A Spatial Difference-in-Differences Approach To Studying the Effect of Greening Vacant Land on Property Values

Megan Heckert
West Chester University

Abstract

This article details the use of a spatial difference-in-differences approach for measuring the effect of a vacant land greening program in Philadelphia, Pennsylvania, on nearby property values. Vacant land is a ubiquitous problem in U.S. cities, and many have recently begun to explore greening programs as an interim management strategy for vacant lots, in the hopes they will reduce the negative influence of vacancy and help to spur neighborhood development. The methods used here draw on previous approaches to modeling effects of greening on property values but expand on them to explore means of incorporating spatial relationships and allowing for spatial nonstationarity, in which the process being modeled changes across space. Spatial methods were used not only to derive data and choose appropriate observations but also to compare global and local versions of the analysis to assess spatial patterns and differences in outcomes across the study area, ultimately showing that, although greening vacant land increases surrounding property values, it does not do so uniformly across urban neighborhoods.

Introduction

The urban revitalization literature is chock full of ideas for how to improve distressed neighborhoods, but actually determining the effects of interventions has proven to be more challenging. One of the most commonly studied effects is the change in property values; these effects are generally studied because the housing market is seen as a good indicator of the desirability or perceived quality of a neighborhood (Galster, Tatian, and Accordino, 2006). Most methods for making these assessments rely at least in part on hedonic regression models, in which the value of an individual property is seen as reflecting a bundle of values of individual amenities, which would include both characteristics of the property itself and characteristics of the neighborhood, and a regression equation is used to estimate the value of each individual amenity or property characteristic (Rosen, 1974). Thus the effect of proximity to an intervention on property values might be assessed by comparing property values at varying distances to the intervention and checking the coefficient of the distance variable to see if lower distances correspond to higher values.

The difference-in-differences approach is an econometric case-control test that investigates whether an intervention influences an outcome over time by comparing observed differences in a case sample that receives the intervention to observed differences in a control sample that does not. This approach enables isolation of the treatment effect above and beyond any difference that would have been expected regardless of the treatment (Meyer, 1995). The difference-in-differences model specification has been used with hedonic modeling of property values to assess the effects of a range of neighborhood interventions, including new housing development (Ellen et al., 2001; Ellen and Voicu, 2005), establishment of community gardens (Voicu and Been, 2008), and tree planting programs (Wachter and Wong, 2008).

Incorporating Space

When it comes to measuring the effects of interventions such as new housing development or tree planting on neighborhoods, space cannot be ignored. A neighborhood is a fixed location in which residential and commercial buildings, amenities such as parks, and disamenities such as vacant lots all coexist with their distinct spatial relationships informing the influences they have on each other and the neighborhood as a whole. The value of any component of that neighborhood cannot be divorced from the values ascribed to the other components of the neighborhood, because it is reliant in part on those values. This reliance means that space and spatial relations must inform any attempt to measure those effects. That being said, the attempts to measure the effects of intervention cited previously take space into account in only one way—by incorporating the distance to an intervention into the equation to estimate its effect. These attempts often take the form of creating distance bands from the intervention within which a property is or is not located (Ellen et al., 2001; Wachter and Wong, 2008). For example, Ellen et al. (2001) studied the effects of new housing development on existing property values by specifying houses as affected if they were within three different distance bands of the new development—500 feet, 1,000 feet, or 2,000 feet—and designating control lots as outside the distance in question but within the same ZIP Code.

These models are helpful, but they often assume a single model that describes relationships and values that are constant across the entire study area. They are unable to account for potential differences in effects across locations, known as spatial nonstationarity, unless the drivers of those differences are known in advance and can be incorporated into the models as interaction terms. More explicitly, spatial methods are required to find differences that are not known from the beginning or are unable to be incorporated because of a lack of appropriate data.

This article details the methods used to measure the effects of a greening intervention to manage vacant land on surrounding property values in Philadelphia, with a focus on the means by which spatial patterns were assessed. The program, Philadelphia LandCare (PLC), uses a simple greening approach to treat vacant lots by removing any existing trash or debris, bringing in topsoil, planting new grass and a few trees, and erecting a split-rail fence to prevent dumping and give the lot a more managed look. During the first decade of the program, more than 5,000 individual parcels received treatment through the PLC program. For more details of the program, see Jost (2010) and Heckert and Mennis (2012).

For this research, which is described in more detail in Heckert and Mennis (2012), I adapted the difference-in-differences approach for spatiotemporal analysis of changes in property values surrounding treatment and control lots through use of a sampling strategy that ensured control lots mimicked the spatial distribution and economic characteristics of treated lots, while also remaining spatially distant enough to prevent diffusion of treatment, whereby any effect from the treatment might also be demonstrated by the control because of proximity. I modified the approach by creating a geographically weighted variant, using geographically weighted regression to explore geographic variation in the effects of the greening program.

Data and Methods

This analysis relied on four primary spatial datasets and on several supplemental datasets. The primary datasets were (1) data on lots that were treated as part of the PLC program represented as points at the center of each of 747 contiguous groups of lots that were greened together between 1999 and 2006; (2) data on vacant lots in Philadelphia in 2010 also represented as points at the center of each group of contiguous vacant lots; (3) a set of points representing all Philadelphia residential real estate sales valued at more than $1,000 between the years 1999 and 2007, with sales prices adjusted for inflation to 2007 dollars; and (4) boundaries for neighborhood planning districts, breaking the city into seven regions. Additional datasets included shapefiles with locations of commercial corridors and schools and a real estate market typology created by a local community development financial institution, which ranked the 2008 real estate markets in each census block group on a 9.0-point scale from distressed (1.0) to regional choice (9.0). The purpose of each dataset is described in more detail in the following section.

The difference-in-differences specification uses a case-control methodology where each case—in this instance, each lot that was ultimately treated through the PLC program—is matched with appropriate controls—in this instance, lots that could have been treated through the PLC program but were not. Although the data on the PLC program and the vacant land data for Philadelphia both started as data on individual parcels, adjacent parcels were merged together for analysis. This merging was necessary because the PLC program was implemented on groups of adjacent lots that look and feel like a single entity, even if they are technically separate properties, and the effects of two adjacent lots cannot be reasonably separated from each other. Following similar logic, the vacant lots used as controls were combined based on adjacency.

It is very important to note that site selection for the PLC program was by no means random and, thus, controls could not simply be assigned randomly from the universe of all vacant lots. The Pennsylvania Horticultural Society (PHS), which developed and manages the PLC program, describes several criteria that are used to determine lots for inclusion in the program. First and foremost, PHS targets communities with large concentrations of blighted vacant lots—PLC is not a program designed for neighborhoods with strong real estate markets and low populations of vacant lots to choose from. Within target neighborhoods, lots are chosen based on loose criteria intended to identify lots with most potential for effect, so that large collections of lots near schools or commercial corridors are prioritized (Jost, 2010). In an effort to select control lots that were closest in characteristics to the treated lots, I restricted the set of all vacant lots to those located within 500 feet of a school or commercial corridor before selection of control lots.

One challenge for selecting control lots was attempting to prevent diffusion of treatment effects, whereby the effect of a greened lot would also happen and be felt in the area of nearby nongreened lots. The concern here was that a large number of untreated vacant lots are also located in close proximity to treated lots, which is not surprising, given that the program targeted areas with large numbers of vacant lots. Although it was desirable to keep control lots as far as possible from the PLC lots to avoid the possibility that property values surrounding them also increased because of proximity to PLC lots, it was also necessary to ensure that controls were located in similar neighborhoods and thus represented appropriate counterfactuals. To mitigate diffusion of treatment effects, all vacant lots within 250 feet of a treated lot were excluded from the pool of potential control lots. To ensure that controls were in similar neighborhoods to treated lots, the final selection of controls matched each treated lot to three randomly selected controls from the pool, with the controls matching the treated lot in both the section of the city and the real estate typology score for the block group of the lot. Thus the matches did not guarantee that the control lots were in exactly the same neighborhood but required that the control lots face similar real estate market conditions and be in the same relative portion of the city.

For the specification of the difference-in-differences model, the unit of analysis was taken to be the lot, with values assigned to represent the value of residential properties at the location of the lot. Property values were assigned to each lot for each year between 1999 and 2007 based on inverse distance weighting of the price per square foot of the closest 15 properties sold in that year within 500 feet of the lot. This approach essentially calculates a weighted average of those sale prices, with the weightings assigned so that closer properties have higher weights. When fewer than 15 properties sold within 500 feet of a lot, the number of properties included in the calculation was reduced to 10. If fewer than 10 properties sold, the search radius was increased to ensure that at least 10 sales were included.

With sale price per lot as the dependent variable, the specification of the difference-in-differences model was

$$lnV_{it} = \beta_0 + \beta_1 P_i + \beta_2 G_{it} + \beta_3 P_i G_{it} + \beta_4 M_i + \beta_5 S_i + \beta_6 Y_t + \varepsilon_{it} , \qquad (1)$$

where lnV_{it} is the natural log of the average price per square foot of residential real estate near vacant lot i at time t; P_i is a dummy variable set to 1 if lot i is part of PLC or 0 if it is not; G_{it} is a dummy variable set to 1 if time t is post-greening for lot i (for a control lot, this value is set to 1 when the associated treated lot is greened); $P_i G_{it}$ (that is, the interaction term defined as P times G) is a dummy variable set to 1 if lot i is in PLC and time t is post-greening; M_i is a variable encoding the real estate market index value of lot i; S_i is a fixed-effects variable for the Neighborhood Planning District of the city of lot i; Y_t is a fixed-effects variable for year to account for temporal effects; ε_{it} is an error term; and β terms are the coefficients to be estimated by the model.

This model provides an overall assessment of whether property values changed near greened lots in a manner that was different from changes near nongreened lots, but, as previously noted, it calculates a single equation that is assumed to represent the relationship between PLC and property values for the entire study area. I was, however, keenly interested in thinking about whether the PLC program behaved differently in different areas. One way of answering this question would be to split up the observations based on neighborhoods that might be expected to behave differently and to calculate model coefficients for each group separately. I attempted to assess neighborhood differences by running the model in two ways—first, by splitting the lots into planning neighborhoods and, second, by splitting lots into different real estate typology categories. That approach, however, requires some decision to be made in advance about the appropriate means for defining areas that might be expected to behave differently from each other. An alternative to the prescribed approach to splitting observations into neighborhoods is to use geographically weighted regression (GWR) instead of ordinary least squares regression in the specification of the difference-in-differences model.

The GWR model essentially calculates a separate regression equation for each observation in the dataset by calculating coefficients using only a subset of "nearby" observations, which are weighted based on proximity so that nearer observations have higher weights than those that are farther away (Fotheringham, Brunsdon, and Charlton, 2002).

The equation for the GWR model can be specified as

$$lnV_{it} = \beta_0(u_i,v_i) + \beta_k(u_i,v_i)\, x_k + \varepsilon_{it}\,, \tag{2}$$

with (u_i, v_i) representing the coordinates of point i, and x_k representing the k^{th} independent variable in the model. All variables from the original global difference-in-differences model were included in the GWR variant. This model weights observations based on their distance to point i so that it creates a Gaussian weight surface in which closer locations are weighted closer to 1 and farther locations' weights decrease ultimately to 0. The GWR model was run with bandwidths specifying neighborhoods of ½ mile, 1 mile, and 2 miles in radius.

One final step taken after the models were run was the calculation for each lot of the percentage of surrounding lots that had been greened. For each greened lot in the study, the number of lots within 500 feet was counted and the percentage of lots that were greened through PLC was calculated. This measure was then averaged for each neighborhood to create a measure of "concentration of greening" within that area. These values were mapped against the model results as a purely visual assessment of a possible relationship between effects of the program and the structure of its implementation.

Results

The global difference-in-differences model coefficients (exhibit 1) showed that property values surrounding greened lots did increase more than property values surrounding control lots, but much more information was gleaned from the geographically focused models, which showed that this relationship varied considerably over space. In the neighborhood-specific models, only three of seven neighborhoods—Eastern North Philadelphia, West Philadelphia, and Southwest

Philadelphia—actually showed the pattern of increased property values surrounding greened lots, while the other four neighborhoods had coefficients that were not significantly different from 0. The real estate market-based model also showed variations across the city, with distressed markets showing increased property values as a result of the PLC program but transitional and steady neighborhoods showing no effect.

Both of these patterns were further illuminated by the GWR model, which similarly demonstrated wide variation in the effects of the PLC program. Exhibit 2 indicates the results of the GWR model with a 1-mile bandwidth, although the results were consistent at all bandwidths tested. Note that

Exhibit 1

Coefficients of the Global Difference-in-Differences Model

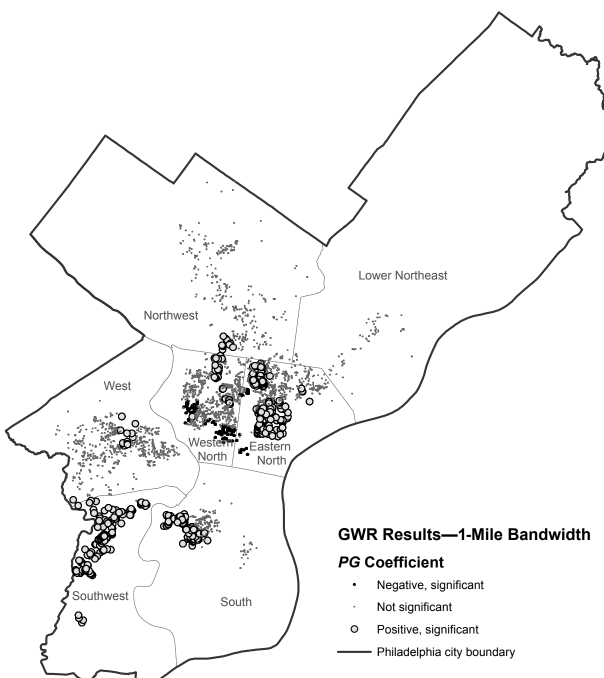
Independent Variable	Coefficient	R^2
P	– 0.084*** (– 8.729)	0.415
G	– 0.013 (– 0.987)	
PG	0.056** (3.100)	

* $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

Notes: t-values are reported in parentheses. N = 26,608.

Exhibit 2

Results of the Difference-in-Differences Geographically Weighted Regression



GWR = geographically weighted regression. PG *= the* PG *term in the model equation.*

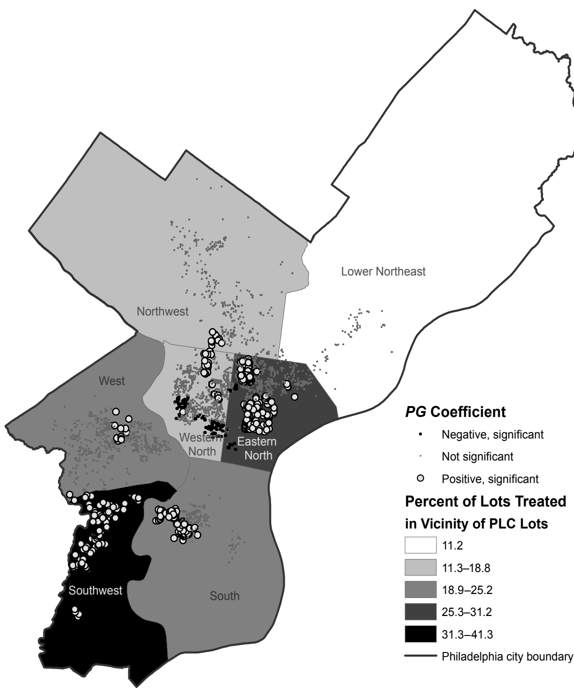
Note: Open circles indicate lots where the PG *coefficient is positive and significant (that is, locations where Philadelphia LandCare raised the nearby residential property values).*

this pattern matches the results of the neighborhood-specific models, with batches of positive coefficients in each of the neighborhoods that showed positive effects of PLC in its own model. The GWR results actually add additional nuance, however, highlighting the potential for variation of effects even within a neighborhood. In particular, additional small clusters of positive coefficients are found in South Philadelphia, Western North Philadelphia, and Northwest Philadelphia, indicating that PLC did lead to increased property values in parts of those neighborhoods, although not consistently throughout them. The GWR results also indicate areas of no effect within the three neighborhoods where the neighborhood models indicated PLC to have positive effects.

Comparison of the various model results with the concentration of greening measure showed that neighborhoods with the most positive effects of greening on property values in both the neighborhood-specific models and the GWR results tended to be those with the higher scores for concentration of greening (exhibit 3), suggesting that areas in which a higher proportion of lots were treated were more likely to see gains in property values. For additional tables, figures, and discussion of the results of this study, see Heckert and Mennis (2012).

Exhibit 3

Results of Geographically Weighted Regression Compared With Neighborhood Greening



GWR = geographically weighted regression. PG = *the* PG *term in the model equation. PLC = Philadelphia LandCare.*

Note: GWR results compared with the concentration of greening in each neighborhood, meaning the percentage of lots surrounding greened lots that also were greened.

Discussion

This analysis ultimately revealed that, although property values throughout the city increased during the study period, properties surrounding treated lots enjoyed a greater increase in value than properties surrounding controls, but it also showed that these effects were not felt evenly across the study area. The local models demonstrated that the effect was more pronounced in some parts of the city than others, a result that may have significant implications for continued implementation of the program.

This study further shows that the difference-in-differences method can be applied in understanding the spatial effects of an intervention—in this case, treatment of vacant lots by greening them—although special consideration must be given to spatial relationships in selecting appropriate controls. The additional use of a geographically weighted variant of the model was a key to generating meaningful results that can be used by program administrators and policymakers in future planning. The use of distinctly spatial methods was crucial throughout the study—first, for appropriate selection of control lots and specification of the initial aspatial model; second, for developing neighborhood and real estate market-specific variants of the model to begin to assess variations across the study area; and, third, in the development and specification of the GWR model, which ultimately provided the most nuanced results. Spatial methods were then able to be used to begin exploring how differences in program implementation in terms of the percentage of lots greened may have contributed to the differences in effects seen across neighborhoods.

The analysis provides direct, robust evidence for a positive change in nearby property values as a result of greening vacant lots while highlighting the importance of using spatial methods. The ability to compare local and global models to determine neighborhood factors that may influence program outcomes provides additional value in helping to target future initiatives to locations where they may be most likely to succeed.

Acknowledgments

The author thanks Jeremy Mennis, the Pennsylvania Horticultural Society, and the Cartographic Modeling Lab at the University of Pennsylvania. This research was supported in part by a Doctoral Dissertation Research Grant from the U.S. Department of Housing and Urban Development and by a University Fellowship from Temple University.

Author

Megan Heckert is an assistant professor in the Department of Geography and Planning at West Chester University.

References

Ellen, Ingrid G., Michael H. Schill, Scott Susin, and Amy E. Schwartz. 2001. "Building Homes, Reviving Neighborhoods: Spillovers From Subsidized Construction of Owner-Occupied Housing in New York City," *Journal of Housing Research* 12 (2): 185–216.

Ellen, Ingrid G., and Ioan Voicu. 2005. Nonprofit Housing and Neighborhood Spillovers. Working paper. Public Law Research Paper No. 05-01. New York: New York University Law School.

Fotheringham, A. Stewart, Chris Brunsdon, and Martin E. Charlton. 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester, United Kingdom: John Wiley and Sons.

Galster, George, Peter Tatian, and John Accordino. 2006. "Targeting Investments for Neighborhood Revitalization," *Journal of the American Planning Association* 72 (4): 457–474.

Heckert, Megan, and Jeremy Mennis. 2012. "The Economic Impact of Greening Urban Vacant Land: A Spatial Difference-in-Differences Analysis," *Environment and Planning A* 44 (12): 3010–3027.

Jost, Daniel. 2010. "Thinking Big... And Small," *Landscape Architecture* March: 1–8.

Meyer, Bruce D. 1995. "Natural and Quasi-Experiments in Economics," *Journal of Business and Economic Statistics* 13 (2): 151–161.

Rosen, Sherwin. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy* 82 (1): 34–55.

Voicu, Ioan, and Vicki Been. 2008. "The Effect of Community Gardens on Neighboring Property Values," *Real Estate Economics* 36 (2): 241–283.

Wachter, Susan M., and Grace Wong. 2008. "What Is a Tree Worth? Green-City Strategies, Signaling and Housing Prices," *Real Estate Economics* 36 (2): 213–239.