehat{\rho}_3)$, and $\widehat{\delta}^a = (\widehat{\delta}_1^{a\prime}, \widehat{\delta}_2^{a\prime}, \widehat{\delta}_3^{a\prime})'$.

Step 2: Use the results in Step 1 to compute the estimates of $y_{j,NT}^*$, $Z_{j,NT}^*$, and $Z_{j,NT}^{*^a}\delta_j^a$, which are written as

$$\begin{aligned} y_{j,NT}^* &= (I_{NT} - \widehat{\rho}_j W_{NT}) y_{j,NT}, \\ Z_{j,NT}^* &= (I_{NT} - \widehat{\rho}_j W_{NT}) Z_{j,NT}, \\ Z_{j,NT}^{*^a} \widehat{\delta}_j^a &= (I_{NT} - \widehat{\rho}_j W_{NT}) Z_{j,NT}^a \widehat{\delta}_j^a. \end{aligned}$$

Step 3: Let $F_{j,NT}^* = [Z_{j,NT}^*, Z_{j,NT}^{*^a}\widehat{\delta}_j^a]$, and $\phi_j' = [\delta_j', \alpha_j]$, equation (18) is simplified as

$$y_{j,NT}^* = F_{j,NT}^* \phi_j + \widetilde{\varepsilon}_{j,NT}, \tag{21}$$

whose parameter vector ϕ_j can be estimated using a 2SLS estimator. Define

$$\widetilde{H} = (X_{NT}, W_{NT}X_{NT}, W_{NT}^2 X_{NT}, M_{NT}X_{NT}, M_{NT}^2 X_{NT}),$$

and let $\widetilde{P}_H = \widetilde{H}(\widetilde{H}'\widetilde{H})^{-1}\widetilde{H}'$, and $\widehat{F}_{j,NT}^* = \widetilde{P}_H F_{j,NT}^*$, then the 2SLS estimator has the following form

$$\widehat{\phi}_{j,2sls} = (\widehat{F}_{j,NT}^{*\prime} F_{j,NT}^*)^{-1} \widehat{F}_{j,NT}^{*\prime} y_{j,NT}^*.$$

To achieve a consistent estimator of $\widetilde{\Sigma}$, let

$$\widehat{\widetilde{\varepsilon}}_{j,NT} = y_{j,NT}^* - F_{j,NT}^* \widehat{\phi}_{j,2sls},$$

then compute

$$\begin{aligned} \widehat{\widetilde{\sigma}}_{jl} &= \frac{1}{NT} \widehat{\widetilde{\varepsilon}}_{j,NT}' \widehat{\widetilde{\varepsilon}}_{l,NT}, \\ j, l &= 1, 2, 3. \end{aligned}$$

which composes the elements of the estimated $\widetilde{\Sigma}$.

Step 4: Stack the equations in (21) over $j = 1, 2, 3$ as

$$y_{NT}^* = F_{NT}^* \phi + \widetilde{\varepsilon}_{NT},$$

where $y_{NT}^* = (y_{1,NT}^{*\prime}, y_{2,NT}^{*\prime}, y_{3,NT}^{*\prime})'$, $F_{NT}^* = \begin{bmatrix} F_{1,NT}^* & 0 & 0 \\ 0 & F_{2,NT}^* & 0 \\ 0 & 0 & F_{3,NT}^* \end{bmatrix}$, and $\phi = (\phi_1', \phi_2', \phi_3')'$.

The variance covariance matrix of $\widetilde{\varepsilon}_{NT}$ is then

$$E\widetilde{\varepsilon}_{NT}\widetilde{\varepsilon}_{NT}' = \widetilde{\Sigma} \otimes I_{NT}.$$

The GS3SLS estimator of ϕ is then obtained as

$$\widehat{\phi}_{3sls} = [\widehat{F}_{NT}^{*\prime} (\widehat{\widetilde{\Sigma}}^{-1} \otimes I_{NT}) F_{NT}^*]^{-1} \widehat{F}_{NT}^{*\prime} (\widehat{\widetilde{\Sigma}}^{-1} \otimes I_{NT}) y_{NT}^*,$$

where $\widehat{F}^*_{NT} = \widetilde{P}_H F^*_{NT} = \begin{bmatrix} \widetilde{P}_H F^*_{1,NT} & 0 & 0 \\ 0 & \widetilde{P}_H F^*_{2,NT} & 0 \\ 0 & 0 & \widetilde{P}_H F^*_{3,NT} \end{bmatrix}$, and $\widehat{\widehat{\Sigma}} = \begin{bmatrix} \widehat{\widehat{\sigma}}_{11} & \widehat{\widehat{\sigma}}_{12} & \widehat{\widehat{\sigma}}_{13} \\ \widehat{\widehat{\sigma}}_{21} & \widehat{\widehat{\sigma}}_{22} & \widehat{\widehat{\sigma}}_{23} \\ \widehat{\widehat{\sigma}}_{31} & \widehat{\widehat{\sigma}}_{32} & \widehat{\widehat{\sigma}}_{33} \end{bmatrix}$.

The small sample distribution of $\widehat{\phi}_{3sls}$ is approximated as follows

$$\widehat{\phi}_{3sls} \tilde{} N(\gamma, [\widehat{F}^{*\prime}_{NT}(\widehat{\widehat{\Sigma}}^{-1} \otimes I_{NT})F^*_{NT}]^{-1}).$$

Step 5: Let $R = \begin{bmatrix} 0_{k_1+3} & 1 & 0_{k_2+3} & 1 & 0_{k_3+3} & 1 \end{bmatrix}_{1\times(k_1+k_2+k_3+12)}$, in which k_1 is the number of exogeneous variables in $X_{1,NT}$, k_2 is the number of exogeneous variables in $X_{2,NT}$, and k_3 is the number of exogeneous variables in $X_{3,NT}$, then the null model H_0 suggests $R\phi = 0$. Thus, the Wald test statistic is written as

$$(R\widehat{\phi})'\{R[\widehat{F}^{*\prime}_{NT}(\widehat{\widehat{\Sigma}}^{-1} \otimes I_{NT})F^*_{NT}]^{-1}R'\}^{-1}(R\widehat{\phi}), \tag{22}$$

which is asymptotically distributed as χ^2 with three degrees of freedom. The null model H_0 would be rejected if the test statistic is greater than the critical value of $\chi^2(3)$ at a given significance level.

6. ESTIMATION RESULTS

6.1. Ordinary Least Squares

We first estimate a simple OLS regression allowing impact of HPA, REO outflow and REO inflow to enter the equations with a time lag. The results of the panel regression for each of our target states can be seen in Figure (4) below. Because all the explanatory variables are demeaned, so the three intercepts are insignificant across. The regression results show that house price, REO outflow and inflow are sensitive to all the lagged HPA, but the signs of the parameters are not consistent over states. The lagged REO outflow and inflow positively influences current REO outflow and inflow in all 4 states, but their behavior in the HPA equation varies across states: lagged REO outflow drives up HPA in AZ and FL, but drives down HPA in MD; lagged REO inflow reduces HPA in AZ and FL, but increases HPA in GA. The variable of sales doesn't behave as expected in the HPA equations for all 4 states: a higher sales tends to lower HPA for AZ, FL, and MD, but has no impact in GA. Cash purchases positively influence HPA in AZ, FL, and MD, but not significant in GA. Short sales negatively impact HPA in FL and MD, but don't have explanatory power in AZ and GA. The impact of short sales in the REO inflow equations is opposite to our expectation: instead of acting as a alternative solution to foreclosure (i.e., REO inflow), a higher value of short sales suggests a higher amount of REO inflow.

6.2. Spatial Model

We present our estimation results of the simultaneous equation system for 4 targeted states respectivly in Figure (5). The neighborhood spillover effect of house price is significant and positive in all 4 states with a spatial coefficient of 0.85 to 0.95 (close to the maximum of 1), which suggests a 1% increase of the weighted average of neighboring counties' house price index leads to almost a 1% increase in the given county's house price index. The REO outflow is positively correlated in space for Arizona, Georgia, and Maryland but not for Florida. Even though the magnitude of the spatial correlation in REO outflow is smaller than that of house prices, the existence of this positive correlation in REO outflow across neigboring counties suggests a possible "yardstick competition" in the decision making procedure of banks holding REOs. Therefore, the release of REO properties in one area might potentially cause a "run" in neighboring areas. If the fear of shadow inventory is real, the scale of the fear should take into account of the spatial spillovers of both HPA and REO outflow because both of them are spatially contagious. Unlike HPA and REO outflow, the spatial effect of REO inflow is not

FIG. 4 OLS Estimation Results of Target States: AZ, FL, GA, and MD

Variables	AZ			FL		
	HPA Equation #1	REO Outflow Equation #2	REO Inflow Equation #3	HPA Equation #1	REO Outflow Equation #2	REO Inflow Equation #3
Intercept	0.0000	0.00	0.0000	0.00	0.00	0.00
	0.00	0.0000	0.00	0.0000	0.0000	0.0000
HPA (t-1)	**0.8759**	**0.03**	**-0.0553**	**0.78**	**0.03**	**-0.04**
	64.73	2.0811	-4.09	91.1321	5.1941	-5.6815
REO Outflow (t-1)	**0.0609**	**0.63**	**0.2432**	**0.05**	**0.62**	**0.15**
	2.16	21.3421	6.85	2.6928	46.1321	8.3223
REO Inflow (t-1)	**-0.0448**	**0.22**	**0.5380**	**-0.05**	**0.22**	**0.47**
	-1.78	8.6242	14.97	-3.4259	17.9465	29.2064
Employment (t-1)	**-0.0130**	0.00	**-0.0146**	0.00	**-0.01**	**-0.02**
	-2.85	-1.0696	-3.89	1.2800	-6.0586	-10.5519
Log Sales (t-1)	**-0.2073**		**-0.2171**	**-0.22**		0.00
	-4.47		-6.54	-7.7962		-0.1417
Log Cash Purchases (t-1)	**0.1945**			**0.30**		
	4.13			9.4794		
Log Short Sales (t-1)	-0.0232		**0.06**	**-0.03**		**0.19**
	-0.91		2.3982	-1.7419		12.1215
R square / equation	0.81	0.69	0.74	0.65	0.69	0.60

Variables	GA			MD		
	HPA Equation #1	REO Outflow Equation #2	REO Inflow Equation #3	HPA Equation #1	REO Outflow Equation #2	REO Inflow Equation #3
Intercept	0.00	0.00	0.00	0.00	0.00	0.00
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
HPA (t-1)	**0.72**	**-0.02**	**-0.05**	**0.85**	**-0.04**	**-0.06**
	62.9236	-2.0782	-5.9259	64.3240	-2.3686	-3.3118
REO Outflow (t-1)	-0.02	**0.59**	**0.19**	**-0.05**	**0.42**	**0.16**
	-1.5963	38.1696	11.8493	-2.4612	16.2254	5.7100
REO Inflow (t-1)	**0.04**	**0.29**	**0.53**	0.00	**0.22**	**0.33**
	2.1093	20.8194	28.7252	0.1375	8.9397	13.2064
Employment (t-1)	0.00	0.00	0.00	**0.02**	**-0.02**	**-0.02**
	1.5462	-1.3994	0.9391	4.0557	-2.8243	-2.5950
Log Sales (t-1)	-0.02		**0.14**	**-0.54**		**-0.73**
	-0.7400		9.4790	-12.4825		-17.5591
Log Cash Purchases (t-1)	0.04			**0.14**		
	1.3010			4.1892		
Log Short Sales (t-1)	-0.02		-0.01	**-0.06**		0.04
	-1.3223		-0.48	-3.6494		1.6191
R square / equation	0.53	0.74	0.67	0.67	0.41	0.48

t-values are provided under the estimates; *p at 10%(t=1.645); **p at 5% (t=1.96); ***p at 1%(t=2.326) straight forward. No impact is found in Arizona and Florida. Though the spatial coefficient of REO inflow is significant in Georgia and Maryland, the directions are opposite. Maryland only allows for judicial foreclosures processed by county court. The positive pattern of REO inflow across counties in MD likely suggestsa competitive behavior of county sheriff. Georgia allows both judicial and non-judicial foreclosures. Therefore, the negative pattern of REO inflow across counties might suggest a substutional effect of short sales and regular sales: more REO inflow in one county means less housing supply, and therefore the housing demand arising from this area are fulfilled by neighboring markets, which in turn reduce the REO inflow there.

In terms of the feedback effect, we expect a higher HPA to increase REO outflow and decrease REO inflow. The estimated coefficients of HPA in the REO outflow equation are consistent with our expectation in Arizona, Florida, and Maryland. A lower HPA tends to discourage REO outflow in these 3 states. But a lower HPA tends to encourage more REO outflow in Georgia. We suspect that the unexpected sign in Georgia has something to do with the persistent rise of its REO outflow during our entire observation time (lack of variations). The feedback effect of HPA in the REO inflow equation has the expected sign in Arizona, Florida, and Georgia: the higher the HPA is, the less the REO inflow; though the impact in Maryland is still negative, but it is insignificant. The impact of REO outflow on REO inflow is almost 1-to-1 for Arizona, Florida, and Georgia; but plays no rolls in MD. The impact of REO inflow on REO outflow has the same direction as the other way around, but with a smaller scale in all 4 states.

Most of the other explanatory variables behave as expected. For instance, the more the cash transactions are, the hotter the market is, therefore, the higher the HPA is. The effect of short sales in the REO outflow equation is positive for some states, and negative for others. We originally treat short sales as a substitute of foreclosure, and thus expect it to have a negative impact on REO inflow. Alternatively, a big volume of short sales might indicate worsening economic conditions in the local market, and therefore more foreclosures happen. So the negative sign of short sales can be interpreted as the impact of macro economic conditions on the housing market (especially when the variable of employment has weak explanatory power in the equations). The sign of total sales is the most difficult to explain. It is used to proximate the liquidity of the market, so a high sale volume suggests a liquid market, and HPA should increase and REO inflow should decrease. It only has the expected sign in the REO inflow equation of MD and HPA equation in FL.

The estimated coefficients of those feedback variables only reveal a instantaneous response of the dependent variable to the feedback variable, but not the total impact of the given feedback variable. This is due to the existence of the simultaneous spatial correlation and the contemporaneous cross-equation interaction. To derive the actual marginal impact of each of the feedback variables, we need to employ the partial derivative of the structural equation system. Based on the reduced form of equation (14), we report the average marginal impact of a structural shock in HPA/REO outflow in Figure (6). A structural shock can happen in one neighboring county, and it can also happen in all neigboring counties. So we present our findings of one neighbor vs every neighbor respectively.

Recall the ijth element in $A_{12}(t)$ indicates the change of HPA in county i due to a 1 unit of structural shock of REO outflow in county j at time t. Because $A_{12}(t) = A_{12}(s)\forall(s = 1, ..., T)$, county i's reponse to county $j's$ shock does not vary across time. The first half of Figure (6) reports the average of the elements in A_{12}/A_{13} $/A_{21}/A_{23}$; the second half of Figure (6) reports the average of row sums of the elements in A_{12}/A_{13} $/A_{21}/A_{23}$. For example, the average change of house price due to a 1 unit increase of REO outflow (i.e., 1% of increase in REO ouflow) is -0.46% in AZ; if the 1% of increase of REO outflow is statewide (i.e., every county has a 1% increase in REO outflow), the average drop of house price is 6.78% in AZ. Among the 4 states, house price response to REO liquidation negatively in AZ and MD. But the increase of REO outflow will not drive down house price in FL and GA, instead, it helps to boost the market. This effect is particularly pronouced in FL: a 0.43% increase of HPI if one county increase the REO liquidation by 1%. The response of REO outflow to a rise in house price is positive in

FIG. 5 GS3SLS Estimation Results of AZ, FL, GA, and MD

Variables	AZ HPA Equation #1	AZ REO Outflow Equation #2	AZ REO Inflow Equation #3	FL HPA Equation #1	FL REO Outflow Equation #2	FL REO Inflow Equation #3
Spatial Lag: lamda	**0.85**	**0.36**	-0.02	**0.95**	0.03	0.04
	15.3530	4.6007	-0.2452	49.9625	1.1285	1.4248
Feedback: HPA		**0.17**	**-0.28**		**0.24**	**-0.26**
		6.5197	-6.6755		15.3927	-14.6962
Feedback: REO Outflow	0.01		**1.21**	-0.01		**1.00**
	0.0408		11.8431	-0.1768		27.0526
Feedback: REO Inflow	-0.12	**0.46**		0.06	**0.91**	
	-0.5623	5.8128		1.0454	33.3660	
Employment	0.00	**0.01**	**-0.01**	0.00	0.00	0.00
	-0.5699	1.7711	-3.0653	0.0472	0.8302	-0.8280
Log Sales	**-0.37**		**-0.20**	**0.10**		**0.02**
	-4.7848		-6.5173	5.3365		2.6410
Log Cash Purchases	**0.33**			0.03		
	6.1977			1.2623		
Log Short Sales	-0.02		**-0.06**	0.00		**0.02**
	-0.3460		-1.6701	0.1581		1.6880
Spatial Error: rho	-0.08	-0.10	-0.04	**-0.13**	**0.15**	**0.11**
	-0.5135	-1.0777	-0.2554	-2.8624	348.6380	53.6393
R square / equation	0.82	0.54	0.59	0.87	0.50	0.55
R square all		0.61			0.61	
J test (Chi Square)		0.0065			0.0037	

Variables	GA HPA Equation #1	GA REO Outflow Equation #2	GA REO Inflow Equation #3	MD HPA Equation #1	MD REO Outflow Equation #2	MD REO Inflow Equation #3
Spatial Lag: lamda	**0.94**	**0.29**	**-0.18**	**0.89**	**0.64**	**0.70**
	30.4926	16.0025	-9.4769	12.6052	7.7309	3.8328
Feedback: HPA		**-0.07**	**-0.09**		**0.26**	-0.13
		-2.3051	-1.9957		3.7740	-1.1176
Feedback: REO Outflow	-0.05		**1.03**	-0.14		0.04
	-0.8010		40.8544	-1.3444		0.2300
Feedback: REO Inflow	0.11	**0.68**		0.08	**0.36**	
	1.3857	38.5821		0.5889	3.7868	
Employment	**0.00**	0.00	0.00	0.00	0.00	0.00
	-2.1631	0.0412	1.3477	-0.0426	-0.5678	0.1903
Log Sales	-0.02		**0.15**	0.00		**-0.19**
	-0.6149		12.3940	0.0548		-3.6516
Log Cash Purchases	0.01			0.04		
	0.3678			1.6527		
Log Short Sales	**-0.06**		**-0.03**	**0.08**		**0.09**
	-4.4075		-1.9831	3.9295		2.4415
Spatial Error: rho	**-0.20**	**0.10**	**0.25**	**-0.22**	**0.02**	**-0.16**
	-4.7421	7.2603	7.4433	-2.9199	0.9727	-9.2183
R square / equation	0.85	0.63	0.62	0.85	0.54	0.61
R square all		0.66			0.67	
J test (Chi Square)		0.6157			0.0183	

t-values are provided under the estimates; *p at 10%(t=1.645); **p at 5% (t=1.96); ***p at 1%(t=2.326)
%critical values for chi-square (dof=3): 10%=6.251; 5%=7.815; 1%=11.345.

FIG. 6 Response to Structural Shock in One Neighbor vs Every Neighbor

Shock in One Neighbor					
	AZ	**FL**	**GA**	**MD**	**Average**
	REO Outflow Shock: 1% Increase in One Neighbor				
HPA Response in %	-0.46	0.43	0.07	-0.09	-0.01
REO Inflow Response in %	0.64	0.48	0.07	0.06	0.31
	HPA Shock: 1% Increase in One Neighbor				
REO Outflow Response in %	0.16	-0.04	-0.12	0.06	0.01
REO Inflow Response in %	0.09	-0.09	-0.11	-0.07	-0.05

Shock In Every Neighbor					
	AZ	**FL**	**GA**	**MD**	**Average**
	REO Outflow Shock: 1% Increase in Every Neighbor				
HPA Response in %	-6.89	28.43	4.13	-2.26	5.85
REO Inflow Response in %	9.56	31.45	4.35	1.36	11.68
	HPA Shock: 1% Increase in Every Neighbor				
REO Outflow Response in %	2.33	-2.68	-7.15	1.41	-1.52
REO Inflow Response in %	1.35	-5.84	-6.73	-1.70	-3.23

FIG. 7 Contemporaneous Response of House Prices to a 1% Increase of REO Liquidation in AZ Counties

	APACHE	COCHISE	COCONINO	GILA	GRAHAM	GREENLEE	LA PAZ	MARICOPA	MOHAVE	NAVAJO	PIMA	PINAL	SANTA CRUZ	YAVAPAI	YUMA
APACHE	-0.52	-0.20	-0.43	-0.14	-0.09	-0.04	-0.03	**-2.34**	-0.16	-0.48	-0.75	-1.06	-0.11	-0.38	-0.16
COCHISE	-0.16	-0.55	-0.34	-0.13	-0.13	-0.08	-0.03	**-2.30**	-0.16	-0.29	-0.96	-1.06	-0.14	-0.38	-0.17
COCONINO	-0.22	-0.22	-0.66	-0.14	-0.09	-0.04	-0.04	**-2.37**	-0.18	-0.37	-0.78	-1.07	-0.11	-0.43	-0.17
GILA	-0.18	-0.22	-0.36	-0.43	-0.11	-0.05	-0.03	**-2.43**	-0.17	-0.32	-0.79	-1.13	-0.11	-0.39	-0.16
GRAHAM	-0.18	-0.26	-0.33	-0.15	-0.45	-0.15	-0.03	**-2.29**	-0.15	-0.32	-0.86	-1.06	-0.12	-0.37	-0.17
GREENLEE	-0.17	-0.24	-0.32	-0.14	-0.34	-0.40	-0.03	**-2.25**	-0.15	-0.29	-0.86	-1.05	-0.12	-0.37	-0.16
LA PAZ	-0.17	-0.21	-0.35	-0.13	-0.08	-0.04	-0.35	**-2.39**	-0.30	-0.29	-0.74	-1.07	-0.10	-0.43	-0.22
MARICOPA	-0.18	-0.23	-0.37	-0.14	-0.09	-0.05	-0.04	**-2.58**	-0.17	-0.32	-0.82	-1.20	-0.12	-0.41	-0.18
MOHAVE	-0.18	-0.22	-0.40	-0.14	-0.09	-0.04	-0.08	**-2.36**	-0.48	-0.32	-0.77	-1.07	-0.11	-0.44	-0.19
NAVAJO	-0.27	-0.21	-0.42	-0.14	-0.09	-0.04	-0.03	**-2.36**	-0.17	-0.64	-0.77	-1.07	-0.11	-0.39	-0.16
PIMA	-0.17	-0.29	-0.35	-0.13	-0.11	-0.06	-0.03	**-2.35**	-0.16	-0.30	-1.12	-1.10	-0.17	-0.38	-0.18
PINAL	-0.17	-0.22	-0.34	-0.14	-0.09	-0.05	-0.03	**-2.50**	-0.16	-0.30	-0.80	-1.42	-0.11	-0.38	-0.17
SANTA CRUZ	-0.16	-0.28	-0.32	-0.12	-0.10	-0.05	-0.03	**-2.28**	-0.15	-0.28	-1.08	-1.05	-0.45	-0.36	-0.17
YAVAPAI	-0.18	-0.22	-0.41	-0.14	-0.09	-0.05	-0.04	**-2.42**	-0.19	-0.32	-0.78	-1.10	-0.11	-0.69	-0.17
YUMA	-0.17	-0.23	-0.35	-0.13	-0.09	-0.05	-0.05	**-2.42**	-0.20	-0.30	-0.82	-1.11	-0.12	-0.39	-0.47

AZ and MD, suggesting REO holders will respond to higher prices by realeasing more REOs, particularly in AZ.

We also present a scenario analysis: what will happen to the house prices in all the counties in Arizona if there is 1% positive REO outflow shock in a given county in AZ. See Figure (7) for the answers. Among all the counties, Maricopa and Pinal (Phoenix metropolitan area) are the most influential counties. A 1% increase of REO outflow shock in Maricopa reduces house price in the other local markets by 2.38% in average. The marginal impact of REO outflow in the other counties other than Maricopa and Pinal is only 0.26% in average.

7. CONCLUSION AND POLICY CONSIDERATIONS

In this paper we have analyzed the impact of the liquidation of REO properties on house prices. The results from our estimated simultaneous equation system imply that there is a substantial impact on house prices when REO properties are sold into the market. We used the estimated coefficients from our model to calculate how much lower prices would be across the state of Arizona if Maricopa County (Phoenix) were to increase REO liquidation by 1%. Our results indicate that house prices would fall by 2.58% in Maricopa county and by 2.25% to 2.43% in the rest of the state.

We found not only strong spillovers in house prices, but also substantial evidence of direct spillover effects of REO outflow across county lines. Because of the sensitivity of house prices to neighboring counties, factors that drive down house prices in one neighborhood are unlikely to be contained in just that neighborhood. The spatial correlation of house prices is well known in the housing literature. Our findings of the "yardstick competition" behavior of banks, who hold REOs in their portfolio, is particularly interesting. If the fear of shadow is real, we need to be better prepared because the shadow from neighbors will overcast as well.

We also found that the presence of investors and other cash buyers helps to mitigate some of the negative impact from REO liquidation on home prices. Increased shares of cash transactions help to support home prices. We also found that the short sales can impact house prices in both directions. As an alternative to foreclosure, short sale can potentially increase house price; while as an indicator of a stressed market, its effect on house price can also be negative.

Our results show that the eventual winding down of the substantial REO stock will have strong negative impact on house price growth for Arizona and Maryland. Florida and Georgia will see more house price appreciation even after an increased REO liquidation.

On the policy front, efforts to prevent foreclosure can help to reduce the pressure on house prices. As we would expect, REO inflow has a large positive impact on future REO outflow. Any policy that would help stem the tide of foreclosures would certainly help homeowners. We also saw that increased short sales can help to reduce the strain on local house prices for some states.

So should we fear the shadow? Though we have not found much evidence of the direct impact of REO outflow on HPA, the shock response table of Figure (6) indicates the weight down on house price is siginificant in AZ, modest in MD; but the opposite in GA and FL. In other words, it depends on the market. REO liquidation itself does not directly influence house price, but through other indirect channels, such as its impact on REO inflow, and the spatial interactions of HPA/REO inflow/REO outflow, the aggregated effect on house price is determined by the functions the other channels. Because of the localness of housing market, the shadow impact indeed varies across markets.

8. REFERENCES

REFERENCES

[1] Anselin L., (2008). "Spatial Hedonics," Mills TC, Patterson K (eds) Palgrave Handbook of Econometrics: Volumn 2, Applied Econometrics. Palgrave MacMillan, Basingstoke.

[2] Anselin L., and Kelejian, H. H., (2003). "Testing for Spatial Error Autocorrelation in the Presence of Endogenous Regressors," Regional Science and Urban Economics, Volume 33 (1): 27-58

[3] Bajari, Patrick, Chu, Sean, Park, Minjung, (2008). "An Empirical Model of Subprime Mortgage from 2000 to 2007," NBER Working Paper.

[4] Boarnet, M.G., (1994), "An Empirical Model of Intra-metropolitan Population and Employment Growth," Papers in Regional Science, 73 (2): 135-153.

[5] Campbell, J., Giglio, S., and Pathak, P., (2011). "Forced Sales and House Prices." American Economic Review. 101(5): 2108-2131.

[6] Carlino, O.G. and E.S. Mills, (1987), "The Determinants of County Growth," Journal of Regional Science, 27(1): 39-54.

[7] Coulson, E. and Li, H., (2013). "Measuring the External Benefits of Homeownership." Journal of Urban Economics, forthcoming.

[8] Ellen, E., Madar, J., and M. Weselcouch., (2012). "What's Really Happening to the REO Stock? An Analysis of Three Cities: New York, Atlanta, and Miami." Furman Center for Real estate and Urban Policy. Unpublished manuscript.

[9] Ellen, I., Lacoe, J., and Sharygin, C., (2013). "Do Foreclosures Cause Crime?" Journal of Urban Economics. 74: 59-70.

[10] Fingleton, Bernard, and Gallo, Julie Le. (2008). "Estimating spatial models with endogenous variables, a spatial lag and spatially dependent disturbances: Finite sample properties," Papers in Regional Science, Volumn 87: 320-339.

[11] Frame, W., (2010). "Estimatign the Effect of Mortgage Foreclosures on Nearby Property Values: A Critical Review of the Literature." Federal Reserve Bank of Atlanta Economic Review. 95(3): 1-9.

[12] Kiefer, Hua, and Kiefer, Leonard, (2011). "The Co-Movement of Mortgage Foreclosure Rate and House Price Depreciation: A Spatial Simultaneous Equation System," working paper.

[13] Harding, J., Rosenblatt, E., and Yao, V., (2009). "The Contagion Effect of Foreclosed Properties." Journal of Urban Economics. 66(3): 164-178.

[14] Ihlanfeldt, K. and Mayock, T., (2013). "The Impact of REO Sales on Neighborhoods and Their Residents." Unpublished manuscript.

[15] Immergluck, D., (2010). "The Accumulation of Lender-Owned Homes During the U.S. Mortgage Crisis: Examining Metropolitan REO Inventories." Housing Policy Debate. 20(4): 619-645.

[16] Kelejian, H.H., and Piras, G., (2011). "An extension of Kelejian's Spatial J-Test for Non-nested Spatial Models," Regional Science and Urban Economics. 41(3): 281–292.

[17] Kelejian, H.H., (2008). "A Spatial J-test for Model Specification Against a Single or a Set of Nonnested Alternatives," Working paper.

[18] Kelejian, H.H., and Prucha, I.R., (2004). "Estimation of Simultaneous Systems of Spatially Interralted Cross Sectional Equations," Journal of Econometrics 118: 27-50.

[19] Smith, G. and Duda, S., (2009). "Roadblock to Recovery: Examining the Disparate Impact of Vacant Lender-Owned Properties in Chicago." Woodstock Institute.

[20] Steinnes, Donald N. and Walter D. Fisher, (1974). "An Econometric Model of Intra-urban Location," Journal of Regional Science, 14(1): 65-80.

APPENDIX A: ESTIMATION - GS3SLS

In this section, we apply Kelejian and Prucha (2004)'s full information generalized spatial three stage least squares (GS3SLS) procedure on a panel data set.[19] Because we are dealing with a panel data with the existence of spatial fixed effect, we first demean all the dependent and explanatory variables on both sides of the equation system.[20] Before the application of GS3SLS, we also stack the observations across spatial units N and over time periods T as specified in equations (8) and (9).[21]

A.1. Step 1 - Initial 2SLS

We first estimate an initial 2SLS estimator of δ_j ($j = 1, 2$) for each equation respectively, using an instrument $\widetilde{Z}_{j,NT} = P_H Z_{j,NT}$, where P_H as expressed as

$$P_H = H(H'H)^{-1}H',$$

with H defined as

[19] GS3SLS estimator has been proved to have the same asymptotic distribution as the true estimator, assuming some regular conditions hold. These assumptions include zero diagnal entries in W_{NT}; $(I_{2NT} - \widetilde{B}_N^*)$ and $(I_{NT} - \rho_j W_{NT})$ are nonsingular with $|\rho_j| < 1$ for $j = 1, 2$; x_{NT} has full column rank; W_{NT}; $(I_{2NT} - \widetilde{B}_N^*)^{-1}$ and $(I_{NT} - \rho_j W_{NT})^{-1}$ are bounded uniformly in absolute value; etc.

[20] See Baltagi (2001) for details. The entries of ys, and xs take the values of the deviation from the average over time.

[21] As pointed out by Gebremariam(2007), because of complication created by the introduction of W_{NT}, conventional rank and order condictions for identification do not apply to the spatial simultaneous equation system. Thus, the spatial system can be treated as a special case of nonlinear simultaneous equation system (i.e., nonlinear endogenous variables), whose identification conditions are satisfied by our SESSAR model.

$$H = (X_{NT}, W_{NT}X_{NT}, W_{NT}^2 X_{NT}),$$

where $X_{NT} = (X_{1,NT}, X_{2,NT})$ including all the exogenous variables.[22] We have $\widetilde{Z}_{j,NT} = (\widetilde{y}_{j,NT}, X_{j,NT}, \widetilde{\overline{\overline{y}}}_{j,NT})$, where $\widetilde{y}_{j,NT} = P_H y_{j,NT}$, $X_{j,NT} = P_H X_{j,NT}$,[23] and $\widetilde{\overline{\overline{y}}}_{j,NT} = P_H \overline{y}_{j,NT}$. The 2SLS estimator of δ_j is

$$\widetilde{\delta}_{j,2sl} = (\widetilde{Z}'_{j,NT} Z_{j,NT})^{-1} \widetilde{Z}'_{j,NT} y_{j,NT},$$

and the 2SLS residuals are

$$\widetilde{u}_{j,2sl} = y_{j,NT} - Z_{j,NT}\widetilde{\delta}_{j,2sl}.$$

We then apply Moran I test described in Anselin and Kelejian (1997) for the above residuals, both equations reject the null hypothesis of no spatial autocorrelation in $\widetilde{u}_{j,2sl}$ at a significant level of 0.01. Hence, we proceed with our estimation on the spatial autoregressive parameter ρ_j specified in equation (12).

A.2. Step 2 - Spatial Autoregressive Parameter ρ_j

Based on the moment functions of $\varepsilon_{j,NT}$ ($j = 1, 2$) as in (13) and the relationship of $u_{j,NT}$ and $\varepsilon_{j,NT}$ as in (12), we have

$$\begin{aligned}
E(\frac{\varepsilon'_{j,NT}\varepsilon_{j,NT}}{NT}) &= E(\frac{u'_{j,NT}u_{j,NT}}{NT} + \rho_j^2 \frac{\overline{u}'_{j,NT}\overline{u}_{j,NT}}{NT} - 2\rho_j \frac{u'_{j,NT}\overline{u}_{j,NT}}{NT}) = \sigma_{jj}, \\
E(\frac{\overline{\varepsilon}'_{j,NT}\overline{\varepsilon}_{j,NT}}{NT}) &= E(\frac{\overline{u}'_{j,NT}\overline{u}_{j,NT}}{NT} + \rho_j^2 \frac{\overline{\overline{u}}'_{j,NT}\overline{\overline{u}}_{j,NT}}{NT} - 2\rho_j \frac{\overline{\overline{u}}'_{j,NT}\overline{u}_{j,NT}}{NT}) = \frac{\sigma_{jj} tr(W'_{NT}W_{NT})}{NT}, \\
E(\frac{\varepsilon'_{j,NT}\overline{\varepsilon}_{j,NT}}{NT}) &= E(\frac{u'_{j,NT}\overline{u}_{j,NT}}{NT} + \rho_j^2 \frac{\overline{u}'_{j,NT}\overline{\overline{u}}_{j,NT}}{NT} - 2\rho_j (\frac{\overline{u}'_{j,NT}\overline{\overline{u}}_{j,NT}}{NT} + \frac{\overline{u}'_{j,NT}\overline{u}_{j,NT}}{NT})) = 0,
\end{aligned}$$

where $\overline{u}_{j,NT} = W_{NT}u_{j,NT}$, $\overline{\overline{u}}_{j,NT} = W_{NT}^2 u_{j,NT}$, and $\overline{\varepsilon}_{j,NT} = W_{NT}\varepsilon_{j,NT}$. Denote

$$\begin{aligned}
\alpha_j &= [\rho_j, \rho_j^2, \sigma_{jj}]', \\
g_{j,NT} &= \frac{1}{NT}[\widetilde{u}'_{j,2sl}\widetilde{u}_{j,2sl}, \widetilde{\overline{u}}'_{j,2sl}\widetilde{\overline{u}}_{j,2sl}, \widetilde{u}'_{j,2sl}\widetilde{\overline{u}}_{j,2sl}], \\
G_{j,NT} &= \begin{bmatrix} 2\widetilde{u}'_{j,2sl}\widetilde{\overline{u}}_{j,2sl} & -\widetilde{\overline{u}}'_{j,2sl}\widetilde{\overline{u}}_{j,2sl} & NT \\ 2\widetilde{\overline{\overline{u}}}'_{j,2sl}\widetilde{\overline{u}}_{j,2sl} & -\widetilde{\overline{\overline{u}}}'_{j,2sl}\widetilde{\overline{\overline{u}}}_{j,2sl} & tr(W'_{NT}W_{NT}) \\ \widetilde{u}'_{j,2sl}\widetilde{\overline{\overline{u}}}_{j,2sl} + \widetilde{\overline{u}}'_{j,2sl}\widetilde{\overline{u}}_{j,2sl} & -\widetilde{\overline{u}}'_{j,2sl}\widetilde{\overline{\overline{u}}}_{j,2sl} & 0 \end{bmatrix},
\end{aligned}$$

where $\widetilde{\overline{u}}_{j,2sl}$, and $\widetilde{\overline{\overline{u}}}_{j,2sl}$ are transformed forms the 2SLS residuals following a similar fashion of $\overline{u}_{j,NT}$ and $\overline{\overline{u}}_{j,NT}$. The generalized moments estimator of (ρ_j, σ_{jj}) can be obtained through the minimization process of

$$(\widehat{\rho}_{j,gmm}, \widetilde{\sigma}_{jj,gmm}) = \arg\min(g_{j,NT} - G_{j,NT}\alpha_j)'(g_{j,NT} - G_{j,NT}\alpha_j).$$

[22] In our specification of $X_{1,NT}$ and $X_{2,NT}$, there exist deplicate exogenous variables, we therefore exclude the repeated ones to avoid multicollinearity and have $X_{NT} = (Unem, Inc, Permit, Mrt, Bkt, Dti)$.

[23] P_H is the projector of $X_{j,NT}$.

A.3. Step 3 - GS3SLS

The estimated ρ_j from the previous step is used to perform a Cochrane-Orcutt-type transformation to account for the spatial dependence in the disturbances. Denote

$$\begin{aligned} y^*_{j,NT} &= y_{j,NT} - \widehat{\rho}_{j,gmm} W_{NT} y_{j,NT}, \\ Z^*_{j,NT} &= Z_{j,NT} - \widehat{\rho}_{j,gmm} W_{NT} Z_{j,NT}, \end{aligned}$$

then estimate the transformed equation of

$$y^*_{j,NT} = Z^*_{j,NT}\delta_j + \varepsilon_{j,NT}, \tag{23}$$

using a 2SLS estimator as

$$\widehat{\delta}_{j,2sls} = (\widehat{Z}^{*\prime}_{j,NT} Z^*_{j,NT})^{-1} \widehat{Z}^{*\prime}_{j,NT} y^*_{j,NT},$$

where $\widehat{Z}^*_{j,NT} = P_H Z^*_{j,NT}$ with $P_H = H(H'H)^{-1}H'$. To achieve a consistent estimator of Σ, let

$$\widehat{\varepsilon}_{j,NT} = y^*_{j,NT} - Z^*_{j,NT}\widehat{\delta}_{j,2sls},$$

then compute

$$\widehat{\sigma}_{jl} = \frac{1}{NT}\widehat{\varepsilon}'_{j,NT}\widehat{\varepsilon}_{l,NT}, \text{ for } j,l = 1,2,$$

which composes the elements of $\widehat{\Sigma}$. We now stack the equations in (23) over $j = 1,2$ as

$$y^*_{NT} = Z^*_{NT}\delta + \varepsilon_{NT},$$

where $y^*_{NT} = (y^{*\prime}_{1,NT}, y^{*\prime}_{2,NT})'$, $Z^*_{NT} = \begin{bmatrix} Z^*_{1,NT} & 0 \\ 0 & Z^*_{2,NT} \end{bmatrix}$, and $\delta = (\delta'_1, \delta'_2)'$. Because $E\varepsilon_N(t)\varepsilon_N(t)' = \Sigma \otimes I_N$,

$$E\varepsilon_{NT}\varepsilon'_{NT} = \Sigma \otimes I_N \otimes I_T = \Sigma \otimes I_{NT}.$$

The GS3SLS estimator of δ is then

$$\widehat{\delta}_{3sls} = [\widehat{Z}^{*\prime}_{NT}(\widehat{\Sigma}^{-1} \otimes I_{NT}) Z^*_{NT}]^{-1} \widehat{Z}^{*\prime}_{NT}(\widehat{\Sigma}^{-1} \otimes I_{NT}) y^*_{NT},$$

where $\widehat{Z}^*_{NT} = P_H Z^*_{NT} = \begin{bmatrix} P_H Z^*_{1,NT} & 0 \\ 0 & P_H Z^*_{2,NT} \end{bmatrix}$. The small sample distribution of $\widehat{\delta}_{3sls}$ is approximated as follows

$$\widehat{\delta}_{3sls} \,\tilde{}\, N(\delta, [\widehat{Z}^{*\prime}_{NT}(\widehat{\Sigma}^{-1} \otimes I_{NT})\widehat{Z}^*_{NT}]^{-1}). \tag{24}$$