Foreclosure externalities: Some new evidence *

Kristopher Gerardi
FRB of Atlanta

Eric Rosenblatt
Fannie Mae

Paul S. Willen
Boston Fed and NBER

Vincent W. Yao
Fannie Mae

February 27, 2012

DRAFT

Abstract: A recent set of influential papers has argued that residential mortgage foreclosures reduce the sale prices of nearby properties. This paper revisits this issue using a more robust identification strategy combined with new data that contain information on the location of properties secured by seriously delinquent mortgages and information on the condition of foreclosed properties. In our baseline specification, we find that properties within 1/16 of a mile of (1) a seriously delinquent property, (2) a bank-owned property, (3) a property sold by the bank in the last year and (4) a property sold by the bank more than a year ago sell at 2.8%, 3.3%, 2.4% and -0.2% discounts respectively. In other words, the measured effect of foreclosures on prices appears long before the bank forecloses and ends about a year after the banks sells. The estimates are very sensitive to the condition of the distressed property with a positive correlation existing between house price growth and foreclosed properties identified as in "above average" condition. We argue that the most plausible explanation for these results is an externality resulting from reduced investment by owners of distressed property. Our analysis shows that policies that slow the transition from delinquency to foreclosure exacerbate the negative effect of mortgage distress on house prices.

JEL Classifications: G21, K11, R31
Keywords: foreclosure, mortgage, judicial, power of sale, right to cure

---

*Thanks to Lauren Lambie-Hanson and Chris Foote for helpful conversations. The views expressed in this paper are those of the authors and not the official position of Fannie Mae or any part of the Federal Reserve System. Contact information respectively: Kristopher.Gerardi@atl.frb.org; eric_rosenblatt@fanniemae.com; Paul.Willen@bos.frb.org; vincent_w_yao@fanniemae.com.

1 Introduction

Many of the policy responses to the worst housing bust in the United States since the Great Depression have been rationalized by the belief that residential foreclosures generate significant social costs in the form of negative externalities on neighboring properties and on municipalities more broadly. One particular externality that has been the focus of attention for both researchers and policymakers is the effect that foreclosures exert on the market value of non-distressed properties. Recent empirical research on this topic has found that foreclosed properties seem to have large, negative impacts on the sale prices of nearby properties. Given the sheer magnitude of foreclosures that currently characterize the U.S. housing market, many market observers are very concerned that they could prevent a housing market recovery from materializing, and in the worst case, possibly cause another significant decline in housing prices going forward. Thus, policies like the Obama Administration's Home Affordable Mortgage Program (HAMP) have been enacted that attempt to facilitate loan modifications as alternatives to foreclosure. In addition, various policies have been enacted that explicitly delay the foreclosure process to provide borrowers with more time to find ways to avoid foreclosure.[1]

While a number of papers in the literature have presented empirical evidence linking the presence of foreclosures to the market values of neighboring properties, the so-called contagion effect of foreclosures, few studies have attempted to distinguish between the various channels through which such an effect might occur. This is a significant hole in the literature as the ap-

[1]One example is the various right-to-cure policies that have been enacted at the state-level that force lenders to wait a specified number of days before initiating foreclosure proceedings on delinquent mortgage borrowers. Another example is the various foreclosure moratoria that have been imposed at the state-level throughout the recent housing crisis. California imposed a 90-day moratorium that went into effect on June 15, 2009. South Carolina (through decisions by the state supreme court) has enacted two foreclosure moratoria – one in 2009 and another in 2011.

propriate policy responses are likely to differ dramatically depending on the particular channel through which this effect operates. For example, if the relationship is driven by a supply effect whereby a glut of foreclosures that come onto the market exerts competitive pressure on nearby non-distressed properties currently on the market, then a sensible policy prescription to stabilize housing prices might be to either prevent the initiation of the foreclosure process and engage in efforts to modify delinquent mortgages, or to drag out the foreclosure process in order to prevent an overabundance of properties from coming onto the market at the same time. Alternatively, if the negative relationship is due to a disamenity effect whereby distressed properties are not properly maintained and have the effect of decreasing the demand for house purchases in the surrounding neighborhood, then a sensible policy might be to shorten the foreclosure process in order to decrease the time that it takes to find a new homeowner to live in the house who has the financial means and incentives to properly invest in the maintenance of the property. Finally, if the causal relationship that previous studies have claimed to identify does not actually run from foreclosures to prices, but rather from prices to foreclosures, then there is no externality to address, and a sensible policy might be to simply ensure that there exists a proper safety net for households that are evicted from their homes through the foreclosure process.

This paper begins to fill this hole in the literature using new data and a more robust empirical identification strategy. Unlike previous studies, our data contain information on the location of properties at various stages of distress, from minor delinquency all the way through the foreclosure process to lender ownership and sale to a new homeowner. In addition, for a subset of the sample, the data include information about the condition of foreclosed property. This information, along with the empirical identification strategy allows us to significantly narrow the interpretation of the contagion effect of foreclosure. We argue that the most plausible explanation for the contagion effect is an externality resulting from reduced investment by owners of

DRAFT

distressed property.

The existing literature has typically estimated some variant of the following regression

$$\log(P_{it}) = \alpha + \beta X_{it} + \gamma NF_{it} + \varepsilon_{it} \quad (1)$$

where P_{it} is the sale price of property i in period t, X_{it} is a vector of controls, and NF_{it} is a measure of the number of properties that experience some type of foreclosure event within a certain distance of property i in some window around period t. There are substantial differences in the types of foreclosure events, the distances, and the time windows that previous papers have focused on, but in general, they have all found negative estimates for γ, the conditional correlation between the sale price of a non-distressed property and the number of nearby foreclosures.

This paper also estimates a variant of equation (1), but with some important differences from the previous literature. The first difference is in the measurement of NF_{it}. Whereas previous analyses have focused on a *flow* measure of foreclosed properties, this analysis employs multiple measures of the *stock* of distressed properties. For example, whereas Immergluck and Smith (2006) and Campbell, Giglio, and Pathak (2011) measure the number of *transitions* of property from serious delinquency into lender ownership, we focus on the number of *outstanding* minor delinquencies, the number of *outstanding* seriously delinquent properties (SDQs), the number of *outstanding* lender-owned properties ("real estate owned," or REO), and the number of *outstanding* properties recently sold by the lender to arms-length buyers. This approach partly reflects the opportunities provided by the data. Whereas previous studies have used public records data which, almost by definition, identify transitions and not stocks, our data include both information from public records and more detailed information on mortgage performance from what we call the "proprietary data" provided by a national mortgage insurer. For every mortgage in the proprietary dataset, we know the street address of the property, the payment history of the mortgage, the exact timing of the foreclosure auction, and the sale of the property to an arms-length buyer.

But the second, and more important reason to focus on stocks and not flows is that for many of the theories of why foreclosures might affect prices, the inventory is what matters and not the flow. For example, many have argued that borrowers facing foreclosure have little reason to invest in their properties, which could generate negative externalities in the neighborhood, and depress nearby home values. But, the approaches used in the previous literature only roughly approximates the number of nearby properties in distress at the time of the sale. For example, counting foreclosure process initiations over the last 18 months prior to a sale (as in Schuetz, Been, and Ellen (2008)), only works if foreclosure timelines do not differ substantially over time or across jurisdictions. If, for example, foreclosure timelines before the crisis rarely exceeded 18 months and after the crisis almost always do then the Schuetz, Been, and Ellen (2008) measure will systematically understate the growth in the stock of distressed properties.

As we discuss in the Conclusion, our focus on the stock or inventory is important for policy reasons. If one interprets equation (1) causally, then flow measures can lead to erroneous inference. For example, suppose that all distressed properties exert downward pressure on prices due to investment externalities, but that equation (1) is estimated using only transitions into foreclosure. Because foreclosure transitions are highly correlated with the number of outstanding distressed properties, one would find a significant, negative correlation between the sale price of a non-distressed property and the number of surrounding properties transitioning into foreclosure. Based on such results, one might conclude that implementing a foreclosure moratorium would increase house prices. However, such a conclusion would be wrong. Delaying transitions into foreclosure does not reduce the total number of distressed properties, which is what exerts downward pressure on prices according to the true model. Indeed, over time, delaying foreclosures with-

DRAFT

out stopping transitions into delinquency would increase the total number of distressed properties and thus lower prices.

Consistent with such a theory, we find that properties in all stages of distress exert downward pressure on nearby home values. Estimating a variant of equation (1) we find estimates of γ that are smallest in absolute value for the number of nearby minor delinquencies and larger for the number of properties with more serious delinquencies. Our estimate of γ peaks in absolute value when the lender owns the property, then falls after the sale out of REO to an arms-length buyer, and finally reaches zero approximately one year after the REO sale.

The second innovation, which is discussed in more detail in Section 2, is the manner in which we attempt to control for unobserved heterogeneity across properties. Unobserved heterogeneity is a serious issue in this context, as it is well known that foreclosures are generated by falling house prices, so that any unobserved factor that causes a decrease in house prices and thus an increase in foreclosures will lead to simultaneity bias and erroneous inference. To deal with this issue, we estimate a version of equation (1) that controls for previous sales of the same property *and* contains a set of highly geographically disaggregated fixed effects (at the census block-group level). Thus, our estimates of γ in equation (1) reflect differences in price growth across properties bought in the same year, and both sold in the same year, say 2009, within the same census block-group (CBG). We argue that this identification strategy is largely immune to issues of reverse causality and simultaneity bias.

The final major innovation in the analysis is the fact that the dataset includes information on whether a seriously delinquent property is vacant and on the condition of lender-owned properties. We find that the estimate of γ in equation (1) is more negative for both vacant properties and lender-owned properties in "below average" or "average condition," while the estimate for lender-owned properties in "above average" condition is significantly positive.

DRAFT

In Section 4 we provide an interpretation of these results. We evaluate three possible explanations: 1.) unobserved relative demand shocks that drive prices down and result in some foreclosures; 2.) foreclosures generating increased relative supply and driving prices down; 3.) an externality of reduced investment by distressed borrowers in the delinquency phase and financial institutions in the lender-ownership phase. Given the data and the limited theory, it is impossible to establish anything conclusively. However, we argue that the weight of the evidence points to the third explanation. Both of the first two explanations require that there be distinct within-CBG micro-markets not generated by the externality from the foreclosures themselves. Given the small size of CBGs, this seems unlikely. In addition, the evidence from the regressions that incorporate information on the condition of foreclosed properties is inconsistent with the supply explanation: a reasonable hypothesis is that the above average properties should generate more competition for non-distressed sales than the properties in poor condition.

The paper proceeds as follows. Section 2 contains a discussion of the empirical approach with an extensive discussion of both the empirical model and the data. Section 2 includes an extensive discussion of the existing literature and how this paper fits into it. In Section 3, we report the regression results and discuss potential interpretations in Section 4. Finally, the conclusion contains a discussion of the policy implications of the analysis.

DRAFT

2 Empirical Methodology and Data

The main focus of this paper centers around the estimation of a regression specification that considers a sample of properties $i \in I$, located in geography $g \in G$, all sold at time T and purchased at various times t in the past. The baseline specification, shown below, is a regression of individual property price growth between periods t and T on the number of nearby distressed properties at time T and a set of controls:

$$\log(P_{igT}/P_{igt}) = \alpha_{gt} + \beta X_{iT} + \sum_{d \in D} \gamma_d N_{iT}^d + \epsilon_i \tag{2}$$

In equation (2) P_{igT}/P_{igt} is price growth from purchase to sale, α_{gt} is a full set of location×time period fixed effects, and X_{iT} is a vector of hedonic controls measured at time T. The variables of interest are counts of distressed property of type $d \in D$ around property i at the time of sale T. Note that d's can differ in both the type of distress, REO versus serious delinquency (SDQ) for example, as well as in the distance from the sale of property i. An example of d is the number of properties in REO inventory between 330 and 660 feet from the sale of property i at time T.

We draw the reader's attention to some important differences between our basic specification shown in equation (2) and the general regression specification employed by the previous literature shown in equation (1). The first difference is that we use the repeat-sales approach to control for time-invariant heterogeneity across properties. Much of the previous literature has estimated hedonic models in which the dependent variable is the logarithm of the sale price at time T, and the set of control variables includes characteristics of property i at time T. The advantage of using the repeat-sales approach is that while a hedonic model usually controls for only the characteristics of the property that are contained in the tax assessor's data base, the repeat-sales model in some sense controls for the previous sale price, which in principle captures a lot of relevant information, including everything from period detail to water views and southern exposures. It is important to stress that the repeat-sales model only addresses time-invariant characteristics of the property, and thus cannot help with reverse causality, as foreclosures are, by nature, time-varying. That said, hedonic models as they are typically implemented only use current information about the characteristics of

DRAFT

the property, and thus only control for time-invariant factors as well.[2]

The second important distinction from the previous literature is the inclusion of location×time period fixed effects, α_{gt}. Only a few previous papers in the literature have included any type of geographic controls, and none have included geographic fixed effects measured at the level of disaggregation used in this analysis, the CBG (for example, Campbell, Giglio, and Pathak (2011) includes a set of census tract fixed effects). Figure 1 provides an example of the breakdown of census tracts, census block-groups, and census blocks for the city of Cambridge, Massachusetts, which is a small city of about 100,000 people, which neighbors the city of Boston to the north. Cambridge is made up of 32 census tracts, and each census tract typically includes 2 or 3 census block-groups.[3] Thus, a typical CBG is a very small geographic area, is composed of a relatively homogeneous housing stock, and a relatively homogeneous population with respect to ethnic and economic characteristics. Thus, as we argue below, a CBG is likely smaller in geographical terms than what we typically think of as a local housing market.

The combination of the repeat-sales model with the CBG×purchase year fixed effects means that we are identifying γ in equation 2 using variation in the price appreciation of properties in the same CBG that were bought in the same year and sold in the same year. In other words, to explain a significant coefficient estimate associated with the presence of nearby distressed property, one must come up with a story about why properties within the same CBG with a large concentration of nearby distressed property, appreciated differently from properties elsewhere in the same CBG over the same time interval with a smaller concentration of nearby distressed property. For example, the fact that properties on the main street in a given CBG are in

[2]For example, the dataset used by Campbell, Giglio, and Pathak (2011) only contains information from the most recent assessor's files.

[3]There are over 200,000 CBGs in the United States, with each group generally containing between 600 and 3,000 people. They are subsets of census tracts, which contain between 1,500 and 8,000 people.

DRAFT

higher demand and thus more valuable than properties off the main street with less nearby foreclosures in the same CBG, would not generate a significant negative estimate of γ. The difference in demand stemming from this within-CBG location difference is present in both times t and T, and thus is subsumed by expressing the dependent variable as the difference in sale prices. Rather, one would need to tell a story about why prices *fell* between times t and T in one area of a CBG relative to another area in the same CBG.

Using within-CBG variation to identify the effect of nearby distressed property on price appreciation also significantly alleviates concerns about reverse causality. Figure 2 provides an example of how using variation across geographies could cause one to mistakenly conclude that there was a causal effect of nearby foreclosures on prices when the true causal effect actually goes from prices to foreclosures. In the example, we assume that we have data on foreclosures and prices in two separate geographic areas, tract A and tract B, as shown in the top panel. We assume that price appreciation is constant for all properties within the same geography, but that price appreciation is significantly lower in tract B compared to tract A. Within each tract, foreclosures are randomly located so there is, by construction, no causal effect of nearby foreclosures on prices. Thus, separate within-tract regressions of sale prices on the number of foreclosures within some radius would correctly yield a γ of zero. In the bottom panel of Figure 2 we look at the same data but ignore the geographic differences. A regression of sale prices on the number of nearby foreclosures now incorrectly yields a negative relationship between price growth and foreclosures. This simple example illustrates how reverse causality can conflate identification and in the analysis below, we show that the estimates of γ in equation (2) are indeed quite sensitive to the inclusion of and the aggregation level of geographic fixed-effects.

In addition to the these important differences from the previous literature, there are a few other aspects of our baseline specification displayed

DRAFT

in equation (2) that need to be discussed. Like Harding, Rosenblatt, and Yao (2009), in our baseline specification, the various measures of nearby distressed properties are defined as the *difference* in the number of properties over the repeat-sale period (the interval between time t and time T). For example, if we consider a repeat sale of a property purchased on July 21, 2004 and sold in April 3, 2009, one measure of N^d_{iT} would be the difference in the REO inventory within 1/16 of a mile of the property on those two dates. This is important because it is not uncommon for a single property to go through foreclosure multiple times over the span of a few years and we would underestimate the impact of nearby distressed properties on non-distressed prices if we failed to take that into account.

Finally, in all of the regressions we adopt a parsimonious approach and use the unweighted number of each type of distressed property within a given radius of the non-distressed repeat-sale. This assumption is common in much of the previous literature with the exception of Campbell, Giglio, and Pathak (2011), who use a couple of different weighting schemes.[4]

2.1 Comparison with earlier work

As we mentioned in the introduction, this paper builds on an extensive existing literature on foreclosure externalities. Narrowly, there are a series of papers that estimate equation (1) starting with Immergluck and Smith (2006) and including: Schuetz, Been, and Ellen (2008), Rogers and Winter (2009), Harding, Rosenblatt, and Yao (2009), Lin, Rosenblatt, and Yao (2009) and Campbell, Giglio, and Pathak (2011). All of these papers use flow measures of foreclosure-related distress as the right-hand-side variables of interest. More

[4]See the discussion on page 2125 in Campbell, Giglio, and Pathak (2011) for a detailed description of their weighting scheme. However, the authors show that their results are largely unchanged by using an unweighted approach. Furthermore, since theory does not provide much guidance on the appropriate weighting scheme to use and since our distance measures are approximate, we are concerned that any inference drawn from a complicated weighting scheme may be potentially misleading, and thus choose to estimate unweighted regressions.

broadly, a much older literature has estimated almost exactly the same hedonic regressions, but with other events not related to foreclosure that might affect local house values. We begin this section with a detailed discussion of the recent literature, as it is more related to our current analysis, and then provide a brief discussion of the older literature.

Although previous studies have used the repeat-sales specification and controlled for geography at a relatively disaggregated level, no analysis has done both at the same time. Harding, Rosenblatt, and Yao (2011), to our knowledge, is the only paper to estimate equation (1) using a repeat-sales specification. They estimate separate regressions by metropolitan statistical area (MSA) but do not control for geography within the MSA. This effectively means that they are comparing price growth and nearby foreclosures for non-distressed repeat-sales across entire MSAs, and thus their estimates are prone to the same identification issues that we discussed above in the context of Figure 2. There are strong within-MSA patterns in price growth with much sharper price declines in poorer neighborhoods and locations further from the city center, and our results below confirm that more disaggregated geographic controls generate a major reduction in the estimate of γ. Campbell, Giglio, and Pathak (2011), to our knowledge, is the only study in the literature that includes a disaggregated set of geographic controls. Campbell et al. uses a hedonic model and includes census-tract$\times$year controls, which as we saw from Figure 1 are slightly more aggregated than the CBG controls that we employ in this paper. However, as we show below, the difference in estimates of γ using CBG versus census tract controls is small, which suggests that the census tract is a sufficiently small geography to eliminate the influence of unobserved heterogeneity in the estimation of equation (1). The other studies mentioned above all use hedonic models with either no disaggregated geographic controls or fairly broad ones.

As we noted above in the introduction, the most important difference between our specification and all previous work is the use of stock measures

DRAFT

of distressed property rather than flows. For example, Immergluck and Smith (2006) count foreclosure deeds in the two years prior to the sale of non-distressed property, Schuetz, Been, and Ellen (2008) count the number of foreclosure initiations, known as *lis pendens* filings in New York, in the 18 months prior to the sale, and Harding, Rosenblatt, and Yao (2009) construct a series of measures of foreclosure deeds in 3 month intervals before and after the sale. In contrast, we look at, for example, the number of properties in REO at the time of the sale and the number of properties in serious delinquency at the time of the sale. To the extent that what we care about is the number of distressed properties nearby at the time of the sale, the other methods implicitly assume that the distressed properties have a hazard of leaving distress of exactly one at some point.

To see the difference, consider the baseline specification in Campbell, Giglio, and Pathak (2011) in which the authors count all properties for which foreclosure proceedings are completed (i.e. a foreclosure auction takes place) in the year prior to the sale of a nearby non-distressed property. Effectively, they assume that a property that was foreclosed on more than one year in the past plays absolutely no role whatsoever in the pricing of a nearby property. One might argue that exactly the opposite is true: the properties that produce the most blight, and which may be most likely to adversely impact surrounding values are the properties that lenders cannot sell. To make matters worse, the potential bias induced by measuring flows instead of stocks is likely not constant over time or across locations.[5] Foreclosure timelines differ widely across states and have slowed considerably through the recent boom/bust cycle, especially in states that require judicial review.[6]

[5] In Table A.12 of the Internet Appendix, Campbell, Giglio, and Pathak (2011) do employ an alternative definition of foreclosure that is very similar to our measure of foreclosure inventory for a subsample of their data that includes only properties in the city of Boston. They find a significant, negative coefficient estimate for this variable. It is hard to compare magnitudes across studies because Campbell, Giglio, and Pathak (2011) use a hedonic rather than a repeat-sale approach.

[6] See Gerardi, Lambie-Hanson, and Willen (2011) for a discussion of foreclosure time-

In contrast to the bulk of the foreclosure externality literature, Campbell, Giglio, and Pathak (2011) focus on the *difference* between two γs: a γ_B associated with foreclosure completions that occur in the year prior to a sale and a γ_A associated with foreclosure completions that occur in the year after a sale. The authors argue that γ_A represents the causal effect of prices on foreclosures writing that, "To the extent that house prices drive foreclosures, low prices should precede foreclosures rather than vice versa," and argue that $\gamma_B - \gamma_A$ therefore represents the causal effect of foreclosures on prices. Although they find that $\gamma_B - \gamma_A$ is negative for their whole sample, which includes single-family, multi-family, and condominiums, for the single-family residential properties that we and all previous researchers focus on, they find that $\gamma_B - \gamma_A$ is very close to zero and statistically insignificant.[7] Taken literally, the conclusion of the paper would be that foreclosures of single-family properties have no effect on the prices of other single-family properties. However, Campbell, Giglio, and Pathak (2011) do find and report significant γ coefficients both before and after the foreclosure and a more plausible interpretation of their empirical results is that γ_A measures foreclosure externalities that occur before the foreclosure is completed and not the causal effect of prices on foreclosures. Indeed, as we show below, properties with seriously delinquent mortgages for which the foreclosure process has not yet been completed or has not yet been started negatively impact the sale prices of nearby non-distressed properties. Many of the nearby foreclosure auctions that occurred in the year after the non-distressed sale (the measure used by Campbell et al. to proxy for simultaneity bias) were likely in a state of serious delinquency at the time of the sale. [8]

lines across and within states.

[7]See Table A-19 of the Internet Appendix to Campbell, Giglio, and Pathak (2011).

[8]Recently, Hartley (2011) uses a similar difference-in-difference identification strategy as Campbell, Giglio, and Pathak (2011) to measure the effect of nearby foreclosures on non-distressed property values. In addition, he uses data on both single-family and multi-family property foreclosure notices to try to distinguish between disamenity and supply effects. Hartley (2011) argues that his results support a supply effect rather than a disamenity

Until the mid-2000s, studies that estimated hedonic price regressions similar to equation (1) largely ignored foreclosures because, up to that point, foreclosures were not a major issue. Two topics of focus in the early literature were the presence of sex offenders and subsidized housing programs.[9] We focus on the latter because it is, in fact closely related to the topic of this paper. Many early studies had attempted to calculate whether subsidized housing raised house prices by using aggregated data. Galster, Tatian, and Smith (1999) developed a methodology to use transactions-level data to measure the impact of Section 8 housing[10] on the sale prices of neighboring properties. They compared the sale prices of properties within 500 feet of a Section 8 site before and after the site transitioned to Section 8 and assumed that the difference in sale prices measures the treatment effect. Many studies in the literature subsequently used this methodology, which is very close to the strategy used by Campbell, Giglio, and Pathak (2011) that we discussed above. The most relevant to our analysis is Schwartz et al. (2006), which

---

effect. However, it is difficult to compare his analysis to ours, as his data only encompasses the Chicago area, while ours incorporates the 15 largest MSAs. In addition, there are at least two reasons to be somewhat skeptical of his interpretation. First, Hartley finds a significant effect of foreclosure filings within 0.05 miles of a non-distressed property sale, but basically no effect of foreclosures between 0.05 and 0.10 miles. There is no reason to expect a property that comes on the market 200 ft. from another home to have a significantly stronger supply effect compared to a home that comes on the market 200-500ft away. Normally we think of a local housing market as corresponding to a school district, or at least larger than a block or two, so that if the externality was truly a supply effect, we would not expect a discontinuity at a distance of 0.05 miles. In addition, one of the identifying assumptions in the analysis is that distressed multi-family homes and distressed single-family homes are characterized by similar depreciation rates. This may be a tenuous assumption, as many states have laws that prevent a landlord from neglecting the property and ignoring a renter's repair requests. In addition, to the extent that more "strategic" defaults occur among multi-family property owners, lack of sufficient funds to maintain the property before delinquency might not be as big of an issue, and thus we might expect distressed multi-family properties to be in better condition, on average, than distressed single-family properties.

[9] Linden and Rockoff (2008) estimates a version of equation (1) using data on the registration of sex offenders at particular addresses.

[10] Section 8 refers to the eighth section of the Housing Act of 1937, which authorized payments of rental housing assistance to landlords on behalf of low-income households.

DRAFT

used both highly disaggregated location variables, and, for one specification, used the repeat-sales method rather than a hedonic specification to control for property characteristics.

All of the cited papers used similar methods regressing log price or price growth on some measure of distressed property within a given radius. The only paper to deviate substantially is Rossi-Hansberg, Sarte, and Owens III (2010) who estimated a version of equation (1) to measure the effects of a program in Richmond, VA to subsidize investment in properties in disadvantaged neighborhoods. They diverged from the literature by using a semiparametric approach to estimate a pricing surface for all locations in their subject area and then looked at how the investment affected the surface.

In comparing the investment externality literature and the foreclosure externality literature, a key difference is the permanence of the effect. With some possible exceptions, serious delinquency and REO status are temporary states whereas investment is more long-lived. Thus, for the investment externality question, the distinction between flows and stocks is likely much less important than for the foreclosure externality question.

2.2 Data

Our sample of repeat-sales includes all pairs of non-distressed transactions on single-family residential properties in the 15 largest MSAs pulled from public records purchased from a national data aggregator. The sample is restricted to include transactions in which the first sale in the repeat-sale pair took place after 2001 and the second sale took place between 2006 and 2010 allowing us to study both the pre-crisis and post-crisis periods.[11] We exclude addresses that cannot be geo-coded and transactions for which recorded prices or dates are missing, zero, or located in a thin market which we define as a CBG

[11]The previous literature has, for the most part, used pre-crisis sample periods. The exception is Campbell, Giglio, and Pathak (2011) who do use data between 1987 and the first quarter of 2009, and thus capture a good portion of the crisis period.

DRAFT

in which there were fewer than 5 sales in a year. The final sample contains 958,513 repeat-sale pairs in 15 MSAs, and 16,932 CBGs, as reported in Panel (A) of Table 1.[12] The bottom panel of Table 1 reports the distribution of observations in the repeat-sales sample by the first and second sale years, respectively. There are several notable patterns in the table. The sample of repeat-sales gets smaller over time as national sales volumes fall. In 2006 and 2007, the modal sale occurred 2 years after purchase, falling to 3 in 2008, 4 in 2009 and 5 in 2010. The MSAs with the most observations are Phoenix, AZ, Washington, D.C., and Riverside, CA, which account for 16%, 13% and 10% of the sample respectively. Table 2 shows that the repeat-sales sample includes enormous variation in returns, which is not surprising given that the dataset includes properties purchased in 2001 and sold in 2006 and also properties purchased in 2006 and sold in 2010. The public records data also contain information on basic property characteristics over time, including house size, lot size, property age, and number of bedrooms. These variables are summarized in Table 2.

Using the public records, we can identify the date of the foreclosure deed[13], when the lender records transfer of ownership from the borrower and the REO sale date when an arms-length buyer takes ownership of the property. Using these flows, we can compute foreclosure inventory in a location at any point in time. The final sample contains 1.04 million foreclosure deeds, which we refer to as the REO inventory throughout the paper and 1.15 million REO sales.

[12] Another important difference with much of the previous literature is the national representativeness of our data. Many previous studies focus on a single state or even a single MSA. For example Campbell, Giglio, and Pathak (2011) use data from the state of Massachusetts, Immergluck and Smith (2006) use data from Chicago, and Schuetz, Been, and Ellen (2008) use data from New York City. Harding, Rosenblatt, and Yao (2011), who use data from seven MSAs, is probably the most nationally representative study.

[13] The foreclosure deed either corresponds to the transfer of the property to the lender at auction, or if the auction is successful, the transfer of the property directly to another arms-length buyer. The latter event is significantly less likely to occur than the former, but we are able to distinguish between both events in the data.

To identify seriously delinquent properties, we use two methods. Our main approach is to use proprietary data from a large national mortgage insurer (the "proprietary data" mentioned in the introduction) which contains all of the information in the public records plus a detailed payment history, and allows us to identify the first month in which a delinquent borrower enters serious delinquency (SDQ) , which we define to be 90 days delinquent (typically 3 missed payments). SDQs correspond to the entire period before the foreclosure auction in which the borrower is seriously delinquent, and thus covers both the time before the foreclosure process is initiated on a seriously delinquent borrower, as well as the time between the start of the foreclosure process and the end of the process (the auction).

The data also allow us to identify the cumulative depth of delinquency at any point in time. Our dataset contains 1.12 million SDQs. Because the proprietary dataset does not cover the universe of all homes, we augment it with data from a nationally representative loan-level dataset (the LPS data). With the more representative dataset, we calculate for each state, the distribution of the number of months that it takes for a mortgage to transition from serious delinquency to foreclosure completion (i.e. the foreclosure auction). We then take the 25th percentile of those distributions, and combine them with the information from the public records database on the date of the foreclosure auction to impute SDQ intervals. For example, the 25th percentile for California is 4 months. Thus, for each of the REO properties located in a California MSA in our sample, we assign an SDQ interval corresponding to the 4 months before the foreclosure auction dates. We use the 25th percentile as opposed to the median or average to be conservative, as this means that 75 percent of foreclosures in California had a serious delinquency spell that lasted for more than 4 months. We call this variable "infilled" SDQs. This provides 726,547 additional SDQs. We then combine our infilled SDQs with the SDQs obtained from the proprietary mortgage database to produce a more encompassing SDQ measure.

DRAFT

For our analysis, we divide SDQs into “long SDQs” and “short SDQs” depending on whether the borrower has been delinquent for more than a year or not. For some regressions, we also look at “minor DQs” which we define to be delinquencies of 60 days or less.

Panel A of Table 3 shows that 2/3 of repeat-sales had no distressed properties nearby. Panel B considers differences in the number of nearby distressed properties between the second and first sale in the repeat-sale pair, and shows that a non-trivial number of repeat-sales had *less* distressed property nearby at the time of the second sale with roughly 5 percent of sales occurring near properties with lower REO inventory and fewer sales of REO in the year preceding the sale. Panel C of Table 3 shows that, not surprisingly, the incidence of sales with distressed property nearby has increased significantly. Most dramatically, the proportion of properties with long SDQs nearby rose from less than 2 percent in 2006 to more than 30 percent in 2010, reflecting both the increased hazard that borrowers transition into serious delinquency and delays in the foreclosure process. By 2010, more than half the sales in our sample occurred with at least one form of distressed property nearby.

Finally, panel D of Table 3 displays correlations between our stock measures of distressed property and flow measures. All of the measures are positively correlated, but no two measures have a correlation higher than 0.50, which emphasizes the importance of distinguishing between stocks and flows.

3 Results

Table 4 shows results from our baseline specification. The right-hand-side variables of interest are nearby long SDQs from the proprietary data and three measures from the public records: nearby REO inventory; the number of nearby REOs sold one year prior to the non-distressed sale; and the number of nearby REOs sold 1-2 years prior to the non-distressed sale. For

DRAFT

each variable, we measure the difference over the repeat-sales in the number within 330 feet (1/16 of a mile). Despite the fact that we are using repeat-sales, we control for the possibility that there is systematic variation in price growth across different types of properties by including the characteristics of the property from tax assessment data.[14] In addition, to control for the possibility that prices fell more in more dense areas, we include the number of properties within 1/16 of a mile in our baseline specification.

Column (1) of Panel A in Table 4 displays the results of our basic specification on 2009 data (repeat-sales for which the second sale occurred in 2009). The estimation results for the variables of interest show a basic pattern that is replicated in all of the subsequent regressions: the coefficient estimates associated with the first three stages of the foreclosure process have roughly similar magnitudes. The exception is nearby REO sales that occurred more than one year in the past, which are not negatively correlated with price growth. Columns (2) and (3) of Panel A show that the controls have only small effects, with the addition of the density measure reducing all of the coefficients of interest by a small amount (in absolute value).

The inclusion of CBG×year fixed effects in the baseline specification plays a significant role in the estimation results. Panel C in Table 4 shows that the coefficient estimates associated with nearby distressed property become much stronger with more aggregated geographic×year fixed effects, as the coefficient estimates associated with nearby long SDQs, REO inventory, and REO sales in the previous year approximately double in absolute value when we move from a specification that includes CBG×year fixed effects to a spec-

[14] If the logarithm of the price of a property is a linear function of time invariant characteristics, then as Harding, Rosenblatt, and Yao (2009) show, taking the difference over the repeat-sales will cause the characteristics to cancel out of the equation, and thus one does not need to control for them in the repeat-sales specification. In other words this implicitly assumes that price growth is not a function of the time-invariant characteristics of a property. However, because of preference changes, it may be the case that properties with different characteristics appreciate at different rates. For example, if homes with multiple bathrooms or with granite countertops become more sought after over time, then we would expect those properties to appreciate at a higher rate, all else equal.

DRAFT

ification that completely excludes geographic×year effects. This confirms the intuition from the example that we discussed in Figure 2. In that example we considered a dataset with two distinct geographic areas, where one area is hit by a price-reducing, foreclosure-causing demand shock and the other area is not. A 330-foot disk drawn anywhere in the first area is more likely to contain a foreclosure than one drawn in the second area meaning that even if a foreclosure has no effect on local prices, we will find a correlation between prices and nearby foreclosures. The described pattern could alternatively reflect foreclosures adding supply to the whole market and driving prices down, but the inference that foreclosures drive prices down within 330 feet would still be wrong. The results in Panel C of Table 4 show that it is across-census tract and across-MSA variation in foreclosure density that is driving much of the observed negative correlation of nearby foreclosures and prices at the national level. The estimates for each of the three variables of interest increase substantially in absolute value when we substitute county×year effects for census tract×year effects and increase again when we move from MSA×year effects to eliminating geographic×year effects altogether. The last two columns in Panel C show that the results are little-changed when substituting CBG×year fixed effects for census tract×year effects, which suggests that the census tract is a sufficiently small geography to deal with reverse causality and simultaneity bias in this context. With the exception of Campbell, Giglio, and Pathak (2011), who include census tract geographic controls, all previous attempts to estimate γ omit narrow geographic controls, and thus most likely significantly overestimate the effect.

Panel B in Table 4 shows how the coefficient estimates in the baseline specification change over time. In the panel we estimate the baseline specification of equation (2) separately for repeat-sale pairs in which the second sale took place in each year between 2006 and 2010. The coefficient estimates associated with nearby long SDQs and REOs sold 1-2 years before the non-distressed sale are the largest and smallest respectively in absolute value for four out of the five years. The coefficient estimates associated with nearby REO inventory and REOs sold in the year before the non-distressed sale are not as stable, with the former variable having a similar impact as the long SDQ variable for all years except 2007 when its estimated coefficient is close to zero. The estimated coefficient associated with the latter variable grows over the sample but is consistently smaller in absolute value compared to the estimated coefficients associated with both long SDQs and REO inventory.[15]

In Table 5, we exploit information about the vacancy status of distressed property and the condition of REO property. As discussed in the previous section, for a subset of SDQs, we have information about the vacancy status of the properties, and for a subset of the REO properties we have information about condition. Since the vacancy data, in particular, is only well-populated beginning in 2010, we focus on that year. The results show that the coefficient estimate associated with vacant SDQ property is approximately 70 percent larger in absolute value than the coefficient estimate associated with occupied property (-0.017 versus -0.010). But, perhaps the more significant results apply to the condition of the REO inventory. According to Table 5, the only significantly negative coefficient estimates are associated with REO in below average condition and with REO for which we do not have condition information. The fact that the estimate associated with the missing category is significantly negative likely reflects the fact that most REO is in below average condition. It is also worth noting that the estimated coefficient associated with "above average" REO is significantly positive, which suggests

[15] There is some evidence from the literature that the effect of nearby foreclosures on prices is non-linear, and specifically that it is diminishing in the number of nearby foreclosures. In unreported regressions that are available from the authors upon request, we explored this using a more flexible specification, in which we specified the number of nearby distressed properties as second and third order polynomials, as well as a series of indicator variables for each specific value. Consistent with the findings from the previous literature, we did find evidence of non-linearities as the effect of nearby distressed properties on prices is diminishing in the number of distressed properties. However, all of the results discussed in this section are robust to this more flexible specification, and thus for space considerations we chose to report the simpler linear specifications.

DRAFT

that nearby REO in good condition actually *increases* the sale price of non-distressed properties.

In Table 6 we look at the impact of distressed properties further away from the repeat-sale observations by augmenting the baseline specification for 2009 data with three additional rings of property: $\frac{1}{16}$ - $\frac{1}{8}$ of a mile away, $\frac{1}{8}$ - $\frac{3}{16}$ of a mile and $\frac{3}{16}$ - $\frac{3}{8}$ of a mile away. As previous researchers have found, the negative effect of nearby distressed property drops off very quickly with distance, as the coefficient estimates associated with distressed properties in the third and fourth rings are very close to zero.

In Table 7 we report estimation results in which we distinguish between short and long SDQs. The results in the first two columns of the table suggest that both short and long SDQs have similar effects on nearby non-distressed sale prices, although the coefficient estimate associated with the stock of nearby long SDQs is slightly larger in absolute value than the estimate associated with nearby short SDQs. In column (3) we distinguish between minor DQs, short SDQs, and long SDQs. Minor DQs also have a negative estimated effect on nearby non-distressed sale, which is similar in magnitude to the effect of short SDQs. When we distinguish between minor DQs and SDQs, we see that a relatively large difference emerges between the effect of long SDQs and short SDQS/minor DQs, with long SDQs having a much larger negative impact on non-distressed price appreciation.

In addition, we estimated a series of regressions to address specification issues. At least two issues merit special attention. The first is a potential omitted variable problem since our coverage of SDQs is only partial due to the fact that the proprietary mortgage data does not cover the entire mortgage market. Since non-proprietary SDQs are likely to be correlated with both proprietary SDQs and all other measures of distress including the large number of non-proprietary REOs from our public records database, the omitted variables here could potentially affect all of the estimates of interest. To address this problem, we used the “infill” method, which we discussed in

DRAFT

detail above to construct a dataset of non-proprietary REOs. Comparing column (5) and column (6) of Table 7 illustrates that using this much broader measure of SDQs makes little difference. Overall, the coefficient estimate associated with SDQ increases slightly in absolute value, while the other coefficient estimates slightly decrease suggesting that non-proprietary SDQs are slightly positively correlated with non-proprietary REO and slightly negatively correlated with proprietary SDQs.

The second specification issue is the choice of a repeat-sales specification rather than a hedonic approach. In Table 7, we consider two alternative approaches to estimating γ. In column (8), we show results from a standard hedonic model, which generates some curious results. We find comparably large coefficient estimates (in absolute value) both for the SDQ inventory variable and the number of nearby REO sales in the previous year as we obtained from the repeat-sales specification, but we find a significantly smaller coefficient estimate (in absolute value) associated with nearby REO inventory. As an alternative, we consider a model, displayed in column (7) of Table 7 that goes part way between the two specifications, in which we use the repeat-sales specification but include the *level* of distressed property at the time of the second sale on the right-hand-side instead of the difference between repeat-sales. This hybrid model also generates the odd non-monotonic pattern in coefficient estimates that we found from the hedonic model, which suggests that properties with persistent foreclosure problems or foreclosure problems in the past may influence the results in the hedonic model.

The results from a third alternative specification that we estimate are displayed in Table 8. This alternative specification is motivated by the approach used by Campbell, Giglio, and Pathak (2011) and others in the investment externality literature, as discussed in Section 2, which consists of simply using the flow of properties into foreclosure over some horizon in the future to proxy for the stock of distressed properties. In Table 8, we consider a model in which we construct two different proxy measures of SDQ properties

DRAFT

by counting nearby properties that completed foreclosure in the year *after* the non-distressed property sale or 1-2 years *after* the sale. In addition, we construct a proxy measure for the stock of nearby REO property by counting nearby properties that completed foreclosure in the year and 1-2 years prior to the non-distressed sale. While the results for 2009 and 2010 look generally similar, the results for earlier years are quite puzzling with significant numbers of *positive* coefficient estimates, whereas with our more precise measures of the stock of nearby REO properties and the stock of nearby SDQs, we find consistently negative coefficient estimates. We view these results as illustrative of the importance of measuring the stock of distressed property as opposed to foreclosure flows.[16]

4 Interpretation

In Section 3, we established some empirical facts. Houses that sell very close to all forms of distressed property appear to do so at lower prices than otherwise similar properties in the same CBG that sell without the presence of nearby distressed properties. The effect appears when the borrower becomes seriously delinquent on his mortgage and disappears one year after the lender sells the foreclosed property to a new homeowner in an arms-length transaction. What can explain these empirical observations? There are, in our opinion, three candidate explanations for the measured effect: demand effects, supply effects, and investment externalities. In this section we first focus on supply and demand effects, and then turn to a discussion of invest-

[16] An additional observation from Table 8 is that the difference between the coefficient estimates for foreclosure completions within one year before and after the non-distressed sale, which Campbell, Giglio, and Pathak (2011) use to measure the externality, is significantly negative in 2006-2009 but is zero in 2009 and 2010. As we discussed above, for single-family properties, Campbell, Giglio, and Pathak (2011) find that the difference is approximately zero. However, we do not view the difference between the coefficient estimates to be a good measure of the externality, especially if it is truly a disamenity effect (we discuss this issue in more detail below).

DRAFT

ment externalities.

4.1 Supply and demand effects

The idea that a fall in demand leads to a fall in prices and an increase in foreclosures has strong support in theory and in the data. Default only makes sense for a borrower if he is in a position of negative equity, that is if his mortgage debt exceeds the value of his home, and it is difficult to have negative equity without falling prices. More broadly, models in which households face only limited opportunities to take out unsecured loans yield a relationship between the extent of negative equity and the probability of foreclosure.[17] Studies using micro-level data confirm the theory showing a strong relationship between equity and foreclosure and Gerardi, Shapiro, and Willen (2009) show that the estimated relationship is driven by within-town, within-time variation in equity and therefore is not attributable to causality running from foreclosures to prices. How does this relate to the current paper? All else equal, a negative demand shock in location A relative to location B would lead to a fall in prices in location A and a concomitant increase in negative equity and foreclosures.

The fact that the demand theory *could* explain the observed facts does not necessarily mean that it *does* explain them. One important fact does provide some support for the demand story and that is the results in Table 7 in which we found that γ is significantly negative even for minor delinquencies. One could argue that it is unlikely for significant property depreciation to occur in situations in which borrowers have only missed one or two mortgage payments, and thus, this empirical finding suggests that the estimates of γ may be explained by reverse causality. However, it could also be the case that a borrower who has only missed a few payments may have been in financial duress for quite some time, in which case the lack of investment in property maintenance might be a plausible explanation for the negative coefficient

[17] See Gerardi, Shapiro, and Willen (2007).

estimate associated with minor DQs.

But two additional empirical facts militate are inconsistent with the demand explanation. The first is the fact that the estimated γ is so much smaller for nearby properties sold out of REO more than a year in the past. While it is without doubt possible to construct a model in which demand recovers after the foreclosure, it seems implausible. The second broader issue here is that we are looking at variation within CBGs and any demand story (and any supply story, as well, as we will argue below) must operate at the market level. For the demand theory to make sense, we would have to believe that there are distinct submarkets within a census block group, geographic areas with a "target size," according to the Census, of around 1500 people.[18] Recall that since we are looking at repeat-sales, some shock would have to happen at a given time and impact one part of the CBG more than another. For example, one end of a CBG may have ocean views but to explain our results, it would have to be the case that demand for ocean views rose or fell more over the relevant time period.

A popular alternative to the demand theory is the supply theory which posits that a foreclosure increases the supply of property on the market and drives down prices. Unlike the demand theory, the supply theory is not an obvious consequence of standard economic theory. Normally, when we price long-lived assets like houses, we define supply as all of the assets that exist, not just the ones currently for sale. Foreclosures do not change the number of houses or the quantity of land in a market, so standard models would not predict any effect on prices. Market frictions obviously could change that. When shopping for a home, buyers generally restrict their attention to properties listed for sale and an increase in the number of options gives the buyer more bargaining power. One might then argue that we should focus our attention on the listing of the foreclosed property to identify supply effects. However, both buyers and especially sellers know about foreclosures before

[18] `http://www.census.gov/geo/www/cob/bg_metadata.html`.

DRAFT

they transition to REO and are listed by the lender, since foreclosures are, by law, public information. Thus, we cannot assume that the foreclosure would not affect the listing behavior of other sellers who might, for example, list prior to the foreclosure making the transition to REO in order to exploit the temporarily low supply. In a sense, with forward-looking households, the question would be when the news that the property will go on the market arrives. But, if buyers and sellers are that forward-looking then the whole short-run frictions story has less merit. To make matters more confusing, a troubled borrower facing foreclosure cannot sell his or her home – that is why they are facing foreclosure – so, according to the bargaining story above, serious delinquency should drive down supply and *increase* prices.

Our view is that the results do not provide strong support for the supply theory. The supply story could potentially explain why prices rise after the REO sale: with the property now off the market, prices recover to their pre-delinquency level. However, other facts from the data work against it. The first problem is the evidence on condition. If we thought foreclosed properties were driving down prices by competing with non-distressed sales, then we would expect, at the very least, that the properties in above average condition would have the same effect as properties in below average condition and, indeed, we might even expect the above-average properties to generate even more competition. The second problem with the supply story is the same one that, in our opinion, dooms the demand story: the fact that we are identifying γ using within-CBG variation.

The supply and demand stories are not, of course, mutually exclusive. As mentioned above, Campbell, Giglio, and Pathak (2011) argue that we can interpret the the coefficient estimates associated with foreclosure sales that occurred after the non-distressed sale as a pure measure of the demand effect and the coefficient estimate associated with foreclosures that occur before as the sum of supply and demand effects. Using our measures, we could then argue that the coefficient estimate associated with nearby SDQs is the demand effect and the coefficient estimate associated with nearby REO inventory is the sum of the supply and demand effects. Since the coefficient estimates associated with both nearby REO and SDQ inventories are quite close, the logic of Campbell, Giglio, and Pathak (2011) suggests that there is no supply effect of foreclosures on prices. It is important to stress that, in this respect, our results do not contradict those of Campbell, Giglio, and Pathak (2011). When they look at single-family residential properties, as we do, they also find no difference in the coefficient estimate associated with foreclosure sales before and after the non-distressed sale. Thus, based on the two papers, one might conclude that the demand effect is really the only effect, but we challenge such an interpretation below.

4.2 Investment externalities

The third candidate explanation for the results in Section 3 is that foreclosures lead to an investment externality. According to this theory, neither delinquent borrowers nor lenders have an incentive to maintain the property properly, which leads to physical deterioration of the property that some have labeled a "disamenity," which, in turn, reduces the value of nearby property to potential buyers. The investment disincentive is arguably present both during the SDQ period and when the property is REO. There are two reasons why seriously delinquent borrowers are unlikely to maintain properties. The first is that many of them have suffered cash-flow depleting life events and discount future consumption heavily relative to current consumption. Effectively, this raises the hurdle rate on any investment in the property. The second problem is that many seriously delinquent borrowers expect to lose their homes and thus, the long-term benefits of any investment would accrue to the ultimate owner of the property – the lender.

After the foreclosure, there are good reasons to expect underinvestment as well. Narrowly, the lender does not obtain any consumption benefit from investment. More broadly, the scale of residential lending leads to an information problem. The agent managing the property rarely ever has a large ownership stake in the property. This is obviously true when the agent is acting on behalf of a securitization trust but it is even true when the agent works for a bank that owns the property. The issue is that the amount of profitable investment in the property is a matter of discretion and the owner of the property cannot be sure whether the manager has other incentives. As with all standard asymmetric information problems, the result would be a failure to exploit profitable gains from trade, in this case, investment in the property. Another way to put this is that the optimal mechanism for investment in single-family residential real estate is to sell the property to a small scale investor who internalizes the costs and benefits.

Evidence from the existing literature supports the existence of an investment externality. First, there is ample evidence that foreclosed properties suffer from underinvestment. We will focus our discussion on studies that use data from the current crisis period, but there is an older literature that emerged well before the most recent housing market boom and bust, which we will not cover. Interested readers can find a summary of this older literature in Frame (2010). Pennington-Cross (2006) uses a variant of the repeat-sales methodology on nationally representative data from 1995-1999 and finds that the price of a distressed property appreciates significantly less than the prevailing metropolitan area price index in which the property is located. The average difference is 22 percent, but is highly sensitive to both the length of time that a property spends in the foreclosure process and the length of time that it remains in REO, as well as certain mortgage characteristics and state-level foreclosure laws.

Campbell, Giglio, and Pathak (2011) find a 27 percent foreclosure discount for single-family properties in Massachusetts over a significantly longer period of time, 1987-2007. In addition to looking at foreclosures, the authors identify properties in which the owners have died or declared bankruptcy. For these properties, they estimate a much lower discount: between 3 and

7 percent. Campbell et al. find evidence that the death discount is due to lack of maintenance while the bankruptcy discount is due to liquidity concerns. They find evidence that the foreclosure discount is larger for houses in poorer neighborhoods with less valuable characteristics and interpret this as evidence that the discount is related to concerns about vandalism: Either the properties have been damaged while they are on the market, or the cost of protection against potential vandalism is so high that the mortgage lenders are willing to accept significant discounts to sell them quickly. This difference in the estimated forced sales discounts between foreclosures and deaths/bankruptcies is consistent with the investment externality effect that we discussed above. Since the non-foreclosure related forced sales do not suffer from a separation between physical and legal ownership, they are less likely to experience the misaligned incentives that could impede property maintenance. This could explain the relatively small discount that Campbell et al. find for non-foreclosure related forced sales compared to foreclosures.

Clauretie and Daneshvary (2009) find a foreclosure discount of slightly less than 10 percent in the Las Vegas metropolitan area between 2004 and 2007. The authors estimate a hedonic specification in which they control for property and neighborhood characteristics as well as a subjective measure of the condition of the property. They show that controlling for the condition of the property (as well as a few other factors including spatial price correlation and marketing time) substantially reduces the estimated foreclosure discount. Harding, Rosenblatt, and Yao (2011) find little evidence of a meaningful foreclosure discount using both hedonic and repeat-sale methods. They first identify a sample of repeat-sales where the first transaction in the pair is the purchase of a REO property, while the second is a non-distressed transaction, and a control group of repeat-sales where the first purchase of the pair is a non-distressed transaction, but closely matches the initial purchase in the REO pair in terms of location, time of sale and property characteristics. They compare price appreciation between the two groups, adjusting for time between sales, and find very small differences, which they interpret as evidence against foreclosure “stigma” discounts, and evidence for the existence of permanent unobserved differences in characteristics between REO properties and non-distressed properties.

Second, the literature on subsidized investment, discussed in Section 2.1, shows that investment in property affects the value of nearby properties. Both Schwartz et al. (2006) and Rossi-Hansberg, Sarte, and Owens III (2010) examine programs that encouraged investment in individual properties, in the case of Schwartz et al. (2006), using a methodology virtually identical to what we do in this paper. Combining the findings of the foreclosure discount literature and the investment externalities literature, it would almost be surprising if previous studies had not uncovered foreclosure externalities.

What specific evidence do we find in favor of the investment externalities theory? First, there is the evidence on the condition of REO property. Again, it is hard to reconcile the demand or supply stories with the fact that the coefficient estimate associated with nearby below average REO comes in so far below the coefficient estimate associated with nearby REO in average and above average condition. Second, the investment externalities story can explain why the coefficient estimate is negative even for minor delinquencies and why it disappears after the new arms-length owner has had a chance to invest in the property.

5 Conclusion

In this paper, we document some new facts about foreclosure externalities. We show that houses trade at lower prices when there are homes nearby with delinquent homeowners, when there are homes nearby owned by lenders, and even when there are homes nearby recently sold by lenders in arm’s length transactions. We show that nearby houses trade at lower prices when the lender-owned property is in below average condition and at higher prices

DRAFT

when it is in above average condition. We discuss three possible explanations for the facts, supply effects, demand effects and investment externalities, and argue that the third is the most plausible. But to be sure, the interactions of all three are so complex that no dataset and no model could likely completely rule out an explanation or precisely allocate the observed correlations to one of the three stories.

The policy implications of a strong investment externality effect are quite significant, however, and suggest that the key to minimizing the costs of foreclosure is to minimize the time that properties spend in serious delinquency and in REO. On one hand that implies putting pressure on lenders to sell properties out of REO quickly. On the other hand, and perhaps much less palatably, it implies minimizing the time a borrower spends in serious delinquency, which means accelerating the foreclosure process.

Put another way, our results suggest that delaying the foreclosure process exacts a substantial cost on society as a whole and must be taken into account when making policy. As an example, Massachusetts passed a "right-to-cure" law in 2007, which forced lenders to give borrowers an additional 90 days to cure their mortgage before foreclosure proceedings could start. Gerardi, Lambie-Hanson, and Willen (2011) use a difference-in-difference approach to show that the law did not benefit borrowers in the sense that borrowers subject to the law were no more likely to cure or to renegotiate their loans than borrowers who were not. One might say that the law only failed to produce benefits, but our analysis suggests that it may well have imposed costs on homeowners who lived near borrowers who could take advantage of the law.

DRAFT

References

Campbell, John Y, Stefano Giglio, and Parag Pathak. 2011. "Forced Sales and House Prices." *American Economic Review* 101(5): 2108 – 2131.

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

Frame, W.S. 2010. "Estimating the effect of mortgage foreclosures on nearby property values: a critical review of the literature." *FRB Atlanta Economic Review*.

Galster, G.C., P. Tatian, and R. Smith. 1999. "The impact of neighbors who use Section 8 certificates on property values." *Housing Policy Debate* 10(4): 879–917.

Gerardi, K., A.H. Shapiro, and P.S. Willen. 2009. "Decomposing the foreclosure crisis: House price depreciation versus bad underwriting." *Working Paper*.

Gerardi, Kristopher, Lauren Lambie-Hanson, and Paul S. Willen. 2011. "Do Borrower Rights Improve Borrower Outcomes? Evidence from the Foreclosure Process." Working Paper 17666. National Bureau of Economic Research.

Gerardi, Kristopher, Adam Hale Shapiro, and Paul S. Willen. 2007. "Subprime Outcomes: Risky Mortgages, Homeownership Experiences, and Foreclosures." Working Paper 07-15. Boston, MA: Federal Reserve Bank of Boston.
Available at `http://www.bos.frb.org/economic/wp/wp2007/wp0715.pdf`.

Harding, J.P., E. Rosenblatt, and V. Yao. 2011. "The Foreclosure Discount: Myth or Reality?" Forthcoming Journal of Urban Economics.

Harding, J.P., E. Rosenblatt, and V.W. Yao. 2009. "The contagion effect of foreclosed properties." *Journal of Urban Economics* 66(3): 164–178.

Hartley, Daniel. 2011. "The Effect of Foreclosures on Nearby Housing Prices: Supply or Disamenity?" Working Paper 10-11R. Federal Reserve Bank of Cleveland.

DRAFT

Immergluck, D., and G. Smith. 2006. "The external costs of foreclosure: The impact of single-family mortgage foreclosures on property values." *Housing Policy Debate* 17(1): 57–79.

Lin, Z., E. Rosenblatt, and V.W. Yao. 2009. "Spillover effects of foreclosures on neighborhood property values." *The Journal of Real Estate Finance and Economics* 38(4): 387–407.

Linden, L., and J.E. Rockoff. 2008. "Estimates of the Impact of Crime Risk on Property Values from Megan's Laws." *The American Economic Review* 98(3): 1103–1127.

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

Rogers, W.H., and W. Winter. 2009. "The impact of foreclosures on neighboring housing sales." *Journal of Real Estate Research* 31(4): 455–479.

Rossi-Hansberg, E., P.D. Sarte, and R. Owens III. 2010. "Housing Externalities." *Journal of Political Economy* 118(3).

Schuetz, J., V. Been, and I.G. Ellen. 2008. "Neighborhood effects of concentrated mortgage foreclosures." *Journal of Housing Economics* 17(4): 306–319.

Schwartz, A.E., I.G. Ellen, I. Voicu, and M.H. Schill. 2006. "The external effects of place-based subsidized housing." *Regional Science and Urban Economics* 36(6): 679–707.

DRAFT

Table 1: Distribution of Repeat-Sale Observations By MSA and Year

Panel A: By Metropolitan Statistical Area

	# CBGs	Share of repeat-sale Transactions 2006	2007	2008	2009	2010	Total
Atlanta	1,020	26.8%	23.0%	17.7%	14.9%	17.6%	71,086
Boston	1,029	19.5	21.0	18.1	21.0	20.4	22,897
Chicago	1,139	24.2	19.7	16.9	20.3	19.0	27,853
Detroit	1,060	18.3	20.1	20.9	20.7	20.0	23,300
LV	438	25.6	15.1	17.2	20.5	21.6	71,189
LA	2,228	24.3	16.3	16.1	22.1	21.2	84,552
Miami	825	33.1	19.6	12.8	16.3	18.1	67,240
New York	2,023	21.5	22.3	19.7	20.4	16.2	45,305
Orlando	383	29.3	17.7	13.2	17.9	22.0	29,906
Philadelphia	1,485	24.7	23.4	18.0	18.3	15.8	51,038
Phoenix	1,069	26.1	18.3	17.1	18.8	19.7	154,243
Riverside	885	27.8	14.9	16.2	19.7	21.4	94,435
Seattle	836	28.5	24.6	16.1	15.7	15.1	35,893
Tampa	745	30.0	18.7	14.5	18.1	18.6	54,743
DC	1,767	26.2	19.2	16.7	19.4	18.5	124,833
Total	16,932	252,478	181,551	159,176	181,391	183,917	958,513

Panel B: By Year

Year of Purchase	Year of Sale 2006	2007	2008	2009	2010	Total	%
2001	30,720	19,144	12,250	11,889	11,607	85,610	8.9%
2002	37,613	23,712	14,955	14,718	14,420	105,418	11.0%
2003	49,456	30,052	19,483	19,295	18,725	137,011	14.3%
2004	58,161	37,764	25,184	25,315	24,964	171,388	17.9%
2005	56,562	34,521	34,445	36,495	35,208	197,231	20.6%
2006	19,966	24,242	28,403	32,917	30,600	136,128	14.2%
2007		12,116	14,704	15,142	17,666	59,628	6.2%
2008			9,752	13,126	7,632	30,510	3.2%
2009				12,494	11,004	23,498	2.5%
2010					12,091	12,091	1.3%
Total	252,478	181,551	159,176	181,391	183,917	958,513	100%
%	26.3%	18.9%	16.6%	18.9%	19.2%	100%	

Notes: This table reports the distribution of repeat-sale pairs of non-distressed transactions across the 15 largest metropolitan statistical areas (MSAs) in the United States. The data is drawn from public records purchased from a national data aggregator. The data include all repeat-sale pairs in which the first sale occurred between 2001 and 2010 and the second sale between 2006 and 2010.

DRAFT

Table 2: Summary Statistics of Property Characteristics for Repeat-Sales Observations

	Mean	Std	P90	P95	P99	Max
Price	348	279	622	790	1,400	9,500
Prior	331	252	580	735	1,283	9,775
Appreciation	2.57	43.57	54.97	66.93	89.38	187.18
Holding Period (qtrs)	13.9	8.4	25	29	35	39
Property Chracteristics						
Size	1,970	848	3,090	3,580	4,735	32,222
Lot	12,391	23,073	21,400	40,075	108,900	435,600
Age	24	35	55	66	95	2,011

Notes: This table displays summary statistics of the hedonic variables that are included in the various empirical models discussed in the text. The mean, standard deviation, 90th percentile (p90), 95th percentile (p95), 99th percentile (p99), and maximum of each variable is displayed.

DRAFT

Table 3: Summary Statistics of Nearby Distressed Properties for Repeat-Sales Observations

A: % of sales with # distressed property within 330 feet

		0	1	2	3	4	≥5
# SDQs		65.6	19.6	7.7	3.5	1.7	2.0
	Long SDQs	87.6	8.4	2.4	0.9	0.4	0.4
	Short SDQs	70.7	19.3	6.4	2.2	0.8	0.6
	Minor DQs	30.2	18.8	13.9	10.4	7.6	19.0
REO Inventory		81.3	11.9	3.9	1.6	0.7	0.7
REO Sales in last year		80.4	10.3	3.9	2.0	1.2	2.1
REO Sales 1-2 years ago		86.5	8.2	2.5	1.2	0.6	1.0
Total Distressed		66.7	14.9	6.5	3.7	2.3	6.0

B: % of sales with difference in # distressed property within 330 feet

		< 0	0	1	2	3	4	≥ 5
Long SDQs		1.5	84.5	9.1	2.9	1.1	0.5	0.4
REO Inventory		4.8	77.8	9.3	3.5	1.8	1.0	1.8
REO Sales in last year		5.2	83.0	7.2	2.2	1.0	0.6	0.9
REO Sales 1-2 years ago		0.2	89.3	7.1	2.0	0.8	0.3	0.3

C: % of Sales with $>$ 0...

		2006	2007	2008	2009	2010	Total
SDQs		22.9	22.8	29.9	45.0	55.5	34.5
	Long SDQs	1.7	2.1	6.5	23.7	31.2	12.4
	Short SDQs	22.0	21.7	26.3	33.7	44.9	29.3
	Minor DQs	68.6	68.6	68.8	70.8	72.2	69.8
REO Inventory		7.4	10.2	23.9	27.1	29.7	18.7
REO Sales in last year		3.8	4.1	18.1	38.4	39.5	19.6
REO Sales 1-2 years ago		6.1	3.8	4.5	17.5	37.1	13.5
Total Distressed		11.7	14.5	35.9	54.3	58.8	33.4

D: Correlations

	SDQ inventory	REO inventory	REO sales in last year	REO sales 1-2 year
Long SDQs	1.00	0.23	0.39	0.32
REO Inventory		1.00	0.41	0.30
REO Sales in last year			1.00	0.50
REO Sales 1-2 years ago				1.00

Notes: This table displays summary statistics on the number of nearby distressed properties to the non-distressed repeat-sales observations in the sample. A "short" SDQ is defined as a property in which the borrower has been in serious delinquency for less than one year, a "long" SDQ is defined as a serious delinquency that has lasted for more than one year, and a "minor delinquency" is defined as a property in which the borrower has is only one or two payments behind.

DRAFT

DRAFT

Table 4: Baseline Specification. Dependent variable is price growth between purchase and sale.

Panel A: Baseline Specification	(1)	(2)	(3)
Δ Long SDQ	-0.028	-0.026	-0.029
	(-39.5)	(-36.7)	(-41.1)
Δ Foreclosure Inventory	-0.033	-0.031	-0.033
	(-52.1)	(-49.5)	(-51.6)
Δ REO sold in last year	-0.024	-0.023	-0.025
	(-68.9)	(-66.6)	(-71.2)
Δ REO sold 1-2 years ago	0.002	0.003	0.003
	(3.2)	(4.0)	(4.2)
Property Age	0.001	0.001	-
	(31.4)	(33.7)	
Property Age^2	0.000	0.000	-
	(-24.5)	(-26.2)	
Log(Lot size)	0.002	-0.011	-
	(2.6)	(-15.6)	
Log(Living area)	0.082	0.078	-
	(50.5)	(48.0)	
# of properties <330 ft	0.908	-	-
	(28.6)		
R^2	0.71	0.70	0.71

Panel B: Different Years	2006	2007	2008	2009	2010
Diff. in Long SDQs	-0.068	-0.039	-0.039	-0.028	-0.030
	(12.3)	(8.2)	(20.4)	(39.5)	(50.4)
Diff. in REO inventory	-0.042	-0.005	-0.023	-0.033	-0.028
	(26.9)	(4.3)	(34.7)	(52.1)	(44.5)
Diff. in REO sold in past year	-0.006	-0.004	-0.020	-0.024	-0.022
	(7.6)	(4.9)	(33.9)	(68.9)	(60.0)
Diff. in REO sold 1-2 year ago	-0.003	-0.005	0.002	0.002	-0.009
	(4.7)	(5.3)	(2.04)	(3.21)	(24.6)
R^2	0.36	0.43	0.65	0.71	0.70

Panel C: Different Geographic Controls	None	MSA	County	Tract	CBG
Diff. in Long SDQs	-0.050	-0.038	-0.039	-0.029	-0.028
	(54.2)	(45.6)	(48.6)	(39.7)	(39.5)
Diff. in REO inventory	-0.067	-0.047	-0.044	-0.035	-0.033
	(83.1)	(64.1)	(62.1)	(53.5)	(52.1)
Diff. in REO sold in past year	-0.049	-0.037	-0.033	-0.025	-0.024
	(107.7)	(91.1)	(82.3)	(68.9)	(68.9)
Diff. in REO sold 1-2 year ago	-0.007	-0.001	0.004	0.003	0.002
	(8.2)	(1.8)	(4.93)	(3.91)	(3.21)
R^2	0.23	0.47	0.54	0.59	0.61

Notes: This table displays results from the estimation of equation (2. In all regressions reported, the independent variable is price appreciation over the repeat-sale interval. Panel A includes a full set of CBGxyear of purchase fixed effects. Panel B includes the hedonic controls listed in Panel A as well as CBGxyear of purchase fixed effects. In panels A and C the second sale year in the repeat-sale pair is 2009. The numbers in parentheses are t-statistics.

Table 5: Effect of Property Condition and Vacancy Status based on 2010 Sample.

Diff. in SDQs by Vacancy Status	
Occupied	-0.010
	(11.0)
Vacant	-0.017
	(17.6)
Missing	-0.010
	(19.5)
Diff. in REO in Inv by Property Condition	
Below average	-0.026
	(5.9)
Average	-0.003
	(1.7)
Above average	0.022
	(4.1)
Missing	-0.014
	(21.0)
Diff. in REO sold in past year	-0.016
	(38.3)
Diff. in REO sold 1-2 year ago	-0.007
	(25.4)
R^2	0.68

Note: All regressions include the specification in column (1) of Panel A in Table 4 and a full set of CBG × year of purchase dummies. The top panel in the table displays results for nearby SDQ inventory by vacancy status, while the bottom panel displays results for REO inventory in different states of upkeep. Numbers in parentheses are t statistics.

DRAFT

Table 6: Effect of REO/SDQs by Distance based on 2009 Sample

	Difference in number of: Long SDQs	REO inventory	REO sold in: Last year	1-2 years ago
0-100 yards	-0.015	-0.014	-0.010	-0.004
	(15.8)	(16.2)	(19.8)	(4.6)
101-200 yards	-0.007	-0.007	-0.003	0.001
	(12.1)	(12.7)	(9.8)	(.9)
201-300 yards	-0.001	-0.002	-0.001	0.002
	(4.9)	(7.8)	(3.2)	(8.6)
301-600 yards	-0.002	-0.001	-0.001	0.001
	(21.0)	(8.1)	(8.6)	(11.75)

Note: The regression includes the specification in column (1) of Panel A in Table 4 and a full set of CBG × year of purchase dummies. Numbers in the parenthesis are t statistics. The R^2 of the regression is 0.77.

Table 7: Alternative Specifications

	Baseline (1)	(2)	(3)	(4)	(5)	(6)	RHS in levels (7)	Hedonic (8)
Minor DQs			-0.017					
Short SDQs		-0.021	-0.016					
Long SDQs	-0.028	-0.026	-0.025				-0.022	-0.020
All SDQs					-0.024			
All SDQs with infill						-0.024		
REO inventory	-0.033	-0.032	-0.032	-0.035	-0.032	-0.029	-0.015	-0.008
REO sold in past year	-0.024	-0.023	-0.024	-0.026	-0.023	-0.021	-0.023	-0.022
REO sold 1-2 year ago	0.002	0.003	0.002		0.003	0.003	-0.011	-0.008
R^2	0.71	0.71	0.71	0.71	0.71	0.71	0.83	0.70

Note: All regressions include the specification in column (1) of Panel A in Table 4 and a full set of CBG × year of purchase dummies. All reported coefficients are significant at the 1% significance level. Numbers in the parenthesis are t statistics.

DRAFT

Table 8: An alternative way to measure properties at different stages of the foreclosure process.

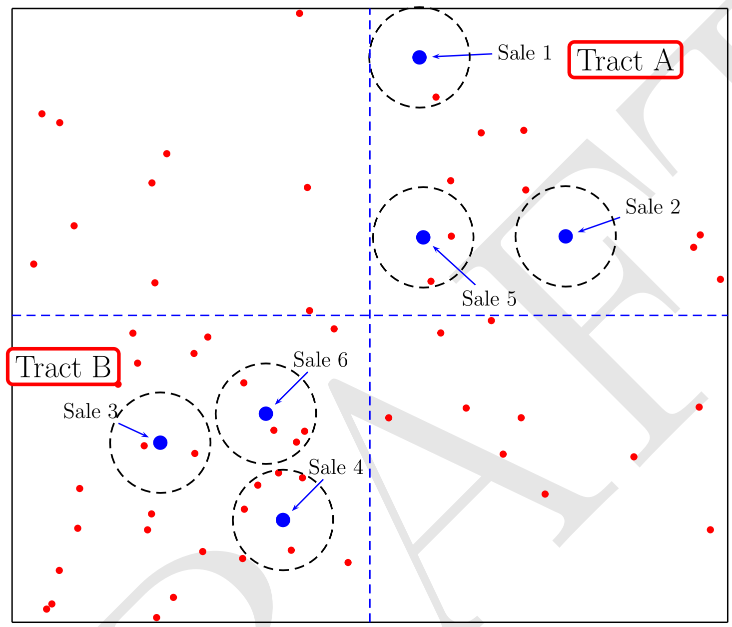
	2006	2007	2008	2009	2010
Δ F 1-2 years after	0.041	0.027	-0.004	-0.021	-0.031
Δ F in year after	0.005	0.012	0.003	-0.018	-0.023
Δ F in year before	-0.030	-0.011	-0.028	-0.020	-0.023
Δ F 1-2 years before	-0.020	-0.022	-0.018	-0.021	-0.021
R^2	0.38	0.45	0.65	0.71	0.70

Note: This table reports results from repeat-sale regression specifications similar in spirit to the regressions estimated by Campbell, Giglio, and Pathak (2011). The variables of interest are the number of nearby (within 330ft) foreclosure completions in the year before the sale, the number of nearby foreclosure completions in the year after the sale, and the number of nearby foreclosure completions between 1 and 2 years before and after the sale. Campbell, Giglio, and Pathak (2011) measured the externality produced by nearby foreclosures by taking the difference between the first two variables. The regressions include hedonic controls and a full set of CBG × year of purchase dummies. Numbers in parenthesis are t statistics.

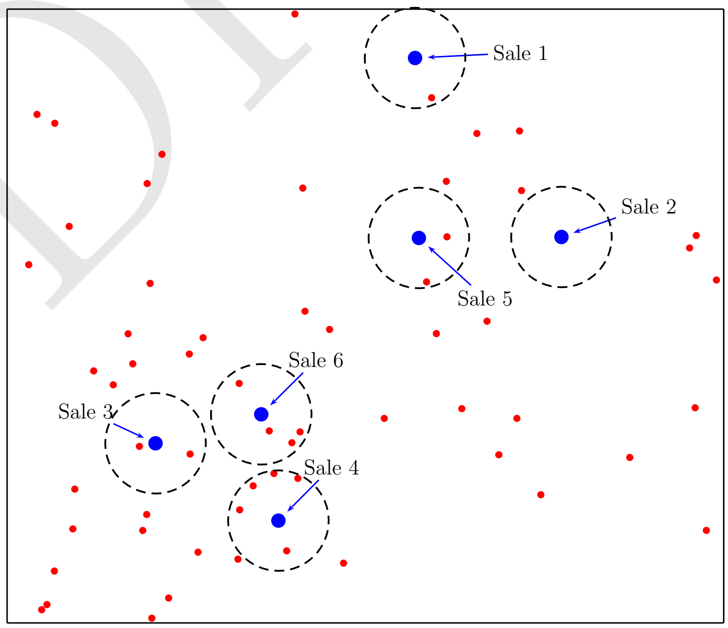
DRAFT

Figure 1: Importance of Geographic Controls

Map of Foreclosures and Sales

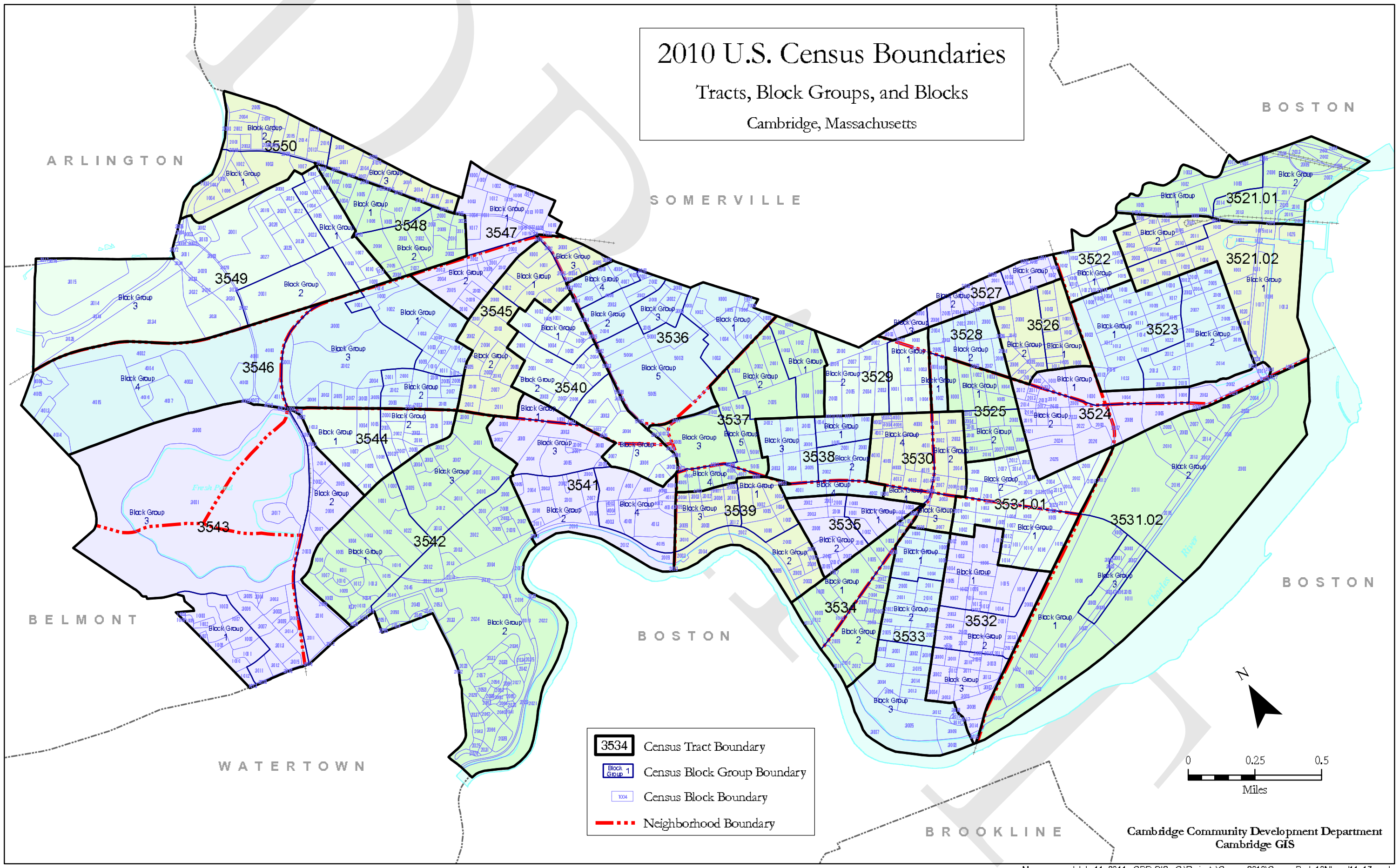


Map of Foreclosures and Sales *without* Census Tract Controls



DRAFT

Figure 2: Census Tracts and Blocks in Cambridge, MA. 2010 Census.



DRAFT