A Variance Decomposition Approach for Bounding Neighborhood Effects

January 11, 2019

Abstract

Political scientists have pointed to neighborhood contexts as an important factor explaining individual voter behavior. This paper presents a variance-decomposition method for placing bounds on the impact of neighborhood effects on political attitudes. The method can be used to complement other empirical approaches used by the discipline in the study of neighborhood effects, as it provides better external validity than the typical experimental design, and less restrictive assumptions than most observational studies. Then, using data from the universe of voters in North Carolina, we present two applications that indicate that neighborhoods have little or no effect on political behaviors. We estimate that neighborhood effects explain at most 1.9% of the total variation in individual turnout decisions, and 4% of the observed variation in party registration, after accounting for the race of households. In both applications, household-specific traits are a much more relevant factor determining those individual outcomes.

Word Count: 6,811

Introduction

It is well documented that individual political behaviors are in various ways the product of their social context (Huckfeldt, Johnson, and Sprague 2004, Rolfe 2012, Sinclair 2012). Among the environments that have been explored, residential neighborhoods have been paid substantial scholarly attention (Gimpel, Dyck, and Shaw 2004, Cho, Gimpel, and Dyck 2006, McClurg 2006, Barber and Imai 2016). In this paper we outline a method for placing an upper bound on the influence of neighborhoods on political attitudes. Then, using data from over six million voters, we provide evidence that neighborhoods do not play a significant role in either individual turnout or party registration decisions.

A lengthy body of scholarship points to neighborhoods as a locus of voter engagement. Experimentally induced awareness of neighbors' turnout behavior, for example, has been shown to directly increase the propensity to vote (Gerber and Green 2000, Gerber, Green, and Larimer 2008), and generate indirect spillover effects within social networks (Sinclair, McConnell, and Green 2012). Observational work similarly suggests an important role for neighborhoods, highlighting the relationship between various characteristics, for example their racial mix (Marschall and Stolle 2004, Cho, Gimpel, and Dyck 2006, Enos 2016), partisan composition (Huckfeldt, Plutzer, and Sprague 1993, Gimpel, Dyck, and Shaw 2004, Barber and Imai 2016), or local economic conditions (Leigh 2005, Bisgaard, Dinesen, and Sønderskov 2016, Bellettini, Ceroni, and Monfardini 2016), and individual voter behaviors and attitudes.

These empirical approaches require a trade-off between often implausible experimental perturbations to neighborhood characteristics and the, understandably, difficult to meet identifying assumptions required of observational work. Both, however, are likely to exaggerate the effect of neighborhoods. On the one hand, interventions that artificially transmit information about pro-social behaviors may overstate the degree to which social pressures cause individuals to take on costly acts like voting. On the other hand, selection on observables assumptions may conflate an unobservable tendency towards homopholy with observable features of neighborhood networks.[1]

As an alternative, we outline an approach that allows us to place an upper bound on the impact of neighborhood characteristics. The downside of the method is that it cannot distinguish between multiple neighborhood traits that could conceivably affect individuals' behaviors. For example, a neighborhood's racial mix and its public transportation network might both impact individual turnout decisions. Our method would not distinguish these effects. The upside of our approach, however, is that it allows us to characterize the largest possible impact that the totality of these neighborhood features have on individuals' choices or attitudes. If this upper bound is large, our approach is not very informative. However, when the upper bound is small, it establishes an effectively null impact of neighborhood effects.

Of course, our approach is not assumption free. To obtain our upper bound we must assume that similar households are more likely to live alongside each-other in similar neighborhoods, i.e., there is a tendency towards neighborhood homophily. In other words, we assume that individuals do not sort into neighborhoods with those that are dissimilar to them. Given the substantial evidence that in the United States people live alongside others with similar cultural, social, and political preferences – potentially the consequence of intentional sorting – we view this as a a fairly benign assumption (Bishop and Cushing 2008, McDonald 2011, Sussell 2013, Motyl et al. 2014, Gimpel and Hui 2015).[2]

[1]Similarly, parallel trends assumptions are likely to conflate changes in measureable neighborhood contexts with unobservable changes driving both contextual and behavioral changes.

[2]Our assumptions is particularly mild compared to other selection on observables assumptions that require researchers to condition on all factors – observable and unobservable – that explain selection into treatment. Instead, to place our bounds we require researchers to make a *directional* assumption

Under the same assumption, we also show that researchers can place both upper and lower bounds on the impact of household characteristics. Doing so allows us to evaluate the relative impact of neighborhood and household features on political behavior. Similarly, we describe how individual covariates can be incorporated to place even more precise bounds on household and neighborhood features by accounting for traits affect political behavior, and also may drive sorting of individuals into households and neighborhoods.

To demonstrate the method we provide two related applications. First, we focus upon the impact of neighborhoods on individuals' turnout decisions. Here, we establish fairly strong evidence that neighborhoods are unrelated to the individual choice to turn out at the polls. In each national election between 2012 and 2016, neighborhood characteristics can only explain between 1.6 and 1.9% of the total variation in individual turnout. Comparing our estimated bound on neighborhood effects to the influence of household characteristics, we find that our upper-bound estimate of household effects is between 14 and 21 times larger than that on neighborhoods. Furthermore, we explore heterogeneity across neighborhood population density, share of non-white voters, median age, and income inequality. In general, the null relationship between neighborhoods and turnout is consistent and small across types of neighborhood.

Second, we estimate the upper bound of neighborhood effects on voters' registration decisions. Here, we find some evidence that neighborhoods may have some impact on registration decisions, estimating that at most 11% of the variance in party-registration can be explained by neighborhood effects. This upper bound is nearly 4.5 times smaller than our estimated upper bound on household effects. What is more, when we conduct the same exercise after accounting for individuals sorting by race, our upper bound estimate of neighborhood effects is reduced by more than half, now

about selection. In our context, we view this as empirically and theoretically well-founded.

explaining just over 4.5% of the total variation in observed partisanship.

This suggests that neighborhoods have very little impact individuals' party membership, after adjusting by the voter's race. Furthermore, it provides suggestive evidence that racial sorting by neighborhood explains a significant share of the decision to become a partisan. The United States' history of institutionalized racial segregation indicates that sharp racial boundaries persist in constraining people's neighborhood choices. The persistent impact of housing discrimination, racial segregation, and Jim Crow affect the spatial distribution of party preferences through their lasting influence on the spatial distribution of racial groups.

These findings are robust to both numerous definitions of neighborhood and operationalizations of party affiliation. In our online appendix, we provide results showing that our ceiling estimates of neighborhood effects remain qualitatively unchanged if we define neighborhood using the census block, tract, or arbitrarily defined spatial units (page f). Moreover, we demonstrate that our findings remain unchanged if we focus upon alternative measures of party registration, including focusing upon independent registrants, as well those of both main parties (page e).

The remainder of the paper proceeds as follows. In the next section we outline how we exploit a variance decomposition method to place bounds on neighborhood and household effects. Then, we detail our application and data. The following two sections present the estimation results from this exercise, placing an upper bound of neighborhoods on turnout and party registration decisions, followed by the conclusion.

Variance Decomposition and Neighborhood Effects

In this section we outline our approach for placing an upper-bound on the generalized influence of neighborhoods.[3] The desired estimate requires two assumptions. The first is that the outcome of interest is a linear function of household and neighborhood characteristics.

Assumption 1. Linearity.

$$y_{nhi} = \alpha' x_{nh} + \beta' z_n + u_{nhi} \tag{1}$$

Where for every individual i, in household h, and neighborhood n, the outcome, y_{nhi}, can be decomposed according to the equation above. Here, x_{nh} represents all observable and unobservable characteristics that are common to household members, and that influence the outcome of interest. In turn, z_n represents a vector of all observable and unobservable neighborhood characteristics; and u_{ihn} is a vector of individual-specific shocks that are orthogonal to both the traits shared by the household and neighborhood characteristics. Neighborhood effects are broadly defined here as the vector of coefficients β. This implies that there could be multiple, and potentially opposing, neighborhood traits that simultaneously influence individual outcomes, and could be the cause of such 'effects'.

Any attempt to examine neighborhood effects by estimating equation 1 inevitably stumbles upon one major challenge: we never observe all relevant variables that influence the household's neighborhood choice. If any unobserved variable is correlated

[3]The methodology follows from Solon, Page, and Duncan (2000) and has been applied primarily in two literatures in economics: (1) the study of effects of growing up in different neighborhoods on future wage outcomes (Page and Solon 2003, Oreopoulos 2003), and (2) the study of teacher impact on educational outcomes (Hoffmann and Oreopoulos 2009).

with the outcome of interest, which is often the case when actors are strategically selecting into neighborhoods, we cannot estimate the causal neighborhood effects (β) without bias.

Even if we could observe exogenous variation in some neighborhood characteristics that 'matters', researchers may still be interested on the overall, aggregate influence of social contexts like neighborhoods.[4] In other words, the existing research on neighborhood effects typically attempts to answer the following question: which characteristics of a neighborhood affect political behavior, and how? In contrast, the broader question we attempt to answer is: do neighborhoods influence political behavior at all? The latter is not only an interesting substantive question, but also one that can be answered with much less restrictive assumptions than the former requires, whether or not the entire set of relevant characteristics of neighborhoods are observed.

With equation 1, we derive the covariance between two individuals in the same neighborhood but from different households, the neighbor covariance, given by:

$$Cov(y_{nhi}, y_{nh'i'}) = Cov(\alpha' x_{nh}, \alpha' x_{nh'}) + Var(\beta' z_n) + 2Cov(\alpha' x_{nh}, \beta' z_n) \qquad (2)$$

The overall influence of neighborhoods is captured by the second and third components of this covariance. Where $Var(\beta' z_n)$ represents the share of variance explained solely by neighborhood effects, and the term $2Cov(\alpha' x_{nh}, \beta' z_n)$ gives the impact of these effects when interacted with household-specific characteristics. The remaining term $Cov(\alpha' x_{nh}, \alpha' x_{nh'})$ is "unambiguously the effect of similar family characteristics, rather than a neighbourhood effect" (Oreopoulos 2003). In other words, it reflects simply how similar outcomes are for individuals in the same household.

[4] Bear in mind that the bias from omitted variables remains a problem even if we replace the vector Z_n with neighborhood dummies. Although the estimates of the individual 'fixed-effects' would reflect the expected increase or decrease in the outcome variable coming from living in each particular neighborhood, they would still be biased.

The neighbor covariance does not disentangle the role of neighborhood effects from the role of similar households sorting into neighborhoods. It nevertheless gives us the maximal possible explanation power of neighborhoods under our second assumption:

Assumption 2. Homophilous Sorting.

$$Cov(\alpha' x_{nh}, \alpha' x_{nh'}) \geq 0 \quad \& \quad Cov(\alpha' x_{nh}, \beta' z_n) \geq 0 \tag{3}$$

Assuming that both the first and third terms in equation 2 are weakly positive, gives us an upper bound on $Var(\beta' z_n)$.[5] If either were negative, an estimate of the neighbor covariance would potentially understate the impact of the neighborhood effect. Substantively this assumption reflects the intuition that similar households sort into similar neighborhoods. Or, in other words, that households sort into neighborhoods based on shared preferences.

The major advantage of this approach is that it produces an aggregate measure of how much all neighborhood characteristics matter in determining the outcome, including those that we never actually observe. A limitation, however, is that it gives us only an upper-bound, which is informative only if the estimated value is small relative to the total variance of the outcome. In other words, a very small neighbor covariance would definitively indicate that neighborhoods do not account for much of the variation in the outcome of interest. In contrast, a large covariance would be inconclusive, as it could also be a consequence of sorting of like-minded households into the same neighborhoods.

[5] By construction $Var(\beta' z_n)$ is always nonnegative.

Neighborhood vs. Household Effects

To better gauge the substantive magnitude of the total impact of neighborhoods, one might want to compare the upper-bound for neighborhood traits to the effect of household traits. We show that under the same set of assumptions, bounds on household effects can be characterized. To start, we define the family covariance, the covariance in outcomes across individuals within the same household. This is given by:

$$Cov(y_{nhi}, y_{nhi'}) = Var(\alpha' x_{nh}) + Var(\beta' z_n) + 2Cov(\alpha' x_{nh}, \beta' z_n) \tag{4}$$

Notice that the covariance above is identical to the neighbor covariance, with the exception of the first term. Here the term $Var(\alpha' x_{nh})$ represents the influence of all characteristics that are common to the individual's household, which could be interpreted as the aggregated household effects. Under Assumption 2, the family covariance is the upper bound of household effects. When it is larger than the neighbor covariance, it can be interpreted to mean that outcomes are better explained by household effects than by the effect of neighborhood features.

Under the same assumption as before, we can also estimate a lower bound for the influence of household effects. This is given by the difference between equation 2 and equation 4, as shown below:

$$Var(\alpha' x_{nh}) - Cov(\alpha' x_{nh}, \alpha' x_{nh'}) \tag{5}$$

This term is positive when individuals within the same household are more similar than those within the same neighborhood, which is shown to be the case in both our empirical applications. Again, the usefulness for applied researchers of these bounds are limited by their magnitude. A large range provides little information about the relative impact of household characteristics. On the other-hand, tight bounds, suggest neighborhood effects play little role in determining the outcome.

Increasing Precision with Covariates

Researchers frequently observe additional individual characteristics that influence both outcomes and selection into neighborhoods. We show that one can obtain a more precise estimate of the upper-bound derived by accounting for these covariates. As a running example, assume we observe the race (r_{nhi}) of each individual. In Equation 6 we decompose our outcome into the effects of race ($\gamma' r_{nhi}$), and an orthogonal component e_{nhi}.

$$y_{nhi} = \gamma' r_{nhi} + e_{nhi} \tag{6}$$

This orthogonal component still includes other household characteristics, neighborhood characteristics, and a true individual-specific shock as shown in Equation 1. It could be written as $e_{nhi} = \alpha' x_{nh} + \beta' z_n + u_{nhi}$. Again, because of omitted variables, estimating this equation directly would result in a bias. Still, we can derive the new neighbor covariance as follows:[6]

$$Cov(y_{nhi}, y_{nhi'}) = Cov(\gamma' r_{nhi}, \gamma' r_{nh'i'}) + Cov(e_{nhi}, e_{nh'i'}) + 2Cov(\gamma' r_{nhi}, e_{nh'i'}) \tag{7}$$

The first term in the equation represents the share of the variation in outcomes directly explained by the covariate (race, in this case). The combination of the second and third terms are roughly equivalent to the upper-bound on neighborhood effects derived before, but now they exclude any effects coming from the individual's race.[7]

[6] By analogous derivation we can also obtain the family covariance.

[7] This is in fact a slightly more conservative estimate. Assuming here that w_{nh} is simply the vector x_{nh} excluding race, the second and third terms can be written as: $Cov(\alpha' w_{nh}, \alpha' w_{nh'}) + Var(\beta' z_n) + 2Cov(\alpha' w_{nh}, \beta' z_n) + 2Cov(\gamma' r_{nhi}, w_{nh'}) + 2Cov(\gamma' r_{nhi}, z_n)$. Although the term $2Cov(\gamma' r_{nhi}, x_{nh'})$ does

The advantage is that all three main components of the neighborhood covariance can be estimated separately, which gives us a more precise estimate of the upper-bound in question. We illustrate the usefulness of this specific adjustment in our second empirical application.

Estimation Procedure

Following Solon, Page, and Duncan (2000), for N neighborhoods, H_n households in each neighborhood n, and I_{nh} individuals in household h, we estimate the total variance of the residualized outcome variable as follows:

$$\hat{\sigma}^2 = \frac{\sum_{n=1}^{N} \sum_{h=1}^{H_n} \sum_{i=1}^{I_{nh}} y_{nhi}^2}{\sum_{n=1}^{N} \sum_{h=1}^{H_n} I_{nh}} \tag{8}$$

For household h in neighborhood n with I_{nh} registered voters, the number of different pairs of individuals within the household is given by $P_{nh} = \frac{I_{nh}(I_{nh}-1)}{2}$. Accordingly, the family covariance from Equation 4, for household h, can be estimated as:

$$\hat{fc}_{nh} = \frac{\sum_{i \neq i'}^{P_{nh}} y_{nhi} y_{nhi'}}{P_{nh}} \tag{9}$$

We estimate the overall family covariance in the sample by taking the weighted average of the household-specific covariances over all households (H_n), in all neighborhoods (N), as follows:

$$\hat{fc} = \frac{\sum_{n=1}^{N} \sum_{h=1}^{H_n} W_{nh} \hat{fc}_{nh}}{\sum_{n=1}^{N} \sum_{h=1}^{H_n} W_{nh}} \tag{10}$$

Where W_{nh} is the weight assigned to each household covariance. The simplest version of this estimator assigns equal weights to all households. Solon, Page, and Duncan

not measure any neighborhood effects, it is likely to be positive if similar individuals sort into the same households, and it cannot be estimated separately.

(2000) argue that this estimator is inefficient because it underweights households containing more information (i.e., families with more members). In our baseline specification we follow their approach and weight households by the square root of household size in order to avoid overweighting larger households. In the appendix (page c) we also provide the results of our first empirical application for two alternative weighting schemes: equal weights or weighted by household size. We also provide ninety-five percent confidence intervals for our estimated covariances using bootstrap with 200 repetitions.

Similarly, the total number of distinct pairs between individuals in households h and h', in neighborhood n, is given by $I_{nh}I_{nh'}$. Thus, the neighbor covariance for this specific pair of households is given by:

$$\hat{nc}_{nhh'} = \frac{\sum_{i=1}^{I_{nh}} \sum_{i'=1}^{I_{nh'}} y_{nhi} y_{nh'i'}}{I_{nh} I_{nh'}} \tag{11}$$

Again, we can estimate the overall average neighbor covariance by taking the weighted average of the household-specific covariances over all households in the sample, as shown in the equation below, where HH_n is the number of distinct pairs of households in each neighborhood. We follow the same weighting scheme as before. Ninety-five percent confidence regions are estimated by bootstrap.

$$\hat{nc} = \frac{\sum_{n=1}^{N} \sum_{h=1}^{HH_n} W_{nhh'} \hat{nc}_{nhh'}}{\sum_{n=1}^{N} \sum_{h=1}^{HH_n} W_{nhh'}} \tag{12}$$

Data for the Applications

We provide a pair of applications using the universe of registered voters in North Carolina,[8] which we obtained from the State Bureau of Elections's voter registration file. These data were downloaded from the NBSE's website[9] in April 2017 and describe the party registration, self-identified race, turnout in recent election, and address of over 6 million registered voters.[10]

Since it is the smallest unit of geography for which the census department records measures of income and inequality, we treat the block-group as defined in the 2010 census as our primary measure of neighborhood. In our sample, there are 6,107 block groups for which data is available. In Table 1 we provide descriptive statistics for individual voters by party registration. Democrats are more likely to be female, live in a densely populated poor neighborhood with a larger share of black neighbors, and are also considerably more likely to be black. On the other hand, they are less likely to turnout to vote than Republicans.

In Table 2 we provide descriptive statistics at the neighborhood level. The average neighborhood has 862 voters distributed in 340 households, 40% of which are registered Democrat and 30% Republican. It has median income of US$48,000 dollars, and a 32% share of non-white voters. Between 60 and 70% of registered voters in a neighborhood turnout to vote in presidential elections of 2012 and 2016, while only 42% turned out in the 2014 midterm elections.

[8] By construction, we can only analyze the turnout of registered voters with this data, and therefore our results do not speak to the probability of citizens registering to vote. This problem has been recently examined in Nyhan, Skovron, and Titiunik (2017).

[9] http://dl.ncsbe.gov/

[10] We focus upon North Carolina for two reasons. First, it is one of the few states who record and release data on the racial backgrounds of voters. Second, among these states, North Carolina is the only one that is roughly demographically representative of the national electorate in terms of party-registration. The other states that record and make available the race of registered voters are South Carolina, Georgia, Louisiana and Alabama. Pennsylvania collects the data but does not disclose it.

Table 1: Descriptive Statistics for Individual Voters

Variable	*Democrats*	*Republicans*	*Other*
Age (mean)	49.8	50.1	43.4
Female, share	0.563	0.490	0.473
White, share	0.457	0.942	0.758
Black, share	0.465	0.017	0.111
Voted in 2012, share	0.662	0.675	0.509
Voted in 2014, share	0.457	0.492	0.331
Voted in 2016, share	0.702	0.765	0.645
Observations	2,003,023	1,665,670	1,595,205

Table 2: Descriptive Statistics for Neighborhoods

Variable	*Mean*	*S.D.*	*Min*	*Max*
Households with Voters	340.3	214.1	2.0	2891.0
Registered Voters	861.7	541.2	8.0	7088.0
Pop. Density, km2	514.0	663.4	0.0	9121.3
Gini	0.405	0.077	0.090	0.803
Income, US$ ’000	48.3	25.3	2.5	250.0
Non-White, share	0.319	0.269	0.000	1.000
Democrats, share	0.405	0.182	0.000	0.926
Republicans, share	0.303	0.150	0.006	0.757
Independent, share	0.292	0.073	0.060	0.675
Voted in 2012, share	0.614	0.089	0.000	0.818
Voted in 2014, share	0.422	0.102	0.000	0.786
Voted in 2016, share	0.691	0.088	0.077	0.900

Note: Includes 6,107 census block-groups.

Application 1: Neighborhood Effects and Turnout

In our first application, we estimate the upper bound for the influence of neighborhoods on turnout in the elections of 2012, 2014, and 2016. A substantial body of scholarship links different dimensions of neighborhoods to individual level turnout choices. Direct social pressure from neighbors (Gerber and Green 2000, Gerber, Green, and

Larimer 2008), out group or in group motivations driven by the presence or absence of co-ethnics (Cho, Gimpel, and Dyck 2006, Enos 2016) and co-partisans (Gimpel, Dyck, and Shaw 2004), and even the physical geography of neighborhoods have each been proposed to impact turnout (Brady and McNulty 2011). In our sample, we find no evidence that neighborhoods have a meaningful aggregate influence on turnout. Furthermore, when we explore heterogeneity in our upper bound across different types of neighborhood, we find that this weak relationship is consistent across broad classes of neighborhood.

We treat a dummy that takes on a value of one if a given voter voted in the election and zero otherwise as our main outcome of interest.[11] Measures of uncertainty are obtained by bootstrapping our estimates using 200 repetitions, drawing samples with equal number of neighborhoods with replacement, and randomizing over neighborhoods.

Results for North Carolina Turnout

In Table 3 we decompose the total variance of turnout into two components: family covariance and neighbor covariance. The family and neighbor correlations are obtained by dividing the covariances by the total variance. This reflects the share of total variation in turnout that is explained by the covariance between household members and neighbors, respectively. The upper bound on the explanatory power of neighborhood effects is given by the neighbor correlation. Similarly, the lower bound for the explanatory power of household effects is given by the last row, labeled *Difference*.[12]

In the three elections under study we establish two distinctive patterns. First, neighborhood effects account for at most between 1.6 and 1.9% of the total variation

[11]Across specifications we adjust this variable for the influence of gender and age by first regressing our outcome on these covariates. Our outcome variable is always the residual of this initial regression.

[12]It is the family correlation subtracted by the neighbor correlation.

Table 3: Decomposition of the Variance in Turnout

	2012	2014	2016
Total Variance	0.199	0.207	0.194
	(0.000)	(0.000)	(0.000)
Family Covariance	0.063	0.082	0.053
	(0.000)	(0.000)	(0.000)
Neighbor Covariance	0.003	0.004	0.004
	(0.000)	(0.000)	(0.000)
Family Correlation	0.316	0.394	0.274
Neighbor Correlation	0.016	0.019	0.019
Difference	0.300	0.374	0.254

Note: General elections in 2012 and 2016, midterms in 2014. Family and neighborhood correlations are calculated by dividing the value of the corresponding family or neighbor covariance by the total variation in y, given in the first row of the table for each election.

in individual turnout. This indicates that, on average, neighborhood-specific characteristics explain only a tiny share of the variation in this outcome. We re-emphasize that this estimate cannot be direct applied to the causal effects of specific neighborhood traits on turnout (e.g. the effect of racial diversity). However, it nevertheless informs causal studies that examine specific neighborhood characteristics. For example, the magnitude of our upper bound is small but in line with the 'cost of voting' effects of rearranging polling station locations estimated by Brady and McNulty (2011). As for studies that find significantly higher magnitudes for singular neighborhood effects, our low upper bound suggests that the experimental variation used to estimate them is relatively rare in occurrence, as in (Enos 2016), and in practice explains little of the aggregate turnout figures.

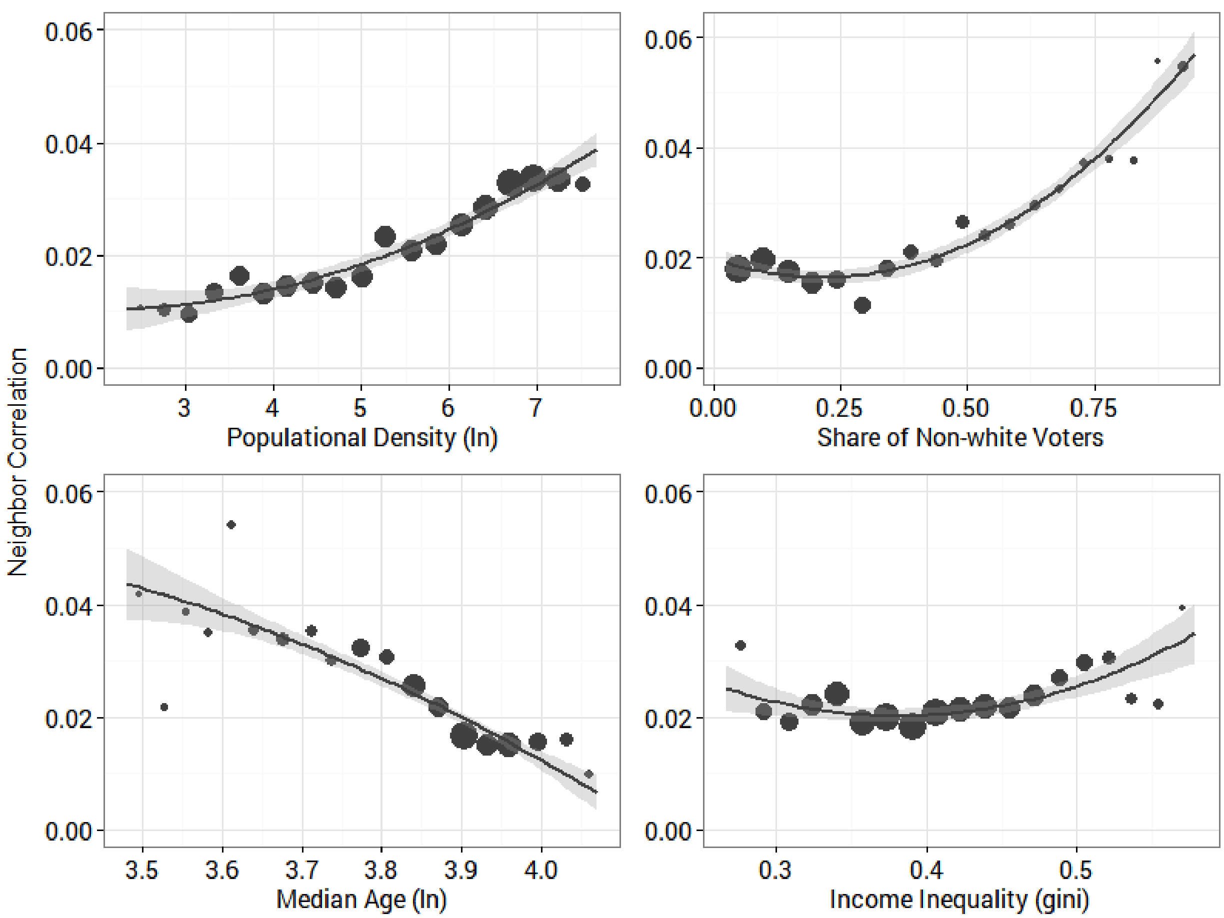
The second pattern we establish is the high lower bound for the explanatory power of household effects on turnout. Accordingly, household-specific traits explain between 25 and 37% of the total variation in the individual's decision to vote. It is also interesting that this estimate is somewhat higher in the midterm election of 2014, suggesting that voter mobilization is even more dependent on household background in elections that, in general, elicit less popular appeal. This high lower bound is not surprising since families share both genetic and social traits that contribute to a high within-household correlation. Indeed, a wide range of empirical studies give evidence of assortive mating based upon political attitudes and preferences (Alford et al. 2011, Huber and Malhotra 2012, Klofstad, McDermott, and Hatemi 2013). Spouses' political attitudes display correlations that are as strong or stronger than nearly all other social and biometric inter-spousal traits (Alford et al. 2011). Similarly, the persistence of political attitudes from parents to children is one of the most well documented regularities in the study of political behavior (Jennings, Stoker, and Bowers 2009, Jennings and Niemi 2015).

Heterogeneity of the Upper Bound Estimates

To provide some additional insight on the nature of neighborhood effects and their overall influence in turnout, we explore how our upper bound estimate varies across four neighborhood features: population density, racial composition, median age and income inequality. Using the 2016 election data we plot these bounds in Figure 1. Across each feature, the range of bounds remains quite small, with maximal estimates just under 6% of the total variation in turnout and minimal estimates just under 1%.

The first plot (top-left) indicates that turnout for individuals living in more dense neighborhoods is more subject to the influence of neighborhood effects. This is in line with the argument that highly urbanized, dense, neighborhoods typically face more fluctuations in the cost of voting due to factors like traffic or waiting times at the polling station. A similar result is found in the second plot (bottom-left) which indicates that neighborhood factors play a larger role in younger neighborhoods. Perturbations to neighborhood costs are unlikely to significantly affect the turnout decisions for groups that have a high baseline turnout rate, which is the case of the older population. Accordingly, the young population seems to be marginally more sensitive to fluctuations in the underlying incentives to turnout.

Figure 1: Upper Bounds by Neighborhood Traits in the 2016 Election



Note: The 6,103 neighborhood-level observations are grouped in 20 different groups for the purpose of plotting the points, according to neighborhood-specific value for the variable shown in the x-axis. The size of each point is proportional to the number of neighborhoods within that group. The lines show a quadratic fit.

It is also seems that turnout in these neighborhoods is more sensitive to higher diversity in terms of race. The third plot (top-right) shows that as the share on non-white voters in a neighborhood increases, so does our upper bound estimate. However, it should be noted that there are very few voters in neighborhoods with greater than 50% non-white voters – the region where neighborhood effects are largest. Moreover, these tend to be the neighborhoods with the greatest population density. Finally, the fourth plot (bottom-right) shows that the income profile of neighborhoods plays virtually no role in determining the extent to which neighborhood effects influence turnout.

All-in, this examination of the heterogeneity in our estimates indicates that even in the neighborhoods that show a higher degree of influence in the turnout of its residents, neighborhood effects explain at most around 5% of the variation in this variable. Nevertheless, this suggests that further research on specific effects of neighborhoods on turnout would benefit from focusing on more urbanized, younger, and racially diverse neighborhoods, which seem to be the most sensitive to shifts in the incentives to vote.

Application 2: Neighborhood Effects and Partisanship in North Carolina

In our second application we estimate an upper bound for the influence of neighborhood effects on the choice of partisanship for registered voters. More than ever, Republican and Democrat serve as social identities similar to ethnicity or religion in terms of their impact on attitudes not directly related to politics. Partisanship of this sort creates an out-group about whom survey respondents increasingly ascribe undesirable attributes (Iyengar and Westwood 2015, Levendusky et al. 2018). As partisanship has becomes a more salient trait of neighborhoods, it seems reasonable to evaluate the influence that neighborhoods may have on the individual voter's decision to register as Democrat or Republican.

Again, we use the same North Carolina data as in the last section. However, in this section we highlight how the inclusion of observable individual characteristics that influence partisanship choice, race in this case,[13] helps us to obtain a significantly more precise estimate of the upper-bound, as explained on page 9.

Again, we define neighborhood as the 6,107 census block groups, and treat a dummy that takes on a value of one if a given voter is a registered Democrat as our main outcome of interest. In the appendix (Table A.3, page e) we provide estimates where we treat Republican registration or third-party registration as the outcome.[14] Under these alternative specifications the upper bound estimates for the effects of sorting are even smaller. As such, the baseline specification presented in this paper is the most conservative estimate of this upper bound.

Results for North Carolina Partisanship

We present the estimation results on Table 4. In column (1) we decompose the total variance of the outcome into two components: family covariance and neighbor covariance. This column represents exactly the same decomposition done in the last application, and shown in Table 3. Accordingly, we find that neighborhood effects explain at most 10.9% of the variation in partisanship in a neighborhood while household effects explain at least 36.0% of the variation in the same outcome (e.g. the the lower bound on the household effect).

Following Equation 6, we further decompose partisanship into one component explained by the individual's race, and the residual, which is explained by other household and neighborhood characteristics, and an i.i.d. shock. Column (2) presents the decomposition of variation in party registration that is explained by race. Race alone

[13]There is substantial evidence that race is central to partisanship in the United States (Hutchings and Valentino 2004).

[14]In this Table, we treat voters registered for parties other than the Democrat or Republican parties as independent.

Table 4: Decomposition of the Variance in Party Registration

	y	$\gamma' r$	e	$2Cov(\gamma' r, e)$	
	(1)	(2)	(3)	(4)	(3)+(4)
Total Variance	0.233	0.055	0.178		
	(0.001)	(0.001)	(0.000)		
Family Covariance	0.109	0.047	0.055	0.008	0.062
	(0.000)	(0.001)	(0.000)	(0.000)	
Neighbor Covariance	0.025	0.015	0.005	0.006	0.011
	(0.001)	(0.000)	(0.000)	(0.000)	
Family Correlation	0.469	0.200	0.236	0.033	0.268
Neighbor Correlation	0.109	0.062	0.022	0.024	0.046
Difference	0.360	0.138	0.214	0.214	0.222

Note: In the first column (y), we decompose the total variance in party registration into the components explained by family and neighborhood. In columns 2-4, we further decompose the total variance components into the share of party registration explained by race ($\gamma' r$), the share explained by all other factors that are orthogonal to race (e), and their covariance ($2Cov(\gamma' r, e)$). Family and neighborhood correlations are calculated by dividing the value of the corresponding family or neighbor covariance by the total variation in y, given in the first cell of the table.

explains roughly one-quarter of the total variation in party choice. What is more, the covariance of race between neighbors accounts for 43% of our estimated upper bound on neighborhood effects.

Columns (3) and (4) help us understand the consequences of adjusting for race. In column (3) we decompose the remaining variation in all unobserved factors that affect partisan choice and are orthogonal to race. In column (4), the effects of these unobserved factors when interacted with the individual's race. It is easy to see that these columns together now represent a more precise, albeit still conservative, estimate of the upper bound on the influence of neighborhood effects on partisanship, after excluding the direct effect of the voter's race. The sum of these columns is shown in column (5).

After adjusting for race, the estimated upper bound on the explanatory power of neighborhood effects falls to 4.6% (from 10.9% before). We emphasize that the 4.6% point estimate pertains to the case in which partisan geographical sorting in nonexistent, i.e., a world where the individuals' choices of neighborhood are completely uncorrelated with their decision to register for a given political party. Given that this is a highly implausible scenario, we are confident to say almost none of the observed variation in partisanship can be explained by neighborhood effects, at least in the context of North Carolina.

The lower bound for the explanatory power of household effects also falls after the adjustment, but it remains high at 22.2% of the total variation in partisan choice. As in the first application, these results emphasize the role of household unique characteristics as one of the main determinants of political attitudes in this context. In all, our findings suggest that uncovering the social conditions that drive the construction of homophilous household units in respect to political preferences is a promising avenue for future research.

Conclusion

In this paper we have presented a method for placing bounds on the influence of neighborhood effects on individual political behaviour. The method relies on decomposing the variance of the outcome of interest into the covariance between members of the same household, and residents of the same neighborhood. Then, using data from the universe of voters in North Carolina we provided two applications that indicate neighborhoods have little or no effect on political behaviors.

In our applications, we find no evidence that neighborhood characteristics affect turnout decisions, estimating that, at most, between 1.6 and 1.9% of the total variation in individual turnout decisions is driven by these features. Furthermore, we find at most 11% of the observed variation in party registration can be explained by neighborhood characteristics. Once we account for race this upper bound falls by more than half to just over 4%. Additionally, we find that household-specific characteristics explain party choice far better than neighborhood features, indicating that the underlying social factors that lead individuals to sort into households with similar political views play a much more substantial role in partisan choice than spatial sorting. This suggests that uncovering the social conditions that drive the construction of homophilous household units in respect to political preferences and behaviors is a promising avenue for future research.

Our methodology, rather than identifying specific characteristics of neighborhoods that influence individual behavior, examines the aggregate explanation power of neighborhoods effects on the outcomes of interest. Thus, rather than being a substitute for the experimental and observational research methods commonly used to study effects of neighborhoods in the political science literature, we recommend that researchers use it as a complement. Accordingly, because households seldom choose a neighborhood at random, any study that examines effects of neighborhood characteristics with observational data is likely to be subject to omitted variable bias, and typically relies on unrealistic 'selection on observables' assumptions. In contrast, our estimated upper bound relies on two very benign assumptions, and can be applied whether or not the researcher observes all neighborhood traits that impact behavior.

As for experimental approaches, they are often plagued by external validity, given that they typically rely on artificial perturbations to neighborhood characteristics that are rarely observed in the aggregate data. Thus, sometimes strong effects found for any singular neighborhood trait might overstate the relevance of neighborhoods in the overall variation of the outcome variable. Even though the method in this article does not allow us to disentangle the specific causal effects of any singular neighborhood trait, it provides a useful tool to inform experimental results of the actual broader, aggregate impact of neighborhoods in the outcome of interest.

Finally, this method is also flexible enough to be applied to various definitions of neighborhoods. It is easy to see that the framework can be employed whenever the outcome variable can be plausibly written as a linear function of characteristics of spatially nested units of analysis. For example, if the researcher is interested in comparing the relative influence of 'state-effects' and 'district effects' on a certain political behavior, the methodology can be applied to estimate an upper bound based on the covariances between (1) individuals in the same state but different districts (state covariance), and (2) individuals living in the same district (district covariance).

References

Alford, John R, Peter K Hatemi, John R Hibbing, Nicholas G Martin, and Lindon J Eaves. 2011. “The politics of mate choice.” *The Journal of Politics* 73 (2):362–379.

Barber, Michael and Kosuke Imai. 2016. “Estimating neighborhood effects on turnout from geocoded voter registration records.” *Princeton, NJ: Mimeo. http://goo.gl/qPPLZz* .

Bellettini, Giorgio, Carlotta Berti Ceroni, and Chiara Monfardini. 2016. “Neighborhood heterogeneity and electoral turnout.” *Electoral Studies* 42:146–156.

Bisgaard, Martin, Peter Thisted Dinesen, and Kim Mannemar Sønderskov. 2016. “Reconsidering the Neighborhood Effect: Does Exposure to Residential Unemployment Influence Voters’ Perceptions of the National Economy?” *The Journal of Politics* 78 (3):719–732.

Bishop, Bill and Robert Cushing. 2008. “The big sort.” *Why the Clustering of America is Tearing Us Apart. NY: Houghton Miffl in* .

Brady, Henry E. and John E. McNulty. 2011. “Turning Out to Vote: The Costs of Finding and Getting to the Polling Place.” *The American Political Science Review* 105 (1):115–134.

Cho, Wendy K. Tam, James G. Gimpel, and Joshua J. Dyck. 2006. “Residential Concentration, Political Socialization, and Voter Turnout.” *Journal of Politics* 68 (1):156–167.

Enos, Ryan E. 2016. “What the Demolition of Public Housing Teaches Us about the Impact of Racial Threat on Political Behavior.” *American Political Science Review* 60 (1):123–142.

Gerber, Alan S. and Donald P. Green. 2000. “The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment.” *American Political Science Review* 94 (3):653–663.

Gerber, Alan S., Donald P. Green, and Christopher W. Larimer. 2008. “Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment.” *American Political Science Review* 102 (1):33–48.

Gimpel, James G., Joshua J. Dyck, and Daron R. Shaw. 2004. “Registrants, Voters, and Turnout Variability across Neighborhoods.” *Political Behavior* 26 (4):343–375.

Gimpel, James G and Iris S Hui. 2015. “Seeking politically compatible neighbors? The role of neighborhood partisan composition in residential sorting.” *Political Geography* 48:130–142.

Hoffmann, Florian and Philip Oreopoulos. 2009. "Professor Qualities and Student Achievement." *The Review of Economics and Statistics* 91 (1):83–92.

Huber, Gregory and Neil Malhotra. 2012. "Political sorting in social relationships: Evidence from an online dating community." In *Annual Meeting of the American Political Science Association, New Orleans, LA*.

Huckfeldt, Robert, Paul E Johnson, and John Sprague. 2004. *Political disagreement: The survival of diverse opinions within communication networks*. Cambridge University Press.

Huckfeldt, Robert, Eric Plutzer, and John Sprague. 1993. "Alternative contexts of political behavior: Churches, neighborhoods, and individuals." *The Journal of Politics* 55 (2):365–381.

Hutchings, Vincent L and Nicholas A Valentino. 2004. "The centrality of race in American politics." *Annu. Rev. Polit. Sci.* 7:383–408.

Iyengar, Shanto and Sean J Westwood. 2015. "Fear and loathing across party lines: New evidence on group polarization." *American Journal of Political Science* 59 (3):690–707.

Jennings, M Kent and Richard G Niemi. 2015. *Political character of Adolescence: The Influence of Families and schools*. Princeton University Press.

Jennings, M Kent, Laura Stoker, and Jake Bowers. 2009. "Politics across generations: Family transmission reexamined." *The Journal of Politics* 71 (3):782–799.

Klofstad, Casey A, Rose McDermott, and Peter K Hatemi. 2013. "The dating preferences of liberals and conservatives." *Political Behavior* 35 (3):519–538.

Leigh, Andrew. 2005. "Economic voting and electoral behavior: How do individual, local, and national factors affect the partisan choice?" *Economics & Politics* 17 (2):265–296.

Levendusky, Matthew, Christopher McConnell, Yotam Margalit, and Neil Malhotra. 2018. "The Economic Consequences of Partisanship in a Polarized Era." *American Journal of Political Science* 62 (1):5–18.

Marschall, Melissa J and Dietlind Stolle. 2004. "Race and the city: Neighborhood context and the development of generalized trust." *Political behavior* 26 (2):125–153.

McClurg, Scott D. 2006. "Political disagreement in context: The conditional effect of neighborhood context, disagreement and political talk on electoral participation." *Political Behavior* 28 (4):349–366.

McDonald, Ian. 2011. "Migration and sorting in the American electorate: Evidence from the 2006 Cooperative Congressional Election Study." *American Politics Research* 39 (3):512–533.

Motyl, Matt, Ravi Iyer, Shigehiro Oishi, Sophie Trawalter, and Brian A Nosek. 2014. "How ideological migration geographically segregates groups." *Journal of Experimental Social Psychology* 51:1–14.

Nyhan, Brendan, Christopher Skovron, and Rocío Titiunik. 2017. "Differential Registration Bias in Voter File Data: A Sensitivity Analysis Approach." *American Journal of Political Science* 61 (3):744–760.

Oreopoulos, Philip. 2003. "The long-run consequences of living in a poor neighborhood." *The quarterly journal of economics* 118 (4):1533–1575.

Page, Marianne E and Gary Solon. 2003. "Correlations between brothers and neighboring boys in their adult earnings: The importance of being urban." *Journal of Labor Economics* 21 (4):831–855.

Rolfe, Meredith. 2012. *Voter turnout: A social theory of political participation*. Cambridge University Press.

Sinclair, Betsy. 2012. *The social citizen: Peer networks and political behavior*. University of Chicago Press.

Sinclair, Betsy, Margaret McConnell, and Donald P Green. 2012. "Detecting spillover effects: Design and analysis of multilevel experiments." *American Journal of Political Science* 56 (4):1055–1069.

Solon, Gary, Marianne E Page, and Greg J Duncan. 2000. "Correlations between neighboring children in their subsequent educational attainment." *Review of Economics and Statistics* 82 (3):383–392.

Sussell, Jesse. 2013. "New support for the big sort hypothesis: an assessment of partisan geographic sorting in California, 1992–2010." *PS: Political Science & Politics* 46 (04):768–773.

Appendix for Online Publication

The appendix provides Tables illustrating the results for both applications under slightly different empirical specifications. All details are provided in the Table notes.

Contents

List of Tables

Table A.1: Decomposition of the Variance in Turnout with Alternative Weights

	2012	2014	2016
		Weights: Equal	
Total Variance	0.199	0.207	0.194
	(0.000)	(0.000)	(0.000)
Family Covariance	0.071	0.093	0.058
	(0.000)	(0.000)	(0.000)
Neighbor Covariance	0.003	0.004	0.004
	(0.000)	(0.000)	(0.000)
Family Correlation	0.358	0.447	0.297
Neighbor Correlation	0.017	0.020	0.021
Difference	0.341	0.426	0.276
		Weights: Neighborhood Size	
Total Variance	0.199	0.207	0.194
	(0.000)	(0.000)	(0.000)
Family Covariance	0.016	0.014	0.022
	(0.004)	(0.005)	(0.007)
Neighbor Covariance	0.003	0.004	0.003
	(0.000)	(0.000)	(0.000)
Family Correlation	0.079	0.070	0.116
Neighbor Correlation	0.015	0.018	0.018
Difference	0.064	0.051	0.098

Note: General elections in 2012 and 2016, midterms in 2014. Family and neighborhood correlations are calculated by dividing the value of the corresponding family or neighbor covariance by the total variation in y, given in the first row of the table for each election.

Table A.2: Party Registration Results by Weighting Scheme

	(1)	(2)	(3)
Total Variance	0.233	0.233	0.233
	(0.001)	(0.001)	(0.001)
Family Covariance	0.109	0.112	0.053
	(0.001)	(0.000)	(0.016)
Neighbor Covariance	0.025	0.025	0.026
	(0.001)	(0.001)	(0.001)
Family Correlation	0.469	0.480	0.227
Neighbor Correlation	0.109	0.106	0.113
Difference	0.330	0.374	0.114

Column (1) represents our base case scenario, where the weight put on households (for family covariance) and household-pairs (for neighbor covariance) is the square root of their respective number of observations. Column (2) is the specification with equal weights, and Column (3) the specification that uses the number of observations for each houshold or household-pair (i.e. it overweights larger households). As before, both family and neighborhood correlations are calculated by dividing the value of the corresponding family or neighbor covariance by the total variation in y, shown in the first cell of the table.

Table A.3: Party Registration Results by Different Outcomes

	(Democrat)	(Republican)	(Other)
Total Variance	0.233	0.233	0.233
	(0.001)	(0.000)	(0.000)
Family Covariance	0.109	0.097	0.052
	(0.000)	(0.000)	(0.000)
Neighbor Covariance	0.025	0.017	0.004
	(0.001)	(0.000)	(0.000)
Family Correlation	0.469	0.419	0.222
Neighbor Correlation	0.109	0.071	0.017
Difference	0.330	0.348	0.205

The first column is our base case scenario where the outcome variable is a dummy that equals one if the party of choice in Democrat. The second column does the same exercise for Republicans, and the third for any other choice (i.e the vast majority of others are independent). As before, both family and neighborhood correlations are calculated by dividing the value of the corresponding family or neighbor covariance by the total variation in y, shown in the first cell of the table.

Table A.4: Party Registration Results by Different Definitions of Neighborhoods

	Block-Group (Census)	Tract (Census)	Small Grid (2 sq miles)	Large Grid (12 sq miles)
Total Variance	0.233	0.233	0.233	0.233
	(0.001)	(0.001)	(0.001)	(0.001)
Family Covariance	0.109	0.109	0.109	0.109
	(0.000)	(0.001)	(0.001)	(0.001)
Neighbor Covariance	0.025	0.012	0.029	0.024
	(0.001)	(0.001)	(0.001)	(0.002)
Family Correlation	0.469	0.468	0.469	0.468
Neighbor Correlation	0.109	0.054	0.126	0.102
Difference	0.330	0.414	0.343	0.366

The first column is our base case scenario where neighborhoods are the Census block-groups. The second column shows the results for Census tracts and the third and fourth columns show the results for neighborhoods defined over arbitrary squares of size 0.02 x 0.02 degrees (2 square miles), and of size 0.05 x 0.05 degrees (12 square miles), respectively. As before, both family and neighborhood correlations are calculated by dividing the value of the corresponding family or neighbor covariance by the total variation in y, shown in the first cell of the table.