Fels Institute of Government
University of Pennsylvania

Estimating the Economic Value and Neighborhood Impact of Tax-Delinquent Properties in Philadelphia

January 28, 2013

Kevin C. Gillen, Ph.D.
Senior Research Consultant
Fels Institute of Government
University of Pennsylvania
3814 Walnut Street
Philadelphia, PA 19104
gillenk@upenn.edu
215-573-9567
http://www.fels.upenn.edu/

Objective: As part of an investigative project on tax-delinquent properties in Philadelphia, this author was asked by the project's lead journalist to undertake some empirical analysis of these properties. Specifically, I was asked to answer two questions:

- What is the current market value of these delinquent properties?
- What effect—if any—does the presence of tax delinquent properties have on the value of their neighboring properties?

The purpose of this document is to describe how these two questions were answered, including the data used and the empirical methodology that was deployed. Although the econometric techniques used in this project are fairly complicated, this document is written for a general audience. Hence, I focus on the intuition of the methodology rather than the technical and mathematical details. But, any further questions about these details are welcome to be directed to the author, whose contact information is at the beginning of this document.

Task 1: What is the current market value of these properties?

Data: Data on tax-delinquent properties was provided to the author by Patrick Kerkstra, who obtained it from the City of Philadelphia's Department of Revenue on May 31, 2012. The file contained detailed property-level information on 102,787 delinquent properties in Philadelphia, including the property's unique parcel number, dollar amount of delinquent taxes and the number of years delinquent. This data was then merged with OPA's property file based on each property's unique parcel number in order to obtain the physical characteristics of each property, such as its square footage, lot size, age, number of stories, type of exterior and physical condition. This data was then geo-coded with the assistance of ArcMap GIS to assign a unique latitude and longitude to its location, and compute locational attributes of each property, such as its distance to the CBD (city hall) and proximity to amenities such as transit or parks. Lastly, this data was merged with data from the most recent decennial Census, based upon the Census Tract of each property, in order to assign neighborhood characteristics to each property, such as the Tract's median household income and vacancy rate.

Data on recent home sales in Philadelphia was obtained by combining deeded title transfers from the City's Department of Records with tax roll data from the City's Office of Property Assessment, merging on the common variable: the unique parcel ID. The sales data was extensively cleaned to remove sales between family members, blanket sales, nominal sales, transactions where the buyer or seller was not a private individual (e.g. bank sales, sheriff sales, sales where either the grantor or grantee was an LLC or LP, etc.), and any sales where the price did not reflect a true "market" value for homes in their neighborhood. Sales were also dropped if they had missing, erroneous and/or implausible values (e.g. extremely large or small square footages). This data was then processed according to the same procedures as the delinquency data: geocoded in ArcView and locational attributes computed, and then merged with Census data. The sales dataset contains 29,597 transactions, over the years 2010-2012, and effectively covers the universe of all arms-length home sales in Philadelphia with a sufficient representative presence in every one of the City's neighborhoods.

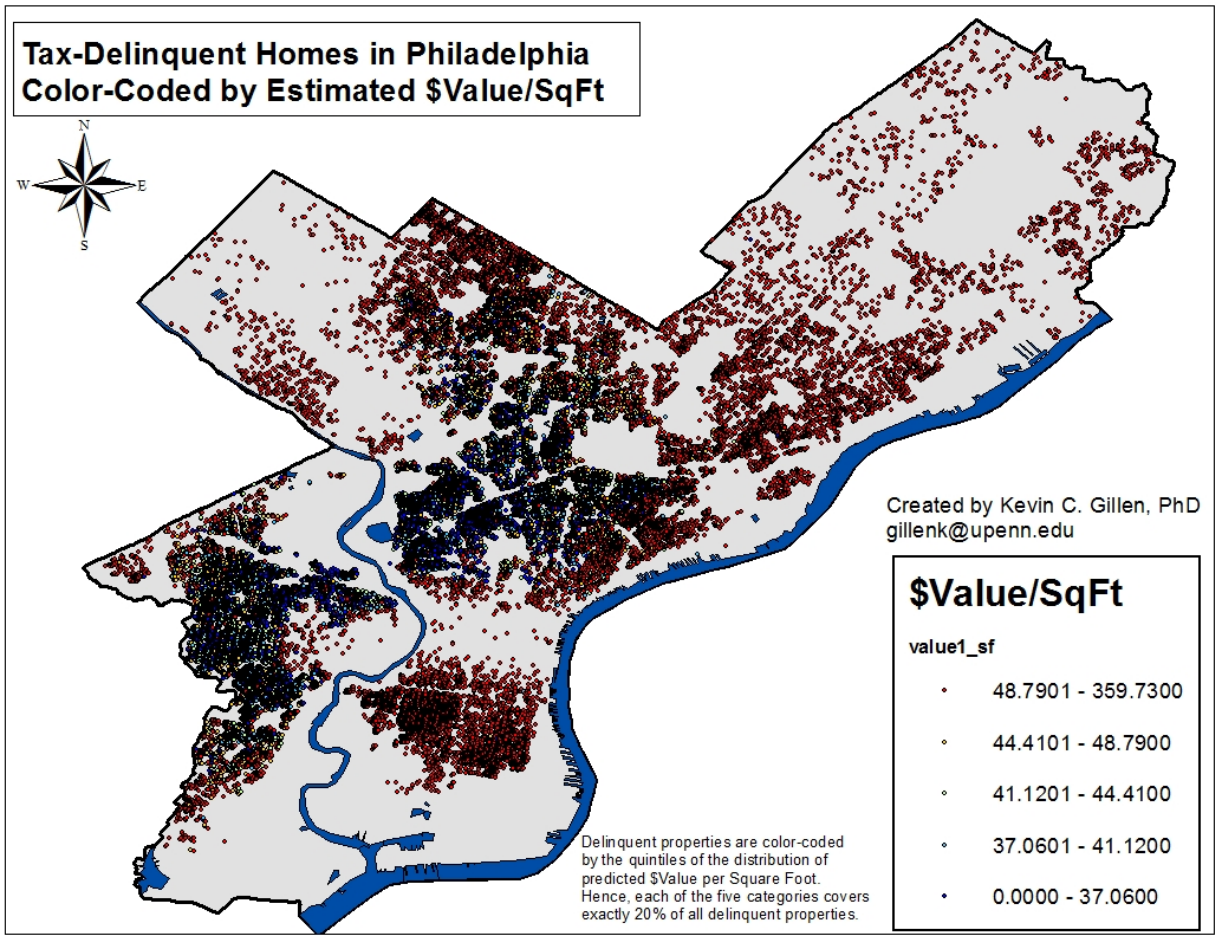
Methodology: Market values for the delinquent properties were derived by first estimating a regression-based pricing model using the home sales data, and then applying this model to the delinquent properties. The model used is known as a "hybrid hedonic regression". The advantage of using a regression model—as opposed to just summary statistics on neighborhood house prices—is that it effectively "controls" for a host of factors which affect a home's value, so that the underlying true appreciation rate of house values over time is effectively isolated and accurately measured. Along with a set of control variables in the regression specification, there is a set of time variables denoting the year and quarter that each property sold. It is from the estimated coefficients on these time variables that the price indices are obtained which can then adjust past home sales to current values (i.e. as-of year-end 2012).

Hedonic regressions are commonly used tools in the field of urban economics. The first hedonic regression of house prices was estimated by Rosen (1974). Using data on house sales, the author estimated a regression of house price on each dwelling's structural and locational characteristics. The estimated coefficients on these characteristics gave the marginal price of a unit increase in their level. For example, what an extra bedroom, bathroom, fireplace or garage space contributed to a property's overall value in dollar terms. The estimation of hedonic price regressions has since become ubiquitous in the housing economics profession, and these models enjoy widespread use in both academic research and industry practices.

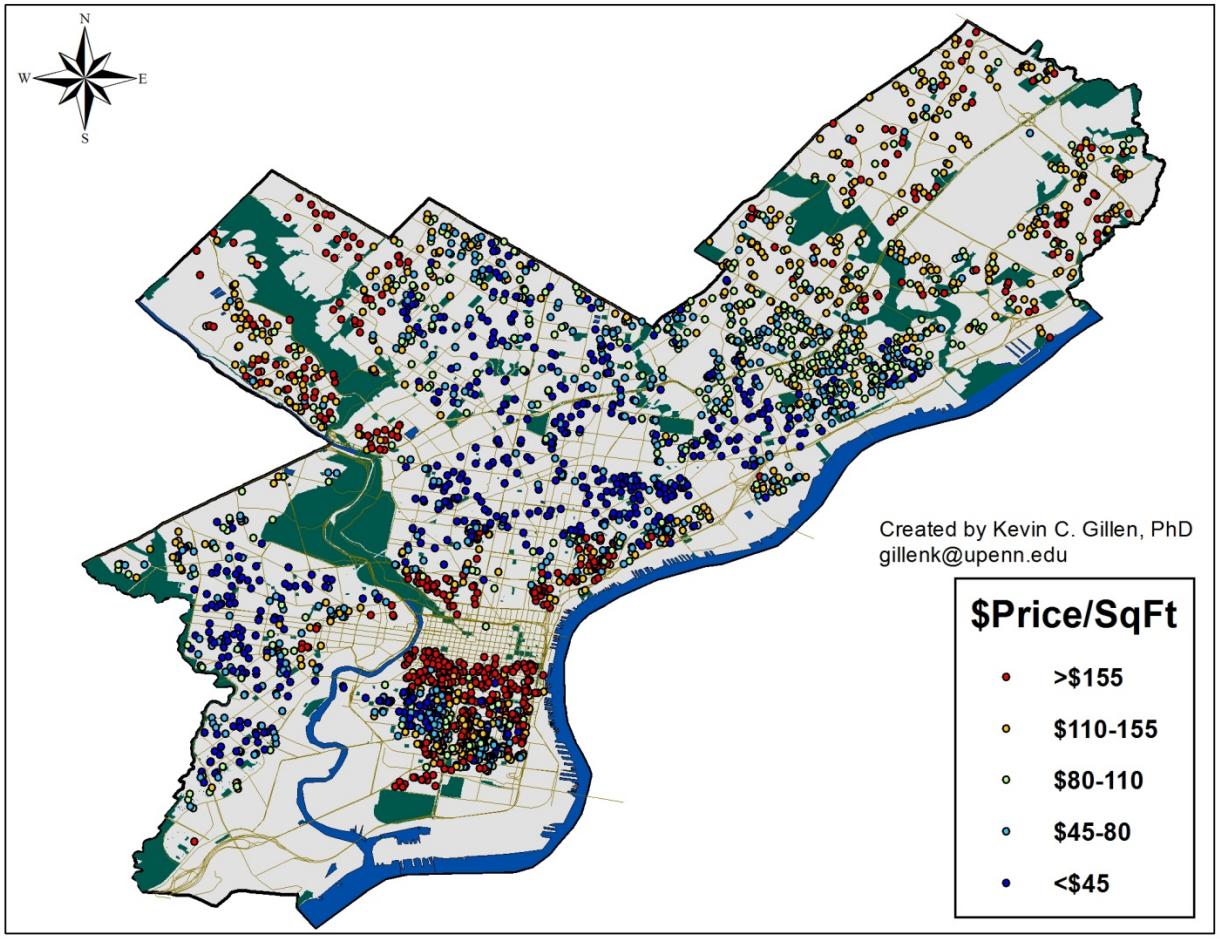
The specific model used in this study regressed the (natural log of) each dwelling's price on the characteristics of the dwelling (e.g. square footage, number of stories, physical condition), the characteristics of the dwelling's location (e.g. distance from Center City, proximity to parks or transit), the tax treatment of the dwelling (e.g. if abated or not), and what year the dwelling transacted in. The regression contains over 50 property-level variables in an attempt to capture as much variation in housing characteristics as possible. A critical set of control variables in the regression are Tract-level variables that are correlated with a neighborhood's overall level of delinquency. These include: the Tract's poverty rate, vacancy rate, median household income, percent of residents with college degrees and percent of the housing stock in below average or distressed condition. If these variables are omitted from the regression, the effect of delinquency becomes significantly over-inflated.

The regression was performed with SAS software, via a procedure known as "Iterative Weighted Least Squares". The resulting regression output is essentially an equation that expresses house prices as a function of their characteristics. The regression has an Adjusted R-squared statistic of 0.76, which implies that 76 percent (out of a possible 100 percent) of the variation in Philadelphia's house prices during 2010-2012 are explained by this particular model. Since housing is a non-standardized product with highly idiosyncratic price behavior, this R-squared would be considered by most housing economists to be quite high, and thus the underlying model to be quite good.

This model was then applied to the portfolio of delinquent properties by plugging in each property's unique characteristics into the regression equation and computing its likely market price as-of year-end 2012. In those few instances where a delinquent property had a missing and/or erroneous characteristic, the median sales price in that property's Census Tract was assigned to it. This procedure was done for approximately 7% of delinquent properties. Lastly, a file of these properties with each dwelling's estimated market value was delivered to Mr. Kerkstra. A map of these properties, color-coded by their estimated price per square foot, is shown below. A map of recent home sales, also color-coded by their price per square foot is also shown below, for comparative purposes.



Home sales in Philadelphia in 2012 Q4.

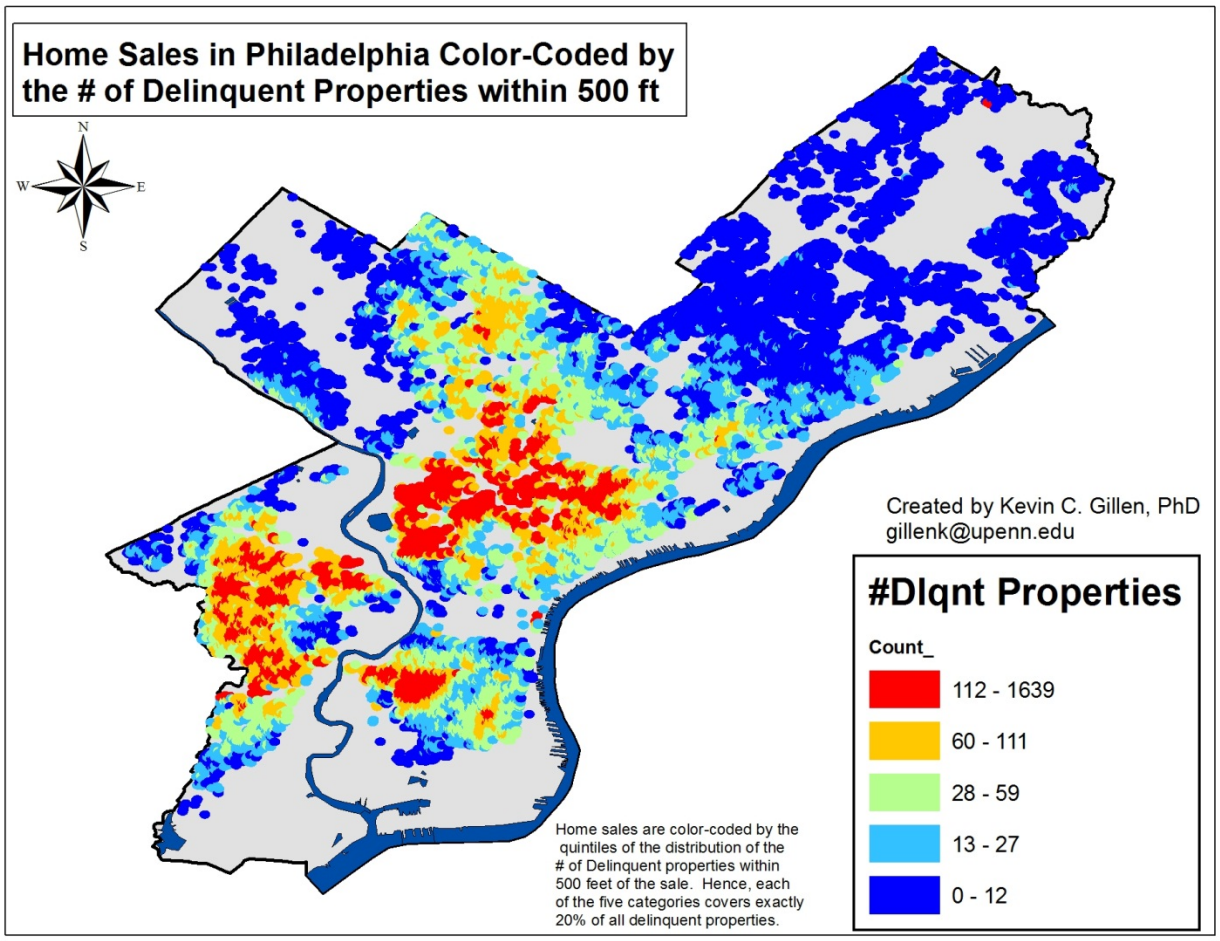


As can be observed between the two maps, the estimated value of the delinquent properties is typically below that of recent home sales: the median value/sqft for the former is ~$43/ft while the median price of recent home sales is ~$95/ft. This is for two reasons. First, the stock of delinquent properties is older, more depreciated (and often abandoned) and disproportionately concentrated in more distressed and lower-income areas of Philadelphia. Thus, we should expect that the value of these properties should, on average, be less than the value of the typical Philadelphia home. Secondly, sales during the 2010-2012 period have been skewed towards the upper end of the market. Tighter credit, a sluggish economy and higher unemployment have all disproportionately fallen on less-educated, less-skilled and less-experienced workers. Since these households are more likely to be lower-income, they are also more likely to live in lower-priced neighborhoods. Moreover, these are also the households who are relatively more constrained from buying a home. Consequently, those home sales that are occurring are more likely to be between relatively more affluent households, who naturally buy and sell in relatively higher-priced locales. This results in house prices being skewed upwards. As a consequence, the average delinquent house value is below the average Philadelphia house value, which in turn is below the current average Philadelphia house price. Thus, the spread between the delinquent values and recent house prices is quite large, but not necessarily unexpected nor a major concern for the purpose of this analysis.

Task 2: What effect—if any—do the presence of tax delinquent properties have on the values of their neighboring properties?

Data: This task used the exact same data as Task 1. However, the data was further expanded by computing the number of delinquent properties within 500 feet of each home sale, and assigning this number to each home sale. This was done in ArcMap by creating 500-foot buffers around each home sale, and then computing the number of delinquent properties in each buffer. A distance of 500 feet was chosen via an iterative regression process that identified this distance as that distance beyond which the effect of delinquent properties began to attenuate (i.e. drop off) sharply. That is, the correlation between house prices and the presence of delinquent properties is quite strong at distances up to 500 feet, but begins to rapidly weaken at longer distances. Perhaps un-coincidentally, 500 feet also happens to be the average length of a Philadelphia city block, which this author takes to support this distance as the appropriate one to use in this analysis.

> *"When it comes to connectivity for instance, much of Philadelphia benefits from a grid of relatively short blocks. Philadelphia's Center City blocks tend to average around 400 to 500 feet, giving pedestrians a substantial degree of freedom in their choice of routes between two points. Dan Burden and Reid Ewing agree that blocks over 500 feet start to push the boundaries of walkability. As Ewing put it, "Generally shorter is better. Four hundred feet probably is O.K., 600 feet is getting a little long, 800 feet is too long because you're requiring people to walk four or five hundred feet to get to an intersection."* -Source: "What makes a city walkable?" By Seth Budick For PlanPhilly, July 2008. http://planphilly.com/articles/2008/07/23/3568



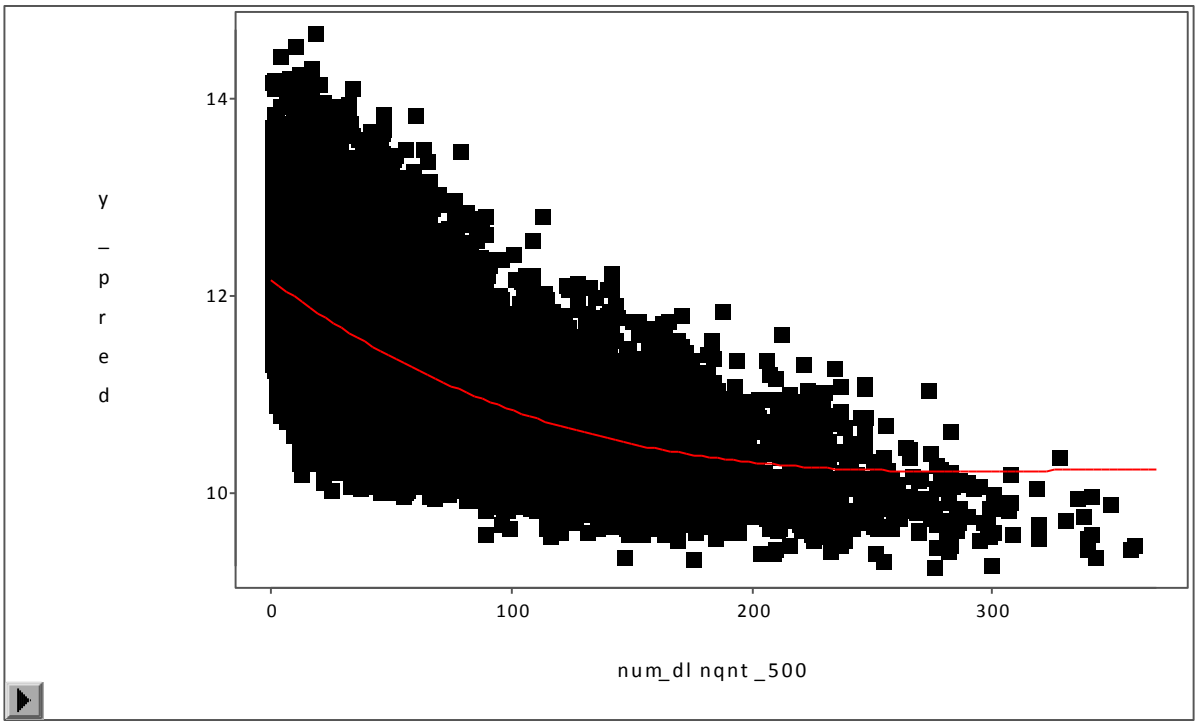
As can be observed from the map, delinquent properties appear to be present in all but the most very affluent of Philadelphia's neighborhoods; i.e. Center City and Chestnut Hill. However, their greatest concentration is in the low-income (and low-priced) neighborhoods of North Philadelphia, West Philadelphia and Point Breeze. In these neighborhoods, it is not uncommon for a dwelling to be within walking distance of more than 100 delinquent properties.

Methodology: It would be tempting to simply add "# of delinquent properties within 500 feet" as a variable to the hedonic regression used in Task 1 and simply re-run the regression. However, this is highly problematic for two reasons:

- The relationship between house prices and the nearby presence of delinquent properties is nonlinear; and
- The direction of causation between property values and the number of delinquent properties goes both ways: although delinquent properties can depress nearby house values, neighborhoods with declining house values can lead many owners to become delinquent on their property taxes.

We address both issues with a rather sophisticated version of our regression model, known as a "hybrid hedonic event-study regression with threshold effects".

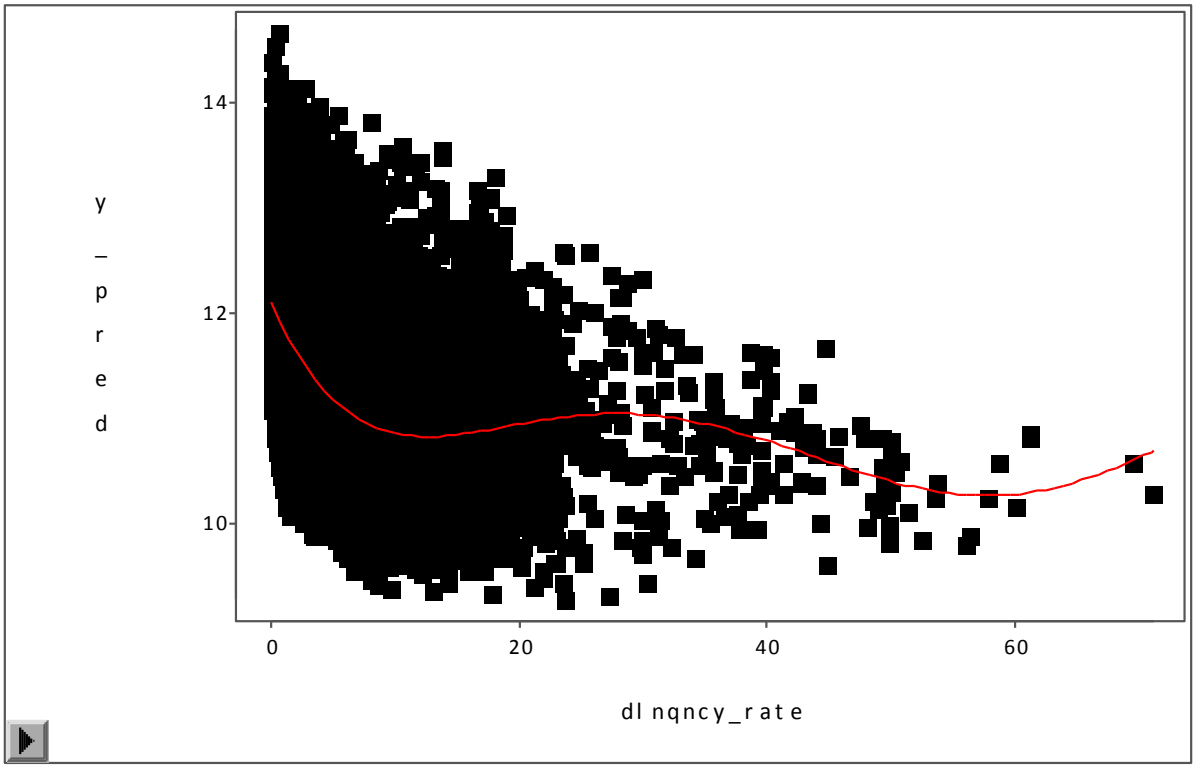
First, we examine the issue of non-linearity. The following scatterplot shows the (log) price of 2010-2012 home sales on the vertical axis and the number of nearby delinquent properties on the horizontal axis:



The red line represents a regression line fit through this data, using a polynomial (i.e. nonlinear) specification. As can be easily observed, the relationship between (the natural log) of house prices and the number of delinquent properties is indeed negative, but it is also nonlinear: house prices decline as the number of nearby delinquent properties increases, but does so at a decreasing rate. This is actually an intuitive result: delinquent properties are indeed associated with lower house values (and vice-versa), but as the delinquency rate increases, the association becomes weaker. Or, to put it more simply, the first set of initial delinquencies can have large downward effects on house prices, but once a neighborhood is saturated with delinquent properties, a few more properties going delinquent has a smaller effect.

A similar, but slightly more complicated relationship is shown if we replace the number of delinquent properties with the delinquency rate within 500 feet[1]:

[1] The delinquency rate was computed as the number of delinquent properties within 500 feet divided by the total number of properties within 500 feet. It was computed in ArcMap by performing a spatial join of the home sales to the file of delinquent properties and to the OPA property file.



Here, the relationship is also shown to be negative and nonlinear, but also non-monotonic: house values actually rise with certain levels of the delinquency rate, which is counterintuitive. We believe this is due to measurement errors induced by the GIS procedure, which sometimes resulted in a delinquency rate exceeding 100%; an obvious mathematical impossibility. Moreover, the use of the delinquency rate in the regression model resulted in a very large adverse effect on house values. For these two reasons, we use the number (rather than rate) of delinquent properties in our final analysis, as we believe it errs on both the side of accuracy and conservatism.

Further exploration of this data revealed there to be certain "threshold effects" of delinquencies. That is, beyond critical levels of the number of delinquent properties, the relationship between house prices and delinquencies exhibited significant shifts. Using an iterative procedure known as nonlinear least squares, these thresholds were identified to be 0-5 delinquent properties, 5-15 delinquent properties, 15-100 delinquent properties and >100 delinquent properties. Variables were then created categorizing the number of delinquent properties within 500 feet, based upon these thresholds.

The second problem of simultaneity was addressed by modifying the regression according to a technique employed by Galster, Tatian and Smith, in a 1999 study examining if the presence of Section 8 households were associated with lower neighboring property values. Here, the authors faced a very similar problem: do low-income (e.g. Section 8) households cause lower neighborhood house values, or do low-priced neighborhoods disproportionately attract Section 8 households? The authors devised an "event study" regression which created variables that measured the level and trend in nearby house prices ***before*** the arrival of Section 8 households, and the level and trend in nearby house prices ***after*** the arrival of Section 8 households. They then tested the statistical significance of the estimated regression coefficients on these variables in order to disentangle the correlation of Section 8 households

with low house prices from any causation that Section 8 households may have on low house prices. In effect, rather than comparing house prices ***near*** Section 8 households to house prices ***far*** from Section 8 households (an apples-to-oranges comparison), the authors compared house prices near Section 8 households ***before*** they arrive in the neighborhood to house prices near Section 8 households ***after*** they arrive in the neighborhood (a more apples-to-apples comparison).

We created very similar variables that measures the level in house prices before nearby properties have gone delinquent and then the level in house prices after nearby properties have gone delinquent. Lastly, we interacted these variables with the threshold effects to measure the change in house prices as a result of the overall delinquency level in the neighborhood. These variables were then added to the same hybrid hedonic regression of the previous section and the regression was re-estimated via the same iterative weighted least squares algorithmic procedure.

Our results were as follows:

1) If there are 5 or less delinquent properties within 500 feet of a home sale, each additional delinquent property is associated with homes being worth 0.218% less than they would be otherwise. However, the t-value for this number was statistically insignificant; i.e. no meaningfully different than zero. This implies that if there are 5 or less delinquent properties nearby a home, then on average there is no effect on a home's value.
2) Beyond the first 5 delinquencies, each additional delinquent property is associated with house values being worth 1.089% less than they would be otherwise. This is the steepest part of the curve, as shown in the previous scatterplots. That is, in excess of 5 delinquencies, nearby property values begin declining rapidly and sharply.
3) Beyond 15 delinquencies, each additional delinquent property is associated with a further 0.451% decline in nearby house values. Note that this is smaller than the declines associated with 5-15 delinquencies. This represents the inflection point of the curve: house values are still declining, but are now doing so at a slower rate.
4) Beyond 100 delinquencies, each additional delinquency is associated with only a 0.33% decline in nearby house values. This is the flattest part of the curve. The neighborhood is now so saturated with delinquencies, any further increases in the number of delinquent properties is negligible.
5) We believe these numbers to be both credible and conservative. We also estimated a simple hedonic regression that omitted the Census-level variables (e.g. vacancy, median income, etc.) and omitted the event study and threshold effects, and found that the effect of vacancy becomes very large if not accurately measure and effectively controlled for. This regression indicated that each additional delinquency was associated with as a much as an 8% decline in nearby house values. The regression results above indicate that, at most, an additional delinquency is only associated with a 1.089% decline in nearby house values.

The final step in our analysis was to compute the individual, and then aggregate effect on the value of the affected housing stock. Using ArcMap, we identified approximately 417,000 single-family homes in Philadelphia that were within 500 feet of at least one delinquent property. We first computed each property's current market value by applying the regression model to the property's characteristics. We then computed its counterfactual value—what the property would be worth if there were zero delinquencies nearby—by applying the above percentage effects to the number of delinquent

properties that are nearby and multiplying through. The difference between the current market value and the counterfactual value represents the associated diminution in value caused by nearby delinquent properties. We then applied the City's property taxation formula to this amount to compute the foregone tax revenue caused by this diminution in value.

We recognize that the computed diminution in value and subsequent foregone tax revenue rests critically on two unrealistic assumptions: a 100% tax collection rate, and 100% accurate and timely tax assessments. Since these are unrealistic expectations for any city, we suggest that our results be interpreted as the upper bound, or "best case" scenario for what outcome could be achieved as a result of increasing collections and reducing delinquencies.

1) The current value of the affected housing stock is computed to be $45.8bn. Since OPA has reported that its estimated value of the City's entire stock of real estate—which includes land and commercial properties—is $95bn, this number would seem to be both plausible and credible.

2) The median value of these affected homes is estimated to be $87,600, and the mean value is $110,000. If anything, this may be on the high side, due to OPA's classification of most properties as being in "average" condition, but still within the bounds of credibility.

3) The median loss in a home's value associated with nearby delinquent properties is estimated to be $15,200, with a mean of $22,800.

4) The total diminution in value to this housing stock is computed to be $9.5bn, which translates into a property tax revenue loss of $297.6m under 2013 property tax rates. That is, the results indicate that if Philadelphia had a 100% collection rate and properties were accurately assessed, these homes would collectively be worth $9.5bn more than they currently are and that the City and School District would collect $297.6m more in property taxes than they currently do.

Email for Kevin Gillen: gillenk@upenn.edu

References

Bailey, M.J., Muth, R.F. and Nourse, H.O. 1963. "A Regression Method for Real Estate Price Index Construction." Journal of the American Statistical Association. 58. 933-942.

Case, Bradford, Henry Pollakowski and Susan Wachter. 1991. "On Choosing Among House Price Index Methodologies." AREUEA Journal. 19. 286-307.

Case, K.E. and Shiller, R.J. 1987. "Prices of Single Family Homes since 1970: New Indexes for Four Cities." New England Economic Review. Federal Reserve Bank of Boston, Sep. 45-66.

Case, K.E. and Shiller, R.J. 1989. "The Efficiency of the Market for Single-Family Homes." The American Economic Review. 79. 125-137.

Court, Andrew T. 1939. "Hedonic Price Indexes with Automotive Examples." Dynamics of Automobile Demand. New York: The General Motors Corporation.

Galster, George C., Peter Tatian, and Robin Smith. 1999. "The impact of neighbors who use Section 8 certificates on property values. " HousingPolicy Debate 10 (4): 879-917.

Genesove, David and Christopher Mayer. 2001. □Loss Aversion And Seller Behavior: Evidence From The Housing Market,□ The Quarterly Journal of Economics. 116(4). 1,233-1,260.

Halvorsen, Robert and Raymond Palmquist. 1980. "The Interpretation of Dummy Variables in Semilogarithmic Equations." American Economic Review. 70(3). 474-475.

Rosen, Sherwin M. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." Journal of Political Economy. 82(1).