Party, Race, and Neighborhood:

An Upper Bound Estimate of Geographic Partisan Sorting

Scott F Abramson* & Anderson Frey †

May 22, 2018

Abstract

This paper evaluates the degree to which partisanship is influenced by the sorting of like-minded individuals into neighborhoods. While many view high levels of contemporary social and political polarization as the consequence of increased geographic concentration of partisans, we place an upper bound estimate on this effect and show that sorting of individuals into neighborhoods explains *at most* 11% of the observed variation in party registration. Once we account for race this upper bound estimate is reduced to just over 4%, indicating that if individuals do sort into neighborhoods, they do so by race. However, given the history of institutionalized racial segregation in the United States we are cautious in interpreting our upper bound estimate as evidence of purposeful racial sorting. Instead, we view our main result as the ultimate effect of a long history of geographic segregation based upon features which now systematically covary with political attitudes.

Word Count: 5,001

*Department of Political Science, The University of Rochester. sabramso@ur.rochester.edu

†Department of Political Science, The University of Rochester. anderson.frey@rochester.edu

1 Introduction

The social divisions that characterize the American electorate exhibit a remarkable degree of partisan homophily (Huckfeldt, Mendez, and Osborn 2004, Mutz 2006). More than ever, Republican and Democrat serve as social identities similar to race, ethnicity, or religion in terms of their impact on attitudes and behaviors not directly related to politics. Partisanship of this sort creates an out-group about whom survey respondents increasingly ascribe undesirable attributes and express negative attitudes towards (Iyengar and Westwood 2015, Levendusky et al. 2015). What is more, the growing isolation of Americans in co-partisan social networks may exacerbate the polarization of attitudes by removing exposure to ideas and information that fail to conform with individuals' priors (Sunstein 2009, Klar 2014).

In this paper we evaluate a form of sorting proposed as a fundamental cause of these broader trends, the sorting of individuals into neighborhoods by party or other social factors that map closely onto partisanship (Bishop and Cushing 2008, Sussell 2013, Gimpel and Hui 2015). Exploiting data describing over 6 million voters, we place an upper bound on the proportion of observed variation in partisanship explained by households sorting into neighborhoods and the proportion explained by individuals sorting into households, respectively. We show that *at most* 11% of the variance in party-registration can be explained by households sorting into neighborhoods, and we find that the underlying social factors that drive individuals to sort into households of similar partisanship are 4.4 times as large as those that drive neighborhood sorting.[1]

When we conduct the same exercise after accounting for sorting by race, our upper bound estimate of neighborhood sorting is reduced by more than half, now explaining just over 4% of the total variation in observed partisanship. What is more, after adjusting for households' race profile, the factors that explain household partisan sorting increase by over thirty percent relative to those that drive neighborhood sorting. This suggests that if individuals do sort into neighborhoods, it is for the most part a consequence of race and not party. Still, we are cautious in interpreting this as evidence of overt racial sorting. Rather, the United States' history of institutionalized racial segre-

[1] For a similar methodological application in political science see Fowler, Baker, and Dawes (2008). For related empirical applications see Solon, Page, and Duncan (2000), Page and Solon (2003), Oreopoulos (2003).

gation suggests that sharp racial boundaries persist in constraining people's neighborhood choices. As such, this result indicates that almost none of the observed variation in partisanship can be explained by partisan geographic sorting. Instead, the persistent impact of housing discrimination, racial segregation, and Jim Crow affect the spatial distribution of political partisanship through their lasting influence on the spatial distribution of racial groups.

The extant empirical evidence on partisan geographic sorting is mixed. On the one hand, survey respondents express a clear preference for neighborhoods that contain a greater proportion of co-partisans (Hui 2013, Gimpel and Hui 2015, Mummolo and Nall 2017). On the other hand, when focused upon observational measures of sorting the empirical results are inconclusive. A number of studies find evidence in favor of partisan geographic sorting, where Democratic and Republican voters select into communities that match with their partisanship (McDonald 2011, Sussell 2013, Motyl et al. 2014). Still others find little evidence of partisan sorting of this sort (Abrams and Fiorina 2012, Mummolo and Nall 2017).

It is important to note that our findings reflect an upper bound on the relationship between partisanship and neighborhood geography. That is to say, we estimate the largest supportable correlation between party registration and residential location. In reality, this correlation reflects a combination of two effects: individuals sorting into neighborhoods and, conversely, neighborhoods having some direct impact on individuals' partisan identification. We place an upper-bound on the sum of these effects and, by construction, an upper bound on each effect separately. In other words, if we assume there is no impact of neighborhood features on individuals' party-identification then we obtain an upper-bound on sorting. Of course, this is a strong assumption and in actuality the true magnitude of the selection effects we are interested in is lower than our upper-bound.

Regardless, in an absolute sense our results indicate that only a very small proportion of party affiliation is explained by the factors that drive sorting by geography. Furthermore, they suggest that this effect is even smaller when compared to the factors that explain within-household similarities in partisan choice. This is not surprising since families share both genetic and social traits that contribute to a high within-unit 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), which is widely supported by the upper bound estimates provided in this article. 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.

Our findings are robust to numerous definitions of neighborhood and operationalizations of party affiliation. In our supplemental appendix we provide results showing that our ceiling estimate of sorting remains qualitatively unchanged if we define neighborhood using the census block, tract, or arbitrarily defined spatial units. 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.

In all, our results speak to a growing literature on the political geography of voting in America. There is abundant evidence showing that political geography in the United States is typified by partisan clusters, with Democrats over-represented in urban settings and Republicans in suburban and rural locations (Reardon and Bischoff 2011, McCarty, Poole, and Rosenthal 2016). This "unintentional gerrymandering" coupled with plurality rule elections for Congress, creates distortions in the mapping of votes to seats (Rodden 2010; 2012, Chen, Rodden et al. 2013) where Democrats are punished for an over-concentration of voters in relatively few, urban, districts. Our results suggest that much of these spatial difference in partisan vote-shares is not so much the direct result of either unintended historical accident or partisan sorting, as has been claimed. Rather, we show more than half of the observed partisan clustering by geography is the consequence of policies that have enforced racial residential segregation.

The remainder of this paper proceeds as follows. In Section 2 we outline our statistical model and assumptions, and show how it can be used to place an upper bound on sorting by neighborhood. In Section 3 we describe our data, showing that our sample is broadly reflecting of broader trends in the populace, and describe the details of the estimation. Next, in Section 4 we provide our baseline estimate of the upper-bound of geographic partisan sorting, and we show that over half of this ceiling is explained by race. Finally, we conclude.

2 Statistical Model

To start we develop a statistical model for estimating an upper bound of partisan geographical sorting.[2] Let y_{nhi} be the party choice of voter i, in household h, and neighborhood n. Assume that the decision to become a partisan is given by the following linear model:

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

Where x_{nh} represents all the observable and unobservable characteristics common to household members that determine both their decision to sort into a specific household and their individual party choice, z_n represents all the observable and unobservable neighborhood characteristics that have an impact on partisanship, and u_{ihn} is a vector of individual-specific characteristics that influence party choice, but are orthogonal to both the decision to sort into a household and to neighborhood effects. We assume that household characteristics ($\alpha' x_{nh}$) and neighborhood features ($\beta' z_n$) have a (weakly) positive correlation. So long as households with similar backgrounds on dimensions correlated with the party choice sort in neighborhoods with common features, this correlation will be non-negative and we can identify the upper bound we seek to characterize.

Still, if we could perfectly observe all relevant variables that influence individual partisanship, we could estimate Equation 1 without bias and establish the direction of potential causal relationships. Since this is not possible, we can only use the total variation in partisanship to establish an upper bound to any existing causal effect. We start by comparing the covariance in party choice across individuals in the same household to the covariance in party choice of individuals within different households, but in the same neighborhood. The covariance between two voters in the same household, the family covariance, is given by:

[2] The framework we adopt has been widely used by labor economists in similar contexts. See, for example Page and Solon (2003).

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

Under the assumption that households sort into neighborhoods based on shared preferences, the covariance in Equation 2 will be positive as $2Cov(\alpha' x_{nh}, \beta' z_n) \geq 0$. Similarly, the covariance between two voters in the same neighborhood but from different households, the neighbor covariance, is given by:

$$Cov(y_{nhi}, y_{nh'i'}) = Cov(\alpha' x_{nh}, \alpha' x_{nh'}) + 2Cov(\alpha' x_{nh}, \beta' z_n) + Var(\beta' z_n) \tag{3}$$

Note that the covariances are identical with the exception of the first term. Thus, if the family covariance is higher than the neighbor covariance, then $Var(\alpha' x_{nh}) > Cov(\alpha' x_{nh}, \alpha' x_{nh'})$. This means that party choice is better explained by household-specific characteristics than geographical sorting, i.e., households sorting into neighborhoods with similar partisan preferences.

The neighbor covariance in party choice has three components, two of which represent the sorting of households into neighborhoods – intended or unintended – and one that is a pure neighborhood effect on partisanship. Our substantive interest is in the effect of individuals sorting into neighborhoods to be near households similar to their own, and is given by $Cov(\alpha' x_{nh}, \alpha' x_{nh'})$. This reflects, for example, a desire to live near families that share similar beliefs, other households of the same race, or more directly individuals that share the same political attitudes.

The second component $(2Cov(x_{nh}, z_n))$ is similar but reflects individuals sorting by neighborhood characteristics that correlate with their household-specific preferences. These could include cultural amenities, access to public transportation or green spaces, and even the level of political activity in the neighborhood. This component would capture the share of partisanship that can be explained by correlation between the household characteristics and these neighborhood traits.

The final component of the neighbor covariance, $Var(\beta' z_n)$, represents the portion of the variation in party choice that is explained by a pure neighborhood effect. That is, this gives impact on party choice of neighborhood features that are not directly correlated with household-specific traits. This term reflects any reverse causal relationship where neighborhoods themselves directly impact the partisan choices.

The estimation of neighbor covariances does not allow us to disentangle the pure neighborhood effect from the effects of sorting. It nevertheless gives us the maximal possible impact of sorting when the pure neighborhood effect is zero (the pure neighborhood effect is a variance, therefore it is always nonzero). It follows that this is an upper bound on the explanatory power of sorting on partisanship.[3]

2.1 Accounting for Race

If we could observe some variable that determines both sorting and partisanship, then we could obtain a more precise estimate of the upper-bound of partisan sorting. Since there is substantial evidence that race is central to partisanship in the United States (Hutchings and Valentino 2004), we directly account for it and provide evidence that race explains a large share of our baseline upper bound estimate of partisan sorting. In Equation 4 we decompose party choice into the effects of race ($\gamma' r_{nhi}$), and an orthogonal component e_{nhi}.

$$y_{nhi} = \gamma' r_{nhi} + e_{nhi} \tag{4}$$

Notice that this orthogonal component still includes household characteristics, neighborhood characteristics, and a true individual-specific shock as in Equation 1. Again, because of omitted variables, estimating this equation directly would result in a biased estimate. Still, we can decompose this equation into household and neighbor covariances as follows:

$$Cov(y_{nhi}, y_{nhi'}) = Cov(\gamma' r_{nhi}, \gamma' r_{nhi'}) + Cov(e_{nhi}, e_{nhi'}) + 2Cov(\gamma' r_{nhi}, e_{nhi'}) \tag{5}$$

$$Cov(y_{nhi}, y_{nhi'}) = Cov(\gamma' r_{nhi}, \gamma' r_{nh'i'}) + Cov(e_{nhi}, e_{nh'i'}) + 2Cov(\gamma' r_{nhi}, e_{nh'i'}) \tag{6}$$

As before, all three components of both neighborhood and household covariances can be estimated.

[3] It also implies the converse, given that the estimation strategy employed here' uncovers neither the existence nor the direction of a potential causal relationship between sorting and partisan choice.

The first term in both equations represents the share of the covariance directly explained by race. The second component is the direct impact of sorting on factors that are orthogonal to race. The third component is the impact on partisanship coming from the covariance between the voter's own preferences that are orthogonal to race, and the race of her neighbors.[4] Together the second and third terms represent the combined impact of sorting and pure effects of households and neighborhoods, respectively.

3 Data and Estimation

We estimated the upper bound on neighborhood sorting using the universe of registered voters in North Carolina which we obtained from the North Carolina State Bureau of Elections's voter registration file. These data were downloaded from the NBSE's website `http://dl.ncsbe.gov/` in April 2017 and describe the party registration, self-identified race, and address of over 6 million registered voters.[5]

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 6,103 block groups for which data is available. In the appendix we provide results for different neighborhood aggregations (census tract, and grids based on latitude and longitude). In Table 1 we provide descriptive statistics 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.

We 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.2) we provide estimates where we treat Republican registration, or third-party registration, as the outcome.[6] Under these alternative

[4]For example, a voter with preference for outdoor activities will potentially interact more with neighbors of a different race, which might further influence its own political preferences.

[5]We focus upon North Carolina for two reasons. First, it is one of the few states who record and make available 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.

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

Table 1: Descriptive Statistics by Party Registration

Variable	*Democrats*	*Republicans*	*Other*
Individual Level Averages			
Age	49.81	50.1	43.38
Female	0.563	0.490	0.473
White	0.457	0.942	0.758
Black	0.465	0.017	0.111
Block-Group Level Averages			
Density	585.1	382.5	544.2
Gini	0.403	0.397	0.398
Income	50.12	59.25	57.35
Share Black	0.309	0.133	0.185
Observations	2,002,972	1,665,636	1,595,136

specifications the upper bound estimates for the effects of sorting are even smaller. As such, the base-case specification presented in this paper is the most conservative estimate of this upper bound. 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.

3.1 Estimation Procedure

Our main outcome variable is a dummy indicating whether the voter is registered Democrat. Across specifications we adjust this variable for the influence of gender and age by first regressing our outcome on these covariates. The residuals of this regression are the actual outcome variable used in the estimation procedure that follows. 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:

$$\hat{\sigma}^2 = \frac{\sum_{n=1}^{N} \sum_{h=1}^{H_n} \sum_{i=1}^{I_{nh}} y_{nhi}^2}{\sum_{n=1}^{N} \sum_{h=1}^{H_n} I_{nh}} \tag{7}$$

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 2, for household h, can be estimated as:

$$\hat{f}c_{nh} = \frac{\sum_{i \neq i'}^{P_{nh}} y_{nhi} y_{nhi'}}{P_{nh}} \tag{8}$$

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:

$$\hat{f}c = \frac{\sum_{n=1}^{N} \sum_{h=1}^{H_n} W_{nh} \hat{f}c_{nh}}{\sum_{n=1}^{N} \sum_{h=1}^{H_n} W_{nh}} \tag{9}$$

Where W_{nh} is the weight assigned to individual household covariances. The simplest version of this estimator assigns equal weights to all households. Solon, Page, and Duncan (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 Table A.1 (appendix), we provide the main results under two alternative weighting schemes, i.e., weighting each household equally or by household size. The baseline family covariances are indistinguishable from the specification with equal weights. However, our baseline estimates are larger than the specification that overweights larger households, most likely reflecting the fact that in very large households there is a greater degree of intra-family variation in partisanship. The total number of distinct pairs between individuals in households h and h', in neighborhood n, is given by $I_{nh} I_{nh'}$. The neighbor correlation for this specific pair of households is given by:

$$\hat{n}c_{nhh'} = \frac{\sum_{i=1}^{I_{nh}} \sum_{i'=1}^{I_{nh'}} y_{nhi} y_{nh'i'}}{I_{nh} I_{nh'}} \tag{10}$$

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, and Table A.1 in the appendix give estimates from the alternative weighting schemes. These are substantively indistinguishable from our baseline estimates.

$$\hat{nc} = \frac{\sum_{n=1}^{N}\sum_{h=1}^{HH_n} W_{nhh'}\hat{nc}_{nhh'}}{\sum_{n=1}^{N}\sum_{h=1}^{HH_n} W_{nhh'}} \tag{11}$$

4 Results

In column y of Table 2 we decompose the total variance of the residualized outcomes into two components: family covariance and neighbor covariance. The family and neighbor correlations are obtained by dividing the covariances by the total variance, and represent the share of total variation in party choice that is explained by the covariance between household members, and neighbors.

The family covariance accounts for 0.47 of the total variation in party registration. By contrast, the neighbor covariance accounts for 0.11 of the total variation. This is effectively the upper bound of partisan sorting into neighborhoods. Moreover, the ratio of these covariances represents the upper bound on the relative importance of these two sets of factors in defining partisanship. Since the covariance across household members is more than four times larger than the covariance across neighbors, it indicates that the household characteristics upon which voters match are far more important in explaining partisanship than geographical sorting.

4.1 The Role of Race

As we have shown in Equations 5 and 6, the variance in party registration can be further decomposed in three components: direct effect of racial sorting, direct effects of other unobserved factors orthogonal to race, and the effect of the interaction between a voter's unobserved factors and the race of her neighbors and vice-versa. This decomposition is given in columns $\gamma' r$, e, and $2Cov(\gamma' r, e)$, respectively.

In the second column ($\gamma' r$) we present the decomposition of variation in party registration that is be explained by race. Race alone explains roughly one-quarter of the total variation in party choice. When we further decomposed the variation in party choice into family and neighbor covariances we show that sorting by race accounts for 57% of our estimated upper bound on the family covariance and 43% of the upper bound or neighborhood sorting.

In the third column (e), we decompose variation in party registration into that predicted by all

Table 2: Decomposition of the Variance in Party Registration

	y	$\gamma' r$	e	$2Cov(\gamma' r, e)$
Total Variance	0.233	0.055	0.178	
	(0.001)	(0.001)	(0.000)	
Family Covariance	0.109	0.047	0.055	0.008
	(0.000)	(0.001)	(0.000)	(0.000)
Neighbor Covariance	0.025	0.015	0.005	0.006
	(0.001)	(0.001)	(0.000)	(0.000)
Family Correlation	0.469	0.200	0.236	0.033
Neighbor Correlation	0.109	0.062	0.022	0.024

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.

unobserved factors orthogonal to race, and in the fourth column ($2cov(\gamma' r, e)$), the effects of the covariance between individual unobserved factors and the neighbors' race, and vice-versa. The sum of the neighbor covariance in the last two columns is the upper bound of partisan sorting effects on party choice adjusted for race. This accounts for just 4.6% of the total variation in outcomes. Finally, after adjusting for race, the estimated upper bound on the ratio of the explanatory power of household and neighborhood sorting increases to 5.8 (from 4.4 before accounting for race).

4.2 Heterogeneity of the Upper Bound Estimates

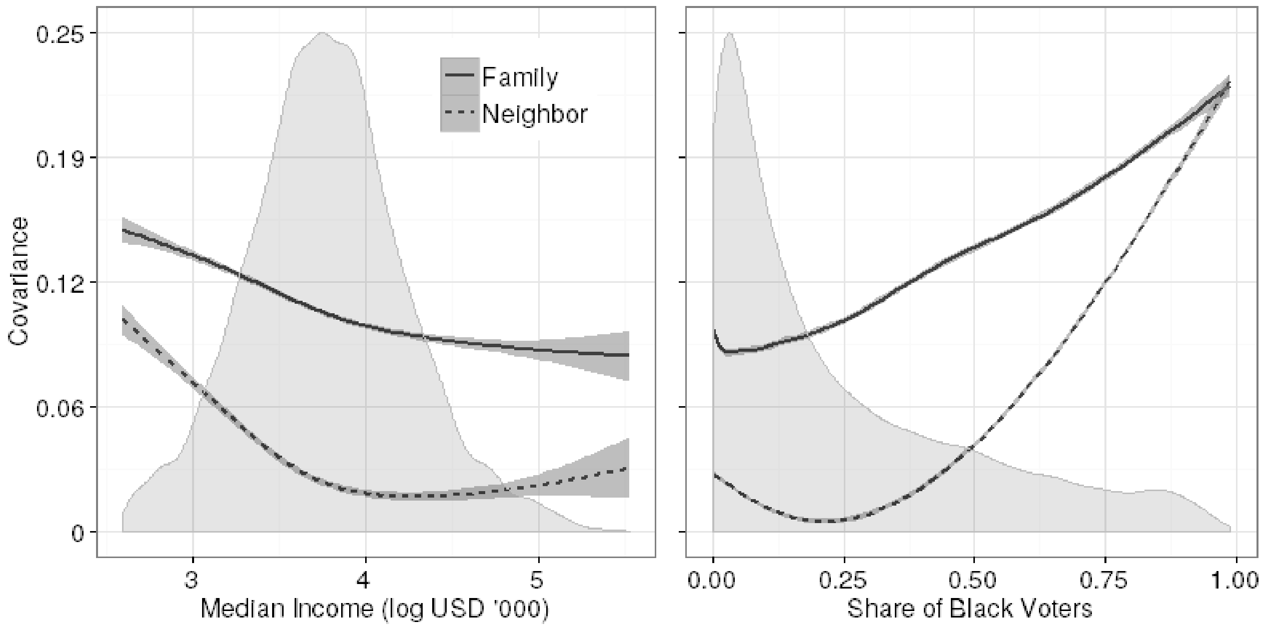
Our estimates cannot distinguish between purposeful sorting based upon partisanship (or other attributes that load onto partisanship) and the unintentional segregation of individuals by party (or attributes that load on to party). Our estimation would represent an upper bound on intentional sorting if we assume that segregation explains none of the observed variation in partisanship. If geographical segregation in fact helps to explain variation in partisanship across neighborhoods, the upper bound on the explanation power of intentional sorting would be even lower.

To provide some insight on the nature of geographical sorting and its influence in partisanship, we explore how our estimate of the upper bound of neighborhood sorting varies across two neighborhood features: racial composition and income. Consistent with the literature showing that race dominates income as a determinant of variation in geographical voting (Hersh and Nall 2016), we show that most of our already low upper bound estimate comes from neighborhoods with a high share of African American voters.

The right plot in Figure 1 shows that our upper bound estimate is increasing sharply in the share of African Americans living in a neighborhood. The estimated upper bound of sorting at 20% African American voters (roughly the proportion of African American voters in the state) is equal to about 2%, but increases to over 24% in neighborhoods that are 99% African American. As for income, the left plot shows that for neighborhoods with income between $12,000 and $35,000, the neighborhood correlation declining from an estimate of 12% of the observed variation in party registration to just over 4%. At average neighborhood incomes above this value the upper bound of sorting remains roughly constant at around 4%.

All in all, it is clear that most of the existent relationship between partisanship and geographical sorting comes from the consistent pattern of neighborhoods with a high share of African Americans having also a high share of registered Democrats. Interestingly enough, although race and income are highly correlated in the context of the sample, the same stark heterogeneity on the upper bound estimate cannot be found for different income profiles of neighborhoods. In other words, we do not observe the same sharp increase in the explanation power of sorting in poorer neighborhoods as we do in African American neighborhoods.

Figure 1: Upper Bounds by Neighborhood Income and Race Profile



Note: This figure plots the natural logarithm of neighborhood average household income against the estimated upper bound of partisan sorting. The shaded area illustrates the distribution on neighborhoods across income or racial profile.

5 Conclusion

In this paper we have provided evidence against the importance of neighborhood sorting in determining partisanship. In all, we find at most 11% of the observed variation in party registration can be explained by neighborhood sorting. Moreover, once we account for race this upper bound falls by more than half to just over 4%. Because of the persistence of institutionalized barriers to racial integration we are reluctant to interpret this as evidence of overt racial sorting. Instead we view our findings as evidence in support the long-lasting impact of policies designed to geographically segregate racial groups. Race, in turn, strongly predicts partisanship, yielding our result. Additionally, we find that household-specific characteristics explain party choice 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.

References

Abrams, Samuel J and Morris P Fiorina. 2012. "The Big Sort That Wasn't: A Skeptical Reexamination." *PS: Political Science & Politics* 45 (02):203–210.

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.

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

Chen, Jowei, Jonathan Rodden et al. 2013. "Unintentional gerrymandering: Political geography and electoral bias in legislatures." *Quarterly Journal of Political Science* 8 (3):239–269.

Fowler, James H, Laura A Baker, and Christopher T Dawes. 2008. "Genetic variation in political participation." *American Political Science Review* 102 (02):233–248.

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.

Hersh, Eitan D and Clayton Nall. 2016. "The Primacy of Race in the Geography of Income-Based Voting: New Evidence from Public Voting Records." *American Journal of Political Science* 60 (2):289–303.

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, Jeanette Morehouse Mendez, and Tracy Osborn. 2004. "Disagreement, ambivalence, and engagement: The political consequences of heterogeneous networks." *Political Psychology* 25 (1):65–95.

Hui, Iris. 2013. "Who is your preferred neighbor? Partisan residential preferences and neighborhood satisfaction." *American Politics Research* 41 (6):997–1021.

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.

Klar, Samara. 2014. "Partisanship in a social setting." *American Journal of Political Science* 58 (3):687–704.

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

Levendusky, Matthew, Neil Malhotra, Yotam Margalit, and Christopher McConnell. 2015. "The Economic Consequences of Political Polarization: Is Partisan Animus Cheap Talk?" *Unpublished manuscript, University of Pennsylvania* .

McCarty, Nolan, Keith T Poole, and Howard Rosenthal. 2016. *Polarized America: The dance of ideology and unequal riches.* mit Press.

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.

Mummolo, Jonathan and Clayton Nall. 2017. "Why Partisans Do Not Sort: The Constraints on Political Segregation." *The Journal of Politics* 79 (1):000–000.

Mutz, Diana C. 2006. *Hearing the other side: Deliberative versus participatory democracy.* Cambridge University Press.

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.

Reardon, Sean F and Kendra Bischoff. 2011. "Income Inequality and Income Segregation 1." *American Journal of Sociology* 116 (4):1092–1153.

Rodden, Jonathan. 2010. "The geographic distribution of political preferences." *Annual Review of Political Science* 13:321–340.

———. 2012. "The Long Shadow of the Industrial Revolution." *Unpublished manuscript. Stanford University Calif* .

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.

Sunstein, Cass R. 2009. *Going to extremes: How like minds unite and divide.* Oxford University Press.

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

Table A.1: 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

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 over-weights 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.2: 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

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.3: 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

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.