Institute for International Economic Policy Working Paper Series
Elliott School of International Affairs
The George Washington University

Testing the Association between Foreclosure and Nearby House Values: Can Differences Deceive?

IIEP-WP-2016-29

Anthony Yezer
George Washington University

Yishen Liu
Fannie Mae

November, 2016

Institute for International Economic Policy
1957 E St. NW, Suite 502
Voice: (202) 994-5320
Fax: (202) 994-5477
Email: **iiep@gwu.edu**
Web: **www.gwu.edu/~iiep**

Testing the Association between Foreclosure and Nearby House Values: Can Differences Deceive?

Yishen Liu* Anthony Yezer†

November 9, 2016

Abstract

Foreclosure externalities, in which the number of recent foreclosures proximate to a housing unit depress its sales price, are well accepted in the literature. The empirical testing that supports this relation is based on highly differenced data designed to isolate the relation between foreclosure intensity and house value by removing the confounding effects of local variation in the price of housing services. The results in this paper demonstrate that differencing over time and space is not a guarantee against misleading results even if the effects of local price variation are removed. Although foreclosure externalities are the specific object of analysis here, the findings likely generalize to a larger body of research where effects of unobserved heterogeneity between treated and controls are "removed" by differencing.

Keywords: Foreclosure; Specification error; Loan-to-value ratio; Externalities.

JEL Classification Numbers: R23, R30, R31.

*Economist, Fannie Mae.

†Department of Economics, George Washington University, address correspondence to yezer@gwu.edu.

1 Introduction

The substantial variation in rates of foreclosure on single family mortgages in the period following 2006 provided a promising opportunity to investigate the possibility that the foreclosure process generates local externalities. Foreclosure rates varied substantially both over time and across areas within a given housing market. Many papers found a negative relation between recent foreclosures and the current sales price of housing proximate to the foreclosure. These papers recognize that local unobserved heterogeneity in housing markets may generate the "treatment" of foreclosure and they adopted clever differencing techniques to eliminate the effects of this correlation on the relation between foreclosure and house price. This paper takes no position on the efficacy of those efforts.

Several theoretical mechanisms have been proposed that could connect foreclosed properties to local house prices. Some of these are real effects. The most direct relation is a real physical externality due to decreased maintenance of the foreclosed property that makes it less attractive to neighbors. Gerardi et al. [15] have called this the "investment effect". To the extent that foreclosure results in eviction or at least forced migration, there might be social effects on a neighborhood. There is evidence that foreclosed properties are more likely to be purchased by investors changing the tenure mix of a neighborhood. If sales prices rise with owner occupancy rates, this "tenure mix effect" could lower prices.[1] To the extent that foreclosure sales are involuntary, they reflect an exogenous increase in local supply which might temporarily lower prices. This has been called the "supply effect". Substantial evidence indicates that sale of REO by lenders occurs at depressed prices, likely reflecting the failure to market the property in an optimal fashion and/or damage done to the unit in the foreclosure process. To result in an externality, there must be a mechanism to translate the lower prices of foreclosure sales to the sales of non-distressed properties. This could occur if there is an "appraisal effect" in which appraisers and buyers fail to distinguish effects of

[1]See Coulson and Li [10] for an attempt to deal with the difficult problems of making this inference about owner occupancy rates and house prices.

property condition on foreclosure sales prices and lower appraisals of non-distressed properties that are well maintained. In sum, previous papers have identified several theoretical mechanisms that could produce empirical results indicating a negative externality associated with foreclosure.

Overall, foreclosure is a deeply endogenous event involving decisions by owners and lienholders and rejection of alternatives such as short sales, refinancing, and loan modification. Effects on nearby properties require a sales and purchase decision by other parties which is also endogenous to neighborhood conditions and future expectations. Based on both a review of the literature and new results reported here, it appears that the investment effect is important but this paper does not claim to identify a precise causal mechanism that may be responsible for any association between foreclosure and sales prices of nearby housing. Indeed, the main contribution of this paper is to question whether foreclosure itself is important in generating what has been called foreclosure externalities.

Past tests of the association between foreclosure and house values have acknowledged difficulties created by potential omitted variables bias. The principle concern in the literature has been the possibility that unobserved location-specific factors that lower sales price may increase the likelihood of foreclosure and thus generate false acceptance of the hypothesis that foreclosure lowers house price. This is a standard concern in models where treatment is selected based on unobservables that can be correlated with the outcome. The literature has addressed this issue by using spatial and temporal disaggregation to remove the correlation between unobserved heterogeneity in the error term and selection into treatment. By saturating a model with time and locational dummies at the sub-neighborhood (census tract or smaller) level, researchers argue that they have removed any correlation between foreclosure and omitted variables influencing housing value. The techniques effectively pair units co-located in the same sub-neighborhood which vary in their proximity to foreclosed units in that same sub-neighborhood during a modest time period.

This approach to dealing with unobserved heterogeneity may succeed in its objective

of removing the effects of selection into treatment. However the estimation technique still involves strong assumptions and it obscures the role of alternative variables in the process being studied. Evaluation of the estimation results is generally based on the implicit assumption that the relation between treatment (foreclosure) and outcome (housing price) is additive independent. If the house value function is not additive independent, then saturation with time and locational dummies at the sub-neighborhood level yields coefficient estimates whose magnitude reflects an average treatment effect of foreclosure that depends on the sample of properties being studied. Empirical research that attempts to deal with unobserved heterogeneity or omitted variables bias by successively differencing observations must rely on this "additive independence" assumption. This issue appears to have gone unnoticed in the literature or at least the research on foreclosure externalities reviewed below makes no serious attempt to argue for additive independence.[2] This omission is common in the literature on other topics where differences in differences or synthetic control estimators are used.

Even if the relation between house value or its logarithm and foreclosure is additive independent, unbiased estimation also requires that any omitted variables not vary spatially at the sub-neighborhood level, and particularly that this variation not be correlated with the included variable, foreclosure in this case. This is generally understood as the selection into treatment problem. But, even if the omitted variables are relatively uncorrelated with the treatment, their omission reduces understanding of the causes of variation in the outcome. This problem will be termed the "micro variation" omitted variable bias problem. Potentially important causes of the outcome are ignored. This is separate from the problem of the issue of selection into treatment based on unobserved heterogeneity. In the context of foreclosure externalities, this means that insofar as practicable given limits of data availability, variables that are important to micro variation of the outcome should be included in the model. In the case of foreclosure externalities, understanding the relation between foreclosure as

[2]The case against additive independence is so obvious that any attempt to consider the problem would quickly cause researchers to realize that it does not hold.

a treatment and local house prices as an outcome is completely reversed as a result of extending the estimation results to include other variables whose micro variation is similar to foreclosure.

This paper identifies variables important to the relation between foreclosure and house prices which illustrate the two problems, additive independence and micro variation. The econometric problems created when additive independence and micro variation are violated are well understood and not an original addition to the literature. The contribution here is to illustrate how radically the interpretation of empirical results can change as a result of recognizing and correcting for them.

The variables routinely omitted in the literature that create the additive independence problem are the number of housing units in the micro area surrounding each observed transaction and measures of the strength of creditors' remedies. Curiously these variables have been omitted in all of the literature reviewed in the next section of this paper in spite of their obvious relation to the effect of additional foreclosures on nearby housing value and the fact that they are easily observed. The importance of counting nearby housing units arises because a foreclosure is surely more consequential for valuation of a nearby property if it is the only unit near the home being sold than if it is one of 100 or more units nearby. Note that this is not a problem of micro variation although nearby housing unit counts may vary within small areas. The appropriate specification involves the ratio of foreclosed units to total units, a relation based on the quotient of the two variables rather than their sum.

Creditors' remedies determine the ease of foreclosure and the ability of lienholders to protect the condition of delinquent properties as they move through the foreclosure process. Cordell and Lambie-Hanson [8] document the substantial variation in foreclosure cost over time and jurisdictions as the housing crisis deepened. There were two reasons for this. Statutory changes weakened remedies and the institutions of the market that administer foreclosure became congested by the volume of activity. The delay in foreclosure substantially lagged recovery in the housing market. This, of course, implies that the size of any foreclosure

externality due to the investment effect should not be constant over space and time as has been implicitly assumed in the literature which postulates that all foreclosures have an equivalent effect on house prices.

The variables exhibiting micro variation that are generally omitted in the literature are counts of both seriously delinquent mortgages and units with high CLTV proximate to the house being sold. Delinquent borrowers or those with high CLTV should exhibit many of the same maintenance issues as foreclosed units. Lack of maintenance could create effects through the same investment effect as foreclosure. Gerardi et al. [15] have added the spatial distribution of delinquencies to a foreclosure model and found that the effects of foreclosure appear to be reduced. More important, they find that serious delinquency depresses surrounding property values independent of foreclosure. Biswas and Davidoff [2] measure the spatial distribution of high CLTV properties around each property sale but use it as an instrument for foreclosure rather than as an independent determinant of property prices.

This paper estimates a standard foreclosure externality model incorporating two variables entering multiplicatively, number of housing units and foreclosure cost (as a measure of effective remedies), and two micro variation variables, proximate high CLTV and delinquent properties. The estimation results confirm that the proper specification of a foreclosure externality equation should be the ratio of foreclosures to housing units, not the level of foreclosures. Perhaps most important, the absolute size of the relation between foreclosure and local house prices appears to vary directly with the difficulty of foreclosure. Foreclosure cost is the primary determinant of the effect of foreclosure on local house prices. Both of the two micro variation variables, the fraction of high CLTV units and serious delinquent units have similar but smaller depressing effects on nearby housing prices given current policies toward creditors' remedies. Estimation results indicate that the relation between house prices and the ratio of foreclosures to housing stock is convex as might be expected.[3]

[3]Some papers have attempted to test for non-linearity in the effects of number of foreclosures on sales price. The results have been rather mixed. Perkins et al. [23] report a threshold effect in which the first two foreclosures have no significant price effect so that the overall relation between foreclosures and sales price is concave.

Overall, the view of foreclosure externalities implied by the approach and results in this paper is substantially different than the previous literature. High CLTV properties that are not in foreclosure produce a negative external effect on local sales prices, likely through the investment effect. A lengthy foreclosure process increases these negative effects. But foreclosure also removes the high CLTV problem. Ultimately foreclosure, even if delayed, cures the externality. Simply put, the current literature has confused the treatment with the disease and failed to consider the manner in which restrictions on the cure (creditors' remedies) inhibit treatment of the problem. Furthermore the specification and interpretation issues identified here may generalize to other empirical work on treatment effects where the assumptions of additive independence and micro variation have not been justified.

2 Literature on Foreclosure and Nearby Property Value

Important characteristics of the foreclosure process

While the empirical literature agrees that there are foreclosure externalities there is a lack of consensus in the literature on the definition of foreclosure. The foreclosure process in states that require judicial foreclosure must begin with court action in which the lienholder sues the borrower.[4] In states that allow statutory foreclosure the trustee acts. Notice must be filed as provided in the deed of trust but there is no need to sue the borrower. In either case, the borrower receives a "foreclosure complaint" documenting that there is a sufficient delinquency in scheduled payments to warrant foreclosure.[5] The borrower may respond to the complaint by curing the deficiency or mounting a defense which requires a trial. Even

[4]These actions are usually unopposed but they are court actions and result in court orders.

[5]In states with judicial foreclosure the lienholder must sue the borrower in court and obtain notice, or lis pendens, which is public information. Where statutory foreclosure is permitted, the trustee sends a notice of delinquency to the borrower who then has a period to cure the deficiency before the property goes to auction. Judicial actions are in the public record and statutory notice of intent to foreclosure must also be filed with the county recorder allowing local vendors to collect and publish this information which is available to investors, realtors, etc. The past standard of 90 day delinquency was extended to 180 days in 2014 by the Consumer Financial Protection Bureau.

in states what require judicial foreclosure, most lienholder motions for summary judgement are granted. If the foreclosure process is not interrupted as discussed in some detail below, it results eventually in a "foreclosure judgement" followed by a "foreclosure auction" where title to the property is transferred. The empirical literature on foreclosure externalities uses either the initial foreclosure complaint or the transfer of title as an indicator of a property experiencing foreclosure. Clearly these are very different events. The former indicates serious delinquency and the latter failure to resolve the delinquency and transfer of title.

Borrowers often respond to the complaint by curing the delinquency. The first recourse is to the lienholder who may be willing to consider a loan modification, forbearance agreement, or payment plan. In some states the borrower may request a mortgage modification mediation hearing. Of course the borrower may self-fund the cure, or refinance the mortgage. Any of these steps reestablishes the borrower's equitable right of redemption. Selling the property is another possibility. In cases of negative equity, refinancing is likely not possible and selling will not cure the entire debt to the lienholder.[6] However, a short sale can be arranged in which the borrower agrees to cooperate with efforts to sell the property and the lienholder agrees to accept the sale proceeds as full payment of principal, interest, and penalties outstanding. In the dataset used here, approximately half of the foreclosure notices do not result in sale of the property.

Other borrower responses are possible. A bankruptcy filing can substantially delay or even reverse foreclosure. The borrower may also invoke the right of rescission under the Truth in Lending Act (TILA). Regardless of delinquency status, if the borrower informs the lender of the intent to rescind based on violation of TILA, this stops the foreclosure process and sets in motion a series of legal options that could permanently impair the lender's security interest in the collateral.[7]

[6] At this point negotiation between lienholder and borrower are likely heavily influenced by whether the state permits deficiency judgments against the borrower. Deficiency judgments are permitted in all three jurisdictions in the data set used here.

[7] This three-year right to rescind under TILA does not extend to home purchase mortgages, i.e. primary purchase mortgages or piggy back seconds, or financing of property which is not a primary residence. Borrowers with a first and non-piggy back second from the same lender can exert an extra measure of influence

If there is no interruption in the foreclosure process, i.e. the borrower does not act to reestablish the right of redemption or surrender the deed in lieu of foreclosure, an auction is scheduled, usually conducted by an officer of the court. The lender generally participates in this auction and, in the absence of other bidders, takes title to the property. Alternatively, title can be transferred to another individual or organization bidding higher than the lender. If the lender takes title, the property is then REO which is sold either immediately or after being held for some period of time.

In some states, borrowers retain a right of redemption even after disposition of the property at auction. This right of redemption may extend for as much as a year, clouds the title available at auction, and may delay final disposition of the property to a willing buyer. Because of the complications associated with this extended right of redemption, the areas selected for study here do not have this provision. In some states, eviction is an issue because the foreclosure judgment converts the borrower into a renter and the eviction from rental units can be protected. In the end, eviction may be accelerated by payment of key money. These details are potentially important and have been embodied in the measure of foreclosure difficulty used here. There can be significant delays in the ultimate disposition of the property to a new owner with clear title and an incentive to maintain, and even upgrade the property. In sum, there are many reasons why the relation between foreclosure actions and condition of the real property collateral may vary across states due to substantial variation in creditors' rights and the performance of the court system.

As noted above, some empirical tests for foreclosure externalities count the foreclosure complaint or lis pendens in the case of judicial foreclosure while others define foreclosure as the transfer of title from the borrower sometime after the complaint, usually at the time of the foreclosure judgment. These two measures are correlated but certainly vary significantly across jurisdictions, as documented later in the data section. To the extent that the link between foreclosure and nearby property values is based on maintenance issues, transfer of

with the lender if they rescind the second. Rescission brings the loan out of technical default until further legal action or negotiation determines the response to the claim of imperfect disclosure under TILA.

title is a more consequential measure because, in almost all cases, it implies negative equity in the unit. There are two reasons for this. First, the borrower's failure to cure through refinancing or from internal funds implies that a high CLTV perhaps approaching unity. Second, borrowers not only lack an incentive to maintain units when title is being transferred completely and permanently, they may damage the unit as they vacate. Borrowers who cure their delinquencies are indicating that they have equity which is worth preserving by curing, maintaining, and not damaging the property. Thus far, empirical studies where foreclosure is based on notice and those requiring evidence of property transfer have produced similar results, i.e. an increase in the number of foreclosures proximate to a unit tends to lower the sales price of the unit. This is unsurprising given the high correlation between the two measures across space and time. Furthermore, studies using data from states requiring judicial foreclosure are similar in their finding of external effects on sales prices to those from states where statutory foreclosure is permitted. Finally the effects of the rights of rescission and redemption have been ignored in the literature. One of the most intensively studied states, Illinois, has a very complex right of redemption provision.

Own-price effects of foreclosure sales

The foreclosure externality literature builds on some papers that have established stylized facts about the depressed sales price of foreclosed properties. Foreclosed properties sell for less than similar surrounding units. Pennington-Cross [22] finds that prices of distressed properties appreciate 22% less than a price index of other units being sold. Clauretie and Daneshvary [7] report a smaller but still substantial foreclosure discount. Campbell et al. [4], not only find a 27% price discount for foreclosures, they argue that part of this is due to the forced nature of the sales process as lenders unload REO. They report price discounts for other situations, such as estate sales, which are grouped under the general category of forced sales. Sumell [26] finds even larger, perhaps 60%, price discounts in inner-city Cleveland. Recently Mian et al. [21] found that, during the 2007 to 2009 portion of the

housing downturn, states permitting statutory foreclosure experienced far more foreclosure and much sharper price declines than nearby states that only allow judicial foreclosure. These effects abated by 2011 to 2013. Lambie-Hansen [20] reports that public complaints about properties undergoing foreclosure are 1.3 to 1.7 times as frequent as otherwise comparable properties. This is a small sample of a literature that documents the existence of a substantial foreclosure discount which arises from the deterioration in condition of housing units as they pass through the foreclosure process. Deterioration may increase with the volume of foreclosure activity and vary across housing markets. All this is consistent with the "investment effect" where lack of maintenance imposes a physical externality on nearby housing and the "appraisal effect" in which comparables include foreclosure sales at depressed prices.

Testing for the existence and size of foreclosure externalities

Testing for an empirical effect of nearby foreclosure on property values has produced a fairly uniform set of conclusions regardless of the definition of foreclosure, initial complaint versus judgment and property transfer. Furthermore, differences in regulatory environment discussed above, including judicial versus statutory foreclosure, recourse versus non-recourse, do not alter findings. Even variation between measuring price effects using classic hedonic versus repeat sales price indexes still produces the same general negative relation between the number of recent foreclosures proximate to a housing sale and the price at which the transaction occurs.[8] By 2010, the literature was large enough to support a review article, Frame [14].

In addition to the general measurement of foreclosure externalities, some studies have attempted to use special situations to produce additional insights regarding the nature of foreclosure externalities. One result, evident in both Hartley [18] and Fisher et al. [13], is that foreclosure externalities are much less important in the multifamily housing stock.

[8]Examples of these studies include Immergluck and Smith [19], Schuetz et al. [25] , Harding et al. [17], Campbell et al. [4], Zhang and Leonard [29].

Specifically effects within a given structure are much larger than spillovers across structures. Whitaker and Fitzpatrick [28] find that vacant properties produce a negative external effect on surrounding property sales similar to that of foreclosed properties. Results in Cui and Walsh [11] suggest that vacancy and foreclosure are associated with higher crime rates.

Gerardi et al. [15] come closest to the tests for foreclosure externalities in this paper. In addition to foreclosure effects on local housing price, they add a measure of the number of delinquent properties and find that both foreclosure and serious delinquency, but not early delinquency, have negative effects on local sales price.[9] They also have some information on property condition for foreclosures and find that the negative foreclosure effect varies inversely with property conditions, suggesting an investment effect.

Groves and Rogers [16] provide strong evidence that foreclosure externalities are produced by the investment effect. In a natural experiment, they test for the size of foreclosure externalities in ordinary housing compared to housing located in residential community or homeowners associations that use covenants in deeds to maintain the physical condition of member properties. They measure both the own-price and externality effects of foreclosure and find that foreclosed properties in associations have very low (3%) own-price and zero external price effects. Non-association property foreclosures in the same area have much larger own-price and significant external price effects.

Size and duration of foreclosure costs

In the discussion of institutional foreclosure, the diversity of the procedures across jurisdictions was noted. Judicial versus statutory foreclosure, right of redemption, eviction procedures, deficiency judgments can all influence the foreclosure process. In addition, the recent crisis introduced requirements that lenders consider modifications through the Home Affordable Modification Program and Home Affordable Refinance Program. Cordell et al. [9]

[9]The delinquency measure is constructed with great difficulty because the state of delinquency is only observed on a local neighborhood basis for properties with mortgages securitized or held by Fannie Mae. Unlike this paper, the number of housing units in each area, high LTV mortgages, and variation in foreclosure time or cost are not part of the analysis.

document the substantial changes in the cost of foreclosure delay that occurred during the housing crisis due both to changing rules and congestion in the courts. Simply put, if foreclosure delay is important for foreclosure externalities, then there is substantial variation over time and location in the property liquidation time. Cordell et al. [9] estimate average liquidation time in judicial (statutory) states at 21 (16) months at the start of the crisis in 2007 and 35 (25) months in 2013. Dagher and Sun [12] have identified a difference in mortgage terms across judicial and statutory states that presumably reflects these differences in costs. Cordell and Lambie-Hanson [8] estimate the difference in foreclosure costs across states and time periods that is used as an index of foreclosure delay in this paper.

3 Additive Independence and Spatial Micro Variation

Omitted variable bias is inevitable in house value equations because of the extreme diversity of the housing stock, particularly the stock of single family residences. This is particularly true because sales price is based on observation of a transaction in which the condition of the housing unit itself is endogenous as sellers may conduct substantial improvements before listing the unit. Another problem is the possibility that valuation of the unit is based on the land rather than structure. Clapp and Bardos [5] and Clapp et al. [6] have established that the existence of sales where the object of the buyer is rebuilding on the site can significantly bias estimation results, particularly the estimated coefficient of structure age. Thus, even with a full set of housing unit characteristics, there is significant potential for unobserved heterogeneity among units because the site value in an alternative use is not observed.

Nearby foreclosure sales are not a physical characteristic of the unit being sold. Their influence on sales price is understood based on Rosen's classic implicit markets paper (Rosen [24]). In the past, this approach has been used to evaluate the implicit valuation of public services (particularly schools), pollution, crime, restaurant quality, and visual amenity. These factors can all generate neighborhood externalities in that they are not under the

control of the homeowner although they may influence the owner's decision to maintain and sell the unit. As a practical matter, many of these neighborhood factors are excluded from empirical work that applies Rosen's implicit markets model because the burden of measuring the quality and proximity of all local amenities so that they could be included explicitly in the analysis would be overwhelming. Instead research, and certainly research on the relation between foreclosure activity and house prices, relies on differences over time and space to remove the effects of omitted variable bias.

The measure of foreclosure activity used here is transfer of title which indicates that an auction or forced sale took place in response to the initiation of the foreclosure process.[10]

The literature on foreclosure and house prices has taken the view that differencing across time and space, dividing the city into very small sub-neighborhoods, eliminates omitted variable bias because there is no omitted variable highly correlated across space and time with foreclosure. One exception to this is recent work by Gerardi et al. [15] who argue that delinquency in mortgage payments is highly correlated with foreclosure because delinquency is necessary for the foreclosure process to begin. They find that both foreclosure and delinquency have statistically significant negative effects on sales prices.

The general point about micro variation of omitted variables made here is that any attempt to argue that the implicit markets model may be applied in estimates where many local amenity variables are omitted must also include a careful examination for omitted variables whose variation over space and time is not eliminated by differencing. Two variables are used to illustrate the micro variation problem here. One is properties with high current loan to value ratio (CLTV) mortgage debt because foreclosures are a subset of this variable and the incentive to maintain a unit is diminished by high CLTV giving rise to a potential external effect on house prices. The other is all properties in some part of the foreclosure process, e.g. properties where a foreclosure certification has been sent. Foreclosure certification is indicative of serious delinquency and is itself a first step to foreclosure which should

[10]See descriptive statistics in the data section for characteristics of foreclosure in the sample used here.

have the same variation over time and space.[11]

The problem created by omitted variables whose potential to bias estimates is not eliminated by successive differencing can be illustrated with a stylized example of the problem of estimating the relation between foreclosure and house value considered there. To simplify the presentation, ignore heterogeneity of the physical characteristics of housing so that variation in sales price is due to time, and location, including the effects of neighborhood amenity variables. In the literature on foreclosure and sales prices, this problem is specified as shown below.

$$V_{ijkt} = \alpha + \beta F_{ijk,t-1} + \theta_{jk} + \psi_{kt} + \epsilon_{ijkt} \tag{1}$$

Here, V_{ijkt} is the logarithm of the sales price of unit i, in neighborhood (census tract or smaller) j, jurisdiction k, at time t, F is the number of foreclosures in i in the time period, $t-1$, (i.e. just before t), θ is any error specific to jk, ψ is a time varying error at the jurisdiction level, and ϵ is an error uniquely associated with unit i. In this specification, θ includes the effects of omitted variables at the neighborhood and jurisdiction level and ψ includes time varying effects at the jurisdiction level. A sub-neighborhood i is associated with each unit and is defined by a small circular area around each unit. There is a distinct sub-neighborhood about each unit. By saturating the specification in Equation (1) with a full set of dummy variables specific to each jk and kt combination and recalling that k is a spatial combination of the j's, any correlation between the regression error and the foreclosure measure is thought to be eliminated, i.e. $E(\epsilon_{ijkt}|F_{ijk,t-1}) = 0$. This leaves the regression estimate of β free of omitted variable bias including the problem of selection into treatment. Effectively, specification with a full set of jk, kt, dummies means that the relation between foreclosure and value is based on differences in the number of foreclosures in the individual sub-neighborhoods within j surrounding two units at a given time. Alternatively it is sufficient to argue that, if there is another variable, say L_{ijkt}, that varies across j at

[11] As noted in Gerardi et al. [15] spatial measures of mortgage delinquency are extremely difficult to construct.

time t, and hence is included in ϵ_{ijkt}, it is uncorrelated with $F_{ijk,t-1}$.

In the case of the relation between foreclosure and sales prices, there are strong theoretical reasons to believe that the number of units with high CLTV also influences sales prices because, like foreclosure, CLTV has a negative effect on maintenance of housing. If the relation between foreclosure and value is based on a localized physical externality due to appearance, CLTV should influence sales price through a similar channel. This implies that Equation (1) be written as:

$$V_{ijkt} = \alpha + \beta F_{ijk,t-1} + \lambda L_{ijkt} + \theta'_{jk} + \psi'_{kt} + \epsilon'_{ijkt} \tag{2}$$

Just as the count of foreclosure sales around each property i is unique to a sub-neighborhood around i, the count of high CLTV units is sub-neighborhood specific. Indeed foreclosed properties are a subset of high CLTV units. Removing the effect of L_{ijkt} from the list of omitted variables modifies θ, ψ, and ϵ as indicated by the primes above these variables in Equation (2). The investment effect suggests that $\beta, \lambda < 0$. Estimating Equation (1) with high CLTV as an omitted variable results in $\epsilon_{ijkt} = \lambda L_{ijkt} + \epsilon'_{ijkt}$ and $E(\epsilon_{ijkt}|L_{ijkt} < 0)$. Obviously, as seems likely, L_{ijkt} and $F_{ijk,t-1}$ are positively correlated this would create a classic problem in which selection into treatment (foreclosure) was correlated with unobserved heterogeneity. The resulting downward bias in estimates of β survives even in a model saturated with dummy variables. This potential bias is not the central point being made here. Failure to consider the potential role of high CLTV changes the entire interpretation of the effects of foreclosure on property values because, one purpose of foreclosure is to eliminate high CLTV properties.

While this discussion was illustrated by high CLTV, the same argument about micro variation issues applies to serious delinquency or other variables that influence maintenance decisions. Even if heterogeneity due to serious delinquency and high CLTV is not sufficiently correlated with foreclosure to bias estimated of the foreclosure coefficient. Furthermore the

need to consider bias due to this type of micro variation can be extended to other applications of the implicit market model. For example, crime could be substituted for foreclosure and police effort for high CLTV and the argument applied to the relation between crime and property values with police effort as the omitted variable.

The second problem with application of the implicit markets model is the assumption of additive independence. Returning to the relation between foreclosures and property values, and alternative specification of Equation (1) is:

$$V_{ijkt} = \alpha + \beta(F_{ijk,t-1}/H_{ijk,t-1}) + \theta_{jk} + \psi_{kt} + \epsilon_{ijkt} \tag{3}$$

where H_{ijkt} is the number of housing units in sub-neighborhood i at time $t-1$. Given that H_{ijkt} is relatively constant over time, if Equation (3) were additive independent in F, omission of H would not bias estimates of β although it is likely that foreclosures are increasing in housing. Because H_{ijkt} can be constructed from the same data sources that provide information on F_{ijkt}, failure to include it in previous research on foreclosure externalities is curious indeed. Some papers include a measure of housing density at the j or k level of spatial disaggregation as an additive independent component of the hedonic equation because density may influence or be influenced by house value.[12] This is evidence that the importance of the assumption of additive independence is not widely recognized.

All of the hypotheses regarding the possible relation between foreclosure and house value presented in the literature imply that the ratio of foreclosures to housing stock rather than a simple count of foreclosures has an effect on values. The supply hypothesis states that foreclosures add to sub-neighborhood supply. The importance of addition of a single unit to supply obviously depends on the size of the market. The tenure effect of a foreclosure

[12]The relation between density and sales price is complex. On the demand side, lower housing density may be the source of positive neighborhood open space externalities. Alternatively, higher density can be the result of higher land values and raise the cost of housing from the supply side. In this particular data, the estimated coefficient of number of housing units inserted as an additive independent variable is not statistically significant because of the saturated locational dummy variables. The fact housing density has no additive independent effect on value in these estimates, has no bearing on its importance for modeling the relation between foreclosure and sales price.

also depends on the number of units in the sub-neighborhood. The "appraisal" hypothesis argues that appraisers are forced to use foreclosures as comparables. Clearly the importance of adding a foreclosure sale to either the supply of units listed or comparables used to appraise depends on the number of alternative units in the area. The relation between the physical externality of an additional poorly maintained unit and the investment effect should fall as the number of well-maintained units in the sub-neighborhood rises. Finally, any measure of the external effect associated with a given fall in value would need to be scaled by the number of units in the sub-neighborhood. Accordingly, the argument for Equation (3) rather than (1) as the appropriate specification to test any of the hypotheses relating foreclosure and nearby property values, is extremely strong.

If (3) is the proper specification and H_{ij} cannot be observed, then the appropriate dummy variable specification of (3) would be:

$$V_{ijkt} = \alpha + \sum_{ij} \beta_{ij} F_{ijk,t-1} D_{ijk} + \sum_{jk} \varphi_{ij} D_{jk} + \sum_{kt} \pi_{kt} D_{kt} + \epsilon_{ijkt} \tag{4}$$

But estimates of (4) would yield estimates of an array of β's with individual i's nested within neighborhood j's. Put another way, when the relation between the variable of interest, F, and the dependent variable is not additive independent, and the relation involves the unobservable, hypothesis testing through successive differencing becomes problematic. Furthermore, investigating the likely hypothesis that equation (3) is not even additive independent in (F/H) cannot even be contemplated unless both F and H are observed and entered into the hypothesis testing directly.[13] Furthermore, empirical results later in this paper will demonstrate that the relation between (F/H) and property value is convex.

The variation over time and jurisdiction in foreclosure cost has also been omitted from previous research on foreclosure externalities. The argument for its inclusion is analogous to that for housing units except that foreclosure cost varies by jurisdiction and over time and

[13]Some papers test for and find non-linearity in the relation between foreclosures and house prices.

would enter as the product of (F/H) and an indicator of foreclosure cost.

The two points, micro variation in high CLTV and serious delinquency, and the specification not being additive independent interact in this case because the effect of high CLTV and serious delinquency should depend on the proportion of units in the sub-neighborhood that have high CLTV or serious delinquency. This results in the specification for testing the relation between foreclosure and property value given below.

$$V_{ijkt} = \alpha + \beta(F_{ijk,t-1}/H_{ijk}) + \lambda(L_{ijkt}/H_{ijk}) + \sum_{jk} \varphi_{ij} D_{jk} + \sum_{kt} \pi_{kt} D_{kt} + \epsilon_{ijkt} \tag{5}$$

Possible omitted variable bias due to correlation between the components of θ_{jt} and F, H, and/or L is then eliminated by saturating the model with time and neighborhood dummies as has been the custom in other studies.

Of course, Equation (5) assumes that β is constant while the investment effect suggests that β varies directly with the difficulty of foreclosure. If restrictions on creditors' remedies in jurisdiction k at time t can be quantified and measured by a variable, R_{kt} , the preferred specification of the test for foreclosure externalities becomes:

$$\begin{aligned} V_{ijkt} = \alpha &+ \beta_1(F_{ijk,t-1}/H_{ijk}) + \beta_2(F_{ijk,t-1}/H_{ijk})R_{kt} \\ &+ \lambda(L_{ijkt}/H_{ijk}) + \sum_{jk} \varphi_{ij} D_{jk} + \sum_{t} \pi_{t} D_{kt} + \epsilon_{ijkt} \end{aligned} \tag{6}$$

This is the general specification of the foreclosure externality equation that will be used here. The size of the pure foreclosure externality is based on estimates of β_1 and the added effect due to elevated foreclosure cost is reflected in β_2. Given that R is an estimate of foreclosure delay and cost, it is anticipated that estimates of β_2 will suffer from attenuation bias.

If Equation (6) is the appropriate specification for testing foreclosure externalities and Equation (1) has been used in the literature, the relation between estimates of β in (1) and β_1 in (6) is problematic. Negative and significant estimates of β in (1) in the literature

have no necessary implications for the sign or significance of β_1 in (6), particularly if R can be measured precisely. If $\beta_2 = 0$, the estimates of β using the specification in Equation (1) reflect an average treatment effect of increases in foreclosures. However, estimates of this average treatment effect will vary with the density of housing units being examined as well as the relation between foreclosures and housing density. In the foreclosure data used here, the foreclosure ratio is a function of the density of housing units.[14] In sum, the information content of estimates of β in Equation (1) regarding the size and significance of a pure foreclosure effect is small.

4 Data Preparation and Description

There is nothing remarkable about the data sources used in this study. The objective is to demonstrate that standard datasets used in previous studies could have been used to conduct the analysis presented here if there had been a concern for problems of additive independence and micro variation. The variables added to the standard specification of Equation (1) were all collected from or constructed using readily available data sources.

The main data source is Corelogic Data. Corelogic (previously known as "DataQuick") is one of several vendors maintaining a large database of property transactions covering substantial areas or, in this case the continental United States. Two data files, the taxation and foreclosure files were provided by Corelogic for five parts of the Washington, D.C. metropolitan area. These areas include the District of Columbia proper, and the counties of Montgomery and Prince George in Maryland, plus Arlington and Fairfax counties in Virginia. The taxation data file includes the most recent tax records (2015 and 2016) along with the most recent two transactions in the past for all residential and commercial properties in the area.

Following the majority of previous studies, this research focuses solely on single family

[14] An added complication is that when squared values of foreclosure ratio are added to Equation (6), the true relation appears to be convex.

units. The arm-length sales can be traced back to 2006, but due to limitation of the foreclosure data explained below, only sales after 2008 are included. Property characteristics (living area, baths, unit age, etc.) were extracted from the most recent tax records. Approximately 1% of sales records were excluded due to missing characteristics. Sales records were also excluded if sales prices were below the 1st or above the 99th percentile in the overall sample. After data de-duplication and cleaning, there were 114,369 single family non-distressed sales ranging from 2008 to 2015 in the final estimation sample. All sales records are geocoded by parcel address for counting nearby units. Summary statistics of these sales prices and housing characteristics are reported in the top panel of Table 1. Insofar as possible, these data were treated in a fashion similar to previous foreclosure externality studies.

The second data file, foreclosure data, provides all historical property-level foreclosure records from initial notice of default to final disposition. Foreclosure units in this study refer to title transfer sales following initial notice of default. Therefore, foreclosure units in this study mostly refer to completed foreclosure proceedings, which is consistent with Campbell et al. [4].[15] The foreclosure disposition date ranges from 2006 to 2015. The common practice in the literature is to count foreclosure units within specified time and distance intervals from recent non-distressed sales to estimate impact of foreclosure. It is generally found that foreclosures from the past 12 months have a significant negative effect on non-distressed sales. Recent studies (Gerardi et al. [15] and Annenberg and Kung [1]) show that REO sales within one half mile have significant impact on sales price. Based on these previous findings, foreclosures from the past year and the year before past year are calculated for each non-distressed sale. In addition, total nearby units within one half mile radius are also computed in order to test additive independence hypothesis laid out in previous section. The total surrounding unit computation is based on the most recent tax records and implicitly assumes units were constant in the study period. As shown in Table 1, total surrounding units vary from 6 to over 4000 within the half mile radius. Regressing

[15] Approximately 94% of these foreclosures are disposed by lenders through REO sales.

surrounding foreclosed units in the past year against total units produces a slope coefficient of 0.008 which is significant at 1% level.[16] Further examination reveals that this relation is in fact highly nonlinear. Figure 1 shows the 95% confidence intervals of all piecewise linear splines of surrounding units.[17] Five knots at 10th, 25th, 50th, 75th, and 90th percentiles are chosen to show the nonlinear relation. Based on the estimation results, there was an unusual concentration of foreclosure units in areas where housing density was lower in the foreclosure sample used here. Of course, this relation could be very different in other studies in the lierature. As argued before, a foreclosure is surely more consequential for valuation of a nearby property if it is the only unit near the home being sold than if it is one of 100 or more units nearby. The estimation results shown in Figure 1 demonstrate that the relation between local housing density and the rate of foreclosure may be complex and non-linear. In addition to foreclosed units, there are also units with initial notice of default but not yet disposed. Such units can either be brought back to current or waiting to be disposed. Because the issuance of default notice is a sign of high delinquency status, these units are termed delinquent units (DLQ units) in this study.

Units with high CLTV could have a negative effect on non-distressed sales similar to foreclosure units. High CLTV is characterized by micro variation in that it is distributed over space and time in a manner similar to foreclosure. But CLTV is not observed for unsold units. An estimate of the number of high CLTV units surrounding each observed sale was constructed using the recently available FHFA zip-level annual repeat sales housing index to estimate high CLTV units.[18] The approach to measuring high CLTV is similar to Biswas and Davidoff [2]. Housing price index series for all 161 zip codes covered in this study are shown in Figure 2. These index numbers indicate housing price appreciation since 1985,

[16]This result is obtained by including census tract FE with census tract clustered standard error. The positive relation however does not depend on inclusion of census tract FE.

[17]The estimating equation with linear splines is specified as follows: $F = \beta_0 + \beta_1 H + \beta_2 (H - H_{10th})_+ + \beta_3 (H - H_{25th})_+ + \beta_4 (H - H_{50th})_+ + \beta_5 (H - H_{75th})_+ + \beta_6 (H - H_{90th})_+ + \theta_j + \epsilon$, where F is number of foreclosure units, H is number of total nearby units, H_{10th} is the 10th percentile of total nearby units, $(H - H_{10th})_+$ is the usual linear spline defined as $(H - H_{10th}) * (H > H_{10th})$, θ_j is census tract FE.

[18]Details of this housing index can be found in Bogin et al. [3]).

the year for which the index number is normalized to 100. Zip code 22201 in Arlington has the highest appreciation rate since the base year as shown by the curve on the very top. Most zip codes experienced housing price appreciation between 2000 and 2007. The mean level housing price growth rate is 117% with a standard deviation of 19%. These same two numbers are -7% and 18% between 2007 and 2015. A value of CLTV in each year is computed as follows for each single family unit included in the taxation data set,

$$\text{CLTV}_{ij,t+\tau} = (\text{Original mortgage amount})_{ijt}/(Price_{ijt}(\text{Appreciation factor})_{j,t,t+\tau}). \quad (7)$$

In Equation (7), CLTV of unit i in zip code j at time $t+\tau$, is calculated by dividing its most recent loan amount at time t by its predicted value at time $t+\tau$. Predicted value is computed by multiplying its sales price at time t by an appreciation factor, which is derived by the zip level housing price index and measures the average price change between t and $t+\tau$ in zip code j. Using this method, the total number of high CLTV units surrounding each observed sale is estimated as units sold in the past five years from the non-distressed property sale date that have an estimated CLTV greater than 90%. Summary statistics of all types of surrounding units are reported in the bottom panel of Table 1. As expected, number of high CLTV units is much greater than number of units in foreclosure. Figure 3 compares the yearly trends of high CLTV, delinquent, and foreclosed units. High CLTV units (blue line) are much more common than delinquent and foreclosed units. The spikes of delinquent units (red line) in 2008 and 2009 are a the result of the housing crisis. Foreclosure sales (green line) remain relatively steady throughout the period. But foreclosure by delinquent year (orange line), which is defined as all subsequent foreclosure disposition after receiving default notice, follows a similar pattern as delinquent units. The difference between foreclosure sales in current year (green line) and foreclosure by delinquent year (orange line) reflects the prolonged foreclosure timeline due to the large volume of foreclosure cases before 2010. The gap between delinquency and foreclosure disposition may reflect weaking of creditors'

remedies and requirements that lenders seriously consider forbearance or loan modification. Figure 3 also shows that the decline of high CLTV units lagged behind that of delinquent and foreclosure dispositions. Foreclosure is one method through which lenders cure problematic mortgages. Following the spikes of foreclosure disposition of units defaulted in 2008 and 2009, high CLTV units gradually reduce to a much lower level from its peak. The estimated high CLTV units are highly correlated with number of surrounding foreclosures. The correlation coefficient between these two variables is 0.65 in the estimation sample. Regressing high CLTV units against foreclosure units in the past year produces a slope coefficient of 4.18, which is significant at 1% level after controlling for sale year and census tract fixed effects. These estimated high CLTV units clearly exhibit "micro variation" at the same level that foreclosures vary and subsequent tests will show that they have a negative relation to nearby sales prices.

Summary statistics relating characteristics of surrounding units as fractions of total units are also reported in Table 1. On average, foreclosure units account only for less than 1% of all nearby units. However this fraction ranges as high as 10% and it has a mass point at zero. On the other hand, surrounding high CLTV units are much more common and can be up to 80% of the surrounding stock in some areas. Table 2 also shows the effect of dropping the additive independence assumption. The distribution of the fraction of nearby units foreclosed differs drastically from a simple count of the number of foreclosures.

5 Empirical Strategy and Estimation results

Testing for additive independence

The first step in the analysis is to estimate a foreclosure externality model following procedures common in the literature. This involves estimating Equation (1) with sub-neighborhoods set at the census tract level and foreclosures within 0.5 miles of a non-distressed property sale assumed to have an additive independent relation to the logarithm of sales price. The

estimated coefficient of foreclosure count in Table 3 is -0.0037 and it is statistically significant. At the mean of the sample, 8 foreclosures per unit sold, this implies that foreclosure externalities were lowering prices by 3% which is a substantial effect. Foreclosure rates were very high during this period and while the effect of an additional foreclosure on sales price is lower than most other studies, the effect at the mean of the sample is typical.

The first modification of Equation (1) relaxes the assumption of additive independence by replacing foreclosure count with the ratio of foreclosures to housing units in the nearby area. The estimated coefficient in Table 3 column 2 is best compared to the coefficient of foreclosure count by noting the effect at the mean of the sample, which is 0.64 multiplied by the estimated coefficient -6.16 for foreclosure ratio. The implied reduction in property value at the mean of the sample is 3.94%. Most importantly when an encompassing test is performed with both the number and ratio of foreclosures to housing units are both forced into the same equation, column 3 of Table 3, the estimated coefficient of foreclosure count is highly non-significant while foreclosure ratio retains its magnitude and significance.[19] As suspected, the local fraction of foreclosed properties rather than the number is related to nearby sales prices and the effect of foreclosure is not additive independent as assumed throughout the previous literature.[20]

Testing the micro-variation hypothesis

The first two columns of Table 4 illustrate the effects of adding the ratio of the number of units with high CLTV to total housing units as an example of an omitted variable whose micro variation is similar to the foreclosure ratio. Addition of the ratio of high CLTV properties appears to have little effect on the estimated effect of the foreclosure ratio. However, given

[19] Vuong's test (Vuong [27]) for non-nested model selection also rejects the null hypothesis that models in columns 1 and 2 in Table 3 are equivalent. In the case of linear models, this likelihood ratio test comes down to a comparison between squared residuals of the two alternative models. This test is in favor of the alternative hypothesis that model in column 2 is preferred to column 1 at a P-value equal to zero up to the third decimal point.

[20] As demonstrated in Figure 1, the relation between foreclosure and housing units is highly nonlinear. The fourth column of Table 3 demonstrates that entering foreclosure and housing units as additive independent variables does not solve the specification problem.

that foreclosure units are also high CLTV units, the addition of high CLTV means that the estimate of the effects of additional foreclosures which are also high CLTV properties has increased significantly.

In columns 3 and 4, squared terms are inserted in the model. In units specification in column 3, although the foreclosure effect is shown to be convex indicating attenuation of foreclosure effect, the estimated coefficient on the squared term does not have any meaningful impact on the marginal effect. The squared term of high CLTV units is essentially zero. In column 4, when foreclosure and high CLTV are measured as fractions of total units, the convex foreclosure effects are both economically and statistically significant. The marginal effect of foreclosure evaluated at the sample mean is -6.70, and the marginal effect of high CLTV is -0.48. Both of these marginal effects represent increases from the effects at the mean of the sample shown in column 2.

Additional results

As stated in previous sections, foreclosure is a complex process and involves multiple participants' decisions at various points in time. In this section, several attempts are made to account for some of these complexities. These results help understand the empirical results here and in the previous literature. Because the foreclosure fraction specification is clearly preferred from results already shown, results based on unit specification are no longer reported. For simplicity and comparability to previous studies, squared terms are also omitted in the results shown below.

Previous research has not attempted to identify "non-distressed" sales that might be short sales. To the extent that short sales are spatially correlated with foreclosures, this could be an element of unobserved heterogeneity that directly biases estimates of the foreclosure externality. Accordingly, a short sale dummy variable, designed to indicate sales that may have been forced short sales, was added to the estimating equation. Lenders often negotiate short sales with borrowers even before the beginning of foreclosure proceedings or before

sending out default notice. But such sales are not directly observed and therefore treated the same as non-distressed sales previously. As a first attempt to measure this variable, short sale indicator is equal to one if the sale price is less than the original mortgage loan amount.[21] As shown in column 1 in Table 5, addition of this variable has no material effect on the estimates of other coefficients. But it is included because it clearly has a significant relation to sales price. The implied 14-15% discount for short sales is sizable but smaller than the usual estimates for the own price effect of a foreclosure sale.

Cordell and Lambie-Hanson [8] report substantial variation in foreclosure cost over time and across jurisdictions as the housing crisis deepened. Foreclosure cost increases with foreclosure timeline. As the institutions dealing with the foreclosure process became increasingly congested and the regulatory environment was modified to make foreclosure more difficult, foreclosure costs rose even in states allowing statutory foreclosure. If there is an investment effect of foreclosure, it should vary directly with foreclosure time and cost. Past studies of foreclosure externalities have generally only allowed dis-aggregation into judicial versus statutory foreclosure states. Precise measure of foreclosure cost (instead of judicial indicator) is taken from Appendix A in Cordell and Lambie-Hanson [8]. These foreclosure costs are plotted in Figure 4. Foreclosure costs vary across the three states in this study, although all three are generally classified as statutory foreclosure states.[22] Not surprisingly, as shown in column 2 in Table 5, most of the foreclosure effect is associated with the interaction term with foreclosure cost. The marginal foreclosure effect evaluated at the mean of foreclosure cost is -5.89. One standard deviation increase of foreclosure cost increases the foreclosure effect by 2.94, which amounts to 49% of the marginal effect.

[21] Alternatively, besides the LTV condition, the time of sale is also restricted to be less than 3 years from loan origination. The results remain very similar with this alternative definition of short sale indicator.

[22] Generally Maryland, Virginia, and the District of Columbia are classified as non-judicial foreclosure states. In Virginia and the District of Columbia this implies that the foreclose procedures can be the responsibility of the trustee if so designated in the trust. However, in Maryland, two methods of foreclosure are possible, power of sale foreclosure in which the trustee sells the property and assent to decree foreclosure where the borrower agrees to allow the court to sell the property. In either case, Maryland requires that the lender or trustee file a lawsuit in court to foreclose. The presence of either of these provisions in the mortgage greatly expedites the judicial process. Thus Maryland is sometimes classified as a statutory foreclosure state and sometimes as judicial.

Most important, the pure foreclosure externality effect, as foreclosure cost approaches zero, is reduced to less than half its former value when the foreclosure cost effect is considered. This implies that most of the foreclosure externality is due to the time that units spend in the foreclosure process rather than the process itself. It is not surprising that the investment effect is largely due to the time period when property rights in the unit are not well defined, or at least in the process of involuntary transfer.

Finally, in column 3 in Table 5, share of nearby delinquent units is added to the model. As expected, the delinquent variable is negative and significant. As argued in previous section, negative and significant estimates of β in Equation (1) in the literature have no necessary implications for the sign or significance of β_1 in Equation (6). This estimation result shows that the pure foreclosure externality effect is first reduced when foreclosure cost is considered and then becomes non-significant when the micro variation of share of delinquent units is added to the estimating equation.[23]

6 Conclusions and implications

This research demonstrates the importance of considering the assumption of additive independence and possible micro variation of other housing finance variables in testing for the relation between foreclosure and local property values. More generally, it illustrates that problems related to failure of additive independence and micro variation should be of at least as much concern as attempts to avoid problems created by selection into treatment based on unobserved heterogeneity in studies using differences in differences, and synthetic controls to identify treatment effects.

Clearly, the appropriate specification for the relation between foreclosure activity and nearby housing price is not additive independent. It is based on the ratio of foreclosures to housing units and allowance for the time to foreclosure must be made. When modeled as

[23]Gerardi et al. [15] report that the effects of their measure of nearby serious delinquency are larger than the effects of REO sales of property values. Unfortunately they do not test the hypothesis that adding delinquency lowers the estimated effect of REO sales on nearby property values.

a fraction of foreclosures the convexity of the effect of foreclosure ratio on property price is apparent. Much more consequential, the role of foreclosure delay in creating the foreclosure externality effect is substantial. This is most important because one reaction to the publication of studies showing large foreclosure externalities has been regulations that raise the cost of foreclosure, a classic case of a regulatory reaction that makes the initial problem worse.

The empirical results also confirm that the fraction of high CLTV properties has a negative relation to nearby sale prices but this does not materially influence the estimated coefficient of foreclosure ratio. Thus the traditional concern that selection into foreclosure is based on high CLTV which has a negative relation to property value does not seem warranted. However, there are major implications of the finding that high CLTV also depresses property values because foreclosure, by recapitalizing the housing unit, reduces the stock of high CLTV properties. Thus foreclosure is a treatment for the effect of high CLTV on nearby housing and without that treatment the stock of such housing would tend to increase. Accordingly, foreclosure may have a temporary negative local house price effect but it is the cure for a potentially long term problem of high CLTV properties. This demonstrates that it is easy to confuse a treatment with a disease, particularly if, in the short run, the treatment makes the patient weaker.

Addition of the ratio of seriously delinquent properties which also micro vary with foreclosure, leaves estimates of the pure foreclosure effect non-significant. This is the classic problem of an omitted variable that is correlated with both the treatment and the outcome. Taken together, the effects of the high CLTV and seriously delinquent ratio variables substantially change the interpretation of foreclosure externalities presented in previous studies.

A more general lesson to be drawn from this exercise is that invoking differences in differences to justify analysis of the effect of a treatment, even when there is no selection problem, requires strong assumptions. Most obvious is additive independence of the treatment effect, or at least caution in interpreting the estimated coefficient of treatment as a number that is applicable to any specific circumstance. Less obvious is the possibility that a more fully

specified model can reveal something about the mechanism underlying the treatment effect that is material to understanding the overall relation between the treatment and social outcomes. In the case of this study, all of the variables added to the model were easily available in most previous studies. The failure to consider the specification used here was based on the conviction that differences in differences do not deceive. Apparently that conviction is misplaced.

References

[1] Elliot Annenberg and Edward Kung. Estimates of the size and source of price declines due to nearby foreclosures. *American Economic Review*, 104(8):2527–2551, 2014.

[2] Arnab Biswas and Thomas Davidoff. How does mortgage distress affect neighboring properties' resale value? evidence from worcester, ma. *Working paper presented at the Midyear Meetings of the American Real Estate and Urban Economics Association*, 2016.

[3] Alexander N. Bogin, William M. Doerner, and William Larson. Local home price dynamics: new indices and stylized facts. *FHFA Staff Working Paper Series*, 16(1), 2016.

[4] John Y. Campbell, Stefano Giglio, and Parag Pathak. Forced sales and house prices. *American Economic Review*, 101(5):2108 – 2131, 2011.

[5] John M. Clapp and Katsiaryna Salavei Bardos. Hedonic pricing with redevelopment options: a new approach to estimating depreciation effects. *Journal Of Urban Economics*, 67(3):362–377, 2010.

[6] John M. Clapp, Katsiaryna Salavei Bardos, and S.K. Wong. Empirical estimation of the option premium for residential development. *Regional Science and Urban Economics*, 42(1-2):240–256, 2012.

[7] Terrence M. Clauretie and Nasser Daneshvary. Estimating the house foreclosure discount corrected for spatial price interdependence and endogeneity of marketing time. *Real Estate Economics*, 37(1):43 – 67, 2009.

[8] Larry Cordell and Lauren Lambie-Hanson. A cost-benefit analysis of judicial foreclosure delay and a preliminary look at new mortgage servicing rules. *Journal of Economics and Business*, 84:30 – 49, 2016.

[9] Larry Cordell, Liang Geng, Laurie S. Goodman, and Lidan Yang. The cost of foreclosure delay. *Real Estate Economics*, 43(4):916–956, 2015.

[10] N. Edward Coulson and Herman Li. Measuring the external benefits of homeownership. *Journal of Urban Economics*, 77:57 – 67, 2013.

[11] Lin Cui and Randall Walsh. Foreclosure, vacancy, and crime. *Journal of Urban Economics*, 87(2):72–84, 2015.

[12] Jihad Dagher and Yangfan Sun. Borrower protection and the supply of credit: Evidence from foreclosure laws. *Journal of Financial Economics*, 121(1):195 – 209, 2016.

[13] Lynn M. Fisher, Lauren Lambie-Hanson, and Paul Willen. The role of proximity in foreclosure externalities: Evidence from condominiums. *American Economics Journal: Economic Policy*, 7(1):119 – 140, 2015.

[14] W. Scott Frame. Estimating the effects of mortgage foreclosures on nearby property values: a critical review of the literature. *Federal Reserve of Atlanta Economic Review*, November:1 – 9, 2010.

[15] Kristopher Gerardi, Eric Rosenblatt, Paul S. Willen, and Vincent Yao. Foreclosure externalities: new evidence. *Journal of Urban Economics*, 87:42 – 56, 2015.

[16] Jeremy R. Groves and William H. Rogers. Effectiveness of rca institutions to limit local externalities: using foreclosure data to test covenant effectiveness? *Land Economics*, 87(4):559 – 581, 2011.

[17] John P. Harding, Eric Rosenblatt, and Vincent W. Yao. The contagion effect of foreclosed properties. *Journal of Urban Economics*, 66:164 – 178, 2009.

[18] David Hartley. The effect of foreclosures on nearby housing prices: Supply or disamenity? *Regional Science and Urban Economics*, 49:108 – 177, 2014.

[19] Dan Immergluck and Geoff Smith. The external costs of foreclosure: The impact of single-family mortgage foreclosures on property values. *Housing Policy Debate*, 17(1): 57 – 70, 2006.

[20] Lauren Lambie-Hansen. When does delinquency result in neglect? mortgage distress and property maintenance. *Journal of Urban Economics*, 90:1 – 16, 2015.

[21] Atif Mian, Amir Sufi, and Francesco Trebbi. Foreclosures, house prices, and the real economy. *The Journal of Finance*, 70(6):2587–2612, 2015.

[22] Anthony Pennington-Cross. The value of foreclosed property. *Journal of Real Estate Research*, 28(2):193 – 214, 2006.

[23] Kristin L. Perkins, Michael J. Lear, and Elyzabeth Gaumer. Concentrated foreclosure activity and distressed properties in new york city. *Urban Affairs Review*, pages 1–30, 2016.

[24] Sherwin Rosen. Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of Political Economy*, 82(1):34–55, 1974.

[25] Jenny Schuetz, Breen Vicki, and Ingrid Gould. Neighborhood effects of concentrated mortgage foreclosures. *Journal of Housing Economics*, 17(4):306 – 319, 2008.

[26] Albert Sumell. The determinants of foreclosed property values: Evidence from inner-city cleveland. *Journal of Housing Economics*, 18(2):45 – 61, 2009.

[27] Quang H. Vuong. Likelihood ratio tests for model selection and non-nested hypothesis. *Econometrica*, 57(2):307–333, 1989.

[28] Stephan Whitaker and Thomas J. Fitzpatrick. Deconstructing distressed-property spillovers: The effects of vacant, tax delinquent, and foreclosed properties in housing submarkets. *Journal of Housing Economics*, 22(2):79 – 91, 2013.

[29] Lei Zhang and Tammy Leonard. Neighborhood impact of foreclosure: a quantile regression approach. *Regional Science and Urban Economics*, 48:133 – 143, 2014.

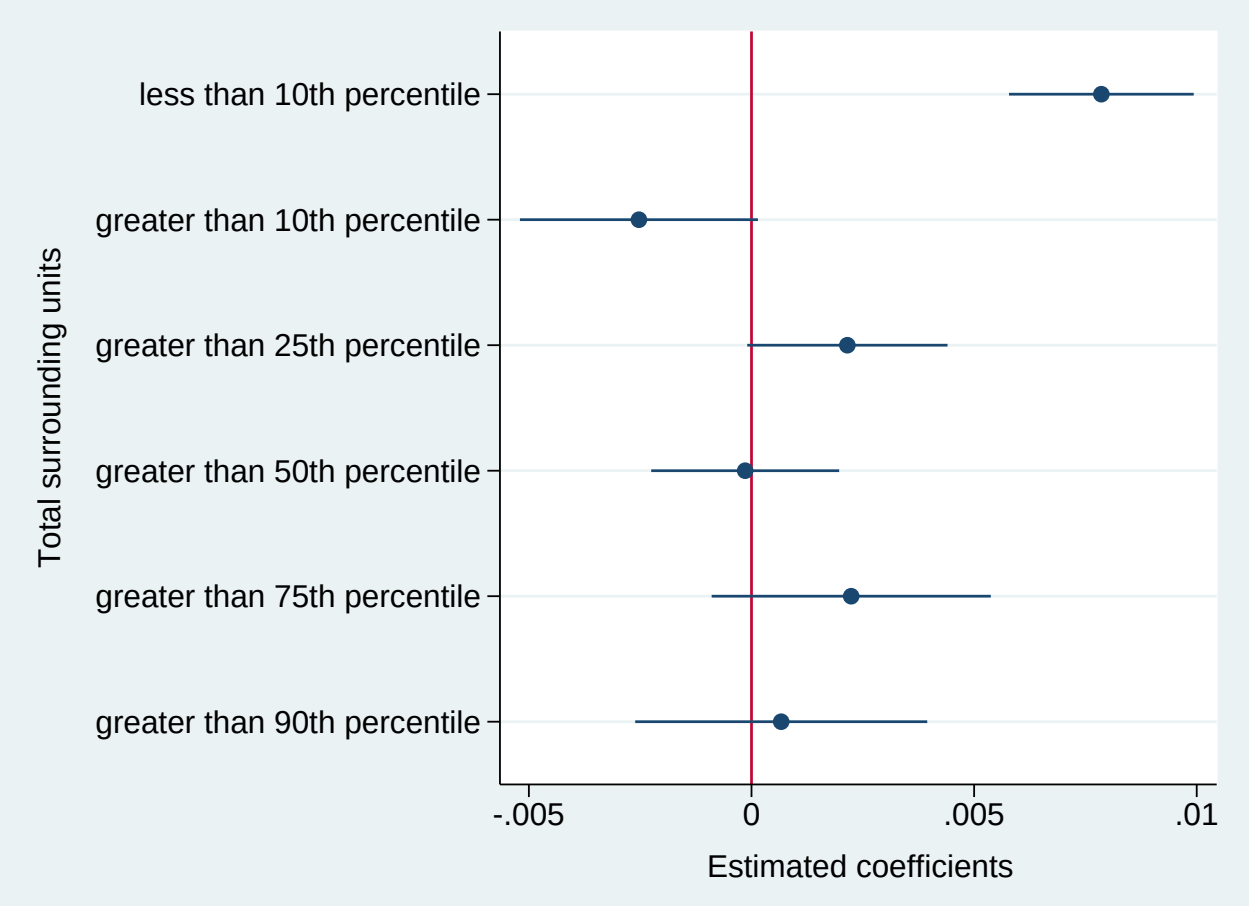

Figure 1: Nonlinear Relation between Foreclosed Units and Total Nearby Units

Note: This figure shows the estimated coefficients and 95% confidence intervals of all surrounding units. The estimating equation with linear splines is specified as follows: $F = \beta_0 + \beta_1 H + \beta_2(H - H_{10th})_+ + \beta_3(H - H_{25th})_+ + \beta_4(H - H_{50th})_+ + \beta_5(H - H_{75th})_+ + \beta_6(H - H_{90th})_+ + \theta_j + \epsilon$, where F is number of foreclosure units, H is number of total nearby units, H_{10th} is the 10th percentile of total nearby units, $(H - H_{10th})_+$ is the usual linear spline defined as $(H - H_{10th}) * (H > H_{10th})$, θ_j is census tract FE.

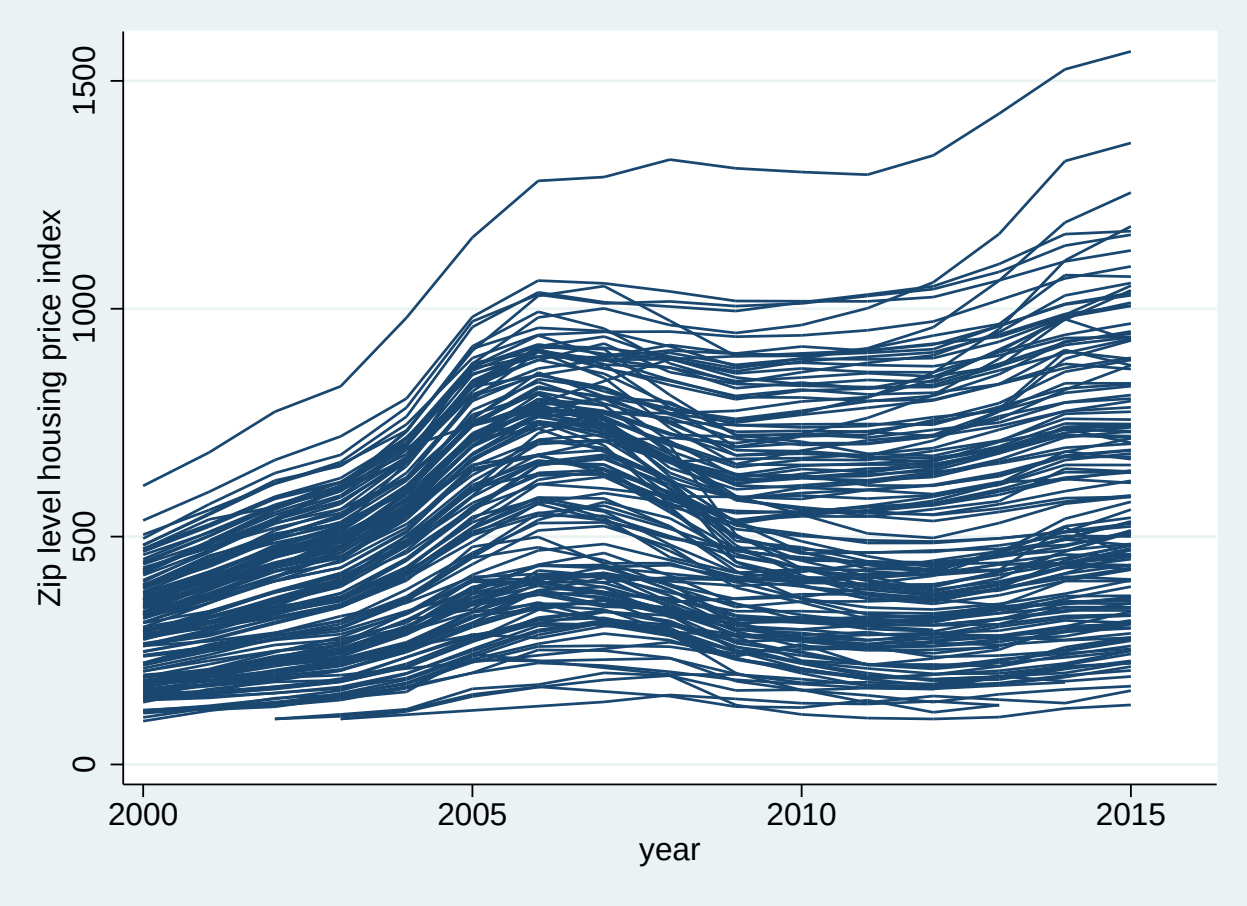

Figure 2: FHFA Zip Code Level Repeat Sale Housing Price Index

Note: This graph shows zip level annual hpi used for producing CLTV for all 161 zip codes used in this study. This index is normalized to 100 in the base year 1985.

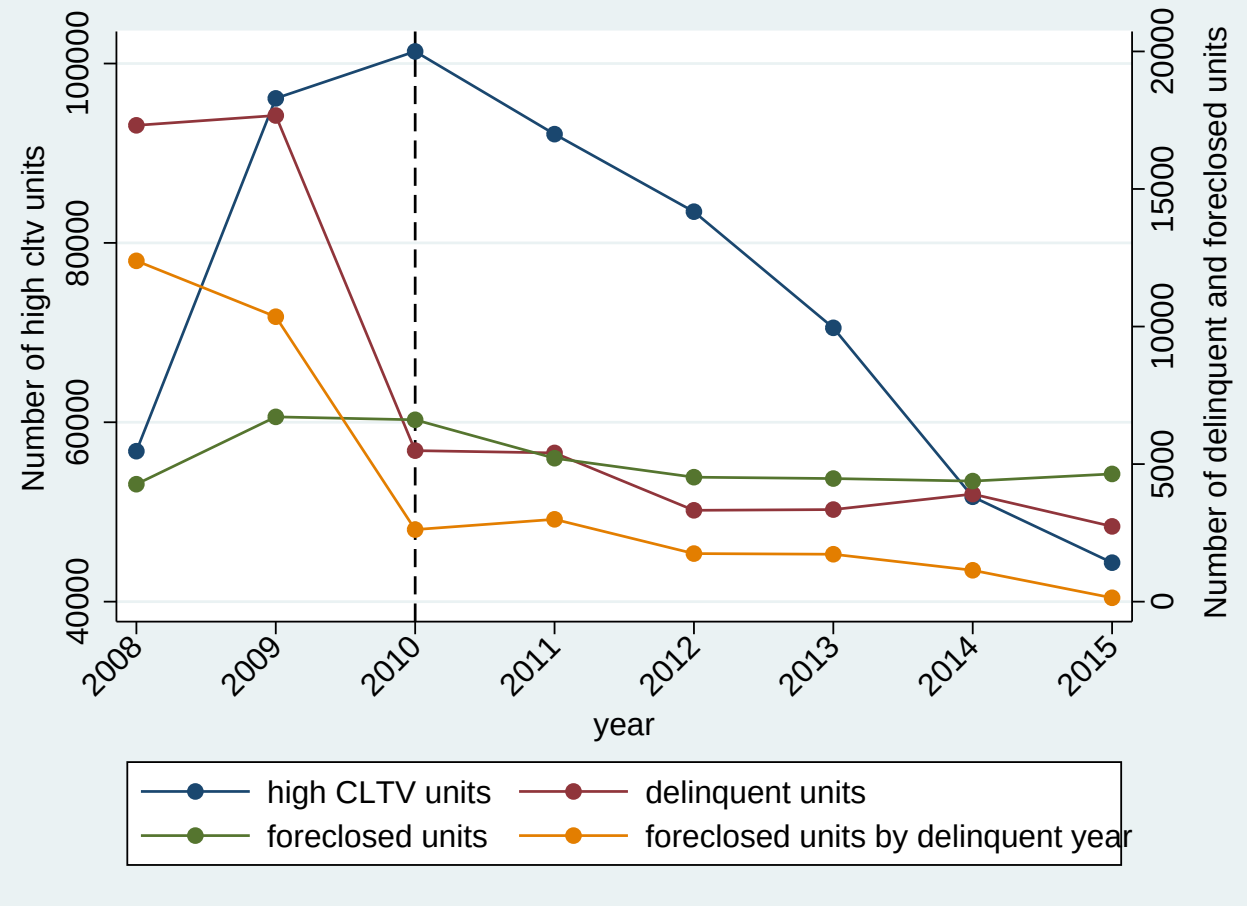

Figure 3: High CLTV, Delinquent, and Foreclosed Units by Year

Note: This figure shows three types of units: (1) "high CLTV units" are derived based on Equation (7) explained in the text, (2) "delinquent units" are properties received notice of default in that year, (3) "foreclosed units" refer to completed foreclosure proceedings in that year, (4) "foreclosed units by delinquent year" refer to all units disposed through foreclosure proceedings after receiving notice of default in that year.

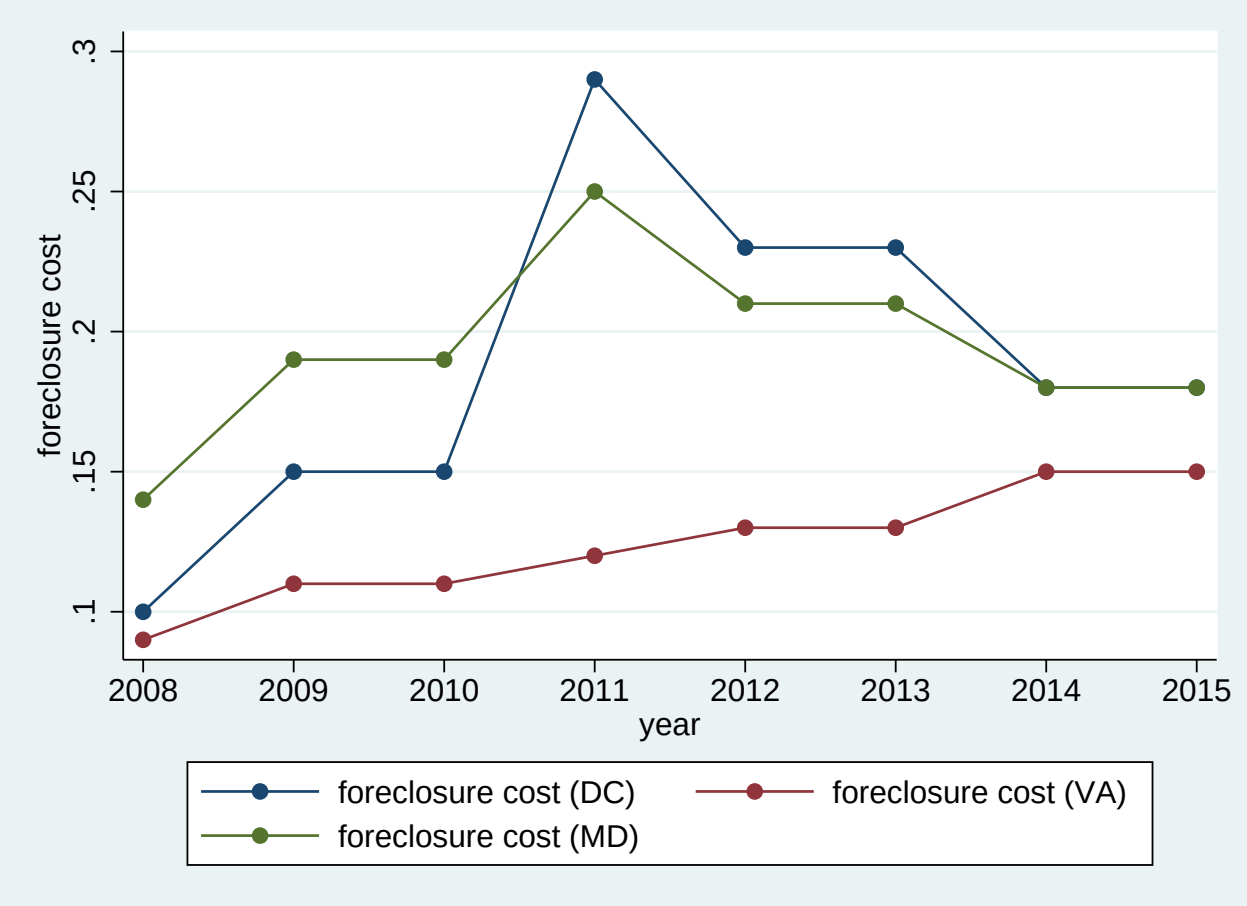

Figure 4: Foreclosure Cost by State and Year

Note: Foreclosure costs are taken from Appendix A in Cordell and Lambie-Hanson [8].

Table 1: Summary Statistics

	Mean	Std.dev.	Min	Max	Observation
Panel 1: Single family non-distressed sales unit information					
Single family non-distressed sales price	462399.30	265641.90	75000.00	2395000.00	114369
Unit age	43	27	1	133	114369
Lot size (square foot)	4436.72	2852.70	1117.00	9998.00	114369
Living area (square foot)	1761.57	681.05	832.00	6114.00	114369
Number of bath rooms	3.05	1.01	1	5	114369
Fireplace indicator	0.65	0.48	0	1	114369
Panel 2: unit counts within one half mile radius					
All surrounding units	1260.16	569.24	6	4215	114369
Foreclosure units in the past year	8.34	9.87	0	178	114369
Foreclosure units in the year before last year	7.03	9.27	0	173	114369
Delinquent units in the past year	11.80	18.18	0	235	114369
High CLTV units (at sale date)	119.91	97.24	0	866	114369
Fraction of foreclosure units in the past year	0.64%	0.65%	0	10.06%	114369
Fraction of foreclosure units in the year before past year	0.54%	0.62%	0	9.86%	114369
Fraction of denquent units in the past year	0.92%	1.21%	0	20.27%	114369
Fraction of high CLTV units (at sale date)	9.84%	7.29%	0	81.70%	114369

Table 2: Distribution of Fraction of Nearby Foreclosures

Number of nearby foreclosures	Mean	Std.dev.	Min	Max	Observation
1	0.12%	0.08%	0.04%	2.56%	13556
2	0.22%	0.13%	0.05%	2.20%	12486
3	0.32%	0.16%	0.07%	3.30%	10837
4	0.41%	0.19%	0.10%	3.08%	9348
5	0.51%	0.23%	0.12%	4.63%	8343
6	0.61%	0.27%	0.15%	3.14%	7359

Note: The summary statistics show the distributions of fraction of nearby foreclosures by the number of nearby foreclosures. The results only include nearby foreclosures up to 6, which account for over half of the non-distressed sales with nonzero nearby foreclosures.

Table 3: Model Selection: Testing Additive Independence Assumption

Dependent variable:	log of sale price			
	(1)	(2)	(3)	(4)
Number of foreclosures in the past year	-0.0037		0.0000	-0.0043
	(0.0002)***		(0.0004)	(0.0003)***
Share of foreclosures in the past year		-6.1590	-6.1143	
		(0.2946)***	(0.6440)***	
Number of total nearby units				0.00005
				(0.0000)***
Log of lot size	0.1442	0.1444	0.1444	0.1445
	(0.0038)***	(0.0038)***	(0.0038)***	(0.0038)***
Log of living area	0.3380	0.3373	0.3373	0.3380
	(0.0078)***	(0.0077)***	(0.0077)***	(0.0077)***
Number of baths	0.0484	0.0484	0.0484	0.0484
	(0.0019)***	(0.0019)***	(0.0019)***	(0.0018)***
Fireplace indicator	0.0534	0.0530	0.0530	0.0528
	(0.0030)***	(0.0030)***	(0.0030)***	(0.0030)***
Unit age	-0.0114	-0.0114	-0.0114	-0.0115
	(0.0005)***	(0.0005)***	(0.0005)***	(0.0005)***
Unit age squared	0.0001	0.0001	0.0001	0.0001
	(0.0000)***	(0.0000)***	(0.0000)***	(0.0000)***
Constant	9.2674	9.2848	9.2848	9.2135
	(0.0529)***	(0.0526)***	(0.0526)***	(0.0542)***
Sale year and jurisdiction FE	Y	Y	Y	Y
Census tract FE	Y	Y	Y	Y
R^2	0.868	0.868	0.868	0.868
Observation	114369	114369	114369	114369

Table 4: Testing for Micro Variation Hypothesis

Dependent variable:	log of sale price			
	(1)	(2)	(3)	(4)
Number of foreclosures in the past year	-0.0034		-0.0042	
	(0.0003)***		(0.0004)***	
Number of foreclosures squared			0.00001	
			(0.0000)***	
Share of foreclosures in the past year		-5.8431		-6.8780
		(0.3302)***		(0.4288)***
Share of foreclosures squared				27.7809
				(5.6355)***
Nunber of high CLTV units	-0.0001		-0.0001	
	(0.0000)**		(0.0001)	
Number of high CLTV units squared			0.0000	
			(0.0000)	
Share of high CLTV units		-0.1251		-0.5764
		(0.0554)**		(0.0756)***
Share of high CLTV units squared				0.9933
				(0.1301)***
Constant	9.3481	9.3038	9.3532	9.3546
	(0.0531)***	(0.0523)***	(0.0529)***	(0.0522)***
Control variables	Y	Y	Y	Y
Sale year and jurisdiction FE	Y	Y	Y	Y
Census tract FE	Y	Y	Y	Y
R^2	0.868	0.868	0.868	0.869
Observation	114369	114369	114369	114369

Note: The same set of control variables is included here. Since the estimated coefficients are so similar to that shown in Table 3, they are not repeated here.

Table 5: Foreclosure Effect by Foreclosure Cost

Dependent variable:	log of sale price		
	(1)	(2)	(3)
Share of foreclosures in the past year	-5.5927	-1.6663	-0.5423
	(03097)***	(0.6820)***	(0.7586)
Share of foreclosures*foreclosure cost		-25.0576	-30.8290
		(4.3647)***	(4.6854)***
Share of high CLTV units	-0.1085	-0.1269	-0.1050
	(0.0542)**	(0.0549)**	(0.0545)***
Short sale indicator	-0.1412	-0.1410	-0.1410
	(0.0056)***	(0.0056)***	(0.0056)***
Share of DLQ units in the past year			-0.5946
			(0.2063)***
Constant	9.3714	9.3749	9.3830
	(0.0528)***	(0.0527)***	(0.0527)***
Control variables	Y	Y	Y
Sale year and jurisdiction FE	Y	Y	Y
Census tract FE	Y	Y	Y
R^2	0.876	0.876	0.876
Observation	114369	114369	114369