On-line Appendix to
"The Long-run Effects of the 1930s Redlining Maps on Children"
Daniel Aaronson, Daniel Hartley, Bhashkar Mazumder, and Martha Stinson

Identification Strategy (more detail)

It is well established in the historical record that the HOLC defined neighborhoods based on the observable characteristics of areas and trends in these characteristics over time. A simple comparison of the outcomes of children growing up in neighborhoods with different HOLC grades might reflect these pre-existing characteristics and trends rather than capture the causal effects of the neighborhood. Therefore, a major challenge to any analysis of the maps is developing an identification strategy that can yield plausible causal estimates.

Our empirical approach follows two of the estimation strategies used in Aaronson, Hartley, and Mazumer (2021; "AHM") and Aaronson, Faber, Hartley, Mazumder, and Sharkey (2021). Both strategies start with the sample of 1940 Census children living in one of the 149 cities for which we have maps. We further restrict the sample to children who lived in the buffer zones on the two sides of the D-C and C-B boundaries. Narrowing the sample to just those living within a few city blocks of a boundary, reduces, but doesn't eliminate, the stark differences in the housing and other characteristics of families by neighborhood grade. For example, these families would have shared many amenities (e.g. transportation, labor markets) in common despite living in different neighborhoods. At the same time, we know from the historical record and from statistical analysis in AHM that these neighborhood characteristics sometimes changed sharply at neighborhood borders, thereby making simple cross-border comparisons fraught with bias.

Method 1: Use of comparison boundaries

Our first strategy uses propensity score methods to create a set of "comparison" boundaries that are weighted to resemble the actual "treated" HOLC boundaries. The intuition behind this approach is that there were likely many "missing" boundaries that the HOLC could have used to separate neighborhoods but did not for practical reasons. For example, buried within a largely homogenous D neighborhood might be a small pocket resembling a C neighborhood based on demographic and housing characteristics. But it may not have made practical sense to classify this smaller area nestled inside a larger neighborhood as a separate neighborhood.[1] Our strategy uses children who grew up in the buffer zones on the two sides of these potential boundaries as a comparison group. We can then compare the cross-boundary *differences* in the treated sample of children to the cross-boundary *differences* in our comparison boundaries. This difference-in-differences strategy allows us to infer the true causal effect of growing up on the lower-graded side of a redlined or yellowlined neighborhood.

To implement this strategy in practice, we overlay a one-half mile by one-half mile grid over each city. We then take all the grid segments that are located within HOLC neighborhoods of the relevant grade. For example, for the estimation of the effects along D-C boundaries, all of the grid segments inside of D areas and C areas are used to construct a comparison sample of boundaries and buffer zones. For each comparison boundary, we randomly assign one side to be the lower-graded side (lgs).[2]

Next, we use the propensity score model below to develop weights that will be used to ensure that, pre-map, comparison group segments are similar to treated boundaries based on observable characteristics:

$$1\{Treated\}_{b,c} = \alpha_c + \sum_{k=1}^{K} \beta_{1910}^{k} z_{b,c}^{k,1910} + \beta_{1920}^{k} z_{b,c}^{k,1920} + \beta_{1930}^{k} z_{b,c}^{k,1930} + \epsilon_{b,c}.$$

Each observation is a border segment b located in city c. α_c is a city fixed effect and $z_{b,c}^{k,t} = x_{lgs,b,c}^{k,t} - x_{hgs,b,c}^{k,t}$ is the gap between variable k on the lower-graded side (*lgs*) versus the higher-graded side (*hgs*). The sample pools treated and comparison borders and $1\{Treated\}_{b,c}$ is an indicator equal to one if the

[1] See Appendix Figure A4 in AHM for a stylized example of this situation. Figure 2 in AHM shows that this is especially plausible for Chicago where there are many broad swaths of red-shaded neighborhoods that likely contained pockets that resembled yellow graded neighborhoods.

[2] We use this approach to make sure that the distribution of boundary differences in our comparison set of borders is representative of the underlying universe of all such boundaries and is not overrepresenting either tail of the distribution. The reweighting of comparison boundaries occurs after the randomization.

border is treated. The variables in k include: share African American, share foreign born, African American population density, White population density, homeownership rate, log house value, log rent and the share of homeowner households that have a mortgage.

We use the results of the propensity score model to create a set of inverse probability weights (IPW) for each individual with the following specification:[3]

$$y_{igbt} = \beta_t 1[lgs]1[treated] + \beta_{lgs}1[lgs] + \beta_{treated}1[treated] + X_{igb1940} + \alpha_b + \epsilon_{igbt}.$$

In this equation, an observation is a person, i, living within ¼ miles of a border segment, b, on either the lower- or higher-graded side, g, in 1940. The sample pools individuals living in the buffers of the actual treated borders along with those living in the buffers of the comparison borders. The index t reflects the fact that the outcomes are measured in either the 1970s (IRS data) or 2000 (Census). $1[treated]$ is an indicator for a treated border, $1[lgs]$ is an indicator for the lower-graded side of the border, α_b is a border segment fixed effect, and $X_{igb1940}$ is a set of covariates from the 1940 Census which were likely determined prior to the HOLC maps. They include the race of the child, whether the child was a teen birth, family size indicators, and mother and father measures for age, marital status, race, foreign born, citizen, Hispanic, and educational attainment.

AHM find strong evidence that this approach largely eliminates any gaps in the cross-boundary differences between the treatment and comparison groups in the share African American, home ownership

[3] The weights are as follows: for the comparison boundaries, $w = \text{pscore}/(1\text{-pscore})$ and for the "treated" boundaries, w = 1. The sample excludes treated borders with a propensity score above that of the maximum comparison border and comparison borders with a propensity score below that of the minimum treated border. Using this procedure essentially "up-weights" comparison boundaries that are most similar to treated boundaries and "down-weights" those that are least similar. As a result, the reweighted comparison borders look more similar to the actual HOLC borders than the unweighted comparison borders do. AHM also show that there is a sizable amount of overlap in the distributions of the propensity scores for the two groups (see their Appendix Figure A6, Panels A and B).

rate, house values, and rents in the pre-period before the maps were drawn which provides support for this identification strategy.[4]

Method 2: Low Propensity Score (LPS) Borders

Our second approach is based on the idea that some borders may have been drawn for more idiosyncratic reasons and did not actually reflect meaningful differences between adjacent neighborhoods of different grades. One way these "misaligned" borders could arise is because the map makers may have simply needed to close a polygon, and the exact street that was chosen to do so was somewhat arbitrary.[5] The way we operationalize this idea is to limit our sample to the subset of actual HOLC borders that appear "idiosyncratic" -- that is, where meaningful pre-existing differences across boundaries do not exist. A nice advantage of this approach is that, by definition, it no longer requires a comparison group based on using a grid to find counterfactual boundaries and therefore provides a second approach that complements our first strategy.

Practically, we use the predicted p-scores from our model and select the actual HOLC borders with a low predicted probability of being chosen (below median p-score) based on our model.[6] AHM show that this approach effectively removes pre-existing cross-boundary differences and trends providing supportive evidence for the validity of this method. The specification for this model is:

$$y_{igbt} = \beta_t 1[lgs] + X_{igb1940} + \alpha_b + \epsilon_{igbt}.$$

[4] We also validated this approach using children 16 years and younger in the 1910, 1920, and 1930 Censuses. In particular, we ran similar regressions but with the following left hand side variables: an indicator for being a Black person, an indicator for living in an owner-occupied household (1920 and 1930 only), the log of rent if living in a rented home, and the log of the home value if living in an owned home. Again, we were largely able to eliminate gaps in these outcomes in the period before the maps were drawn as the estimates in this pre-period were generally economically small and statistically insignificant. The one exception is that the estimate for the 1930 gap in rent along the C-B boundary is -3 percent and statistically significant at the 10 percent level. However, we did not correct for multiple hypothesis testing and this could have been significant at the 10 percent level purely by chance as we ran a total of 15 regressions.

[5] See Appendix figure A4 and the text within AHM for a stylized and actual example of this situation.

[6] Our focus on low-p-score boundaries is akin to the subclassification approach described in Imbens (2015) and Imbens and Rubin (2015).

Heterogeneity of Effects on AGI

We stratified the sample by demographic and socioeconomic characteristics to consider situations that may have led to especially negative outcomes and to begin to collect results that could help uncover mechanisms. To help with precision, we focus on the tax records where our sample is substantially larger. Even still, sample size ultimately limits how much can be inferred about many group differences. We also focus on AGI as our outcome of interest, partly for brevity, but also to highlight some useful differences by the type of income filers.

We begin by looking at individuals who owned a business or had financial income. Specifically, we use an indicator variable for whether individuals filed either a schedule C or schedule SE to proxy for whether they had business income. Similarly, we use an indicator for whether individuals filed a schedule D to capture capital income from sources such as financial or real estate investments. We also construct a separate indicator for those who fit into both categories. Finally, we compare these three groups to those who did not file C, D, or SE schedules.

The results using the treated vs comparison group approach are shown in Figures A2 and A3.[7] We find an especially large negative effect on AGI among those with schedule D capital income. Along D-C boundaries, the magnitude of the effect is -$3,400 for those who only filed a schedule D for capital income, although there is no significant effect for those who file both a business schedule and schedule D. The impact is extremely large along the C-B boundaries: -$10,100 for those who only file a schedule D and -$14,700 for those who filed both a business schedule and a capital income schedule. These results are consistent with individuals experiencing smaller capital income gains from housing investments if they grew up in lower-graded neighborhoods.

[7] We find broadly the same qualitative patterns when using the low propensity score method. However, the magnitudes of the estimates differ in a similar pattern to the full sample shown in the main text's Figure 4.

We now turn to examining heterogeneity based on characteristics from the 1940 Census. Figures A4 and A5 provide treated vs comparison group estimates by homeownership, age, income of parents in 1940, and whether the parents were married and present in the household.[8] The top bar reproduces the overall effect on AGI for each border type that was discussed in the main text's Figure 4. It is important to note that the scales on Figures A4 and A5 are quite different, and the overall negative effects of yellowlining in Figure A4 are more than twice as large as those of redlining in Figure A5.

The first set of results considers whether the parent was a homeowner or renter in 1940. One hypothesis is that children of homeowners might have been more severely impacted by financial disinvestment in a neighborhood due to declining access to credit. We discuss this point in greater detail in the next section. AHM show evidence that the lower-graded side of neighborhoods experienced declines in house values and suffered impacts associated with financial disinvestment such as higher rates of vacancies and dilapidated housing. Indeed, we find that the negative effects on AGI appear to have been larger, albeit not precisely so, for children whose parents were homeowners. For example, along C-B borders, the effect on AGI for children of homeowners was -$7,430 versus -$3,680 for children whose parents did not own their home. However, these differences are not statistically significant at conventional levels.

A growing literature on the effects of early life events has highlighted how the consequences of shocks experienced by young children may be especially severe (e.g. Almond and Currie, 2011, Almond,

[8] The main text briefly discusses differences by race and gender. We also considered differences by proximity to the border and by whether families had migrated between 1935 and 1940, but it was *a priori* unclear what such differences may reveal. Regarding proximity to the border, focusing only on those families living extremely close to the border (within 1/8th of a mile), on either side, might compare more similar families and thus might provide a better causal estimate. At the same time, there could be negative spillover effects for those on the higher-graded side if they lived extremely close to the lower-graded side. Our estimates were consistently higher for those living farther from the border (between 1/8th and ¼ mile) but the magnitudes varied quite a bit depending on our estimation strategy and border type. Regarding movers in the 1930s, those who lived at the same address in 1935 may have been "treated" for longer (depending on exactly when the maps were made in their city) and could have experienced larger effects. On the other hand, families who decided to move from 1935 to 1940 into an area that received a lower HOLC grade may reflect a selected sample. We generally found more negative income effects for children in households that had been in the same address since at least 1935 compared to those whose households had moved between 1935 and 1940. Appendix Figures A6 and A7 show comparable results using the low propensity score approach. These results generally reveal similar patterns.

Currie and Duque, 2017). In the context of neighborhoods, Chetty, Hendren and Katz (2016) found that the Moving to Opportunity program only had effects on the long-run income of children below the age of 13. Therefore, we divided our sample into those who were "young" in 1940 (below the age of 9) and those who were "old" (between the ages of 9 and 16). We found little difference in the estimates between these age groups though the point estimates were consistently larger for the older group particularly when using the low propensity score approach (Figures A6 and A7).

We also consider how the effects might differ based on two measures of the socioeconomic status of the family in 1940, namely the income level of the parents and whether there were two married parents in the household. We created three categories of wage and salary income, those with below median income ("low"), those with above median income ("high") and those with zero or no reported income ("missing"). We find a similar pattern by parental income along both the redlining and yellowlining borders. The effects are least negative for the children with high-income parents, then more negative for the children with low-income parents, and the most negative for children whose parental income is missing in 1940. Unfortunately, the latter mixes business owners and farmers, since their income was not collected in 1940, with those who did not report an income. But there may have been reason that each group would have been hard hit by less access to credit. The maps may have been particularly detrimental to very low-income households that had no earnings. Moreover, it is also possible that the maps negatively impacted access to credit for businesses or further exposed business owners to negative real estate shocks if they owned their premises.

Finally, we see almost no difference in the effect of redlining between children with two parents in the household and those without two parents. The effect of yellowlining is more negative for children with two parents, but the sample size is somewhat small (26,500) for those without two parents, resulting in a wide confidence interval.[9]

Selection into the Tax Sample

One concern is not all children in the 1940 Census who were successfully matched to the Social Security NUMIDENT database and received a personal identification key (PIK) are observed in the tax data. This selection issue could arise from children not filing taxes as adults or from any failure to match the tax records to Social Security records. Across our four estimation samples the baseline rate of being included in the tax records is between 82 and 84 percent, suggesting that we are capturing most of the potential sample. Still, in Figure A8, we show the results of using our statistical model where the outcome is an indicator of inclusion in the tax records and our universe is the sample of all children who were "PIKed." In all four models, we find very small negative point estimates of less than 0.5 percentage points and in only one case is the estimate statistically significant. This suggests that there is no meaningful difference in the effect of being redlined or yellowlined on selection into the tax sample.

Caveats

An important issue is how to interpret our findings considering the paucity of historical information on the use of the maps. As noted earlier, private banks and the FHA drew their own color-coded maps that may have been influenced by the HOLC maps, which were supposed to have been confidential and limited in distribution. Fishback et al (2021) point out that the FHA's discriminatory practices predated the HOLC maps and that the FHA continually updated their maps (which largely no longer exist) over time suggesting that the influence of the HOLC maps may have been minimal. They also find little change in FHA loan insurance provision in three cities during the period in which the HOLC maps were made available to them. If these other maps had causal effects, then our estimates may be picking up some combination of the direct effects of the HOLC maps, along with any overlap from other private or public maps that shared common

[9] We also have a relatively small subsample of 82,000 children who did not live with both parents when estimating this along the D-C boundary compared to 643,000 children who lived with both parents.

neighborhood grades and borders with the HOLC maps. In that case, our estimates should be interpreted as a proxy for the overall effect of the general practice of redlining/yellowlining that arose largely through federal housing policies. On the other hand, this also implies that our estimates may actually be *understating* the full effects of discriminatory housing practices as measurement error could be introduced if some of the HOLC neighborhood grades or borders differed from those drawn in other, perhaps practically more impactful, maps from the FHA or private lenders.

One important downside of our focus only on children living in buffer zones around the borders of HOLC neighborhoods is that our estimates may not generalize to the children who grew up in the interior of neighborhoods. For example, it could be the case that areas in the interior of neighborhoods had higher concentrations of Black families and Black-owned businesses, and that the children growing up in these areas were much more severely impacted than the children growing up near a border and therefore in closer proximity to a higher-graded neighborhood.[10] We believe that this is an important area for future research to consider, although it will be challenging to come up with a research design that can deliver causal estimates. Similarly, there may be other important outcomes that our analysis did not consider due to data limitations such as effects on business ownership, social networks, criminal activity, health, and exposure to pollution.

It is also important to recognize that there were many other overlapping policies and environmental influences that were affecting these neighborhoods both at the time the maps were drawn, as well as in subsequent years. For example, urban renewal policies that started in the 1950s that aimed to clear "slums" and redevelop urban areas and which were sometimes associated with highway construction, may have also impacted the opportunities faced by some of the children in our sample.[11] The boundaries of school districts

[10] We thank Trevon Logan and Sun Kyoung Lee for their comments in this spirit.
[11] See Collins and Shester (2013), LaVoice (2019) and Shi et al (2021) for recent studies of urban renewal.

and school funding could have also been affected by the maps.[12] Our estimates should be viewed as the reduced form effects of all of the possible effects that were associated with redlining and yellowlining.

Finally, given that much remains unknown about the extent to which the information in the HOLC maps were used by private actors or the FHA in their decisions to insure mortgage loans (see section 2), it is important to carefully consider how to interpret our findings. As we discuss earlier, it may be useful to think of our analysis as serving as a proxy for the overall effects of discriminatory policies that targeted urban neighborhoods. If so, our estimates may serve as a *lower bound* of the full effects of such policies. This is because the grade classifications and the borders between neighborhoods chosen by the HOLC maps are unlikely to have always lined up with those chosen by other pertinent actors, adding measurement error and likely attenuating estimates.

Additional references not in main paper

Almond, Douglas and Janet Currie, 2011, "Human Capital before Age Five," In the *Handbook of Labor Economics*, edited by Ashenfelter, O. and Card, D. Vol. 4 (Chapter 15): 1315-1486.

Almond, Douglas, Janet Currie, and Valentina Duque. 2018, "Childhood Circumstances and Adult Outcomes: Act II," *Journal of Economic Literature*, 56(4): 1360-1446.

Collins, William J. and Katharine L Shester, 2013,"Slum Clearance and Urban Renewal in the United States," *American Economic Journal: Applied Economics* 5(1):239–273.

LaVoice, Jessica, 2019, "The Long-Run Implications of Slum Clearance: A Neighborhood Analysis," Working Paper.

Shi, Ying, Daniel Hartley, Bhashkar Mazumder, and Aastha Rajan, 2021, "The Effects of the Great Migration on Urban Renewal," Federal Reserve Bank of Chicago Working Paper 2021-4.

[12] See Lukes and Cleveland (2021).

Table 1: Summary Statistics

Panel A. Income Tax Data from 1974/1979

	Treated vs Comparison						Low Propensity Score					
	D-C			C-B			D-C			C-B		
			Rounded			Rounded			Rounded			Rounded
	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.	N
Adj Gross Income (74/79 avg)	64,110	67,340	725,000	68,940	61,480	305,000	66,050	76,820	133,000	78,600	75,020	55,500
Wage and Salary Inc (74/79 avg)	56,610	47,640	725,000	60,040	43,930	305,000	58,630	44,430	133,000	66,440	55,530	55,500
AGI, No Schedule	54,640	31,660	486,000	57,890	29,440	179,000	56,630	28,560	87,500	60,860	32,960	30,500
AGI, Bus. Schedule (C/SE)	53,100	32,230	70,000	54,910	37,640	31,500	54,750	31,900	13,000	59,260	35,510	5,800
AGI, Capital Inc Schedule (D)	94,130	79,660	107,000	96,110	73,710	58,000	95,090	82,970	21,000	111,500	97,700	11,500
AGI, both Bus. and Cap Inc Sched	94,310	164,600	62,000	103,000	130,000	37,000	96,550	208,500	12,000	115,700	130,400	7,500
Neighborhood Index (Tract)	-0.046	0.76	609,000	0.12	0.64	256,000	0.05	0.65	113,000	0.22	0.59	46,000
College Rate (Tract)	0.14	0.11	609,000	0.16	0.12	256,000	0.14	0.11	113,000	0.18	0.13	46,500
Hish School Rate (Tract)	0.61	0.17	609,000	0.65	0.17	256,000	0.61	0.17	113,000	0.67	0.15	46,500
Employment to Pop Rate (Tract)	0.58	0.06	609,000	0.59	0.06	256,000	0.59	0.06	113,000	0.59	0.06	46,500
Unemployment Rate (Tract)	0.04	0.02	609,000	0.03	0.02	256,000	0.03	0.02	113,000	0.03	0.02	46,500
Poverty Rate (Tract)	0.06	0.06	609,000	0.05	0.04	256,000	0.05	0.04	113,000	0.04	0.03	46,000
Single Headed HH Rate (Tract)	0.11	0.08	609,000	0.10	0.06	256,000	0.10	0.07	113,000	0.09	0.06	46,000
Moved Tract	0.96	0.19	609,000	0.96	0.19	256,000	0.97	0.18	113,000	0.97	0.18	46,500
Moved County	0.51	0.50	609,000	0.54	0.50	256,000	0.56	0.50	113,000	0.62	0.49	46,500
Moved State	0.28	0.45	609,000	0.30	0.46	256,000	0.26	0.44	113,000	0.34	0.47	46,500
Panel B. 2000 Census												
Years of Education	13	2.68	91,000	13.5	2.58	40,000	13	2.63	17,000	14	2.7	7,400
High School or More	0.81	0.39	91,000	0.86	0.35	40,000	0.82	0.38	17,000	0.90	0.29	7,400
Some College or More	0.44	0.50	91,000	0.53	0.50	40,000	0.43	0.50	17,000	0.60	0.49	7,400
College or More	0.20	0.40	91,000	0.26	0.44	40,000	0.2	0.40	17,000	0.34	0.47	7,400

Figure A1

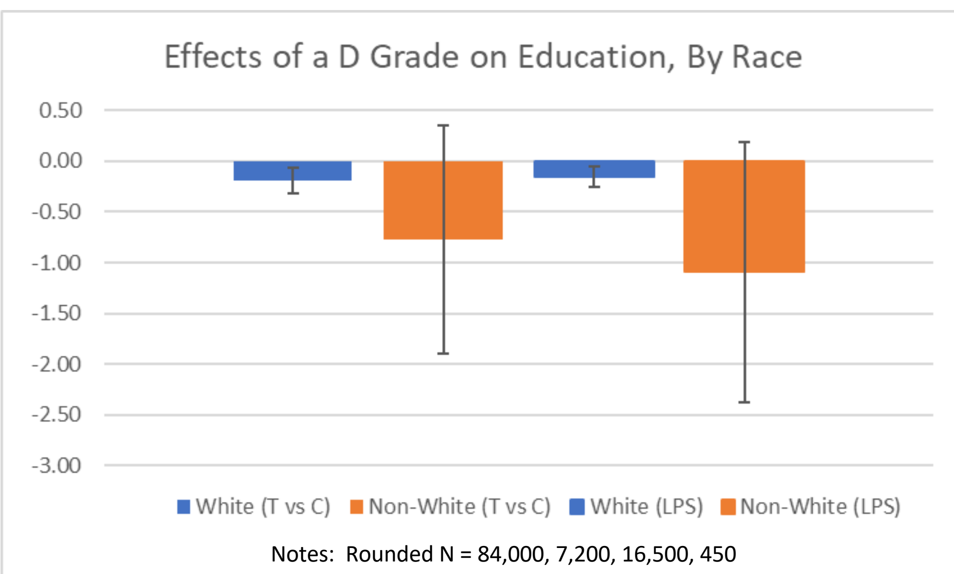


Figure A2

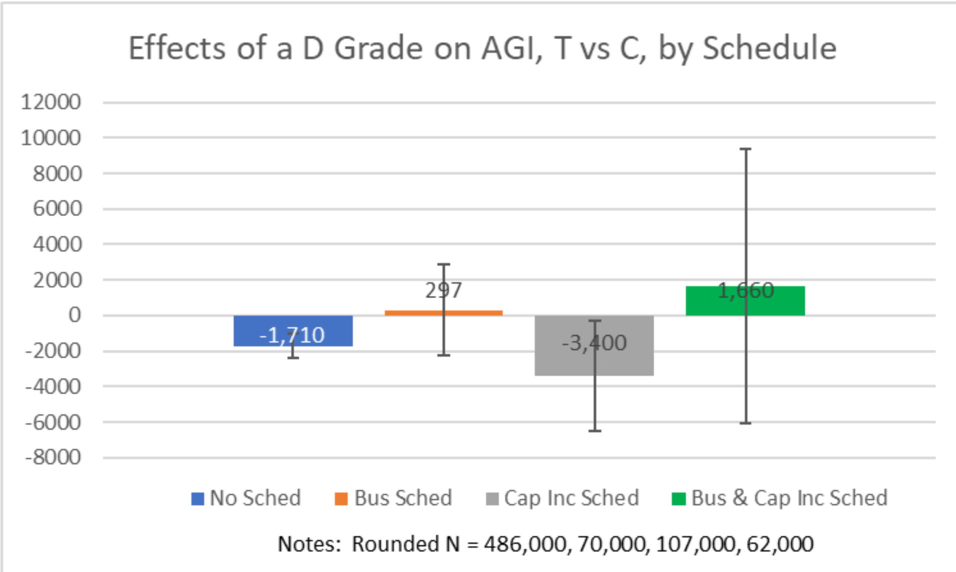


Figure A3

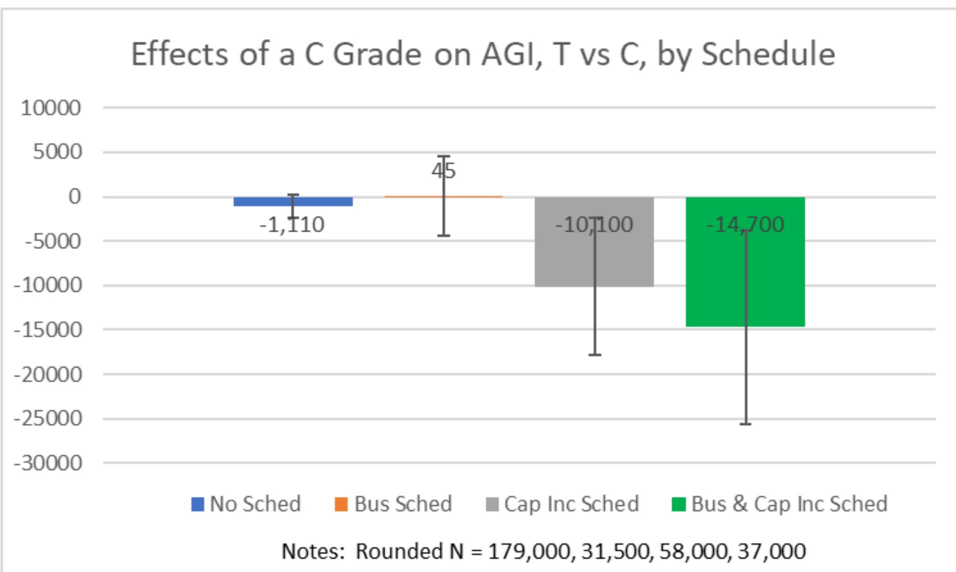


Figure A4

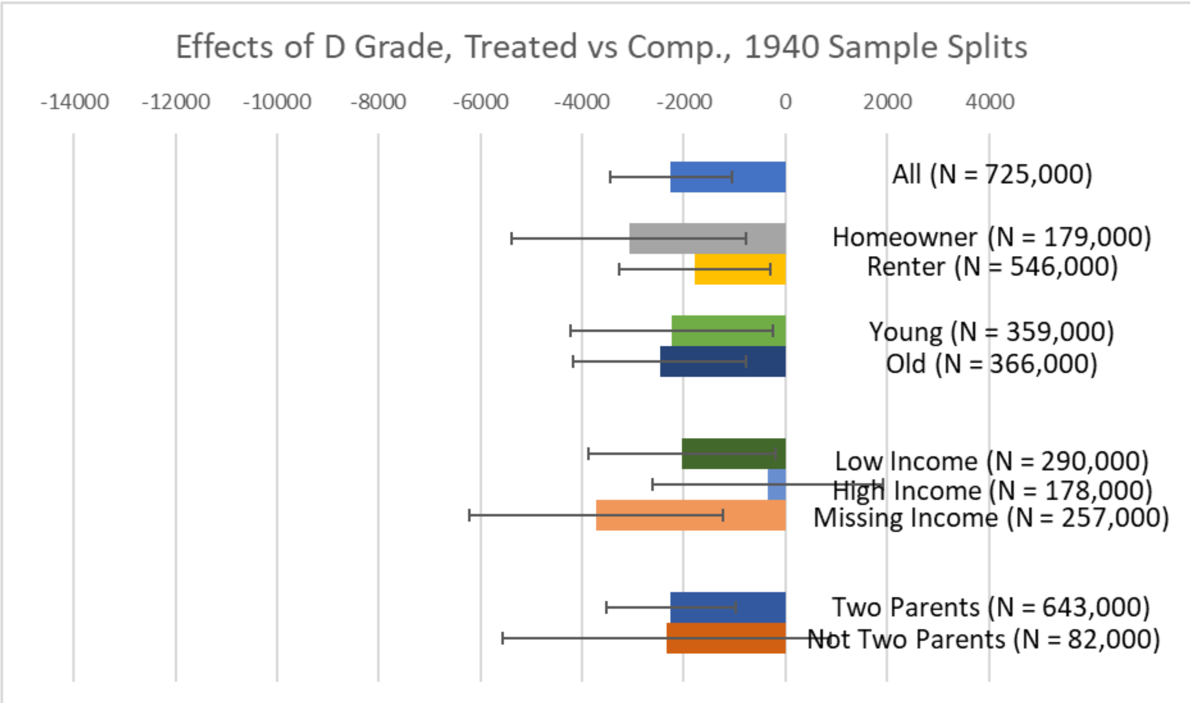


Figure A5

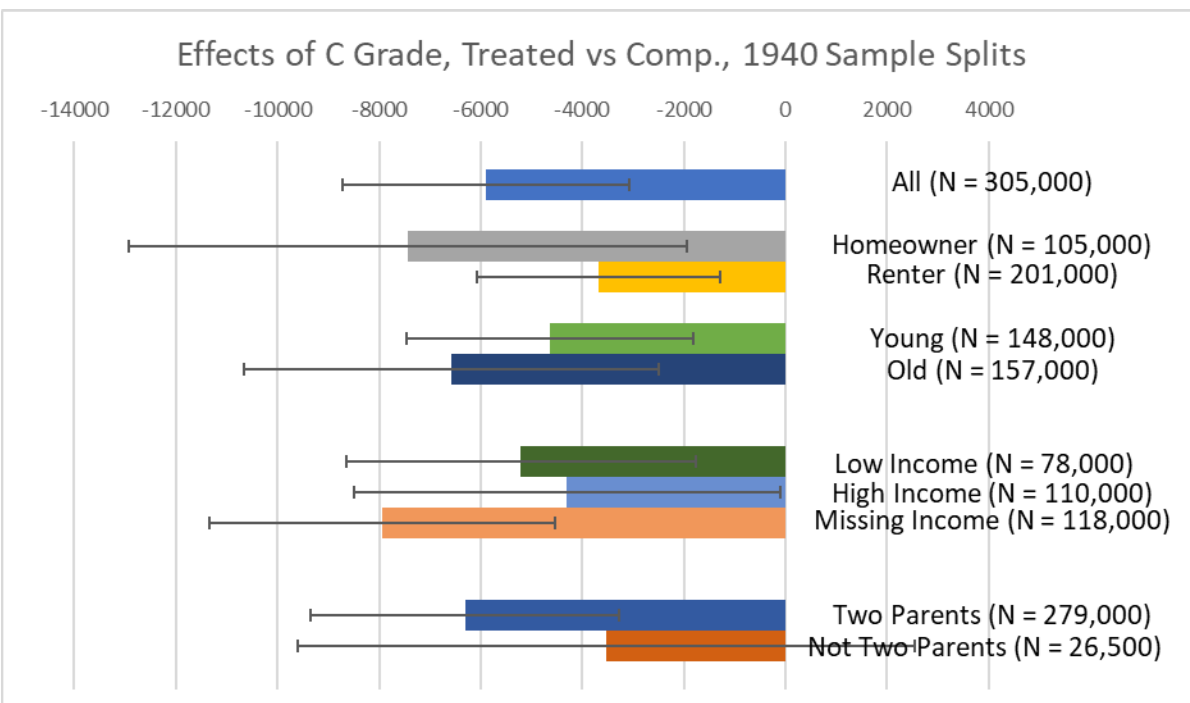


Figure A6

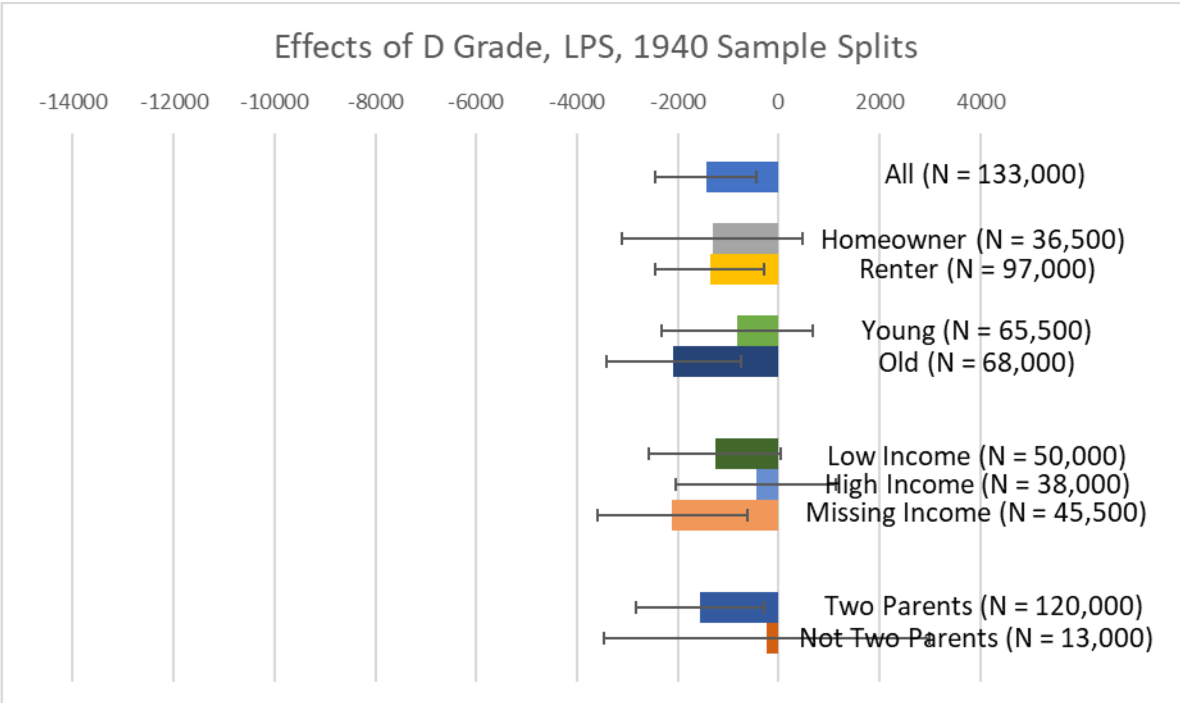


Figure A7

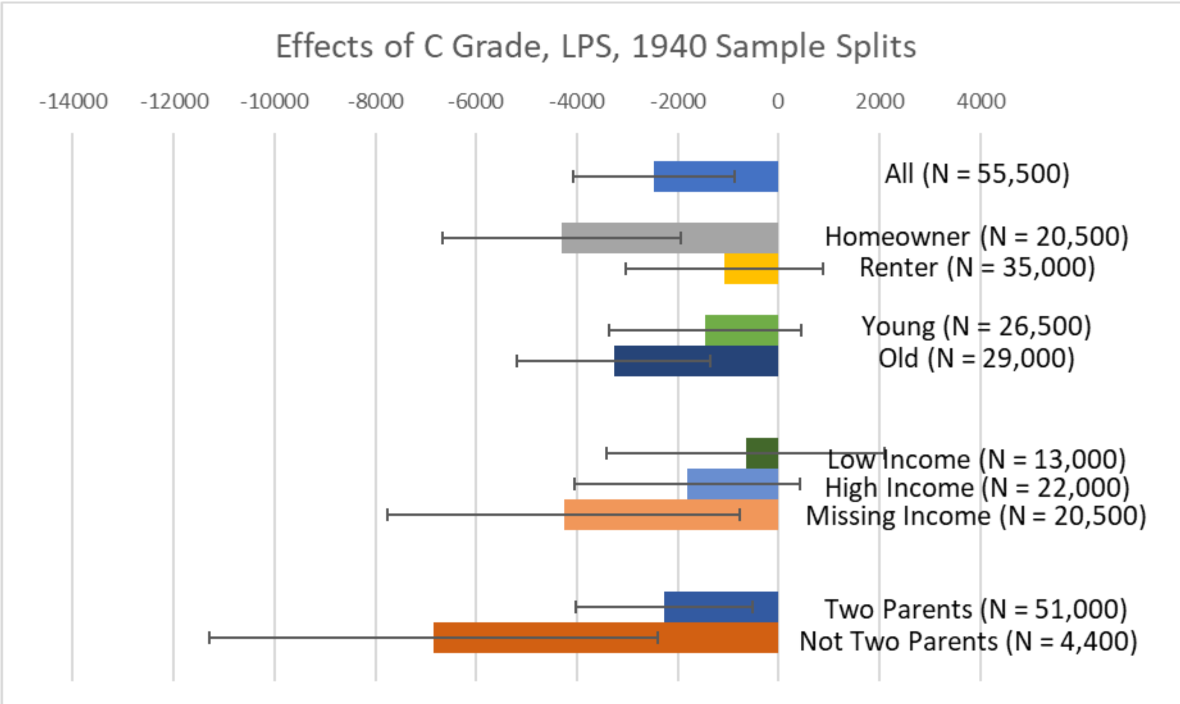
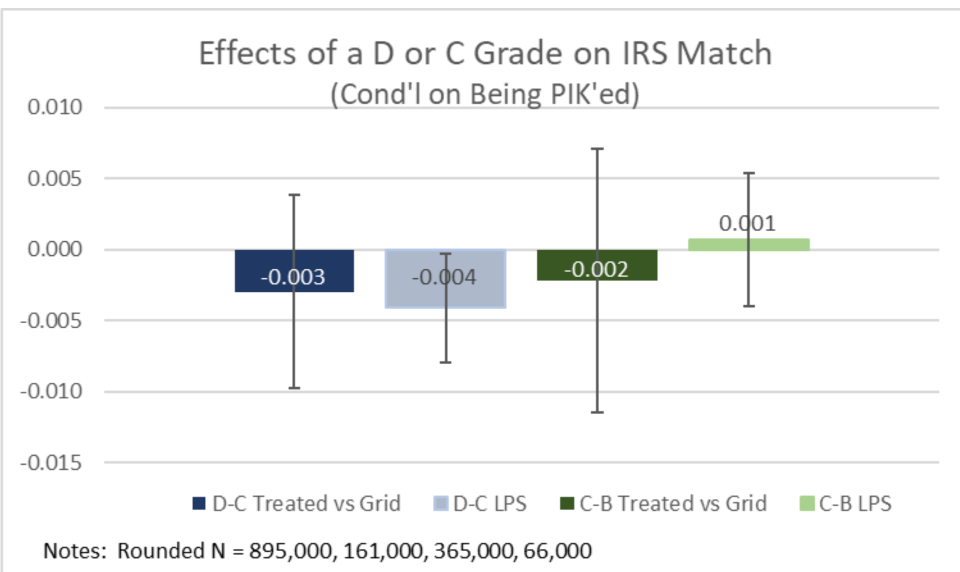

Effects of C Grade, LPS, 1940 Sample Splits
-14000
-12000
-10000
-8000
-6000
-4000
-2000
0
2000
4000
All (N = 55,500)
Homeowner (N = 20,500)
Renter (N = 35,000)
Young (N = 26,500)
Old (N = 29,000)
Low Income (N = 13,000)
High Income (N = 22,000)
Missing Income (N = 20,500)
Two Parents (N = 51,000)
Not Two Parents (N = 4,400)

Figure A8


Effects of a D or C Grade on IRS Match
(Cond'l on Being PIK'ed)
0.010
0.005
0.000
-0.005
-0.010
-0.015
-0.003
-0.004
-0.002
0.001
D-C Treated vs Grid
D-C LPS
C-B Treated vs Grid
C-B LPS
Notes: Rounded N = 895,000, 161,000, 365,000, 66,000