Official Statistics Moderately Overestimate Residential Segregation in the United States*

Jacob R. Brown†
Department of Political Science
Boston University

Christopher T. Kenny‡
Department of Government
Harvard University

Tyler Simko§
Department of Politics
Princeton University

August 30, 2024

Abstract

Residential segregation in the United States is widespread, has persisted over time, and threatens fair economic opportunity and social cohesion. Most commonly used measures of segregation rely on aggregate data that impose arbitrary definitions of local geography. Previous work demonstrates that segregation measures are sensitive to the particular aggregation. Using recent advances in redistricting software, we evaluate the bias and uncertainty created by these measurement choices. We sample alternative Census tract maps that follow Census guidelines, ensuring that maps are contiguous and meet certain population bounds. We then calculate segregation metrics for each map to construct probabilistic distributions of segregation indices. With these data we provide bias-corrected estimates of racial segregation across U.S. cities and quantify the uncertainty induced by aggregation measurement error. The data demonstrate that official Census Tract definitions overstate the degree of racial segregation in the United States. This measurement error is most pronounced in small and medium sized cities and is most severe for two-group, rather than multi-group, segregation measures. We offer these new data as a tool for researchers and demonstrate their potential by re-examining contemporary and over-time segregation.

Keywords segregation • measurement • simulation • Census data • measurement error

1 Introduction

Residential segregation is an enduring feature of American society. Despite a diversifying population, racial segregation persists across American cities (Massey and Denton, 1993; Elbers, 2021), paralleled by increasing economic (Bischoff and Reardon, 2014; Mijs and Roe, 2021) and political (Rodden, 2019; Kaplan et al., 2022; Brown et al., 2024) segregation. Such geographic separation is associated with extensive negative societal outcomes: the perpetuation of economic inequality and racial wealth gaps (Ananat, 2011; Chetty et al., 2019), the propagation of inter-group prejudice through reducing inter-group contact and heightening group threat (Habyarimana et al., 2009; Enos, 2017), the inhibition of equitable distribution of public goods (Alesina et al., 1999; Trounstine, 2016), and threats against the functioning of democratic governance (Putnam, 2007; Chen and Rodden, 2013). Contemporary segregation is a function of historical legacies of de jure and de facto racial segregation (Rothstein, 2017; Logan and Parman, 2017) and economic constraints on housing and mobility (Reardon and Bischoff, 2011).

*Thank you to Cory McCartan.
†Email: jbrown13@bu.edu. Website: https://www.jacobrbrown.com.
‡Email: christopherkenny@fas.harvard.edu. Website: https://christophertkenny.com/.
§Email: tsimko@princeton.edu. Website: https://www.tylersimko.com.

Further, residential segregation is both further exacerbated by and contributes to ongoing housing supply and affordability crises in American metropolitan areas (Trounstine, 2018; Einstein et al., 2020).

Understanding the extent of segregation and its consequences requires precise measurement. However, researchers and policymakers face numerous methodological challenges. Foremost among these challenges is the question of how to carve up geographic space so as to best measure whether populations are unevenly distributed across that space. Segregation can be understood as the uneven distribution (how different do neighborhoods look from each other?) of group populations across a geographic area (Massey and Denton, 1988). Most commonly used measures of segregation rely on aggregate data that impose arbitrary definitions of local geography. This aggregation requirement creates measurement challenges, commonly understood as the Modifiable Areal Unit Problem (MAUP) (Openshaw, 1983) – how to draw the boundaries of a geographic unit and how different decisions change aggregate statistics – and the problem of scale (White, 1983) – how big should a geographic unit be to capture the underlying quantity of interest? Previous work demonstrates that geographic data are sensitive to the particular aggregation used (Gehlke and Biehl, 1934; Jelinski and Wu, 1996; Fotheringham and Wong, 1991).

Many researchers use Census Tracts, defined by the U.S. Census Bureau, to measure segregation. Census Tracts are drawn by Bureau officials to meet a set of pre-specified criteria, including population thresholds and, when first established, “to be as homogeneous as possible” (U.S. Census Bureau, 2023). While these criteria are no doubt reasonable, a simple question arises: what if the Census Tract map had been drawn differently? Would we still observe the same segregation estimate? Census Tracts are often used as proxies for local neighborhoods, but in reality are arbitrary boundaries drawn for statistical purposes. Without some actual definition of real neighborhood boundaries (for example, see Wong et al., 2012), there are many reasonable versions of a Census Tract map that could exist, each providing a slightly (or dramatically) different segmentation of the city. The segregation estimate from an official Census plan is just one snapshot of what segregation looks like under that segmentation. But if other reasonable alternative Census Tract maps yield different segregation estimates, then there is unquantified bias and uncertainty induced by unexamined aggregation choices. While the potential for these problems to distort geographic variables is well documented, and recent work has leveraged new geographic data to make progress on aggregation challenges (Dmowska and Stepinski, 2024; Brown and Enos, 2021), most segregation research continues to rely on metrics where the potential for aggregation measurement error is present and the extent of underlying error is not well understood (Hwang and McDaniel, 2022).

In this article, we diagnose this measurement error by developing new methodological tools that measure segregation across millions of alternative possible Census Tract maps. We evaluate the bias and uncertainty in segregation measures using recent advances in algorithmic redistricting methods and software to redraw geographic Tract boundaries (DeFord et al., 2021; Autry et al., 2019; McCartan and Imai, 2023; Kenny et al., 2024a). We sample alternative Census Tract maps that follow Census guidelines: ensuring that maps are contiguous, respect natural boundaries, and meet certain population bounds. We then calculate segregation metrics for each map to construct probabilistic distributions of segregation indices. These distributions reflect the segregation estimates that we could observe across many reasonable Census Tract definitions. We use these data to provide bias-corrected estimates of racial segregation for every U.S. city with a population of over 100,000 people in 2020, and quantify the uncertainty induced by aggregation measurement error.

Our results demonstrate that official Census Tract definitions overstate the degree of racial segregation in the United States. We find that the most commonly used segregation measure, the Dissimilarity Index, with official Census Tract data overstates Black-White segregation in our sample by 16.8% on average. We calculate this bias by comparing the observed segregation estimates to the averages from our simulated distributions. We evaluate uncertainty by taking the 95% credible intervals from our simulated distributions. This analysis reveals that, on top of the above reported bias, for the typical city the Black-White Dissimilarity index could reasonably be as much as 11.9% higher or lower than the bias-corrected average.

Bias and uncertainty are most pronounced in small and medium sized cities, and our estimates most resemble official statistics in large cities. We find this variation is largely driven by size: smaller cities have fewer Census Tracts, so each redrawn map may dramatically restate how segregated the city looks. By contrast, bias tends to be small and average out across simulations in larger cities with many Census Tracts. The observed bias is also most severe for two-group, rather than multi-group, segregation measures. The H Index, for example, only overstates multi-group segregation by on average 3.8%.

We further describe how our analyses inform understanding of how segregation has changed across recent decades. We conduct equivalent simulation analyses for 2000 and 2010 and find that the bias of official Census statistics is stable across recent decades. However, we also find that official Census statistics slightly exaggerate the extent to which racial segregation has declined since 2000.

We make four key contributions. First, we quantify both the bias and uncertainty in common segregation metrics that result from aggregation measurement error. Previous research demonstrates the potential for aggregation challenges to distort geographic measurement (Gehlke and Biehl, 1934; Openshaw, 1983; White, 1983; Wong, 1997; Jelinski and Wu, 1996; Fotheringham and Wong, 1991), and some studies offer methods to circumvent the Modifiable Areal Unit problem in certain contexts (Hennerdal and Nielsen, 2017), but to the best of our knowledge this is the first study to fully diagnose the extent of this measurement error for the measurement of segregation. Some progress on aggregation issues to date comes from studies with access to individual-level data making them unconstrained by aggregation (or at least able to measure geographic variables across many different contextual definitions) (Östh et al., 2014; Logan and Parman, 2017; Brown and Enos, 2021). Such data, while increasingly available, are expensive and most segregation research must still rely on publicly-available demographics data from the U.S. Census or similar institutions in other countries. Many of the most cited recent studies of segregation's effects – in economics, political science, and sociology – use such aggregate data (for example, Ananat, 2011; Trounstine, 2016; Legewie and Schaeffer, 2016). Other studies make progress on different problems with common segregation metrics, such as aspatiality, lack of decomposibility, and sensitivity to baseline city demographics (Reardon and O'Sullivan, 2004; Lee et al., 2008; Mazza and Punzo, 2015; Roberto, 2018). While these other problems are related, none of their solutions solve problems of aggregation. These modern but less commonly used improvements thus still suffer from the challenges we diagnose in this article.

Second, we show that the U.S. Census Bureau's official Census Tract definitions (moderately) overstate segregation. Rather than represent the typical conclusion from the distribution of the many ways that Census Tract maps could be drawn, official Census definitions provide a picture of segregation in the upper tail of that distribution. The U.S. Census first drew nationwide Census Tracts in 1940, and updates them every ten years by splitting Tracts that grow too big (we fully detail how the U.S. Census draws Census Tracts in Section 4). We suggest this bias is most likely due to how the U.S. Census Bureau explicitly prioritizes population homogeneity when constructing Census Tracts, as they aim to make tracts "as homogeneous as possible with respect to population characteristics, economic status, and living conditions." We show this decision is likely to inflate segregation indices by artificially inducing additional demographic differences between geographic units. Any comparison, therefore, of how different Census Tracts look in terms of racial demographics is using racial demographics to create the unit across which the comparison is being made and as the metric of comparison. Consequentially, the geographic units most used to measure segregation were inadvertently created to maximize it. This is akin to measuring whether different age cohorts have different political attitudes by deciding to define the age cohorts in part based on the homogeneity of their attitudes. Our analyses identifies this discrepancy, and provides recommendations for when, where, and with which metrics these measurement issues are most severe.

Our third key contribution helps to resolve this measurement problem in future research by providing original data that researchers can use to measure segregation and its impacts. We offer nationwide simulated Census Tract plans for every county and every town and city in the United States for the years 2000, 2010, and 2020.

These data are open-source, free to use, and can be readily adapted to simulate other geographic units and contextual variables.

Fourth, we provide a methodological approach with open-source software that researchers can use to diagnose the extent of this bias in their own applications. Based on our findings that bias is largest in small geographies, we particularly recommend researchers use our approach when their research question concerns segregation across small areas. If our pre-provided data does not suit their needs, researchers can use our provided template software to diagnose the sensitivity of their results to alternative ways that geographies like Census Tracts could have grouped the same population in different ways. While we focus on Census Tracts in our application, our software can be used with any geography. This means that researchers can easily adapt our approach to examine the sensitivity of their results to the ways that **any** (Census or non-Census) geographies could have been redrawn, such as voting precincts, school attendance zones, or neighborhoods.

The rest of the paper proceeds as follows. Section 2 describes in greater detail the challenges in accurately measuring segregation. Section 3 formalizes understandings of how aggregation choices leads to bias and uncertainty in segregation metrics. Section 4 explains how U.S. Census Tracts are drawn. Section 5 describes our simulation analyses, including incorporation of previously discussed Census Tract drawing rules, as well as performance statistics and validation. Section 6 presents results from these simulation analyses, comparing official U.S. Census statistics to simulations. Section 7 reports our simulation estimate over-time, examining segregation across 2000, 2010, and 2020. Section 8 summarizes these analyses and concludes with broader takeaways and areas for future research.

2 Segregation Measures are Sensitive to Geographic Aggregation

Measuring segregation requires the distillation of continuous information on the locations of groups in (at least) two-dimensional space into numerical summaries of how unevenly these groups are distributed across that space (Jahn et al., 1947; Gatrell, 1983). Even with complete information on the exact location of all individuals in a city, measuring segregation is difficult, but researchers for the most part must work with incomplete information: with statistical summaries from aggregate units. These data limitations introduce a litany of measurement challenges. Most prominent among these are problems of aggregation, referred to as the Modifiable Areal Unit Problem (MAUP) (Openshaw, 1983), and the problem of scale (White, 1983). The MAUP is the problem of deciding where to draw the boundaries of a geographic unit, and how this decision may impact summary statistics about the unit being drawn. For segregation measures, this means that grouping the same population into different boundaries can alter how segregated groups appear to be across these boundaries. The problem of scale is the question of at what geographic resolution, so to speak, to start drawing the unit in question. In practice, this might manifest as deciding between measuring a contextual variable at the Census Block, Census Block Group, Census Tract, Zip Code, Town/City, County, or Country level (how big a space is the relevant context?). Ideally, researchers hope that the units they choose (or the units they must rely on) reflect real-world neighborhood (or other contextual) definitions that are well-suited to their quantities of interest. In practice, this is often not the case, as most units where researchers can access geographic data reflect largely arbitrary carvings up of geographic space. For example, previous research that has collected data on how people view their neighborhoods argues that Census Tracts are a flawed proxy for subjective neighborhood definitions (Wong et al., 2012; c.f. Velez and Wong, 2017).

Thus, any given map of geographic units reflects one particular aggregation among many options that could have been chosen. To the extent that this set distorts the ground truth, or deviates from a typical division of the geographic space, then the statistics derived from this aggregation will contain measurement error. We focus on diagnosing measurement error from the MAUP in our analysis, rather than testing sensitivity to different scale choices. Census Tracts are roughly constant in size and thus scale is fixed in official statistics. See Lee et al. (2008) for a diagnosis of segregation using different neighborhood units. Previous work shows how MAUP can distort numerical summaries, regression coefficients, and multivariate analyses (Gehlke and Biehl, 1934; Openshaw, 1983; Fotheringham and Wong, 1991). Other studies further assess how the effects of the MAUP vary by the levels of spatial autocorrelation in the data (Jelinski and Wu, 1996; Sang-Il Lee and Griffith, 2019).

Approaches to making progress on the MAUP can generally be organized into three groups. First, geographic research increasingly utilizes data on individuals and their locations to provide more flexible measures of geographic variables (e.g, Östh et al., 2014; Dinesen and Sønderskov, 2015; Enos, 2016; Logan and Parman, 2017; Larsen et al., 2019; Brown and Zoorob, 2022; Brown, 2024). With data on individuals rather than geographic units, these studies can measure geographic variables across essentially any contextual definition to compare how results are sensitive to aggregation choices. Individual data also offer the advantage of incorporating more precise information to construct more specific measures. For example, in their study of partisan segregation Brown and Enos (2021) construct metrics of partisan exposure by identifying each U.S voter's 1,000 nearest neighbors and measuring partisan exposure within this individual-specific definition, further using the distance individuals live from their neighbors to weight exposure by proximity. In recent work on racial segregation, Dmowska and Stepinski (2024) disaggregate Census Block data into monoracial cells and conducts pattern-based analysis of racial clustering using these data. While these approaches offers some advantages, to a certain extent they are not circumventing the MAUP so much as shifting its focus. Rather than the question being whether a geographic unit was drawn correctly, the question now becomes whether the nearest neighbor cutoff was correctly chosen, and whether the proper distance-decay or clustering function was used in modeling exposure. These studies also do not offer clear sense of the extent to which the MAUP is distorted using aggregated geographic data, but rather presumes that the bias is sufficient so as to motivate the individual spatial data approach.

Though analyzing individual data offers an alternative approach to analyzing segregation with some advantages, comprehensive and contemporary data on individual residences is extremely unlikely to be released by the Census to protect the privacy of respondents. This means that researchers, policymakers, communities, and other Census data users will likely always need to use aggregated geographic units in some way. The Census Bureau faces a difficult trade-off between private and accurate data for individual responses, and existing research examines the impact of various privacy protection systems adopted by the Bureau (Kenny et al., 2024b; Hotz and Salvo, 2022; Ruggles and Riper, 2022; Hotz et al., 2022; Kenny et al., 2021; Santos-Lozada et al., 2020; Ruggles et al., 2019). While ongoing conversations discuss how exactly the Census Bureau should navigate this trade-off (Abowd, 2024; McCartan et al., 2023b; Muralidhar et al., 2024; Hotz et al., 2024), future advancements in using individual data are unlikely to solve the aggregation issue we consider here.

A second group of studies tries to avoid the MAUP problem by identifying the "correct" measure of local context for a given application. Generally focusing on individual perceptions of their neighborhoods, these studies survey individuals to measure what they define as their local neighborhood, then use these subjective definitions to measure how demographics within one's subjective neighborhood relate to social and political attitudes (Wong et al., 2020; McCartan et al., 2024). For connecting local context to individual behavior, these studies offer a promising path forward, but they are less useful for studying segregation and its impacts due to the person-specific nature of their data collection and theoretical data generating process.

The third group of studies focuses on the measurement of geographic phenomenon using common aggregate data and try to correct it through modeling or simulation approaches. These studies are distinct from a much longer literature discussed above evaluating sensitivity to the MAUP under different conditions and contexts, focusing on the next step of correcting or finding less sensitive alternatives without requiring completely new data. Given the long-standing intractability of the MAUP, these studies are fewer in number, but some work has made progress. For example, Nelson and Brewer (2017) develop techniques for analyzing and visualizing the spatial relationships between a variable and itself at different aggregation levels to measure clustering of geographic variables across different geographic scales. Related work by Hennerdal and Nielsen (2017) offers alternative multi-scalar approaches to measuring clustering at different geographics scales. While these studies grapple directly with different components of the MAUP, they primarily offer methods for demonstrating the durability of results across different available aggregation units, rather than assessing the measurement error inherent in an available aggregation.

In many applications, aggregation issues like MAUP are ignored because researchers lack proper tools to diagnose how sensitive measurements are to aggregation challenges. Most studies of segregation use racial demographic data from the U.S. Decennial Census or American Community Survey. Local segregation is then calculated by metrics that compare how different racial group populations are distributed across sub-geographies (usually Census Tracts). These metrics focus on what Massey and Denton (1988) refer to as evenness, the dimension of segregation encompassing how equally (or unequally) racial groups are distributed across geographic units. Other dimensions of segregation include exposure (on average how many people of group B live in the same Census Tract as the typical person in Group A), clustering (are homogeneous neighborhoods located close to one another), concentration (how much land area do different groups inhabit), and centralization (how close to the geographic center do minority groups live) (Massey and Denton, 1988). For the most part, however, when researchers measure segregation they focus on evenness metrics (Hwang and McDaniel, 2022). The most common is the Dissimilarity Index, which measures two-group segregation and can be interpreted as what proportion of the minority group in the city would need to move to achieve complete integration between two groups (proportions of group in each unit that match the proportions of those groups in the city as a whole) (Duncan and Duncan, 1955).

Let $\mathbf{D}_j$ be the Dissimilarity Index for a given city j with n Census Tract. Let A_j and B_j be the populations of Group A and B, respectively in the city, and a_i and b_i be the population of Groups A and B in each Census Tract such that $\sum_{i=1}^{n} a_i = A_j$ and $\sum_{i=1}^{n} b_i = B_j$. The Dissimilarity Index is calculated by:

$$\mathbf{D}_j = \sum_{i=1}^{n} \left| \frac{a_i}{A_j} - \frac{b_i}{B_j} \right|$$

The Dissimilarity-Index only calculates segregation between two groups. Given the racial history of the U.S., historically, this index has been commonly applied to measure Black-White segregation. As the U.S. diversifies, particularly with large increases in Hispanic populations, many segregation studies increasingly focus on multi-group segregation (Reardon and Firebaugh, 2002). These indices measure diversity across multiple groups and compare diversity across (typically) Census Tracts. The most commonly used of these is the entropy-based H Index, developed by Theil (1967). This metric measures the entropy score (measuring multi-group diversity) of each Census Tract and then compares the weighted average of Census Tract entropy to the entropy of the city as a whole.

Let E_i be the entropy for neighborhood i, and E_j be the entropy for city j. For each racial group r, let p_{ri} be the population of racial group r in neighborhood i, and let P_{rj} be the population of racial group r in the city. Let p_i be the population of neighborhood i and let P_j be the population of the city, such that $\sum_{r=1}^{R} p_{ri} = p_i$ and $\sum_{i=1}^{n} \sum_{r=1}^{R} p_{ri} = \sum_{i=1}^{n} p_i = P_j$. Neighborhood entropy, city entropy, and the H Index are calculated through the following three equations:

$$\mathbf{E}_i = \sum_{r=1}^{R} \frac{p_{ri}}{p_i} \times \mathbf{ln} \left(\frac{p_i}{p_{ri}} \right)$$

$$\mathbf{E}_j = \sum_{r=1}^{R} \frac{P_{rj}}{P_j} \times \mathbf{ln} \left(\frac{P_j}{P_{rj}} \right)$$

$$\mathbf{H}_j = \sum_{i=1}^{n} \frac{p_i}{P_j} \left(\frac{E_j - E_i}{E_j} \right)$$

In the main text, we primarily focus on the Black-White Dissimilarity index and the H Index measuring multi-group segregation for Black, Hispanic, White, and an Other racial category. In the Supporting Information, we report simulation analyses focusing on other segregation indices, measures of exposure and isolation, and alternative racial category versions of the above two indices.

The MAUP is paralleled by other challenges in measuring segregation. These includes problems of aspatiality, lack of decomposibility, and sensitivity to baseline city demographics (Reardon and O'Sullivan, 2004; Lee et al., 2008; Mazza and Punzo, 2015; Roberto, 2018). Aspatiality refers to the "checkerboard problem" (White, 1983), wherein common measures of segregation compare differences across Census Tracts without considering where Tracts are in relation to each other. The same arrangement of uneven Tracts in any spatial grouping will yield the same Dissimilarity or H Index result. Spatial measures of segregation typically take a Census Tract map, identify the centroid of each Census Tract, measure the distance each Census Tract is from all other Cenus Tracts in a city. With this information, researchers can calculate spatial versions of the segregation indices (including the Dissimilarity and H indices) which incorporate distance relationship, essentially returning a higher segregation metric when groups are uneven and clustered (i.e. homogeneous neighborhoods are spatially close to one another) (Reardon and O'Sullivan, 2004). These spatial measures tackle a different problem than the MAUP, and are still potentially biased by it since they still rely on aggregate units even with spatial information.

Segregation researchers are increasingly interested in developing metrics that are decomposable, that can compare segregation within and across groups or areas, such as different levels of geography (the Dissimilarity Index is not decomposable). These decomposable indices are useful for determining how much more segregated, say, the suburbs are from the city center in a metropolitan region, or whether one group is more unevenly distributed than others in a multi-group context (Roberto, 2024). Such decomposition is a promising area of innovation in segregation research but largely lies outside of the scope of our focus.

Lastly, there are a series of critiques of the Dissimilarity Index and related indices focusing on the sensitivity of the metric to baseline city demographics (Taeuber and Taeuber, 1976; Falk et al., 1978; James and Taeuber, 1985). Meaning, the Dissimilarity Index can result in different numeric outputs in response to redistributing populations across Census Tracts holding city demographics constant and in response to changing city demographics absent changes in within-city relative distributions (Winship, 1978). Despite these critiques, suggested mathematical corrections (Cortese et al., 1976), and alternative proposed indices (Falk et al., 1978) the Dissimilarity Index remains widely used in both scholarly and popular investigations of segregation. As such, we focus on measurement error in the most commonly used measures of segregation. We formalize the conceptualization of this measurement error in the next section.

3 Formalizing bias and uncertainty from aggregation

We propose to consider the MAUP as a sampling problem. There are essentially infinite ways of segmenting a city into roughly equal population areas such as Census Tracts. Each boundary could be drawn a little (or a lot) differently to place different areas in different units. All maps are reasonable draws from the sampling distribution of all possible maps, but some maps are more representative of the typical draw than others.

In this framework, we cannot know from looking at any single Census Tract map whether it is representative of all the other maps that could have been drawn under similar drawing rules, or whether it presents a distorted view of segregation. If we can observe the sampling distribution, however, we can know on average what segregation looks like by taking the average over the sampling distribution. We can also measure how uncertain we should be about this average by looking at how spread out (high versus low variance) the sampling distribution is. Then, we can place any observed Census Tract map within this distribution to see the bias in its estimate of segregation.

Figure 1 illustrates this logic. The figure shows the same geographic space with three different population groups (X's, triangles, and circles). The locations of these groups are fixed, but the top row shows two different versions of drawing a three neighborhood plan to summarize the segregation of the space. The neighborhoods are drawn to be of equal population. In this stylized example we use the two-group Dissimilarity Index, measuring the segregation between triangles and the two other groups (grouped into a single "non-triangle" group). The first neighborhood plan returns a Dissimilarity Index of 0.25, while the second returns a Dissimilarity Index of 0.625.

The bottom row illustrates that these options are two of many potential ways this small space could be reorganized into three groups. The right subplot shows a histogram of the (triangle vs. non-triangle) Dissimilarity Index values from each draw. This distribution reflects all of the possible ways (in expectation) that the space could be divided up under the rule of equal population and three neighborhoods. From this distribution we see that the average Dissimilarity Index observed is much closer to 0.25 than 0.625. Still, the distribution shows substantial variance, with significant mass as much as ten percentage points from the center of the distribution. Therefore, we would conclude, in this example, that our estimate of the true segregation has a large margin of error due to aggregation measurement error.

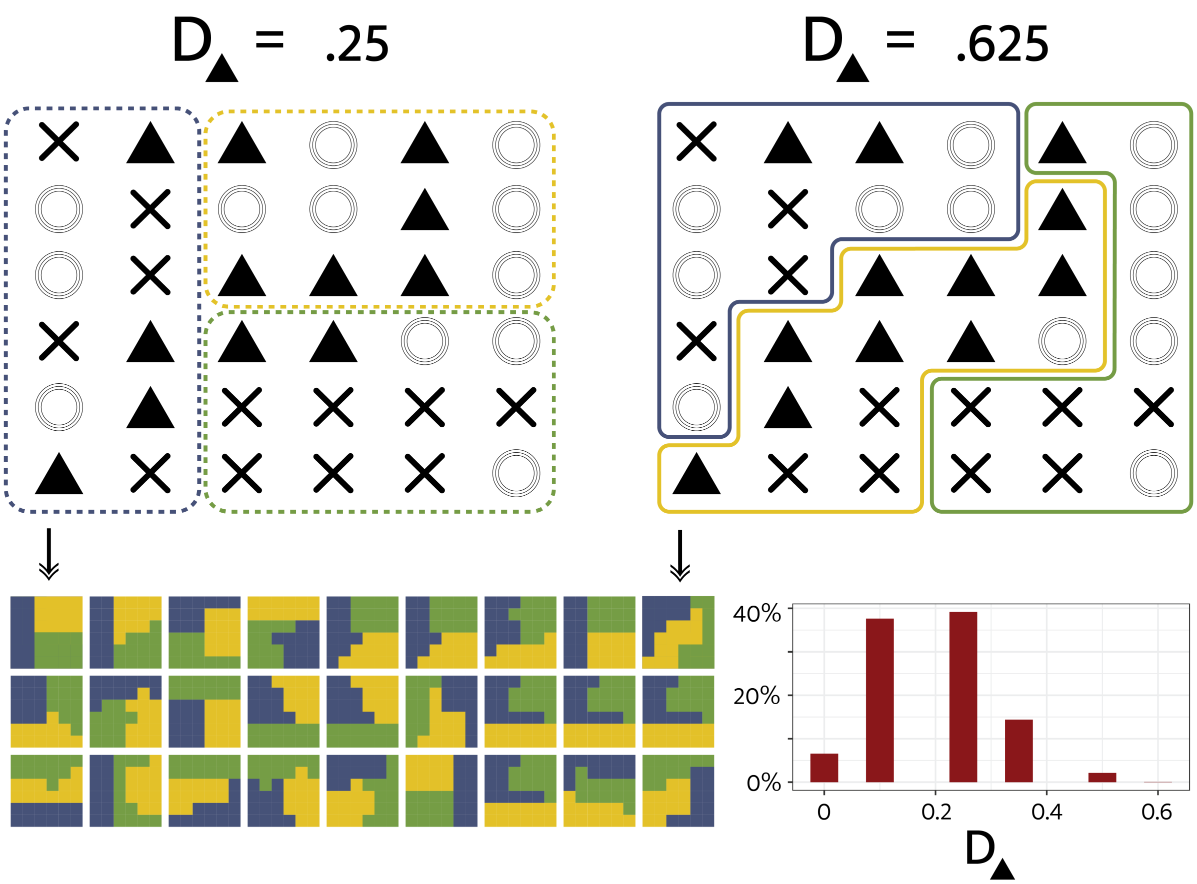


Figure 1: Different neighborhood maps, different segregation estimates

This figure also illustrates our methodological approach: we aim to simulate the Census Tract drawing process to construct representations of the sampling distribution of segregation across many alternative maps. These maps are independently drawn and adhere to Census drawing rules and population targets. This independence ensures that the sample of maps drawn reflects probabilistically the true sampling distribution. Many plans will look quite similar, particularly those closer to the center of the distribution. We calculate the bias of any observed

Census Tract map as its deviation from the sampling distribution of plausible maps. For uncertainty, under the assumption that each plan is independently drawn, we can use the quantiles of the sampling distribution to measure the spread of the distribution.

4 How Census Tracts are drawn

According the U.S. Census Bureau, Census Tracts are "are small, relatively permanent geographic entities within counties (or the statistical equivalents of counties)" (U.S. Census Bureau, 2023).[1] Census Tracts typically contain between 2,500 and 8,000 people, with a target population of about 4,000. Census Tracts were first drawn as far back as the 1890 Decennial Census in a small number of cities, but were not administered at a larger scale until 1940, when the Census began publishing Census Block data (the smallest available U.S. Census unit) for all cities above 50,000 people. This procedure expanded through the 1950-1980 Decennial Censuses. From the 1990 Decennial Census onward, the Census assigned Census Blocks and Tracts covering the the entirety of the U.S..

Census Tracts are not completely redrawn with each Decennial Census. Rather, as Census Tracts grow in population, they are eventually split into different Census Tracts once the populations exceed the target threshholds. Thus, the Census Tracts we observe today are from a path-dependence process that builds on the original definitions. Tracts were defined as early as 1910 for some US large cities like New York City, and were established nationwide throughout the course of the twentieth century (U.S. Census Bureau, 2023). At the time of original definition, Census Tracts were drawn to hit the targetted population and to adhere to "visible" features of the geographic environment such as roads, waterways, railroads, and other major infrastructural barriers. In practice, adherence to these infrastructural features is not always guaranteed as infrastructure and other geographic boundaries change. Census Tracts are nested in the Census geographic "spine," meaning they never cross larger Census geographies like counties or states.

We follow and diagnose compliance with these population targets and infrastructural guidelines when designing our simulations. Our goal is not to follow the data generating process of the Census identically, but rather to adopt a set of rules inspired by the practical limitations the Census Bureau faces that are consistent and reasonable ways of dividing up geographic space. As such, we do not emulate the method of starting with original Census Tract definitions and simulating the splitting process across decennial Censuses. Rather, for any given decennial Census we draw new Census Tract maps under the guidelines described above. This removes any temporal dependence from our distributions and more directly reflects the quantities of interest described in previous sections.

There is one key factor the Census uses when drawing Census Tracts that we do not emulate. Census Tracts were originally drawn to be as "homogeneous as possible with respect to population characteristics, economic status, and living conditions" (U.S. Census Bureau, 2023). Demographics characteristics like race and ethnicity, in effect, were used to construct Census Tracts, with a priority placed on drawing racially distinct Census Tracts. In terms of understanding segregation, this decision is problematic, because it introduces dependence between the Census Tract drawing process and the estimation of how unevenly groups are distributed across Census Tracts. As such, we do not include it in our simulations, and return to this point when considering biases of official Census Tract segregation estimates.

[1] For a complete accounting of how Census Tracts are drawn, see: https://www2.census.gov/geo/pdfs/reference/GARM/Ch10GARM.pdf

5 Simulation details

We simulate alternative Tracts from the smallest geographic unit published by the US Census Bureau: Census Blocks.[2] Census Block data are collected from PL 94-171 Decennial Census Files and contain information on race (7 categories) and ethnicity (Hispanic/not). We supplement these data with information on larger geographies similar to towns: minor civil divisions (MCDs) and Census Places. We collect data separately by state and group the data by counties.

We sample plans using "Merge-Split", a technique for constructing plans according to contiguity, population, compactness, and administrative boundary constraints, but otherwise forming plans randomly, developed by Carter et al. (2019) (based on work by Deford et al. (2019)). Merge-Split was originally created to simulated redistricting plans to assess gerrymandering, under the similar logic that any district plan can be compared to the sampling distribution of reasonable district plans (e.g, Kenny et al., 2023). We implement Merge-Split using the R software `redist` (Kenny et al., 2024a). This software integrates Merge-Split with a Markov Chain Monte Carlo simulation using the Sequential Monte Carlo sampler from McCartan and Imai (2023). This MCMC method edits two districts simultaneously and accepts them probabilistically, iteratively replacing them in the plan. We thin this, taking every k-th plan where k is large relative to the number of districts.

We ensure our simulations closely match Census boundaries in the following ways:

1. **Contiguity**: We draw Tracts from contiguous groups of Census Blocks. Tracts are contiguous shapes composed of Census Blocks. We follow this in our simulations by creating adjacency lists for every Census Block in the United States using (Kenny et al., 2024a). These adjacency lists indicate which Census Blocks are contiguous pairs, and our simulations ensure that only contiguous pairs can be grouped into Tracts. When blocks are completely discontiguous from any other blocks (e.g. islands, river segments), we connect them to their nearest neighbor Census Block.
2. **Geographic Spine**: We ensure our simulated Tracts never cross county lines by simulating each county separately. Tracts must fit into the broader "Geographic Spine," in which Census geographies nest into Blocks, Block Groups, Tracts, Counties, and States.
3. **Population**: The Census generally draws Tracts to have between 2,500 and 8,000 people, with an average of 4,000. We follow this guideline by setting population bounds on our simulated Tracts. For each county, taking the least populated Census Tract in each and set its population multiplied by 0.9 as the lower bound, and take the most populated Census Tract and set its population multiplied by 1.1 as the upper bound.

Further, we check diagnostics to ensure our simulations respect the following Census guidelines:

1. **Boundary Crossings**: We diagnose how often our simulated tracts cross interstate highways (e.g., I-90 or U.S. 1), all major railroads, and waterways defined as rivers, lakes, or channels using the same shapefiles used by the Census. The Census Bureau attempts to draw Tract boundaries that respect visible infrastructural features like major highways and rivers. There is wide variation in how this guideline is adopted across the country due to geographic and infrastructural variation. The Census releases no public information on how this is operationalized in practice, though they say such boundaries are "generally" respected.

[2] Census Blocks are statistical areas defined by visible features such as roads, railroads, waterways, property lines, and municipal boundaries [@censusblocks]. They are not defined by population, but typically contain about 40 people on average although populations can range from 0 to over a thousand people.

2. **Off-Spine Geographies**: We diagnose how often our simulated Tracts split Census Place boundaries, which represent cities, towns, and villages. Although places bear no formal relationship to Census Tracts by being off the main geographic spine, we calculate diagnostics to ensure places are split at similar rates in our simulations compared to enacted Tracts.
3. **Compactness**: We use a "soft," probabilistic constraint to nudge our simulated tracts into compact shapes. Census Tracts generally have "compact," broadly rounded or rectangular shapes. There is no official documentation or criteria that we are aware of as to particular requirements on compactness. These shapes may also be a consequence of other decisions, such as the path-dependency from earlier hand-drawn maps or decisions to follow major infrastructure. Specifically, we use the compactness constraint implemented in (Kenny et al., 2024a), and use the author's recommended value of 1.

We run separate simulations by county[3] and decennial census year (2000, 2010, 2020). For each county, we run between 25,000 and 4,000,000 simulations, increasing simulation iterations based on county size and convergence evaluation. In each iteration, Merge-Split creates a complete Census Tract map for the county, creating the same number of Census Tracts as the official Census Tract plan with population constraints approximately equal to the population range of Census Tracts in the official plan. We run four independent chains. Each chain runs for at least 50,000 steps and stops when converge diagnostics are met. We test both plan diversity (McCartan and Imai, 2023) and $\hat{R}$ statistics (Vehtari et al., 2021) to evaluate convergence and whether the simulation distribution reflects the theoretical sampling distribution of plans.[4] We perform a final thinning step to output 5,000 simulated plans for each county (1,250 from each chain).

We then calculate for each of these 5,000 stored plans the Dissimilarity Index for each pair of racial groups (Black-White, Hispanic-White, Hispanic-Black), the H Index for Black, White, Hispanic, and an Other racial category, and the Isolation Index for each racial group.[5]

To illustrate this output, we focus on the example of Elizabeth, New Jersey. With a population of 137,298, 31 Census Tracts, and 951 Census Blocks, Elizabeth represents a typical small city in our data. Figure 2 plots the official Census Tract map (top left) for Elizabeth in 2020 compared to a randomly sampled simulated plan (top right). Each map plots the two interstate highways that cross through Elizabeth as well as a major waterway. This side-by-side comparison demonstrates how a typical plan from our simulations might differ from the official Census Tract map, organizing Elizabeth into different Census Tract definitions. In this example, the official plan contains more compact Census Tracts than the simulated plan. We also observe how the official and simulated plans do (and do not) respect highway and waterway boundaries. In Elizabeth, the official Census Tract map completely respects the waterway (as in Census Tracts do not cross the waterway), and only partially respects the highways. The simulated plan generally adheres to the waterway but does not as assiduously respect it as the official plan. There are some cases where the simulated plan respects the highways and the official plan does not and many cases where both plans cross the highways.

In the bottom row of the figure, we compare chloropleth maps for the official plan and the simulate plan shading each Census Tracts by their proportion Black population. The official Census Tract map in Elizabeth returns a Black-White Dissimilarity estimate of 0.415, while the simulated plan returns an estimate of 0.350. These numbers can be compared to the complete distribution of Black-White Dissimilarity Index from all 5,000 Elizabeth simulated plans, which are shown in the left panel of Figure 3. The average of this simulated distribution for Elizabeth is 0.373, much closer to the simulated plan in Figure 2 than the official Census Tract plan, which falls in the right tail of the distribution. Figure 3 also shows the simulation distributions for the

[3]For counties consisting of just 1 populated Census Tract, of which there are 225, we do not simulate the Census Tract map but rather return the official Census Tract.

[4]We report convergence statistics in the Supporting Information Section A.

[5]The Isolation Index reflects the exposure of Group A to itself across Census Tracts. Thus, it is a weighted average of the proportion of Group A in each Census Tract, weighted by the number of Group A residents in each Census Tract.

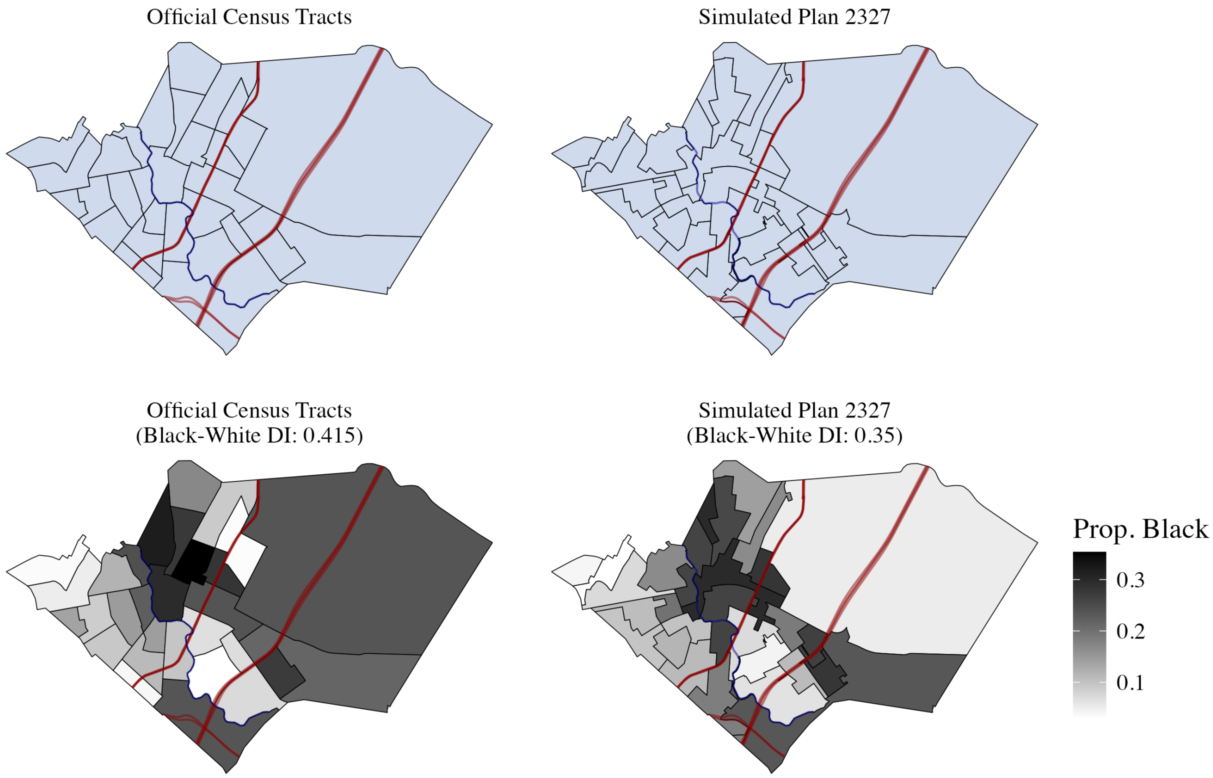


Figure 2: Elizabeth, NJ Official versus Simulated Tract Map

Black Isolation Index and the White Isolation Index. Black and White Isolation are both overstated by the official plan, meaning that across simulations Blacks (or Whites) are more likely to live in Census Tracts with other racial groups than in the official plan. These distributions are all consistent with the general takeaway that the official Census Tract map in Elizabeth, NJ is overstating segregation in comparison to the simulations.

5.1 Our Simulations Closely Follow Census Guidelines

All together, our simulations do a good job at approximating the relevant data generating process for generating reasonable Census Tract plans. To get a more complete picture of how well our simulations match Census rates of respecting road, waterway, and railroad boundaries, in Table 1 we report these statistic for both official and simulated plans for a random sample of 100 counties. We then take random sample of 50 plans from the 5,000 simulated for each county.[6] For each sampled plan and for the official Census Tract plan for the sampled counties, we calculate the proportion of the length of highways, railroads, and major waterways in that county that correspond with a Census Tract border. The table reports these statistics, demonstrating that on average our simulations do a similar but if anything better job – as in have Tract borders coincide with more of highways, railroads, and waterways – than official Census Tract maps.

We also report comparative statistics of compactness, population deviation, and place splitting. Population deviation represents the maximum deviation from having all tracts the same size. So, it is a scaled measure of the largest or smallest tract. Compactness is measured by a graph theoretic measure of compactness which

[6] We calculate overlap statistics for a random sample of cities and plans rather than all plans across all cities due to long computational times to measure these spatial overlays.

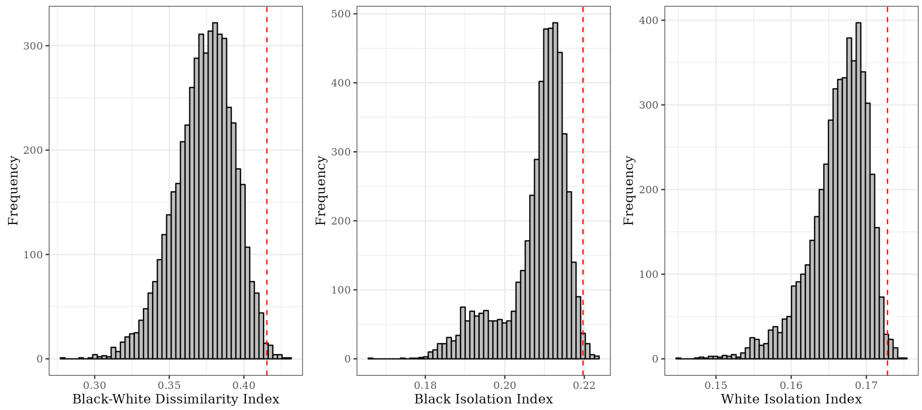


Figure 3: Elizabeth, NJ simulated distribution of segregation indices

Table 1: Comparison of boundary overlap statistics

	Official		Simulation		Difference	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Highways	0.421	0.246	0.471	0.222	−0.045	0.143
Railroads	0.363	0.242	0.461	0.213	−0.096	0.109
Waterways	0.323	0.289	0.367	0.238	−0.041	0.165

measures how many edges in the adjacency graph are kept uncut. Place splits measure the number of Census Places which are split (as in a Census Tract crosses the boundary between two Census Places). Table 2 reports the average levels of these metrics across all official plans and simulated Census Tract plans for counties consisting in 2020 of more than one Census Tract. This comparison shows that our simulation tends to draw tracts with slightly smaller population ranges, draw tracts with substantially similar compactness scores, and split towns at a very similar rate.

6 Assessing Official Census Statistics

Compared to our simulations, we find that official Census Tracts slightly, but systematically, overstate segregation across a range of measures. Figure 4 shows a scatter plot of Black-White Dissimilarity Index estimates from our simulations (x-axis) and the official Tracts (y-axis) for every city in the United States with a population over 100,000 people in 2020. The scatter plot reveals consistent bias across cities, wherein official estimates tend to overstate Black-White Dissimilarity relative to simulation averages. On average, this bias is 4.4 percentage

Table 2: Comparison of compactness, population deviation and place splitting

	Observed (N=2,915)		Simulation (N=14,578,000)			
	Mean	Std. Dev.	Mean	Std. Dev.	Diff. in Means	Std. Error
Dev.	0.635	0.448	0.524	0.360	-0.111	0.008
Comp. Frac.	0.964	0.018	0.953	0.021	-0.011	0.000
Place Splits	4.495	7.907	4.646	8.768	0.151	0.146

points. This means that, on average, official Census Tract maps return estimates of Black-White Dissimilarity that claim that 4.4 more percent of the minority group has to move in order to achieve complete integration than is actually the case. Proportionally, we find official statistics overstate Black-White Dissimilarity by 16.8% on average. This bias is not overwhelming, but still constitutes a substantial distortion from our best estimate of the actual state of Black-White segregation in U.S. cities. What's more, the consistency of the bias across many cities points to systemic patterns.

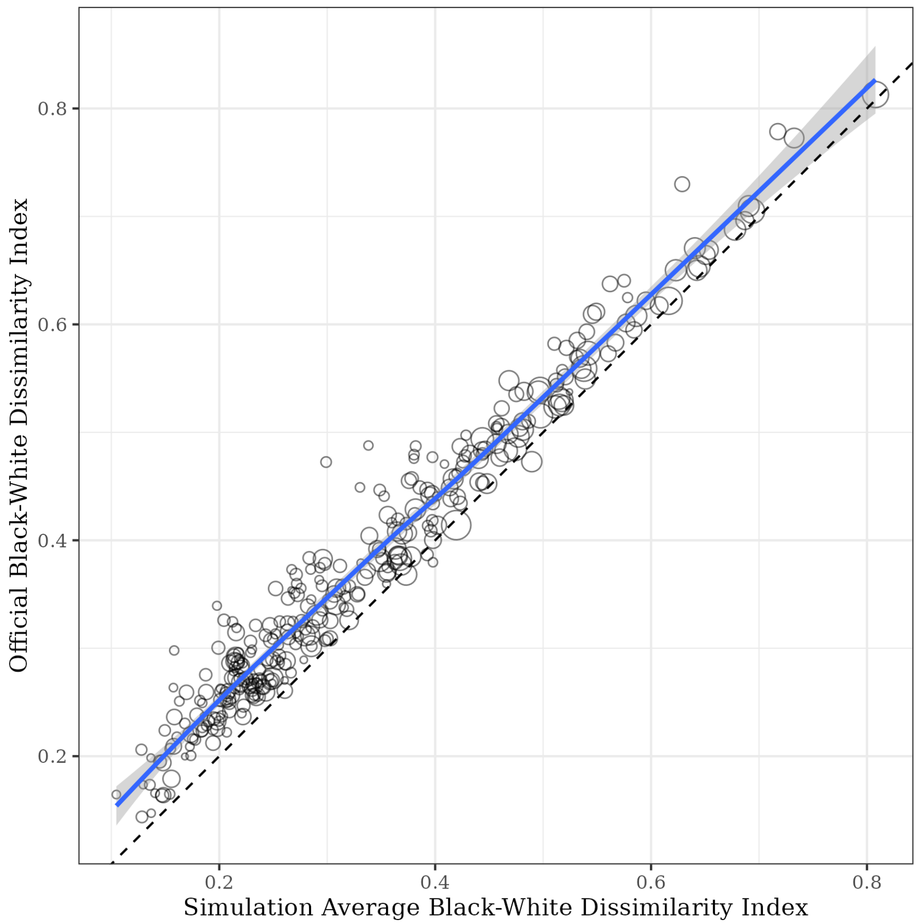


Figure 4: Official Black-White Dissimilarity versus simulated

Figure 5 further unpacks the extent of bias in official Black-White Dissimilarity statistics. Bias is defined for each city as the segregation estimate from official Census Tract definitions minus the average of our 5,000 simulated Census Tract plans for that city. The top left panel (a) plots the histogram of bias across cities above 100,000 in population. This histogram demonstrates that official statistics for almost all cities have biases above 0, and many cities even have biases of 10 percentage points or more. Thus, most cities have some sizeable bias, and some have quite high distortions.

We see clear correlations between the magnitude of bias and city population, shown in panel (d) on the bottom right of Figure 5. Bias is strongly decreasing with logged city population, meaning that bias from aggregation-induced measurement error is predominantly a problem in small to medium-sized cities, and much less severe in the largest cities. This relationship between bias and city size is likely due to smaller cities having fewer Census Tracts. For a city with fewer Tracts it is easier to draw plans that look completely different demographically, whereas in a large city with hundreds of Census Tracts any distortions across Census Tracts are more likely to average out. For example, Richmond, CA, with a population of 116,448 people, has a bias of 17.3 percentage point, a 57.9% over-estimation compared to the simulation average. Whereas Los Angeles, with a population of 3,898,747 has a bias of just 0.6 percentage points, or proportionally 1.0%.

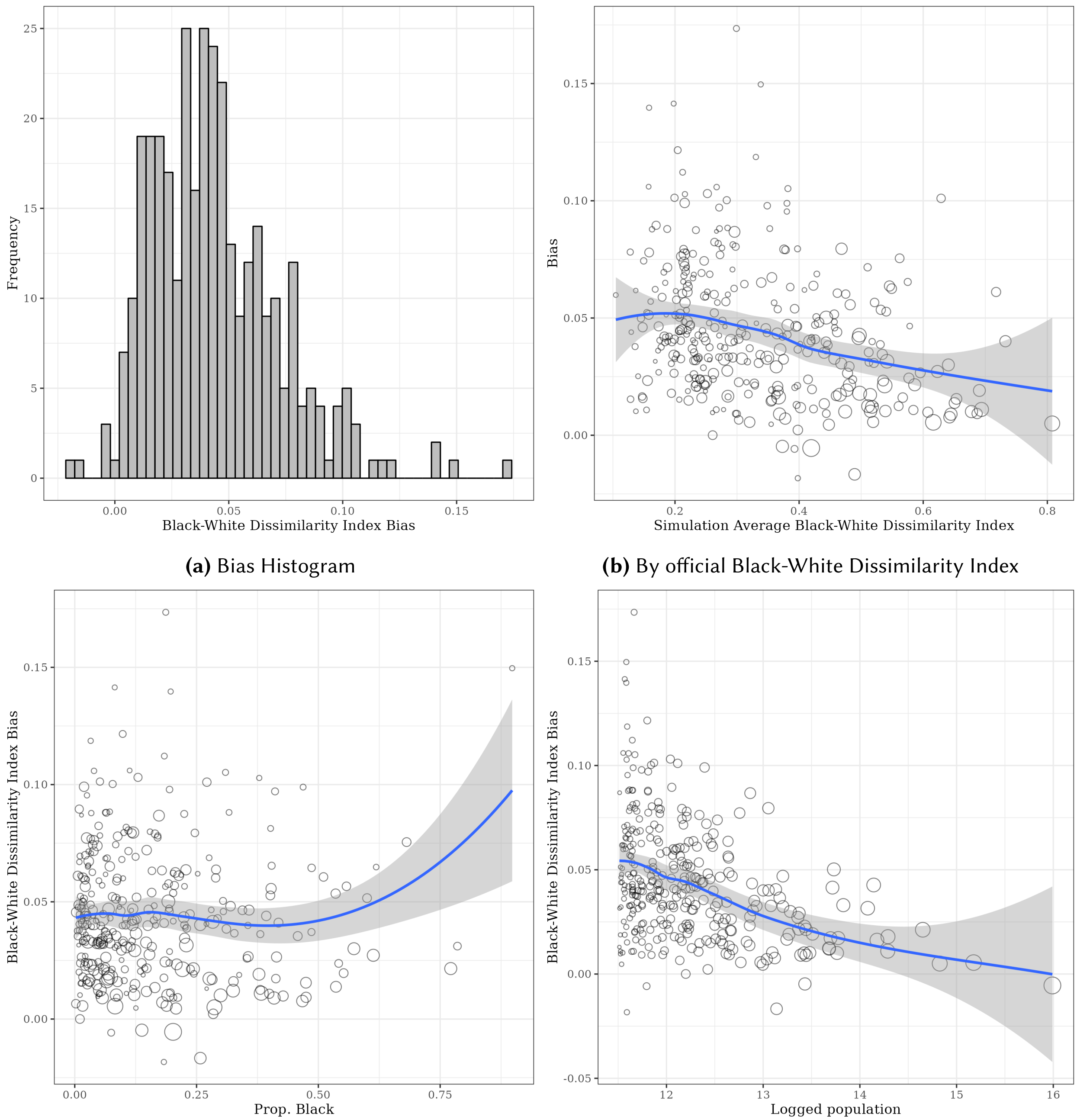
We find some suggestive evidence that official Tract maps overstate segregation more in places that are less segregated, and are more accurate in more segregated cities. Considering the mechanics of the Dissimilarity Index (and other measures of two-group or multi-group evenness), it seems likely that sensitivity to the MAUP may be most severe in more diverse cities, since those cities offer the demographic mix to either create many diverse Census Tracts, or to create many homogeneous Census Tracts that looks quite different from the overall city demographics. However, the higher the (true) segregation of racial groups within the city, the harder it may be to crack those populations into more diverse Census Tracts. Panels (b) and (c) test these hypotheses. Panel (b) in the top right of the figure plots the scatter plot of bias on the y-axis against simulation average Black-White Dissimilarity Index on the x-axis. The scatter plot shows a decreasing relationship between bias and simulation average segregation. Panel (c) plots bias against the proportion of the city's population that is Black, showing a generally flat relationship between bias and proportion Black across most of the data. The line does start increasing after approximately 0.5 proportion Black, though the support for this portion of the scatter plot is only a few cities. Thus bias may be higher in the cities with the largest Black populations, but otherwise does not increase as proportion Black increases.

Next, we examine our estimates of the uncertainty induced by aggregation measurement error for the Black-White Dissimilarity Index. To do so, we take the length of the 95% interval of our simulated Black-White Dissimilarity Index distributions (the 0.975 percentile minus the 0.025 percentiles). The longer this interval, the more uncertainty in our estimates of segregation due to aggregation measurement error. Figure 6 panel (a) plots the distribution of 95% interval length. The average interval length is 6.3 percentage points. This means that our segregation estimates are on average consistent with segregation being approximately 3.1 percentage points higher or lower than the simulation averages. In percentage terms, this translates to segregation that is on average 11.9% higher or lower than the bias-corrected average. The histogram shows that this uncertainty varies across cities, with some cities having intervals close to 0 and some having intervals higher than 10 percentage points.

These uncertainty estimates also reinforce the extent of the bias observed in official Census Tract estimates of the Black-White Dissimilarity index. 76.6% of official Census Tract Black-White Dissimilarity estimates fall outside of the 95% interval for their respective city.

How does uncertainty vary by city characteristics? The logic described above as to how bias varies with city segregation and diversity may hold for uncertainty as well, with wider distributions in cities with more diversity and narrower distributions when segregation is higher. Panel (b) plots interval size by simulation average Black-White Dissimilarity index, showing a slightly increasing relationship for low values of segregation, then a decreasing relationship across most of the support of the data. Panel (c) plots interval size by city proportion Black population, showing a flat relationship for most values of segregation but an increasing one for the largest segregation values. Thus, similar to bias, aggregation-induced uncertainty is less pronounced in more segregated cities, and generally consistent across cities with different diversity, except for cities with the largest Black populations, where uncertainty is largest. We also see decreasing uncertainty with larger city population, as simulations in cities with more people (and more Census Tracts) converge on a narrower range of segregation values.

In the Supporting Information Section B.2, we present scatter plots showing the relationship between interval size and bias across cities. We see a strong positive relationship between these two components of aggregation measurement error. Cities where official estimates are more biased also have substantially more variation in simulation estimates of the Black-White Dissimilarity. This positive correlation demonstrates that the factors



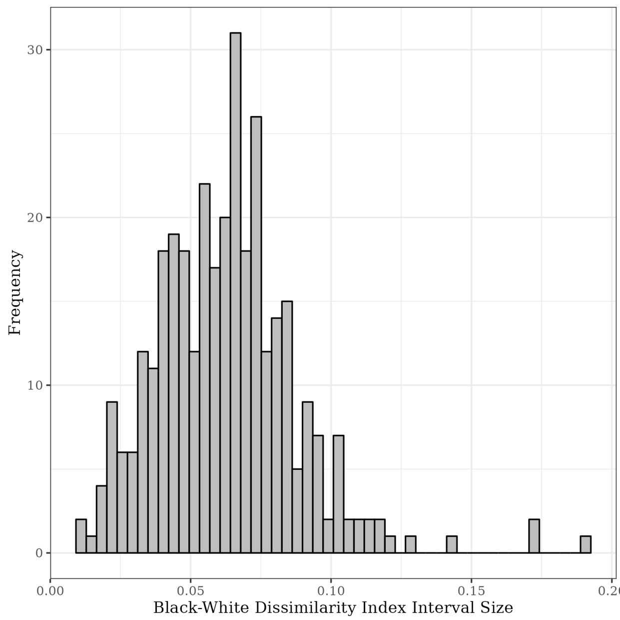
(a) Bias Histogram

(b) By official Black-White Dissimilarity Index

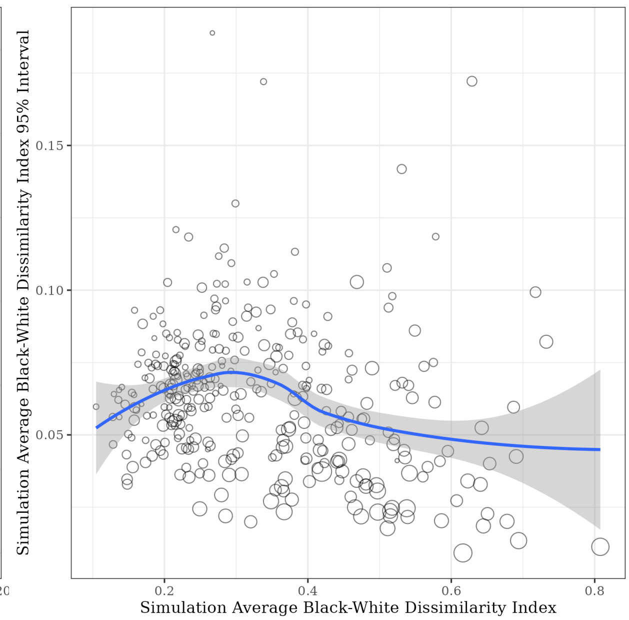
(c) By city proportion Black

(d) By logged city population

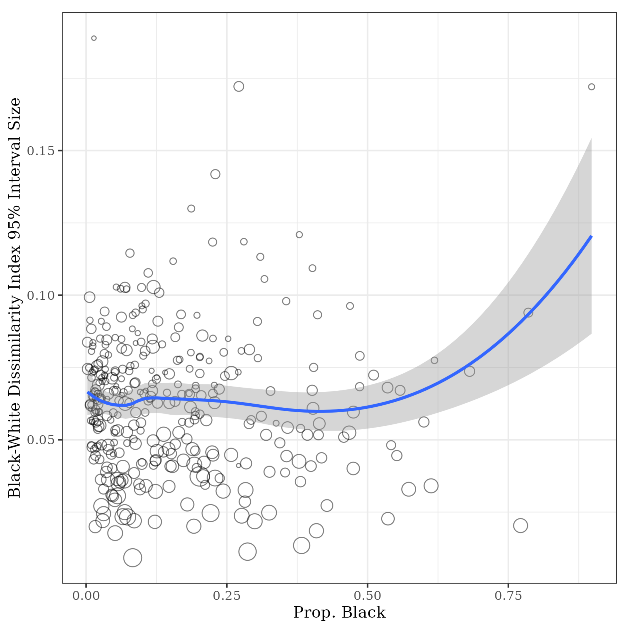
Figure 5: Black-White Dissimilarity Bias



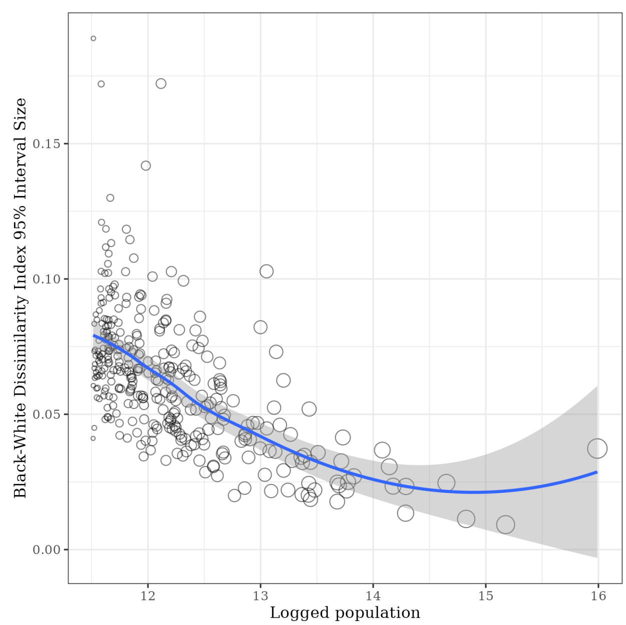
(a) 95% Interval Histogram



(b) By simulation average Black-White Dissimilarity Index



(c) By city proportion Black



(d) By logged city population

Figure 6: Black-White Dissimilarity Uncertainty

driving measurement error from the MAUP work similarly in producing both more biased and more uncertain understandings of segregation.

This first set of results reveals persistent biases in official estimates of segregation between Black and White populations in U.S. cities. In the Supporting Information Section B.4, we show results for the Hispanic-White Dissimilarity index. Official estimates of this Dissimilarity Index overstates Hispanic-White segregation by 3.8 percentage points on average, or by 16.9%. This bias is similar in size to the bias in official Black-White Dissimilarity Index estimates, so the aggregation measurement error in official estimates is not limited to Black-White segregation.

Next, we examine simulation analyses of the H index measuring multi-group segregation. We calculate this index for four racial categories: Black, Hispanic, White, and Other. Figure 7 plots the simulation averages against the official H Index statistics. Here, we see a much less pronounced pattern of bias in official estimates. On average, H Index estimates official Census Tract plans are 0.4 percentage points higher than simulation averages, a bias of just 3.8% on average in percentage terms.

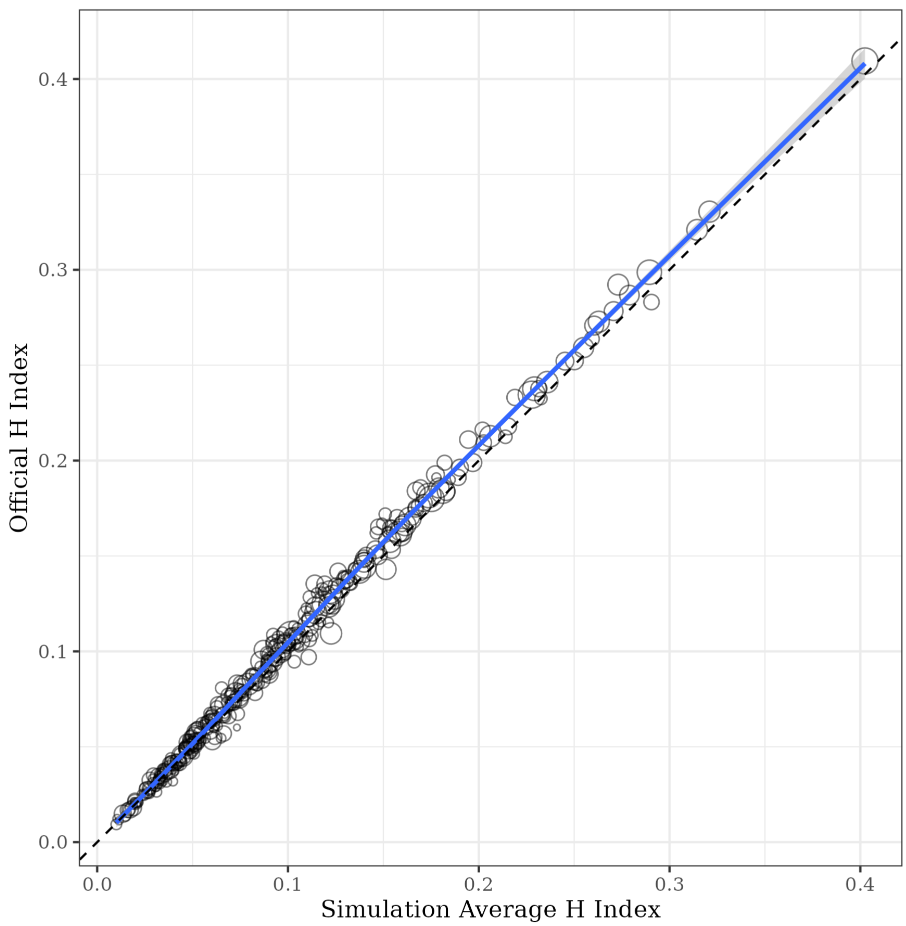
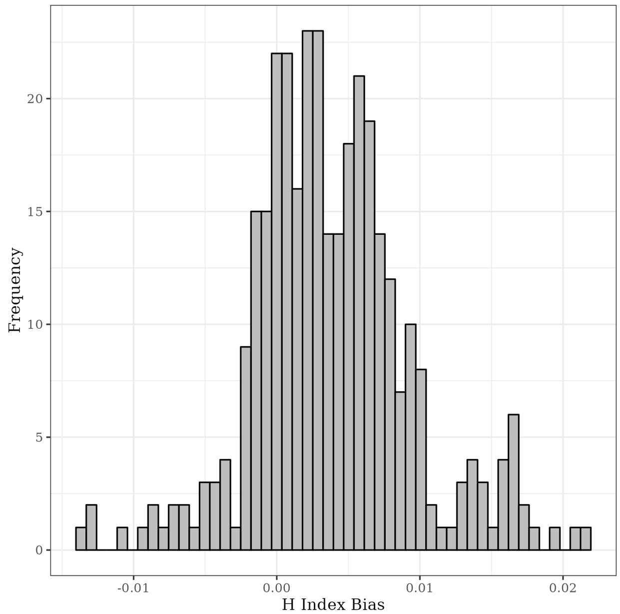
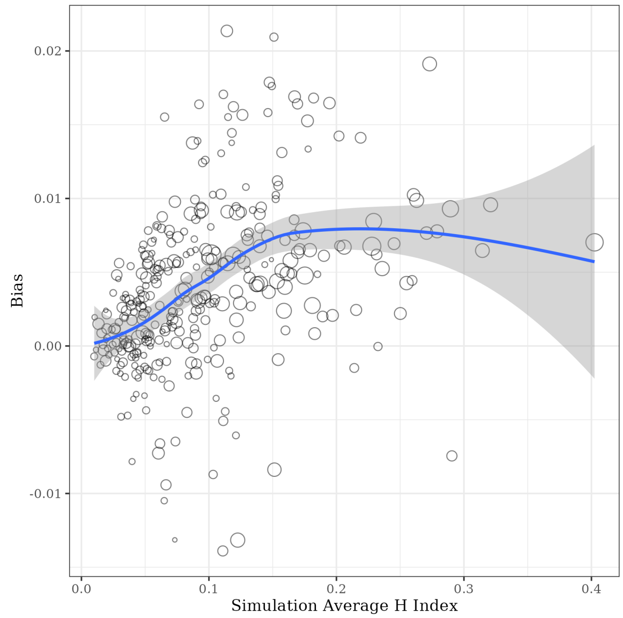


Figure 7: Official H Index versus Simulated

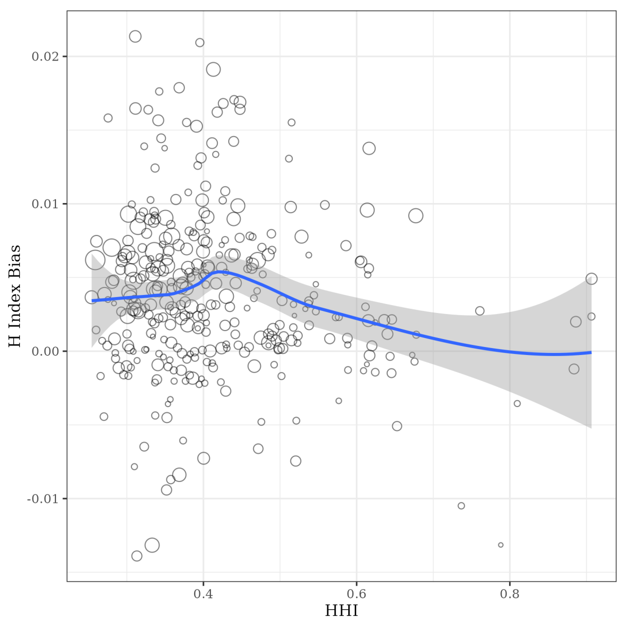
Panel (a) in Figure 7 shows the histogram of H Index bias across cities. The span of the distribution ranges from -1.4 percentage points to 2.1 percentage points. So unlike, the Black-White Dissimilarity Index, where very few cities had a negative bias, for the H Index we observe both positive and negative bias, but overall the bias is centered close to 0. Even in absolute terms, however, official H Index estimates deviate from the simulated averages by just 0.5 percentage points on average, a 5.9% deviation.



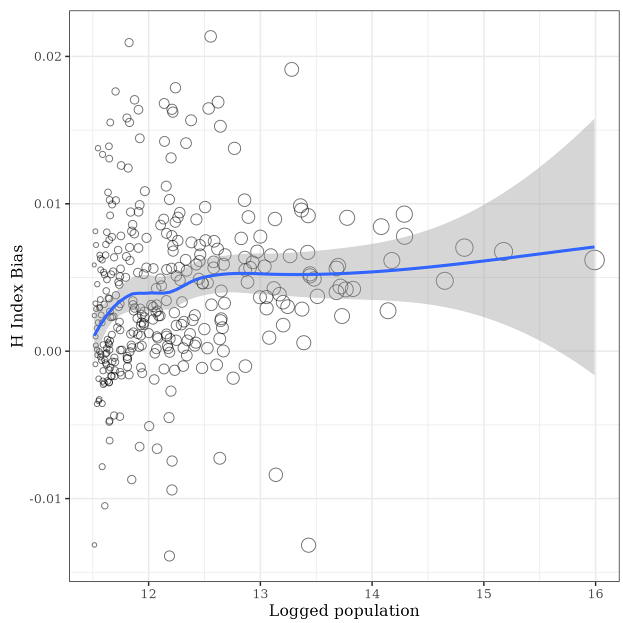
(a) Histogram



(b) By simulation average H index



(c) By city Herfindahl Index



(d) By logged city population

Figure 8: H Index Bias

In panel (b) of Figure 8 we plot H Index bias against the simulation average H Index estimates. Unlike the Black-White Dissimilarity Index, where we saw a decreasing relationship between bias and segregation, here we see an initially increasing relationship before the local average line flattens out. We also see a negative relationship in panel (c) between the city Herfindhal Index (a measure of racial diversity) and bias, so the H Index exhibits less bias in official estimates in more diverse places. The relationship between city population and bias is also muted, with a generally flat line through the scatterplot.

Figure 9 plots assess H index uncertainty. The histogram in panel (a) shows that the 95% interval size ranges from 0.2 to 4.8 percentage points. The average interval length is just 1.2 percentage points. This interval size can be best compared to the Black-White Dissimilarity index in percentage terms, since the H Index and the Dissimilarity Index are not directly comparable. Based on the simulation intervals, the true H Index could be as much as 8.4% higher or lower than the bias-corrected average, whereas the equivalent calculation for the Black-White Dissimilarity Index was 11.9%.

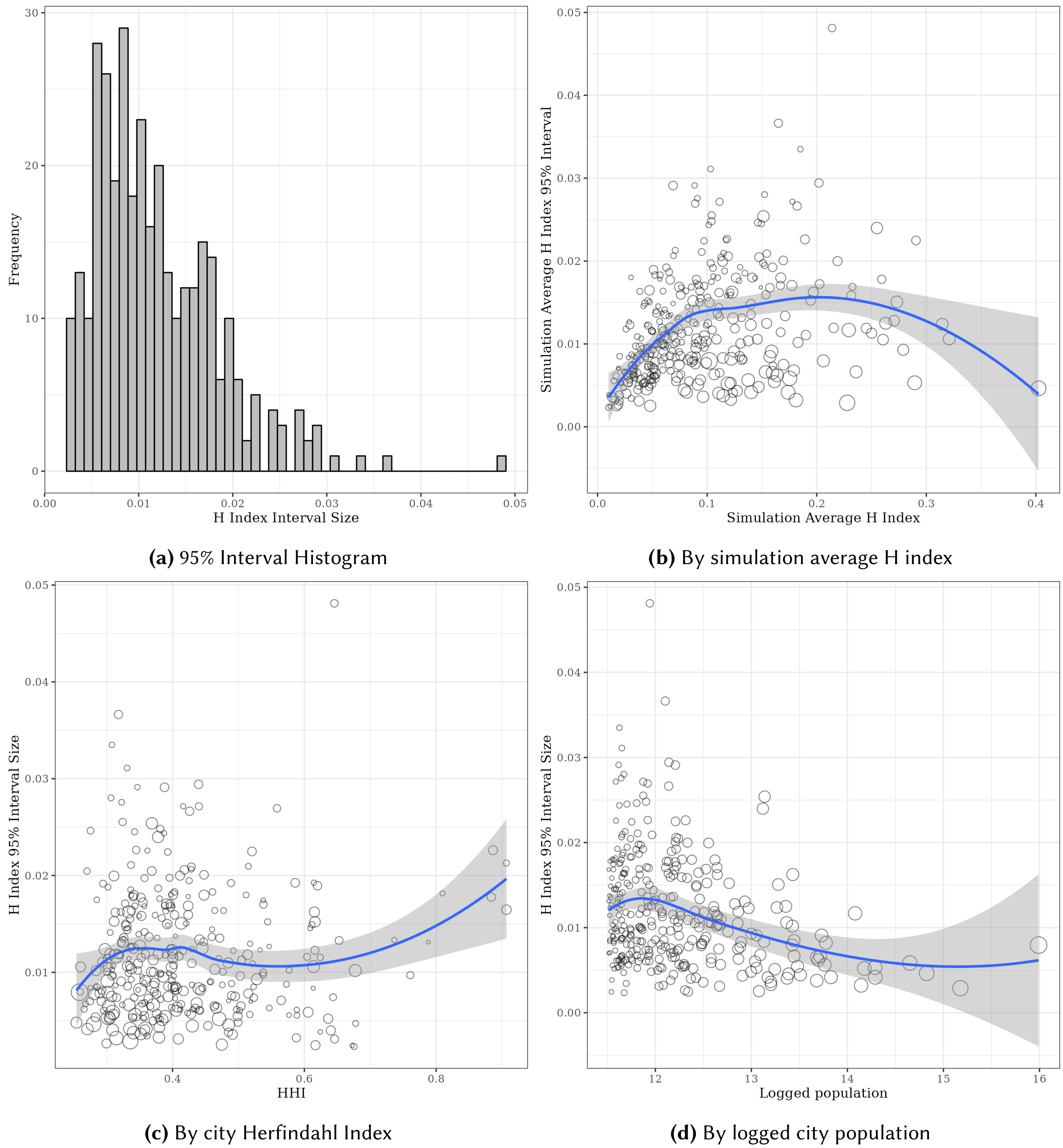
Figure 9 further demonstrates in panel (b) that there is a ∩-shaped relationship between interval size and simulation average H Index, with an initial increasing relationship plateauing across the middle of the x-axis range and then decreasing across the most cities with the highest H Index estimates. Interval size and Herfindhal Index have a roughly flat relationship that becomes increasing in the most diverse cities. Lastly, we observe a negative relationship between interval size and city population, meaning uncertainty is again decreasing with city size.

All together, these results clarify when, where, and for which metrics of segregation the MAUP is creating the most measurement error. Official estimates of the Dissimilarity Index are much more biased and exhibit much higher uncertainty than the H Index. This difference is potentially a function of the different number of groups accounted for in these indices. The Dissimilarity Index by construction considers segregation between two groups, while the H Index is a popular alternative when researchers want to consider segregation across multiple groups at once. It is possible that it is more difficult to draw Census Tract plans that differentiate segregation between more groups. With two groups the dimensionality of the problem is simpler, you can crack or pack one of the groups and put them in isolation or in exposure to the other group. But with multiple groups it may complicate this process and different maps are more likely to yield similar estimates. This logic mirrors the reasoning for why bias and uncertainty are less pronounced the larger a city is. Across many Census Tracts deviations average out, while for cities with fewer Census Tracts each differently drawn boundary is more impactful.

7 Over-time segregation

While the previous section focused on the 2020 Decennial Census, we now investigate how our methodology alters understanding of how segregation has changed over time. To do so, we replicate our simulation analyses above to redraw Census Tract plans for both the 2000 and 2010 Censuses. We then compare, for the 236 cities that had over 100,000 people in 2000, 2010, and 2020, changes in segregation over time in both simulation and official statistics. Figure 10 plots the box plot distributions of simulation (all 5,000 draws for each city) and official (one for each city) segregation index estimates separately for the years 2000, 2010, and 2020.

In each year, the average of official statistics is larger than the average of simulation statistics, particularly for the Black-White Dissimilarity Index. Thus, the bias observed in 2020 is relatively constant across the last three decennial Censuses. We also observe declining segregation from 2000 through 2020, both for official and simulation statistics and for both the Black-White Dissimilarity Index and the H index. Qualitatively, at least, the understanding that racial segregation is declining holds even when accounting for bias and uncertainty from aggregation measurement error. This consistency is likely driven, in part, by the Census Bureau's decision to keep tracks relatively constant over time (even as the underlying population distribution shifts).



(a) 95% Interval Histogram

(b) By simulation average H index

(c) By city Herfindahl Index

(d) By logged city population

Figure 9: H Index Uncertainty

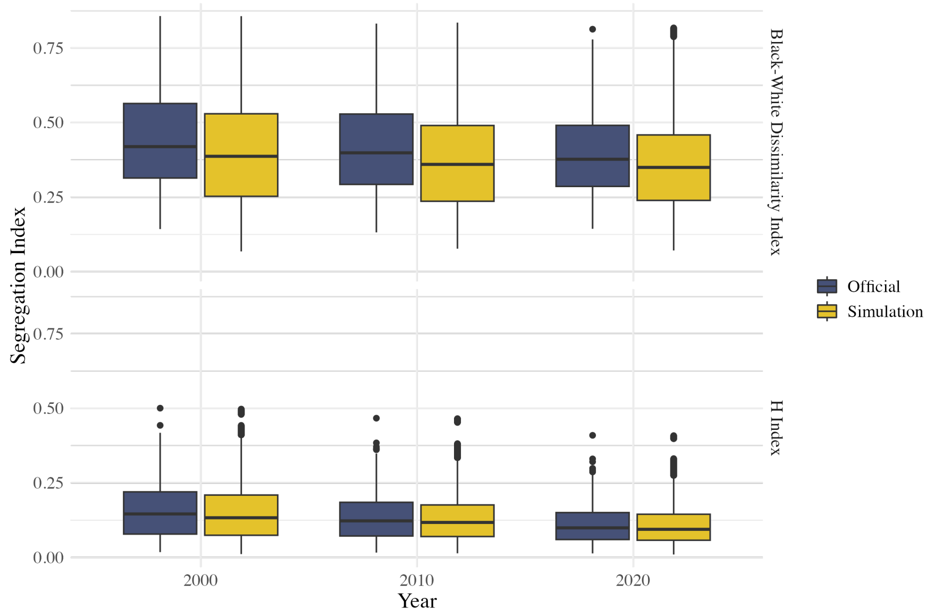


Figure 10: Official vs. simulation segregation, 2000-2020

We use regression models to estimate exactly how much segregation has declined and to see how these estimates vary across official and simulation statistics over time. Specifically, we model a simulated or official plan's segregation estimate as a function of the year, whether the plan is official or simulated, the interaction of these two variables, and fixed effect terms for city. We fit this model to the set of cities with populations over 100,000 in all three decennial Censuses, stacking the data such that each city in each year has 5,001 associated rows (5,000 simulated plans and 1 official plan), treating each plan as an independent observation (as per the independence assumption in our simulation construction).

Let $\mathbf{D}_{jtk}$ be the Black-White Dissimilarity Index for city j in year t from draw k (where k varies from 1 to 5,000 simulated draws and an "official" draw). Let $\sum_{t=2000}^{2020} I(T = t)$ be a series of 3 indicator variables taking values of 1 if a simulated plan is in year t, 0 otherwise. The terms for the year 2000 are the omitted category in the regression estimation. Let Official_{jtk} be an indicator variable taking a value of 1 if plan k is the official Census Tract estimate, 0 otherwise (and thus the plan is a simulation plan). γ_j is the city fixed effect, which constrains estimation to measure over-time and between official versus simulation statistics differences to plans in the same city. We cluster standard errors ϵ_j at the city level to account for correlated uncertainty within cities. We estimate the following linear model:

$$\mathbf{D}_{jtk} = \gamma_j + \sum_{T=2000}^{2020} \beta_T I(T = t) + \theta \text{Official}_{jtk} + \sum_{T=2000}^{2020} \tau_T I(T = t) \times \text{Official}_{jtk} + \epsilon_j \tag{1}$$

From this model, each β_T represents coefficients on whether segregation has increased among simulation estimates between 2000 and 2010, and 2000 and 2020. θ can be interpreted as the different between official and simulation estimates in 2000. The coefficients on the interaction of year and official versus simulation (τ_T) represent the extent to which official statistics differ from simulation statistics in their description of 2000-2010 or 2000-2020 changes. We also estimate equivalent estimates for the H Index, estimating similar equations as above but swapping in the H Index as the outcome.

Table 3: Modeling over-time changes in official vs. simulation segregation

	Black-White Dissimilarity Index (1)	H Index (2)
2010	-0.0260***	-0.0186***
	(0.0031)	(0.0018)
2020	-0.0369***	-0.0432***
	(0.0044)	(0.0027)
Official	0.0457***	0.0069***
	(0.0025)	(0.0007)
2010 × Official	-0.0012	-0.0017***
	(0.0017)	(0.0005)
2020 × Official	-0.0066***	-0.0021***
	(0.0020)	(0.0006)
Observations	3,540,708	3,540,708
R^2	0.95051	0.95181
Within R^2	0.16696	0.48275
City FE	✓	✓

Table 3 reports the estimates from the model. In column (1) which reports coefficients from the model estimating the Black-White Dissimilarity Index, the coefficients for 2010 and 2020 are both negative and significant, showing that segregation has declined in simulation estimates since 2000. The coefficient for 2010 is -2.6 percentage points, while the coefficient for 2020 is -3.69 percentage points. The coefficient for the indicator variable measuring whether a plan is an official plan is 0.0457, meaning that official estimates in 2000 overestimate the Black-White Dissimilarity index by 4.57 percentage points. The interaction coefficient on 2010 × Official is not statistically distinguishable from zero at conventional significance threshholds ($\alpha < 0.05$). The interaction coefficient on 2020 × Official, however, is negative and significant, with a value of -0.66 percentage points. This estimate means that compared to the -3.69 estimated decline in segregation from 2000-2020 from simulation statistics, official statistics report a 0.66 percentage point greater decline. Thus, official statistics overstate the decline in racial segregation from 2000 to 2020 by about 17.9% of the estimated decline after accounting for the MAUP.

For the H index, we observe similar estimates. Per the coefficients for the 2010 and 2020 indicator variables, multi-group segregation declined by 1.86 (2010) and 4.32 (2020) percentage points since 2000. The gap between official Census H Index estimates and simulation estimates in 2000 was 0.69 percentage points, so the H Index at any given point in time has a smaller estimated bias than the Black-White Dissimilarity Index. The interaction coefficients are both negative and significant, wtih estimates of -0.17 percentage points and -0.21 percentage points for the 2010 × Official and 2020 × Official coefficients, respectively. Thus, the decline in the H Index since 2000 is also overstated by official Census statistics, albeit to an even smaller degree than the Black-White Dissimilarity Index.

8 Concluding remarks

Measurement is the foremost challenge of social science research. Complex quantities of interest are difficult to capture using available data, distorting analyses of the drivers and consequences of social phenomena.

Segregation is a prime example of these difficulties: researchers must organize continuous information in geographic space into aggregate buckets that distort true underlying spatial relationships. New methodologies are required to understand the limitations of common metrics, and provide new tools for understanding segregation.

In this article, we investigate how a long-standing problem in geographic research — the Modifiable Areal Unit Problem — impacts the measurement of segregation using Census data. Using innovations in algorithmic redistricting software, we quantify the bias and uncertainty that aggregation choices produce in common segregation metrics that rely on official Census data. We simulate the data generating process for creating Census Tract plans and simulate the sampling distribution of all the likely ways racial populations could be organized in to Census Tracts. From these analyses, we derive bias-corrected estimates of segregation and measure the uncertainty underlying these estimates.

This analysis reveals that official Census Tract definitions yield estimates of segregation – particularly of the Black-White Dissimilarity Index, the most commonly used metric for understanding Black-White segregation in the U.S. – that systematically overstate residential segregation. Based on how the U.S. Census Bureau describes the genesis of Census Tracts, this bias is likely due to the Census prioritizing minority group homogeneity when Census Tracts were first created, and potentially continuing to prioritize racially distinct Census Tracts as it redraws Census Tracts maps every decennial Census (U.S. Census Bureau, 2023). This reasonable strategy is likely useful for Census purposes, such as producing Tracts that might proxy cohesive neighborhood-like communities. The primary purpose of Census Tracts is not to measure segregation. But they remain the primary tool for doing so, and this strategy is concerning from a theoretical measurement standpoint, as the data used to understand segregation is constructed in part based on racial demographics.

How worried should we be about official estimates of segregation? Our analysis reveals that the extent of the bias, while found in most U.S. cities, is moderate in size. Looking at cities with populations over 100,000, we find that the Black-White Dissimilarity Index is biased upward in official Census estimates by 4.4 percentage points, or 16.8%. The Dissimilarity Index can be interpreted as the proportion of the minority group who would have to move across Census Tracts to achieve complete integration. So based on our estimates, 4.4% fewer people need to move to achieve complete integration. While this is good news for citizens and policymakers looking to curb segregation, the bias is not so great as to change the broader understanding that Black-White segregation in the U.S. is high and, despite decreasing across the past two decades, persistent. Our analyses of over-time trends in segregation support the general finding that racial segregation has decreased since 2000, but also shows that it has not declined quite as much as official Census statistics report.

Several of our analyses point to contexts where official segregation metrics exhibit little measurement error. First, we find that bias is most pronounced in small and medium sized cities, and the largest U.S. cities exhibit only small amounts of bias and uncertainty. Measurement error is a much bigger problem in places like Richmond, CA or Elgin, IL than in Los Angeles or Chicago. This relationship with city size is likely due to how consequential each Census Tract decision is relative to the other Census Tracts drawn in a given plan for a given city. In a large city, cracking or packing racial groups in one area is more likely to average out across many many neighborhoods. In cities with fewer Census Tracts, each boundary decision represents a large proportion of the total boundary decisions made, and is thus more consequential. This dynamic has been observed in redistricting research as well, where at small scales there is potential for large distortions, but across large geographic scales biases tend to cancel out (Kenny et al., 2023). This heterogeneity in measurement error by city population informs as to the relationship between the MAUP and the related problem of scale – i.e. the choice of which geographic unit to use a sub-geography – when measuring segregation (White, 1983). The choice of larger (and by definition fewer) units in a city to measure evenness across will increase the potential for different plans to be drawn, and thus for greater aggregation measurement error to emerge. Conversely, a smaller unit means that more units must be redrawn, and aggregation problems are more likely to average. So researchers should be most concerned about aggregation measurement error in contexts where they have few geographic units across which to compare.

We also find that the H Index, a measure of multigroup segregation, shows only small amounts of bias and uncertainty. Official Census estimates for this index show a bias of just 0.4 percentage points, or 3.8% of simulation estimates. The differences between these two measures of segregation require further investigation. But differences may be due to the the higher dimensionality of the estimation when more than two groups are involved, making it harder to draw much different plans in terms of comparative diversity across Census Tracts. In the Supporting Information, we present analyses of single-group Isolation Index estimates, as well as the Hispanic-White Dissimilarity Index, finding bias more in line with the Black-White Dissimilarity Index analyses than the H Index analyses.

In total, our analyses provide a corrective to official statistics but can not be interpreted as fundamentally altering understandings of the current state of racial segregation in the United States. From a research standpoint, however, the consequences may be more severe. If bias is correlated with other variables use in analyses for predicting segregation or measuring its effects, then severe problems may emerge in such analyses (Knox et al., 2022; McCartan et al., 2023a; Egami et al., 2023). We also quantify uncertainty in bias-corrected estimates of segregation that results from aggregation. Even when using our best estimate as to true segregation in a city, the center of the sampling distribution, the spread of the distribution is often wide such that segregation could reasonably be significantly higher or lower. Thus, one contribution of this analysis is quantifying uncertainty that was heretofore not well understood, and is often ignored in applied segregation research. Even classical measurement error – uncorrelated noise in variable measures, the least harmful of measurement errors – is shown to downward bias effect estimates (Angrist and Pischke, 2008). If uncertainty is correlated with variables of interest, then estimation of what drives segregation, or what it influences, will be even further biased.

Fortunately, our simulated segregation data can be used in future research to account for measurement error and better understand segregation and its impacts. We provide these data free of charge and with open-source software to implement simulations in other contexts and with different parameters. In future work, we plan to apply these data to previous studies of the effects of segregation, in order to provide researchers with a template for how to use these data in applied research.

References

Abowd, J. M. (2024). Noisy Measurements Are Important; The Design of Census Products Is Much More Important. *Harvard Data Science Review*, 6(2). https://hdsr.mitpress.mit.edu/pub/vct0o3sw.

Alesina, A., Baqir, R., and Easterly, W. (1999). Public goods and ethnic divisions. *The Quarterly Journal of Economics*, 114(4):1243–1284.

Ananat, E. O. (2011). The wrong side(s) of the tracks: The causal effects of racial segregation on urban poverty and inequality. *American Economic Journal: Applied Economics*, 3(2):34–66.

Angrist, J. and Pischke, J. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Autry, E., Carter, D., Herschlag, G., Hunter, Z., and Mattingly, J. C. (2019). Metropolized forest recombination for monte carlo sampling of graph partitions. *arXiv preprint arXiv:1911.01503*.

Bischoff, K. and Reardon, S. F. (2014). Residential segregation by income, 1970–2009. *Diversity and disparities: America enters a new century*, 43.

Brown, J. R. (2024). Partisan conversion through neighborhood influence: How voters adopt the partisanship of their neighbors. Technical report, Working Paper, Harvard University.

Brown, J. R., Cantoni, E., Enos, R. D., Pons, V., and Sartre, E. (2024). A micro-level analysis of the increase and contributing factors of geographic partisan segregation. Working Paper, Harvard University.

Brown, J. R. and Enos, R. D. (2021). The measurement of partisan sorting for 180 million voters. *Nature Human Behaviour*.

Brown, J. R. and Zoorob, M. (2022). Resisting Broken Windows. *Political Behavior*, 44(2):679–703.

Carter, D., Herschlag, G., Hunter, Z., and Mattingly, J. (2019). A merge-split proposal for reversible monte carlo markov chain sampling of redistricting plans. Technical report, Working Paper, Duke University.

Chen, J. and Rodden, J. (2013). Unintentional gerrymandering: Political geography and electoral bias in legislatures. *Quarterly Journal of Political Science*.

Chetty, R., Hendren, N., Jones, M. R., and Porter, S. R. (2019). Race and Economic Opportunity in the United States: an Intergenerational Perspective*. *The Quarterly Journal of Economics*, 135(2):711–783.

Cortese, C. F., Falk, R. F., and Cohen, J. K. (1976). Further considerations on the methodological analysis of segregation indices. *American Sociological Review*, 41(4):630–637.

Deford, D., Duchin, M., and Solomon, J. (2019). Recombination: A family of markov chains for redistricting. Technical report, Working Paper, Tufts University.

DeFord, D., Duchin, M., and Solomon, J. (2021). Recombination: A family of Markov chains for redistricting. *Harvard Data Science Review*. https://hdsr.mitpress.mit.edu/pub/1ds8ptxu.

Dinesen, P. T. and Sønderskov, K. M. (2015). Ethnic diversity and social trust: Evidence from the micro-context. *American Sociological Review*, 80(3):550–573.

Dmowska, A. and Stepinski, T. F. (2024). Quantification and visualization of us racial geography using the national racial geography dataset 2020. *PLOS ONE*, 19(7):1–19.

Duncan, O. D. and Duncan, B. (1955). A methodological analysis of segregation indexes. *American Sociological Review*, 20(2):210–217.

Egami, N., Jacobs-Harukawa, M., Stewart, B. M., and Wei, H. (2023). Using large language model annotations for valid downstream statistical inference in social science: Design-based semi-supervised learning.

Einstein, K., Glick, D., and Palmer, M. (2020). *Neighborhood Defenders: Participatory Politics and America's Housing Crisis*. Cambridge University Press.

Elbers, B. (2021). Trends in u.s. residential racial segregation, 1990 to 2020. *Socius*, 7:23780231211053982.

Enos, R. D. (2016). What the demolition of public housing teaches us about the impact of racial threat on political behavior. *American Journal of Political Science*, 60(1):123–142.

Enos, R. D. (2017). *The Space Between Us: Social Geography and Politics*. Cambridge University Press, New York.

Falk, R. F., Cortese, C. F., and Cohen, J. (1978). Utilizing standardized indices of residential segregation: Comment on winship. *Social Forces*, 57(2):713–716.

Fotheringham, A. S. and Wong, D. W. S. (1991). The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A: Economy and Space*, 23(7):1025–1044.

Gatrell, A. (1983). *Distance and Space: A Geographical Perspective*. Contemporary problems in geography. Clarendon Press.

Gehlke, C. E. and Biehl, K. (1934). Certain effects of grouping upon the size of the correlation coefficient in census tract material. *Journal of the American Statistical Association*, 29:169–170.

Habyarimana, J., Humphreys, M., Posner, D. N., and Weinstein, J. M. (2009). *Coethnicity: Diversity and the Dilemmas of Collective Action*. Russell Sage, New York.

Hennerdal, P. and Nielsen, M. M. (2017). A multiscalar approach for identifying clusters and segregation patterns that avoids the modifiable areal unit problem. *Annals of the American Association of Geographers*, 107(3):555–574.

Hotz, V. J., Bollinger, C. R., Komarova, T., Manski, C. F., Moffitt, R. A., Nekipelov, D., Sojourner, A., and Spencer, B. D. (2022). Balancing data privacy and usability in the federal statistical system. *Proceedings of the National Academy of Sciences*, 119(31):e2104906119.

Hotz, V. J., Bollinger, C. R., Komarova, T., Manski, C. F., Moffitt, R. A., Nekipelov, D., Sojourner, A., and Spencer, B. D. (2024). The key role of absolute risk in the disclosure risk assessment of public data releases. *Proceedings of the National Academy of Sciences*, 121(11):e2321882121.

Hotz, V. J. and Salvo, J. (2022). A Chronicle of the Application of Differential Privacy to the 2020 Census. *Harvard Data Science Review*, (Special Issue 2). https://hdsr.mitpress.mit.edu/pub/ql9z7ehf.

Hwang, J. and McDaniel, T. W. (2022). Racialized reshuffling: Urban change and the persistence of segregation in the twenty-first century. *Annual Review of Sociology*, 48(Volume 48, 2022):397–419.

Jahn, J., Schmid, C. F., and Schrag, C. (1947). The measurement of ecological segregation. *American Sociological Review*, 12(3):293–303.

James, D. R. and Taeuber, K. E. (1985). Measures of segregation. *Sociological Methodology*, 15:1–32.

Jelinski, D. E. and Wu, J. (1996). The modifiable areal unit problem and implications for landscape ecology. *Landscape Ecology*, 11(3):129–140.

Kaplan, E., Spenkuch, J. L., and Sullivan, R. (2022). Partisan spatial sorting in the united states: A theoretical and empirical overview. *Journal of Public Economics*, 211:104668.

Kenny, C. T., Kuriwaki, S., McCartan, C., Rosenman, E. T. R., Simko, T., and Imai, K. (2021). The use of differential privacy for census data and its impact on redistricting: The case of the 2020 u.s. census. *Science Advances*, 7(41):eabk3283.

Kenny, C. T., McCartan, C., Fifield, B., and Imai, K. (2024a). redist: Simulation methods for legislative redistricting. Available at The Comprehensive R Archive Network (CRAN).

Kenny, C. T., McCartan, C., Kuriwaki, S., Simko, T., and Imai, K. (2024b). Evaluating bias and noise induced by the us census bureau's privacy protection methods. *Science Advances*, 10(18):eadl2524.

Kenny, C. T., McCartan, C., Simko, T., Kuriwaki, S., and Imai, K. (2023). Widespread partisan gerrymandering mostly cancels nationally, but reduces electoral competition. *Proceedings of the National Academy of Sciences*, 120(25):e2217322120.

Knox, D., Lucas, C., and Cho, W. K. T. (2022). Testing causal theories with learned proxies. *Annual Review of Political Science*, 25(1):419–441.

Larsen, M. V., Hjorth, F., Dinesen, P. T., and Sønderskov, K. M. (2019). When do citizens respond politically to the local economy? evidence from registry data on local housing markets. *American Political Science Review*, 113(2):499–516.

Lee, B. A., Reardon, S. F., Firebaugh, G., Farrell, C. R., Matthews, S. A., and O'Sullivan, D. (2008). Beyond the census tract: Patterns and determinants of racial segregation at multiple geographic scales. *American Sociological Review*, 73(5):766–791. PMID: 25324575.

Legewie, J. and Schaeffer, M. (2016). Contested boundaries: Explaining where ethnoracial diversity provokes neighborhood conflict. *American Journal of Sociology*, 122(1):125–161.

Logan, T. D. and Parman, J. M. (2017). The national rise in residential segregation. *The Journal of Economic History*, 77(1):127–170.

Massey, D. S. and Denton, N. A. (1988). The dimensions of residential segregation. *Social Forces*, 67(2):281–315.

Massey, D. S. and Denton, N. A. (1993). *American Apartheid: Segregation and the Making of the Underclass*. Havard University Press, Cambridge, MA.

Mazza, A. and Punzo, A. (2015). On the upward bias of the dissimilarity index and its corrections. *Sociological Methods & Research*, 44(1):80–107.

McCartan, C., Brown, J. R., and Imai, K. (2024). Measuring and modeling neighborhoods. *American Political Science Review*, page 1–20.

McCartan, C., Goldin, J., Ho, D. E., and Imai, K. (2023a). Estimating racial disparities when race is not observed. *arXiv preprint*, page 2303.02580.

McCartan, C. and Imai, K. (2023). Sequential Monte Carlo for sampling balanced and compact redistricting plans. *The Annals of Applied Statistics*, 17(4):3300 – 3323.

McCartan, C., Simko, T., and Imai, K. (2023b). Making differential privacy work for census data users. *Harvard Data Science Review*, 5(4).

Mijs, J. J. B. and Roe, E. L. (2021). Is america coming apart? socioeconomic segregation in neighborhoods, schools, workplaces, and social networks, 1970–2020. *Sociology Compass*, 15(6):e12884.

Muralidhar, K., Ruggles, S., Domingo-Ferrer, J., and Sánchez, D. (2024). The counterfactual framework in jarmin et al. is not a measure of disclosure risk of respondents. *Proceedings of the National Academy of Sciences*, 121(11):e2319484121.

Nelson, J. K. and Brewer, C. A. (2017). Evaluating data stability in aggregation structures across spatial scales: revisiting the modifiable areal unit problem. *Cartography and Geographic Information Science*, 44(1):35–50.

Openshaw, S. (1983). *Concepts and Techniques in Modern Geography*, chapter The Modifiable Areal Unit Problem. Geo Books, Norwich, UK.

Putnam, R. D. (2007). E pluribus unum: Diversity and community in the twenty-first century: The 2006 johan skytte prize lecture. *Scandinavian Political Studies*, 30(2):137–174.

Reardon, S. F. and Bischoff, K. (2011). Income inequality and income segregation. *American Journal of Sociology*, 116(4):1092–1153.

Reardon, S. F. and Firebaugh, G. (2002). Measures of multigroup segregation. *Sociological Methodology*, 32:33–67.

Reardon, S. F. and O'Sullivan, D. (2004). Measures of spatial segregation. *Sociological Methodology*, 34(1):121–162.

Roberto, E. (2018). The spatial proximity and connectivity method for measuring and analyzing residential segregation. *Sociological Methodology*, 48(1):182–224.

Roberto, E. (2024). The divergence index: A decomposable measure of segregation and inequality.

Rodden, J. (2019). *Why Cities Lose: The Deep Roots of the Urban-Rural Political Divide*. Basic Books.

Rothstein, R. (2017). *The Color of Law: A Forgotten History of How Our Government Segregated America*. Liveright Publishing Corporation, a division of W.W. Norton & Company, New York ; London, first edition. edition. HOLLIS number: 990149136710203941.

Ruggles, S., Fitch, C., Magnuson, D., and Schroeder, J. (2019). Differential privacy and census data: Implications for social and economic research. *AEA Papers and Proceedings*, 109:403–408.

Ruggles, S. and Riper, D. V. (2022). The role of chance in the census bureau database reconstruction experiment. *Population Research and Policy Review*, 41:781–788.

Sang-Il Lee, Monghyeon Lee, Y. C. and Griffith, D. A. (2019). Uncertainty in the effects of the modifiable areal unit problem under different levels of spatial autocorrelation: a simulation study. *International Journal of Geographical Information Science*, 33(6):1135–1154.

Santos-Lozada, A. R., Howard, J. T., and Verdery, A. M. (2020). How differential privacy will affect our understanding of health disparities in the United States. *Proceedings of the National Academy of Sciences*, 117(24):13405–13412.

Taeuber, K. E. and Taeuber, A. F. (1976). A practitioner's perspective on the index of dissimilarity. *American Sociological Review*, 41(5):884–889.

Theil, H. (1967). *Economics and Information Theory*. Studies in mathematical and managerial economics. North-Holland Publishing Company.

Trounstine, J. (2016). Segregation and inequality in public goods. *American Journal of Political Science*, 60(3):709–725.

Trounstine, J. (2018). *Segregation by Design: Local Politics and Inequality in American Cities*. Cambridge University Press.

U.S. Census Bureau (2023). Geographic areas reference manual. Technical report.

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., and Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: An improved r^ for assessing convergence of mcmc (with discussion). *Bayesian Analysis*, 16(2).

Velez, Y. R. and Wong, G. (2017). Assessing contextual measurement strategies. *The Journal of Politics*, 79(3):1084–1089.

White, M. J. (1983). The measurement of spatial segregation. *American journal of sociology*, 88(5):1008–1018.

Winship, C. (1978). The desirability of using the index of dissimilarity or any adjustment of it for measuring segregation: Reply to falk, cortese, and cohen. *Social Forces*, 57(2):717–720.

Wong, C., Bowers, J., Rubenson, D., Fredrickson, M., and Rundlett, A. (2020). Maps in people's heads: Assessing a new measure of context. *Political Science Research and Methods*, 8(1):160–168.

Wong, C., Bowers, J., Williams, T., and Drake, K. (2012). Bringing the person back in: Boundaries, perceptions, and the measurement of racial context. *Journal of Politics*, 1(1):1–18.

Wong, D. W. (1997). Spatial dependency of segregation indices. *The Canadian Geographer*, 41(2):128–136.

Östh, J., Malmberg, B., and Andersson, E. K. (2014). *Seven: Analysing segregation using individualised neighbourhoods*, pages 135 – 162. Policy Press, Bristol, UK.

A Simulation statistics

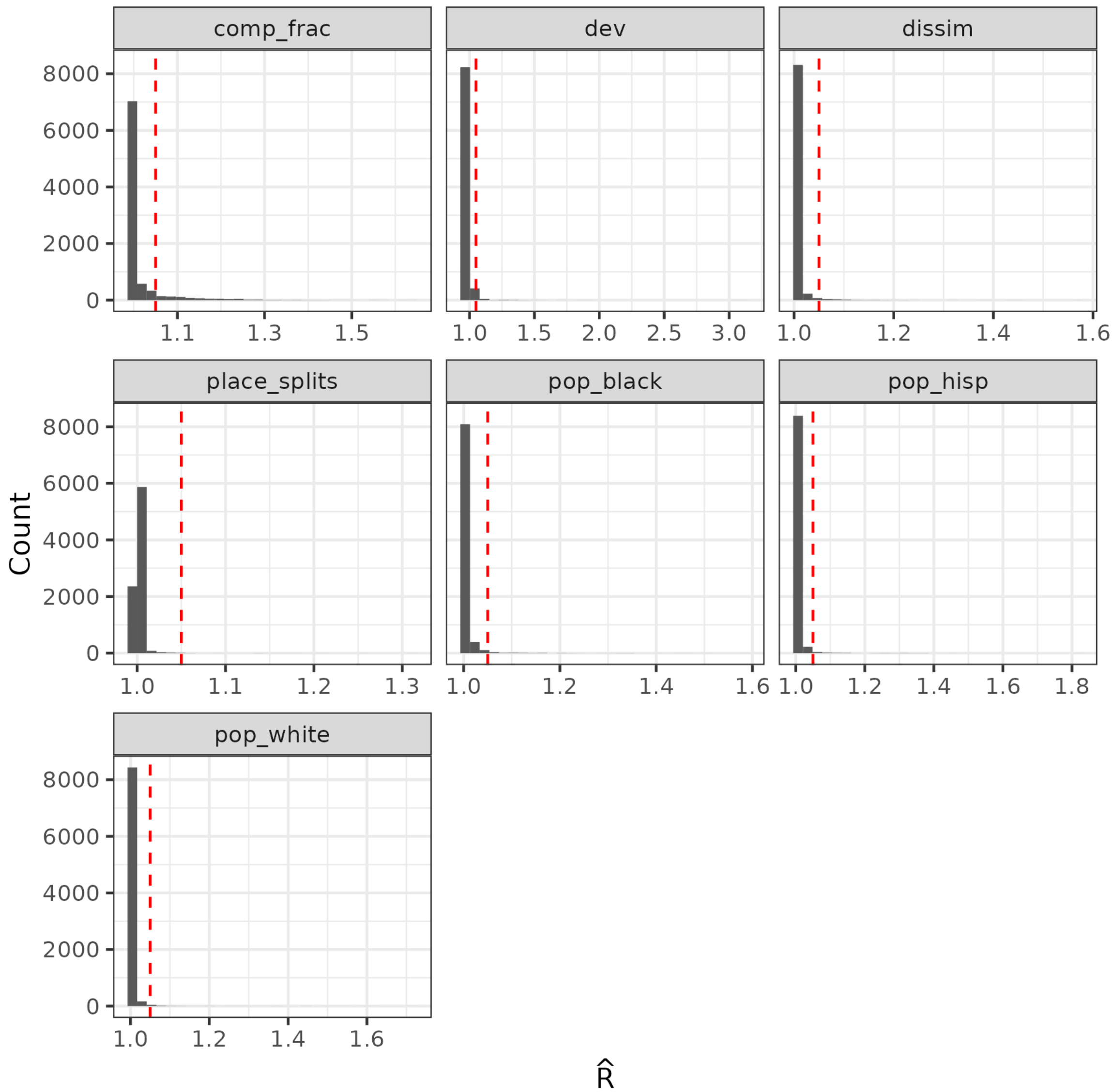


Figure A1: 2020 simulation $\hat{R}$ distributions

B Additional results

B.1 Estimating differences in official versus simulation estimates

Table 4: Comparison of compactness, population deviation and place splitting - 2010

	Observed (N=2,903)		Simulation (N=14,511,000)			
	Mean	Std. Dev.	Mean	Std. Dev.	Diff. in Means	Std. Error
Dev.	0.662	0.535	0.492	0.408	-0.170	0.010
Comp. Frac.	0.972	0.016	0.963	0.019	-0.009	0.000
Place Splits	3.907	7.465	4.208	8.128	0.301	0.139

Table 5: Comparison of compactness, population deviation and place splitting - 2000

	Observed (N=2,932)		Simulation (N=14,660,000)			
	Mean	Std. Dev.	Mean	Std. Dev.	Diff. in Means	Std. Error
Dev.	0.671	0.512	0.479	0.392	-0.192	0.009
Comp. Frac.	0.971	0.017	0.960	0.020	-0.011	0.000
Place Splits	4.416	8.303	4.550	8.597	0.135	0.153

Table 6: Modeling official vs. simulation segregation

	2000		2010		2020	
	Black-White Dissimilarity Index	H Index	Black-White Dissimilarity Index	H Index	Black-White Dissimilarity Index	H Index
	(1)	(2)	(3)	(4)	(5)	(6)
Official	0.0456***	0.0069***	0.0480***	0.0048***	0.0440***	0.0038***
	(0.0048)	(0.0013)	(0.0043)	(0.0011)	(0.0028)	(0.0006)
Observations	1,185,237	1,185,237	1,400,280	1,400,280	1,685,337	1,685,337
R^2	0.98646	0.99515	0.98829	0.99704	0.98419	0.99706
Within R^2	0.00112	0.00022	0.00156	0.00023	0.00126	0.00024
City FE	✓	✓	✓	✓	✓	✓

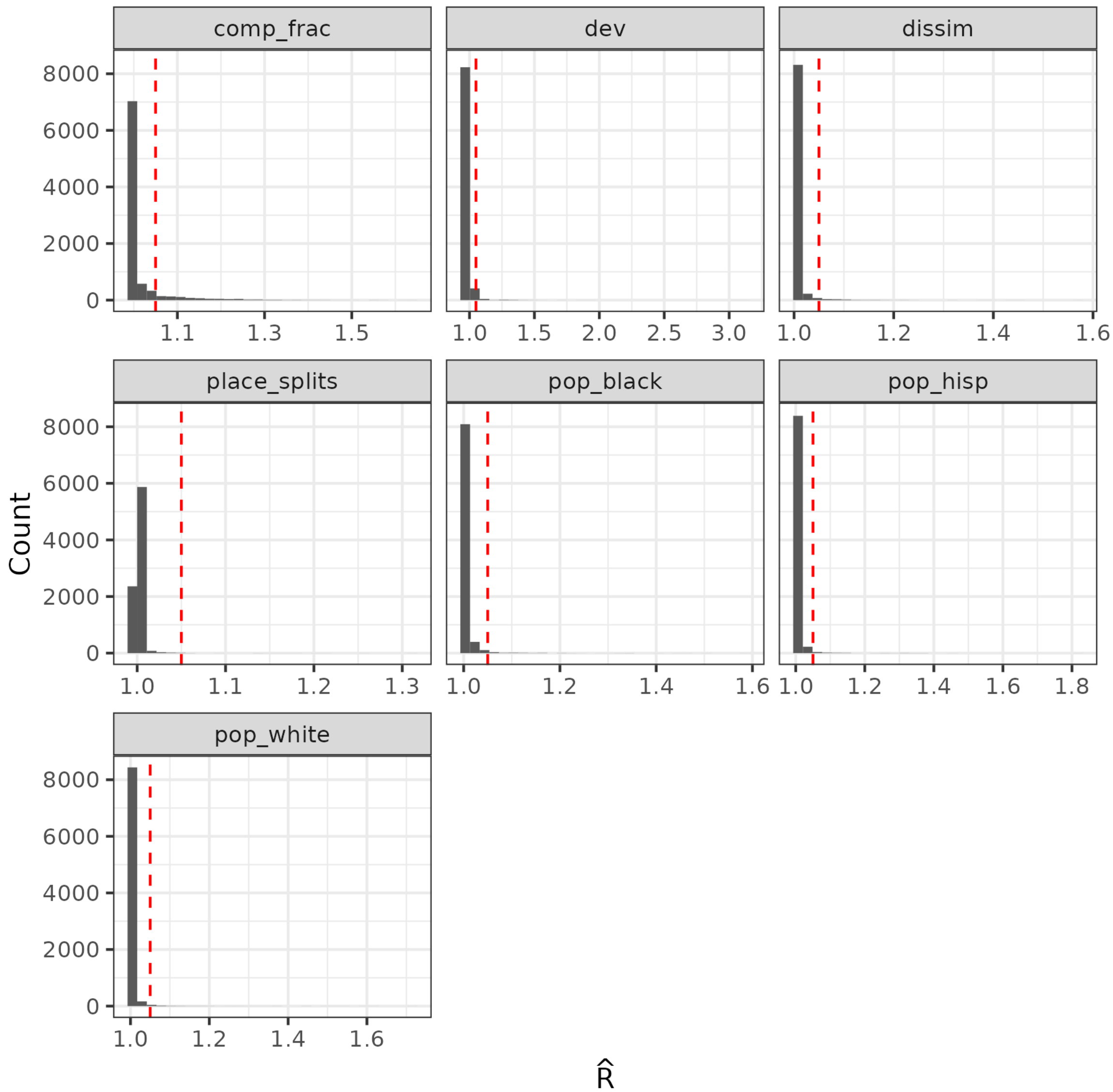


Figure A2: 2010 simulation $\hat{R}$ distributions

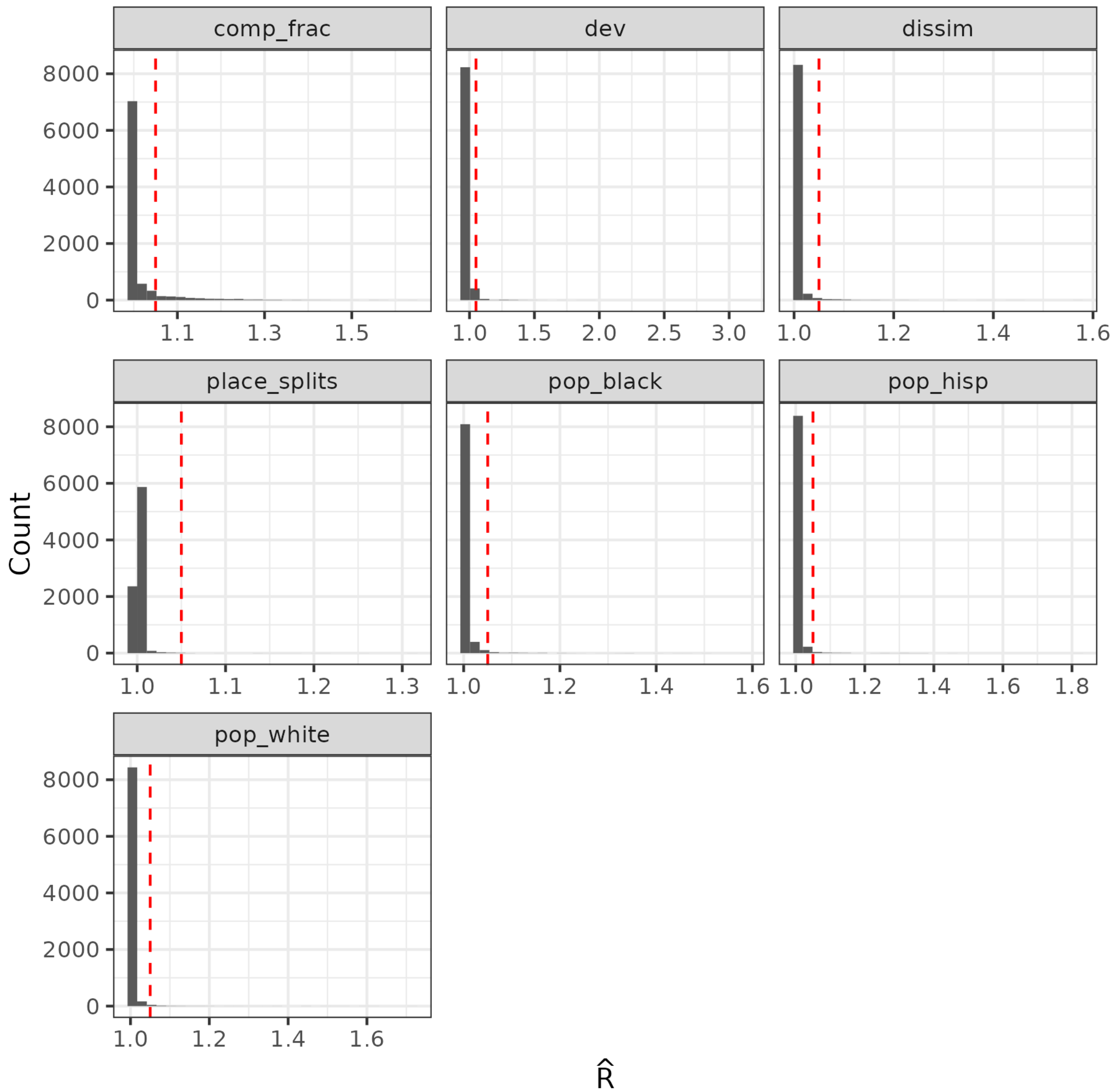


Figure A3: 2000 simulation $\hat{R}$ distributions

B.2 Relationship between bias and uncertainty

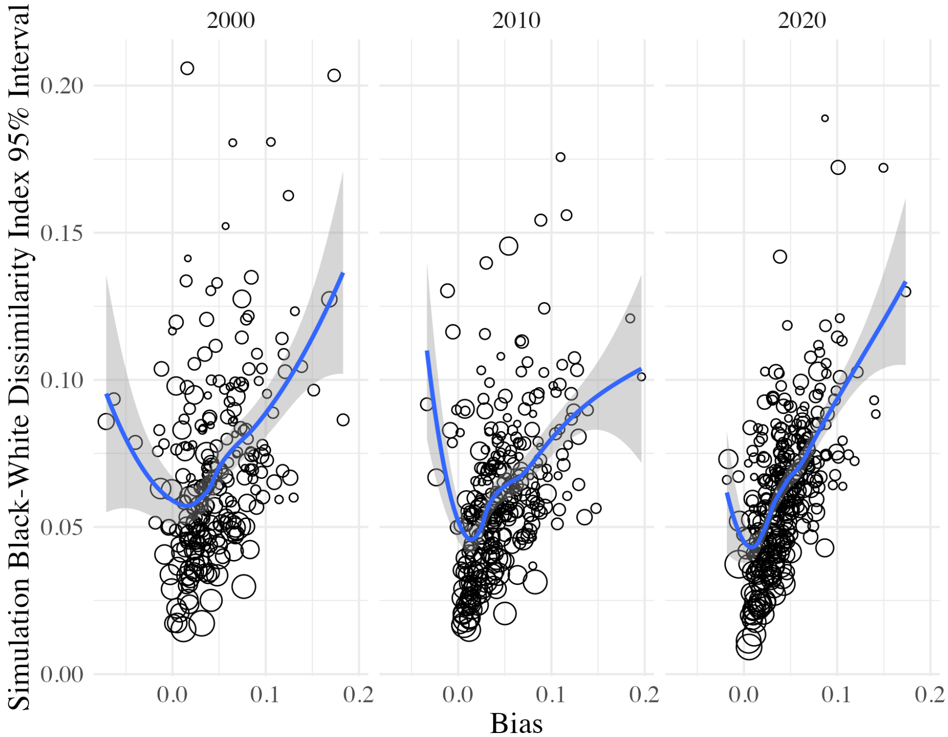


Figure B4: Black-White Dissimilarity Index bias vs. 95% interval length, 2000-2020

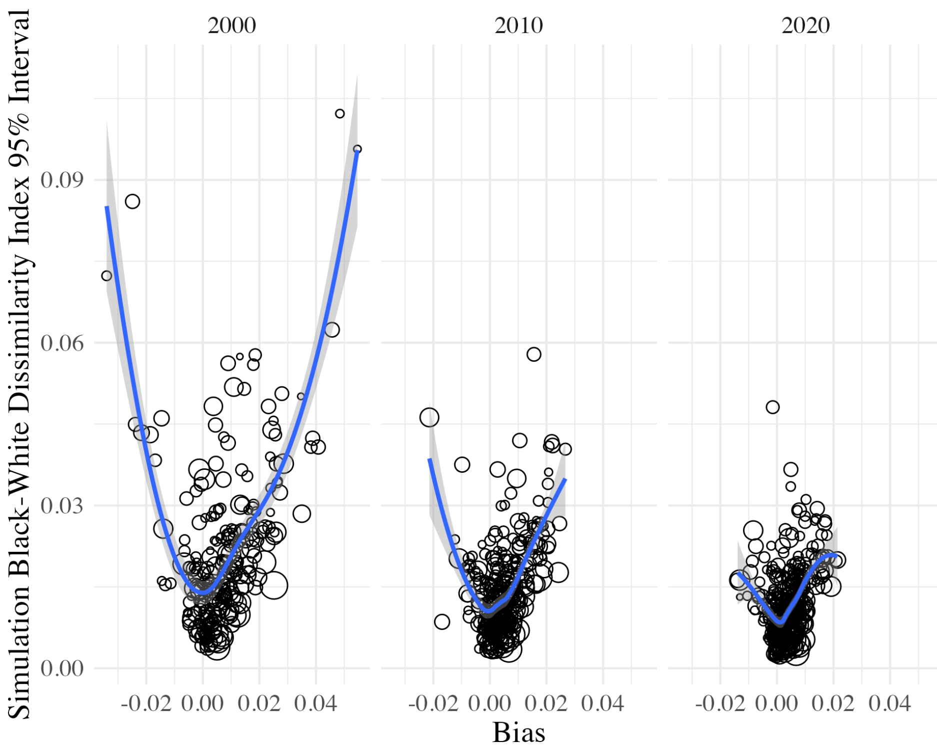


Figure B5: H Index bias vs. 95% interval length, 2000-2020

B.3 Black-White Dissimilarity Index

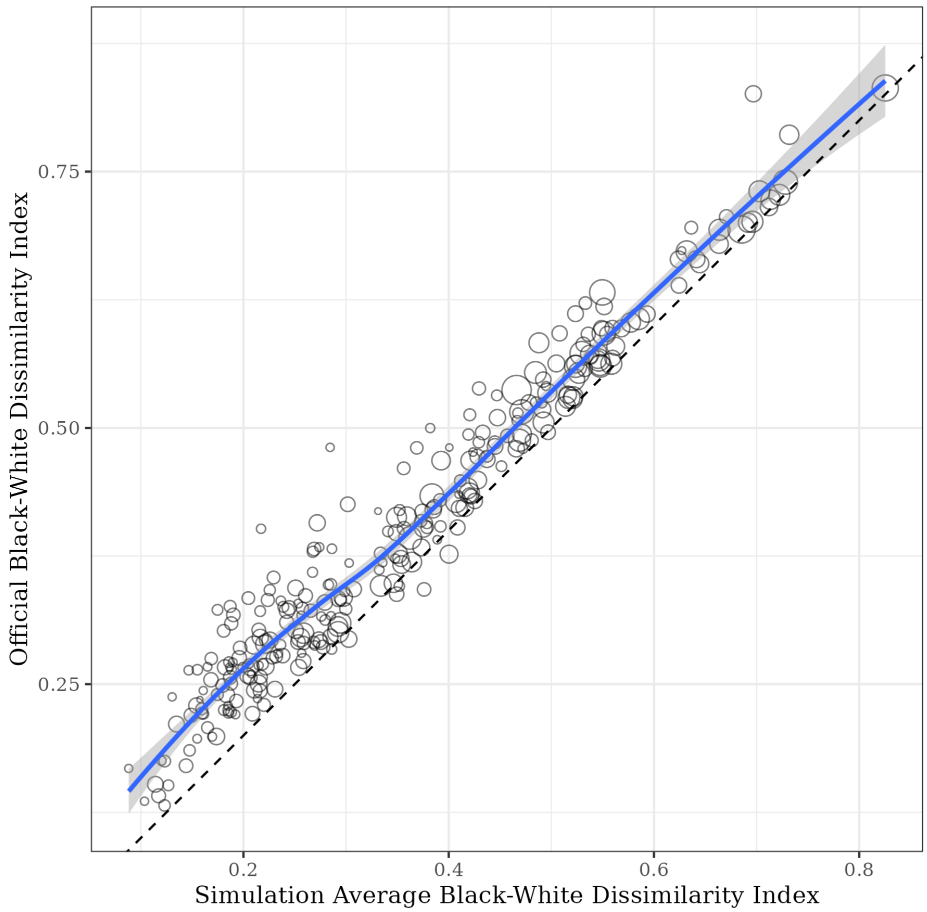


Figure B6: Official Black-White Dissimilarity versus simulated - 2010

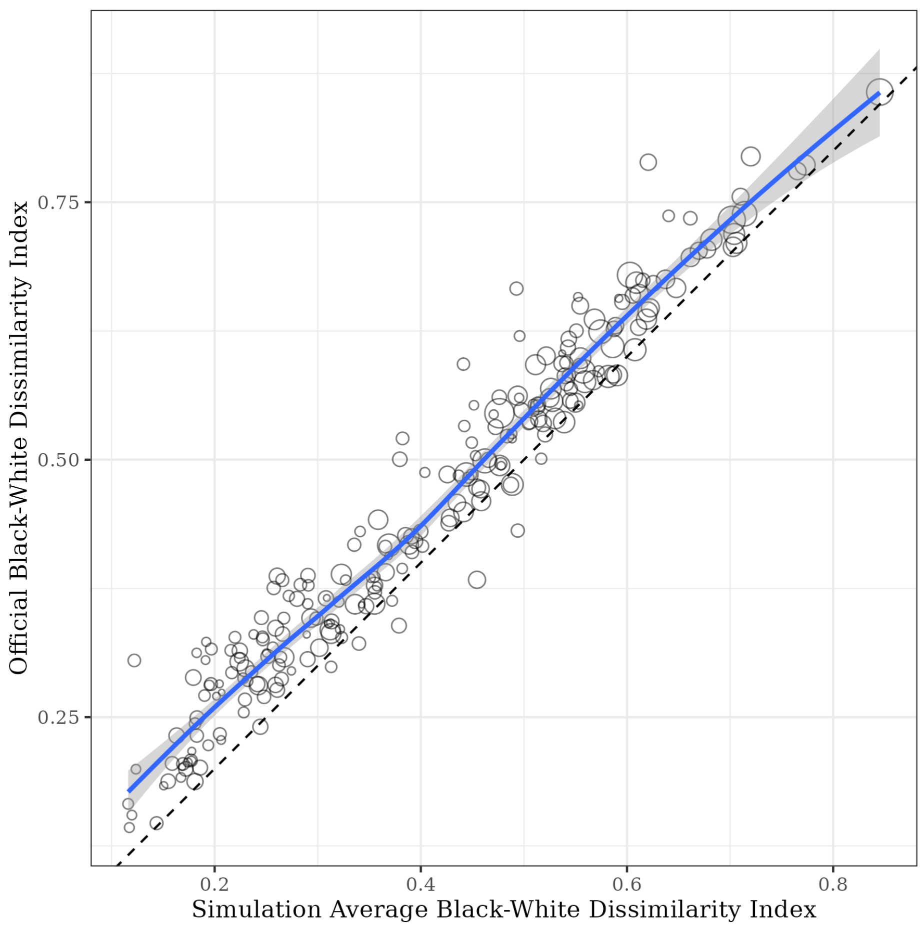
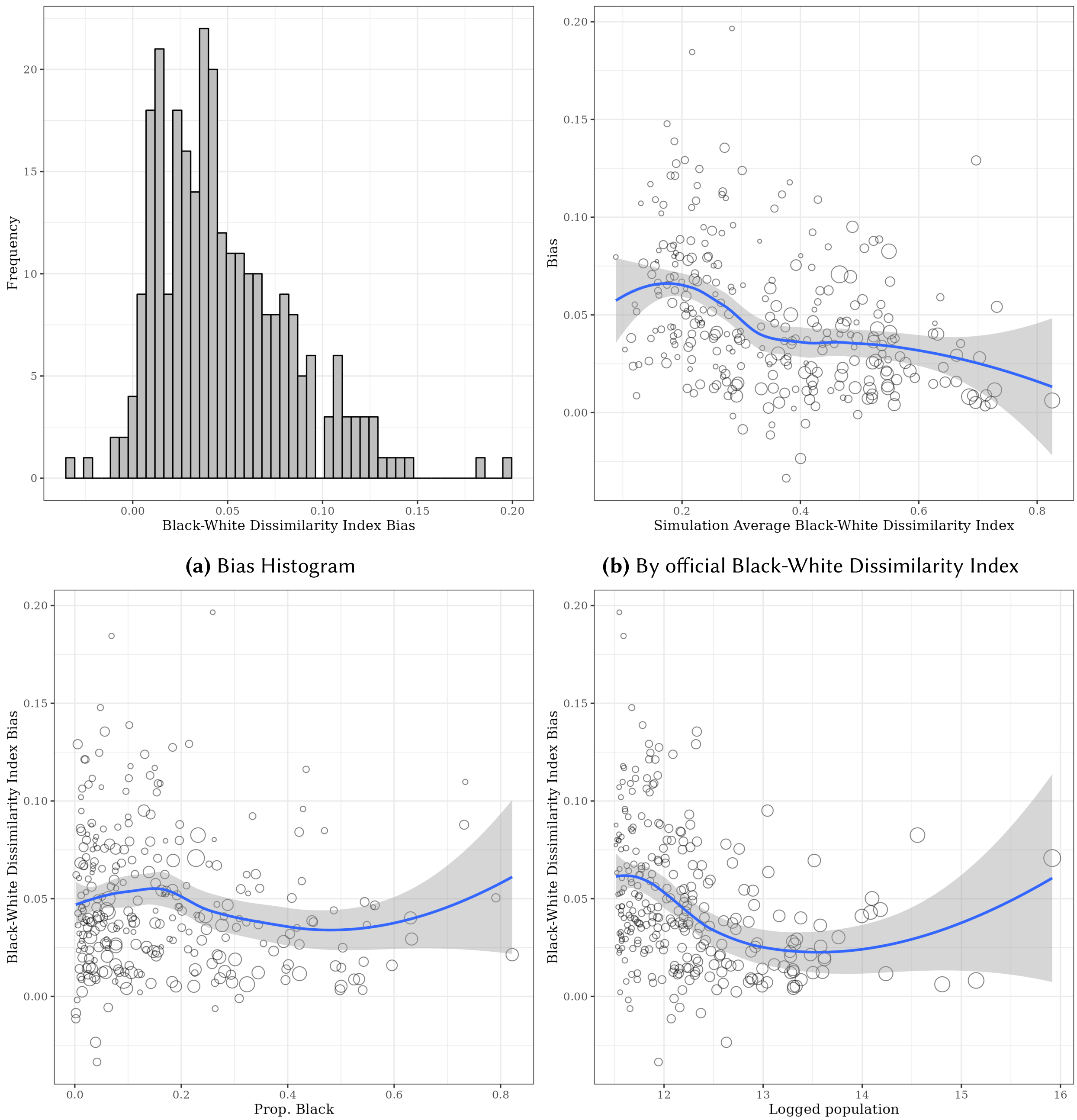


Figure B7: Official Black-White Dissimilarity versus simulated - 2000



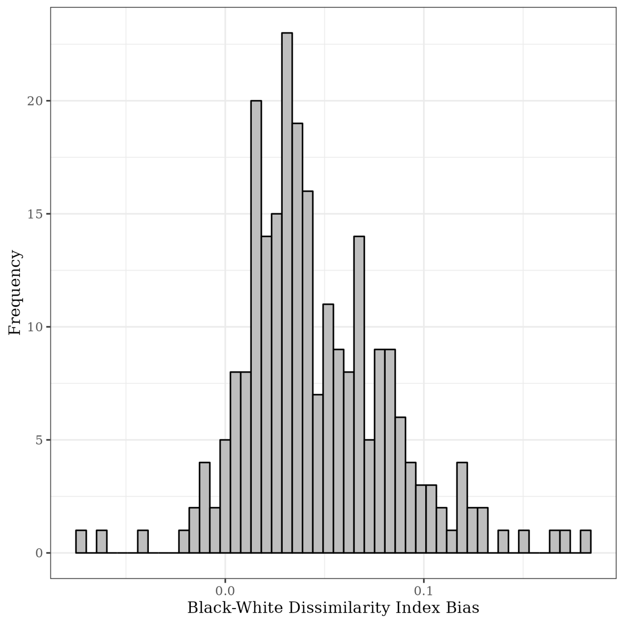
(a) Bias Histogram

(b) By official Black-White Dissimilarity Index

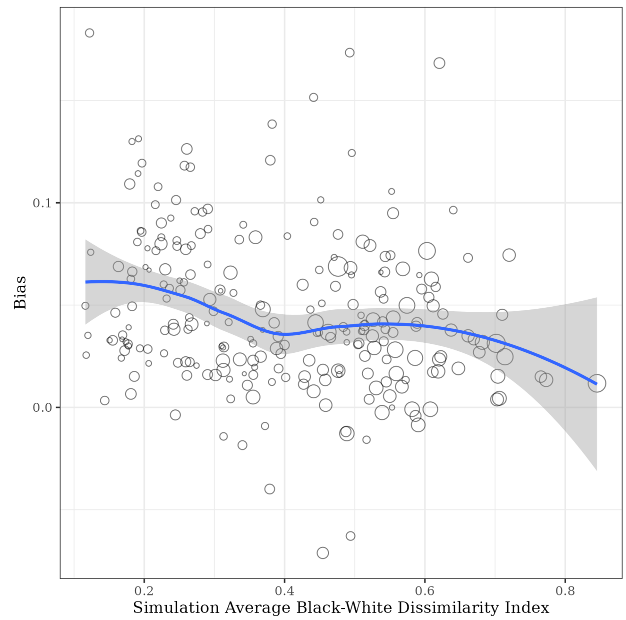
(c) By city proportion Black

(d) By logged city population

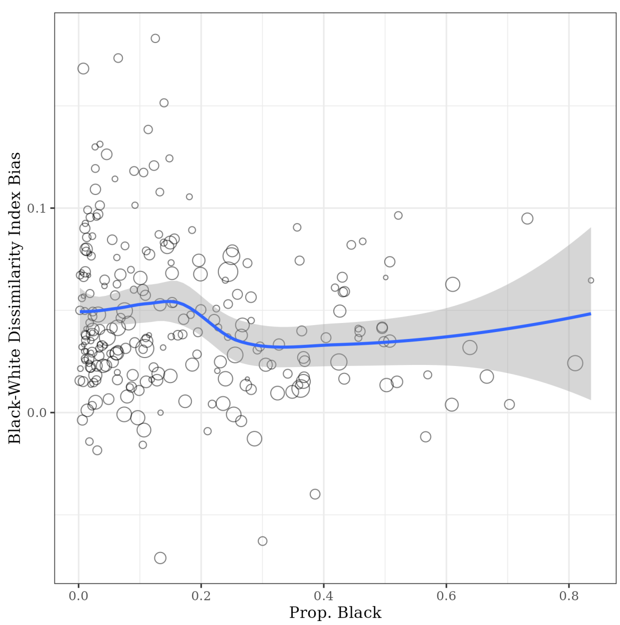
Figure B8: Black-White Dissimilarity Bias - 2010



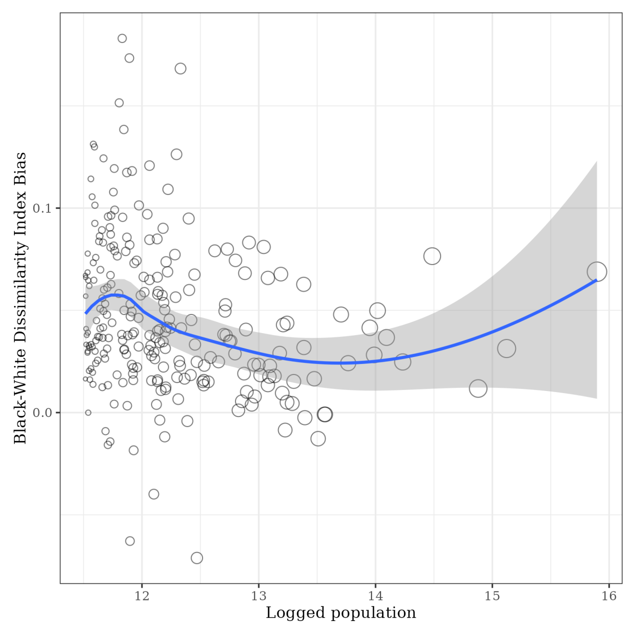
(a) Bias Histogram



(b) By official Black-White Dissimilarity Index

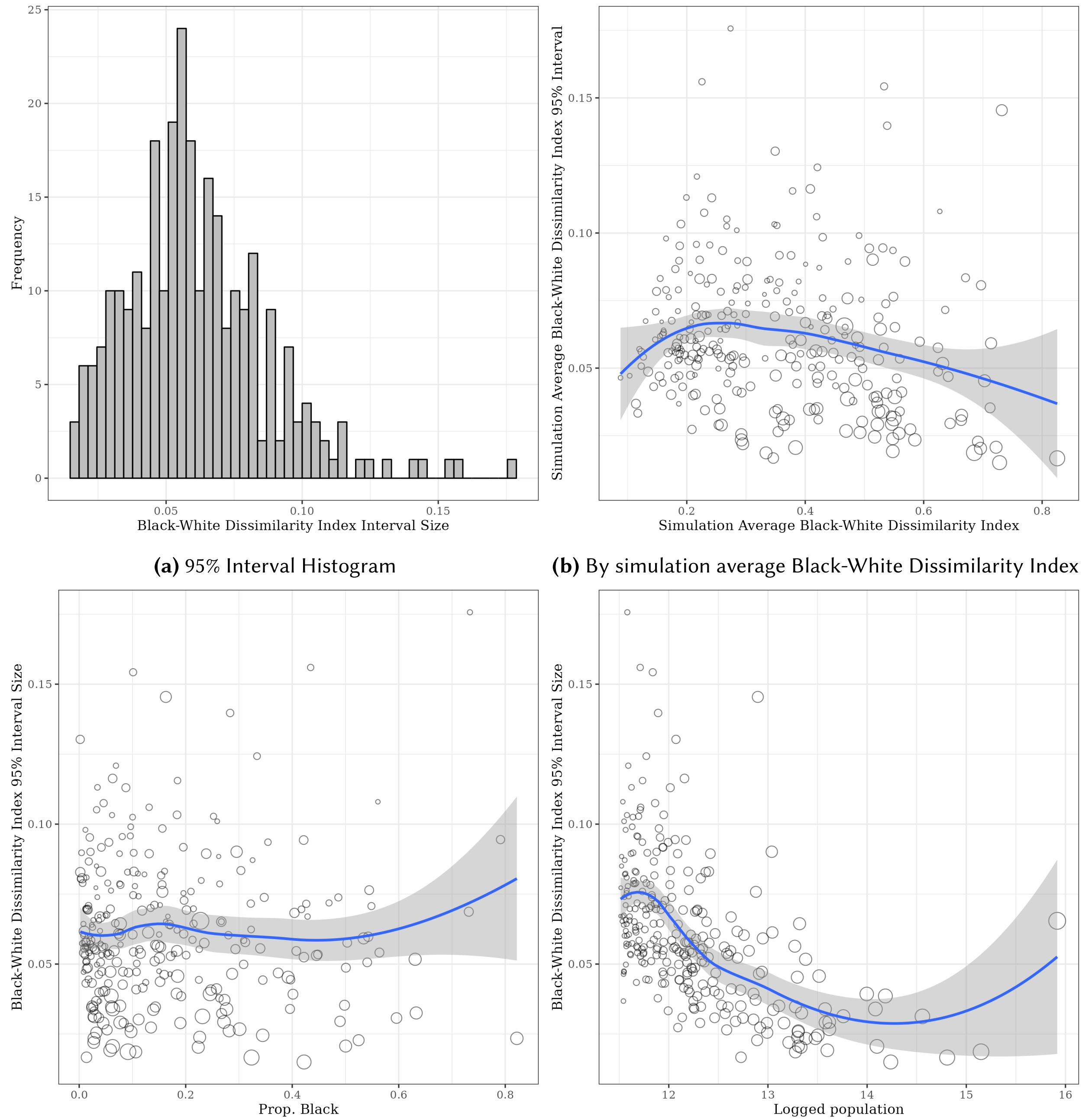


(c) By city proportion Black



(d) By logged city population

Figure B9: Black-White Dissimilarity Bias - 2000



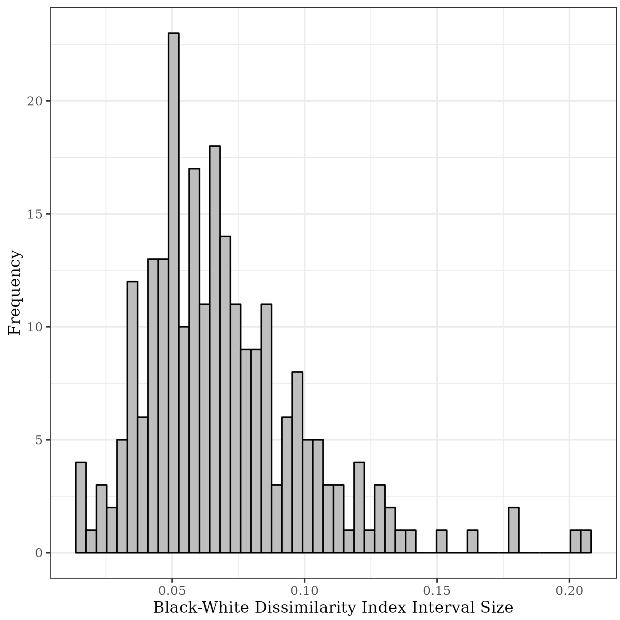
(a) 95% Interval Histogram

(b) By simulation average Black-White Dissimilarity Index

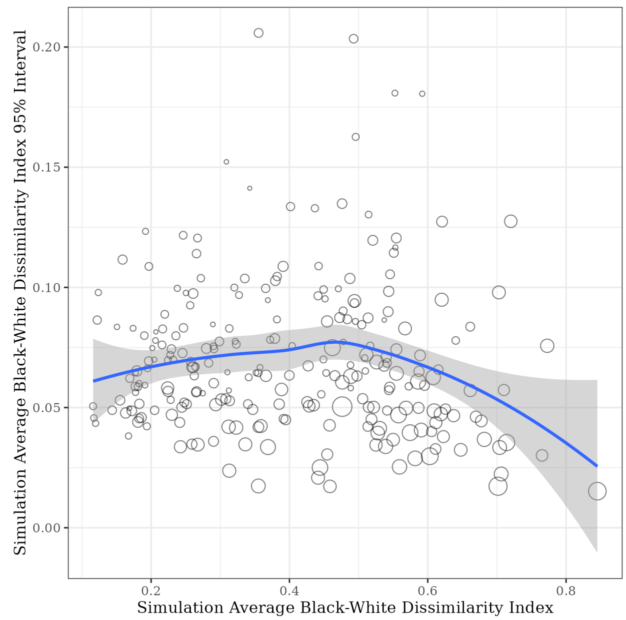
(c) By city proportion Black

(d) By logged city population

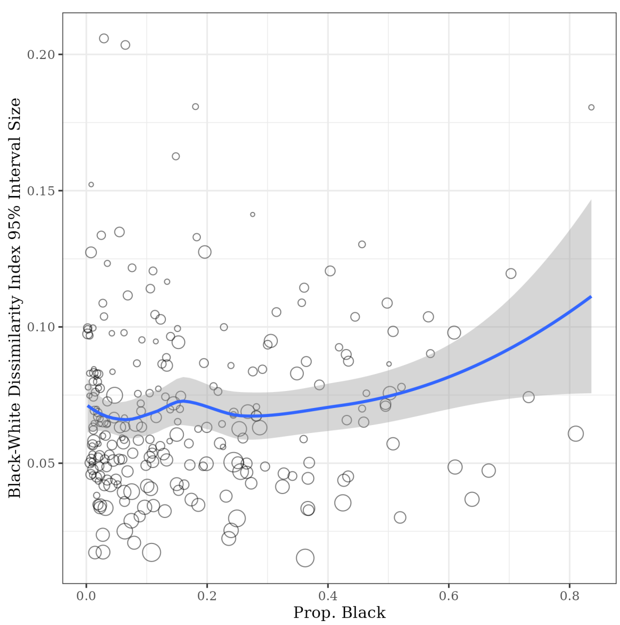
Figure B10: Black-White Dissimilarity Uncertainty - 2010



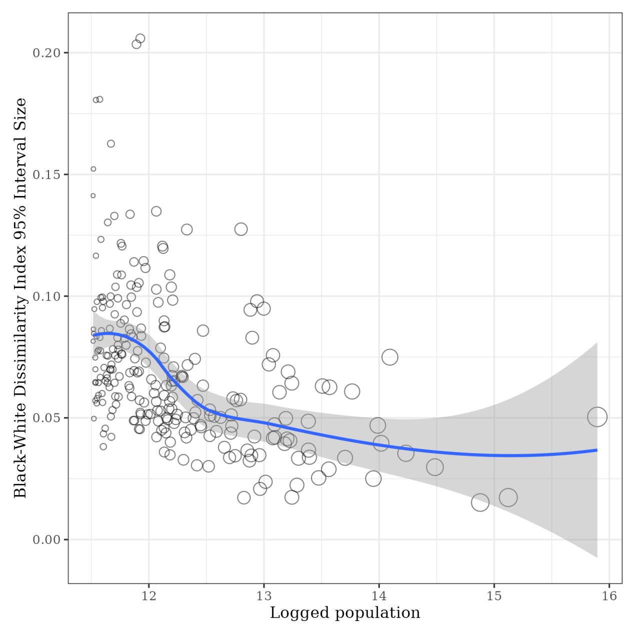
(a) 95% Interval Histogram



(b) By simulation average Black-White Dissimilarity Index



(c) By city proportion Black



(d) By logged city population

Figure B11: Black-White Dissimilarity Uncertainty - 2000

B.4 Hispanic-White Dissimilarity Index

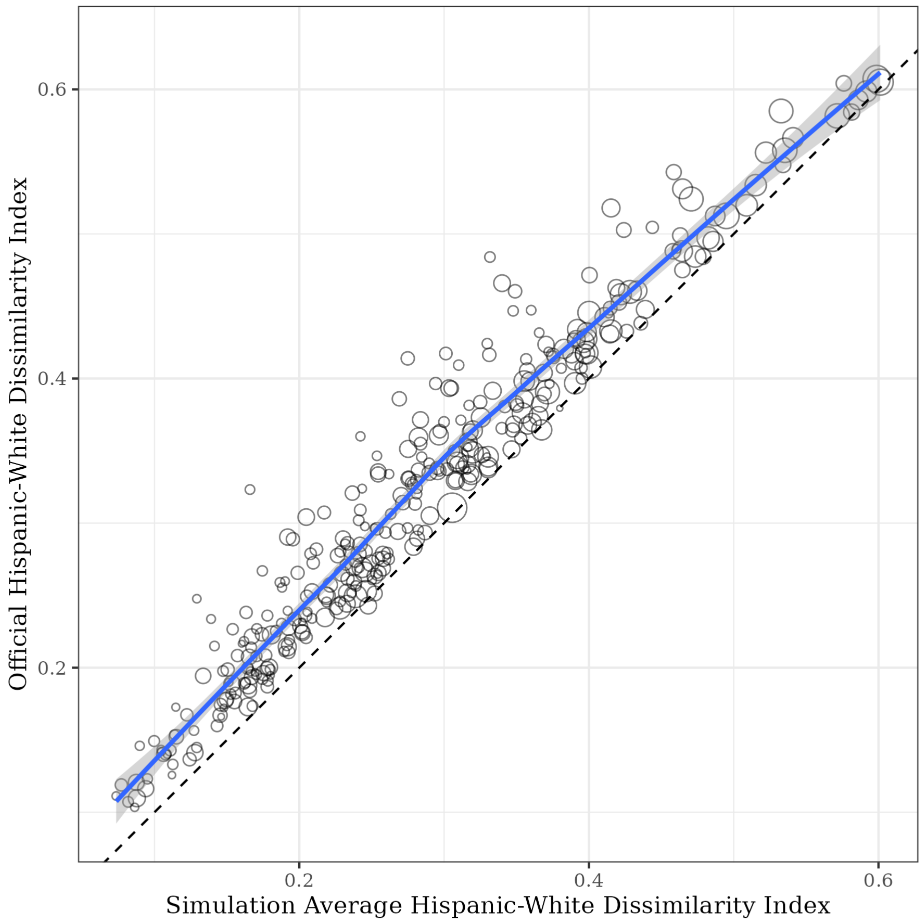


Figure B12: Official Hispanic-White Dissimilarity versus simulated - 2020

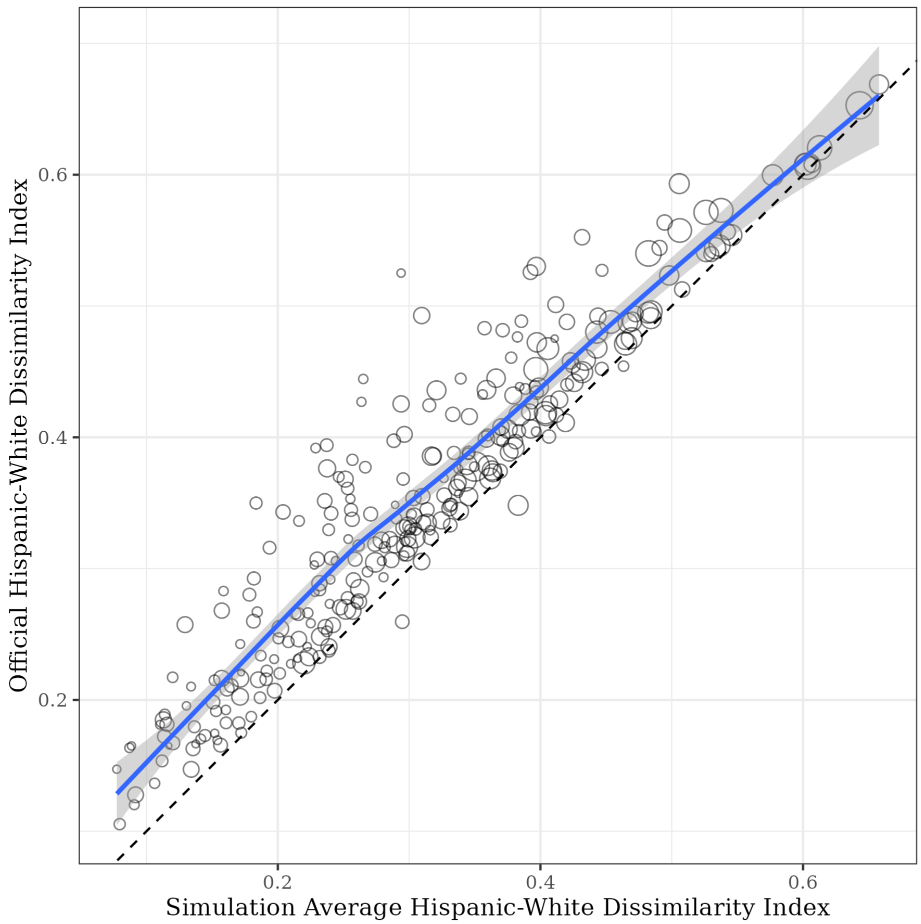


Figure B13: Official Hispanic-White Dissimilarity versus simulated - 2010

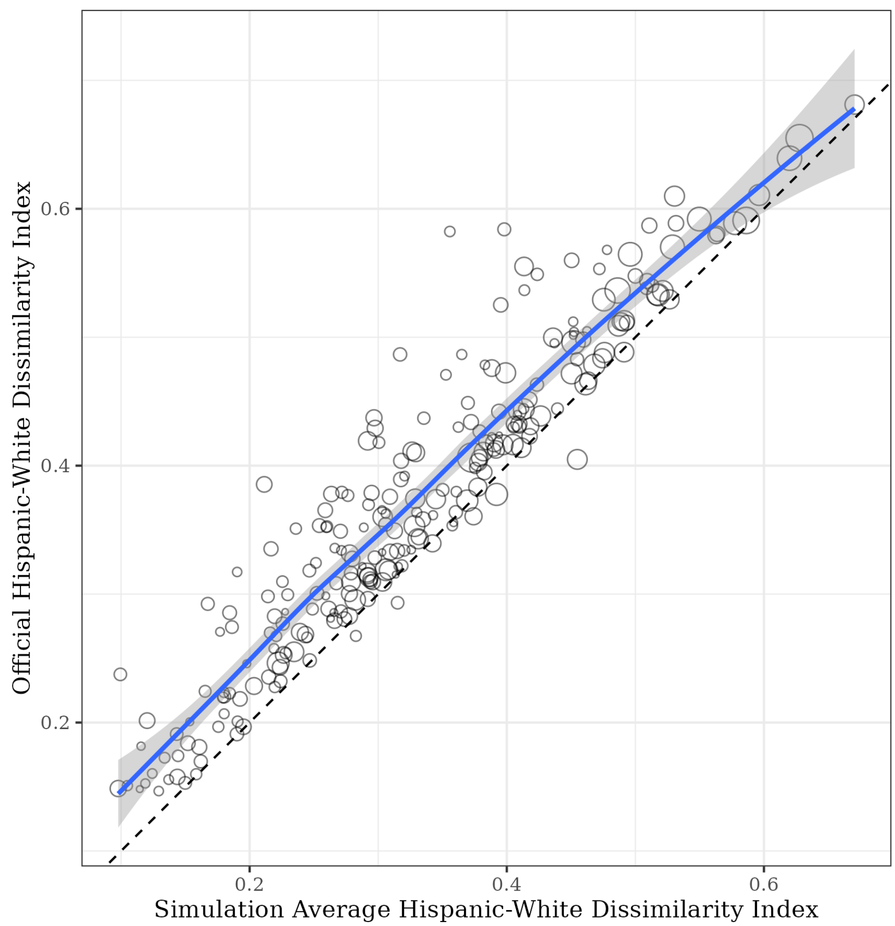
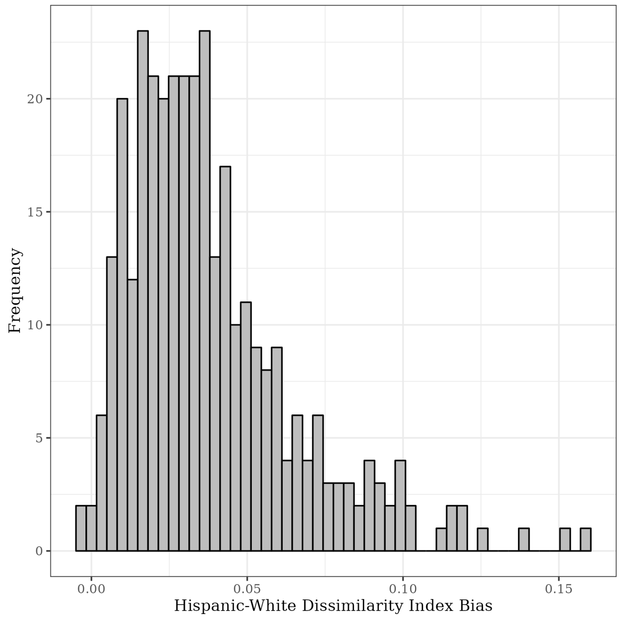
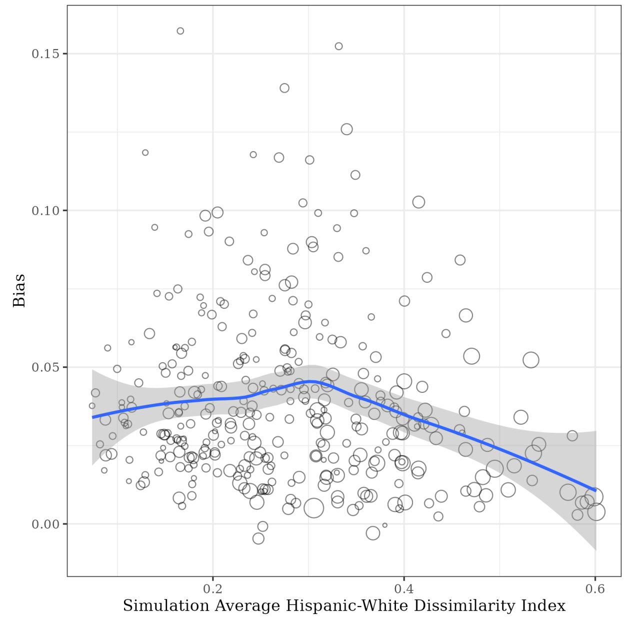


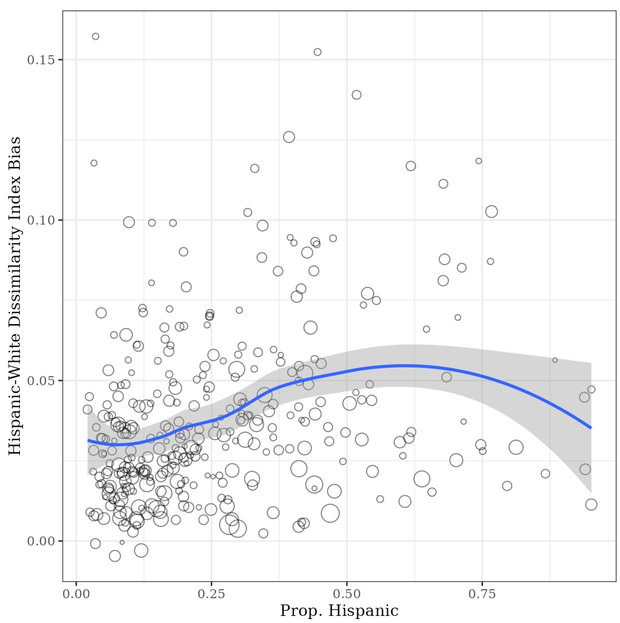
Figure B14: Official Hispanic-White Dissimilarity versus simulated - 2000



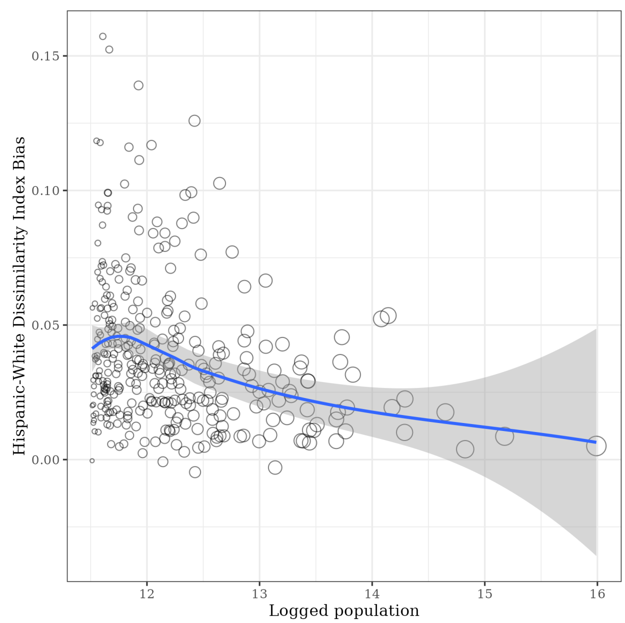
(a) Bias Histogram



(b) By official Hispanic-White Dissimilarity Index

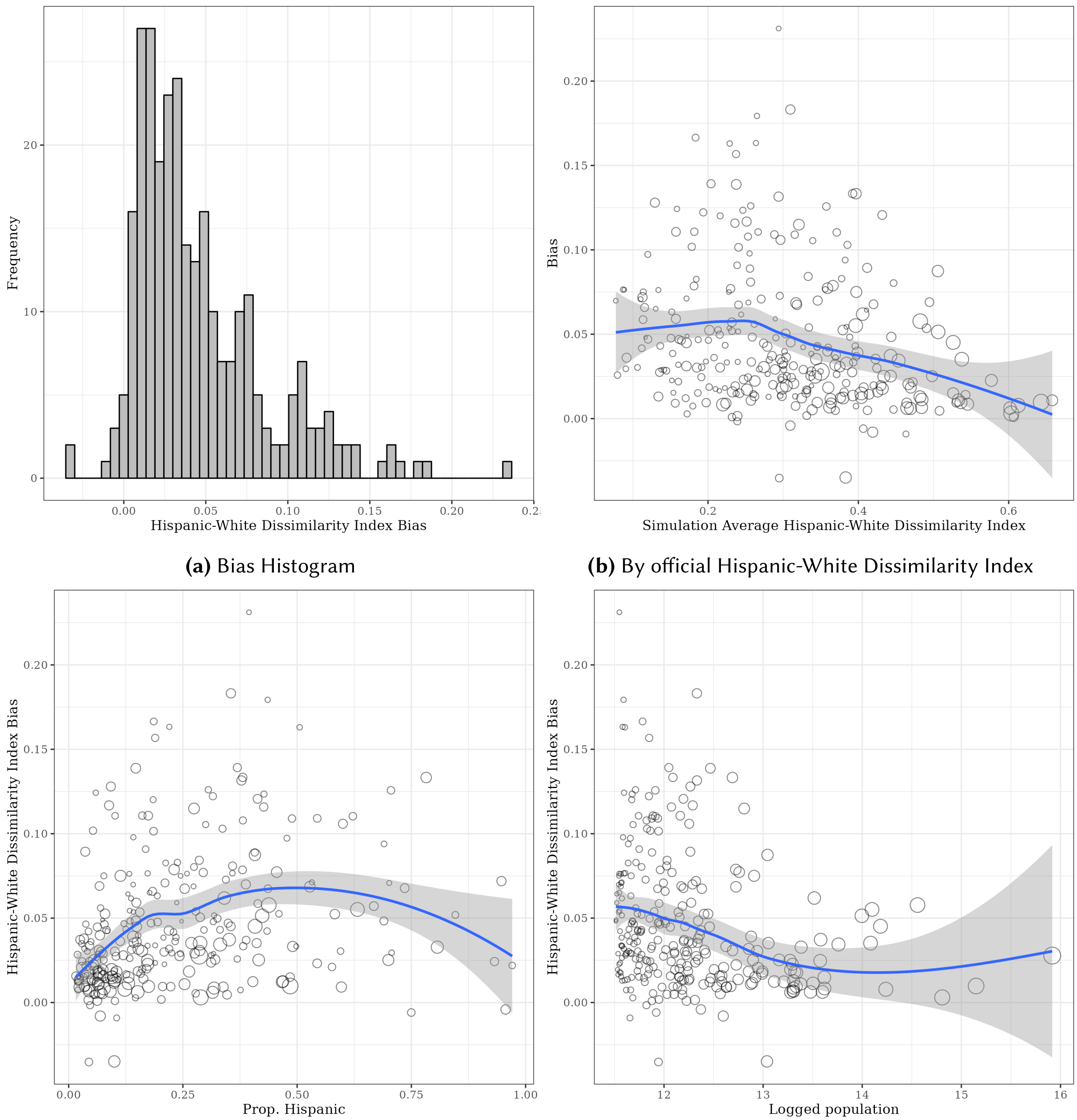


(c) By city proportion Hispanic



(d) By logged city population

Figure B15: Hispanic-White Dissimilarity Bias - 2020



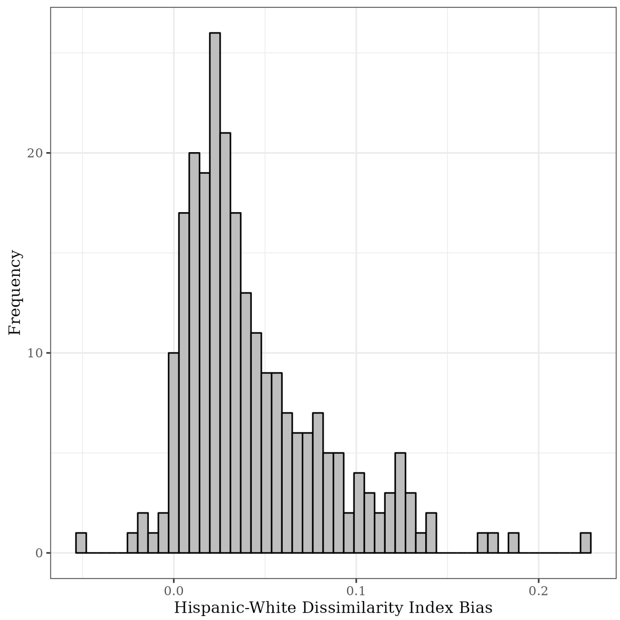
(a) Bias Histogram

(b) By official Hispanic-White Dissimilarity Index

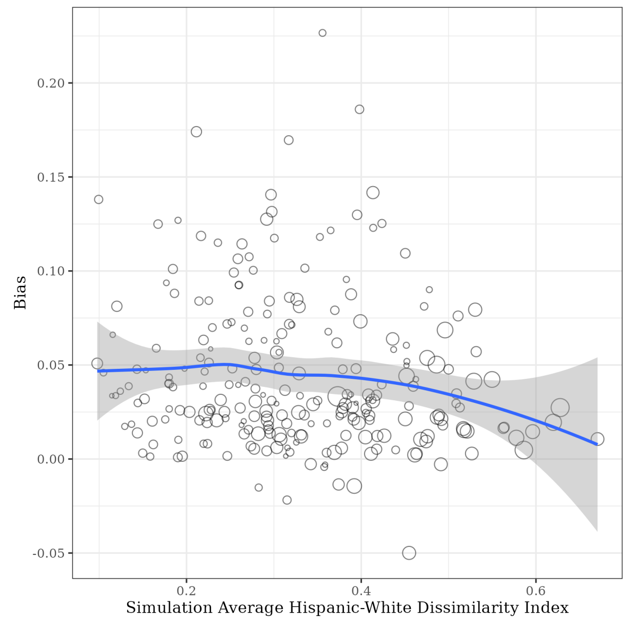
(c) By city proportion Hispanic

(d) By logged city population

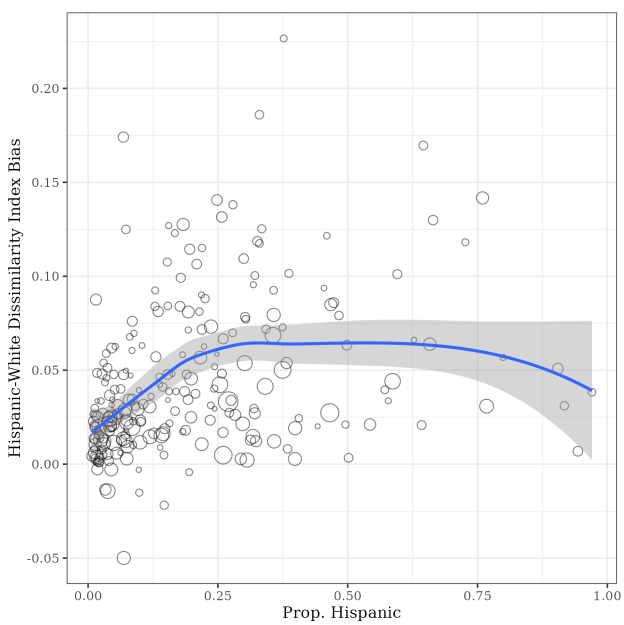
Figure B16: Hispanic-White Dissimilarity Bias - 2010



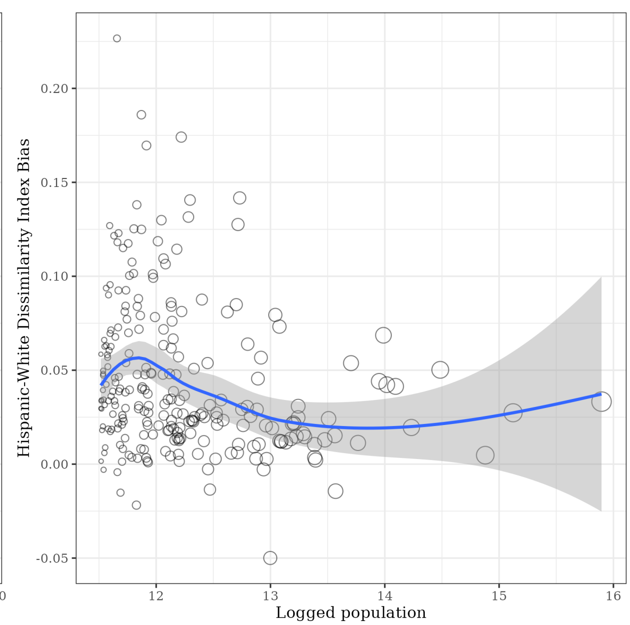
(a) Bias Histogram



(b) By official Hispanic-White Dissimilarity Index

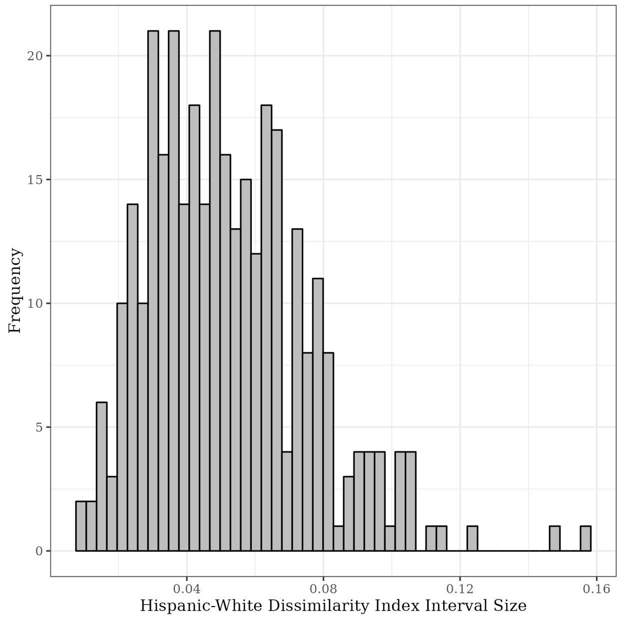


(c) By city proportion Hispanic

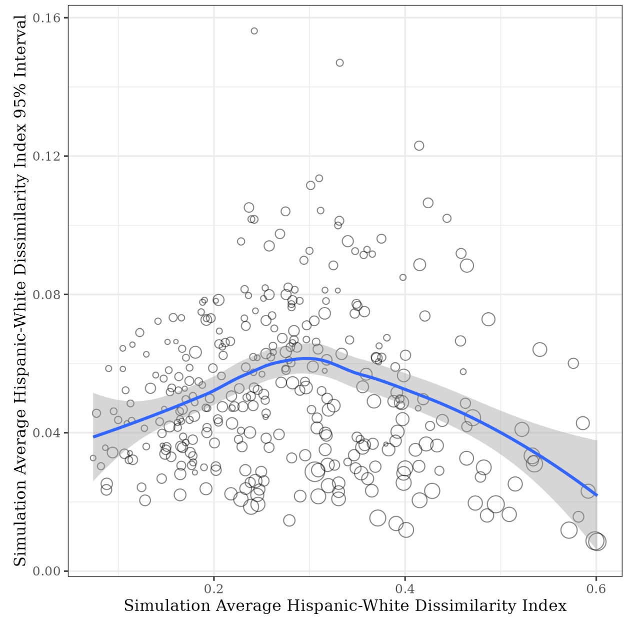


(d) By logged city population

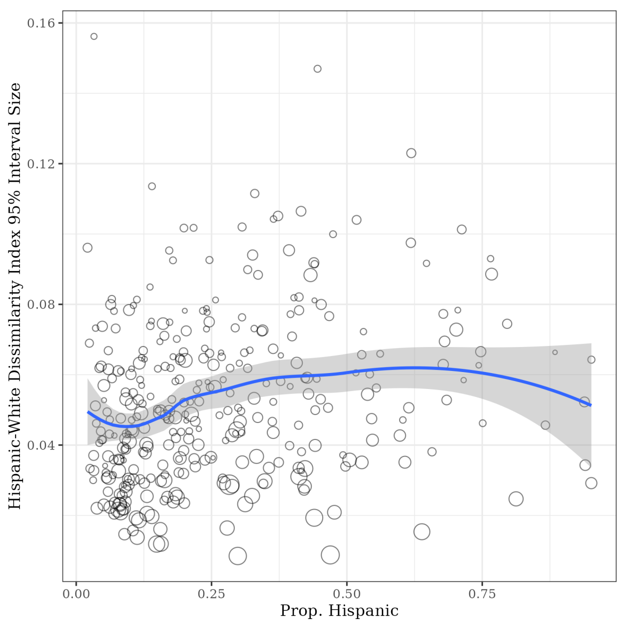
Figure B17: Hispanic-White Dissimilarity Bias - 2000



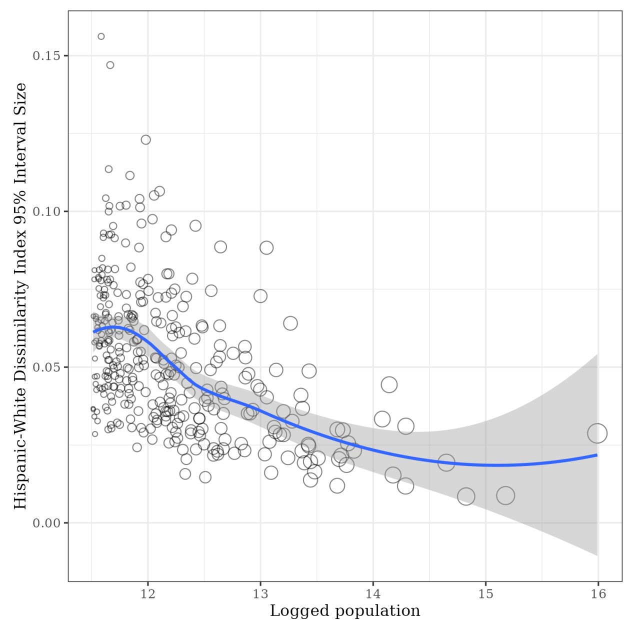
(a) 95% Interval Histogram



(b) By simulation average Hispanic-White Dissimilarity Index

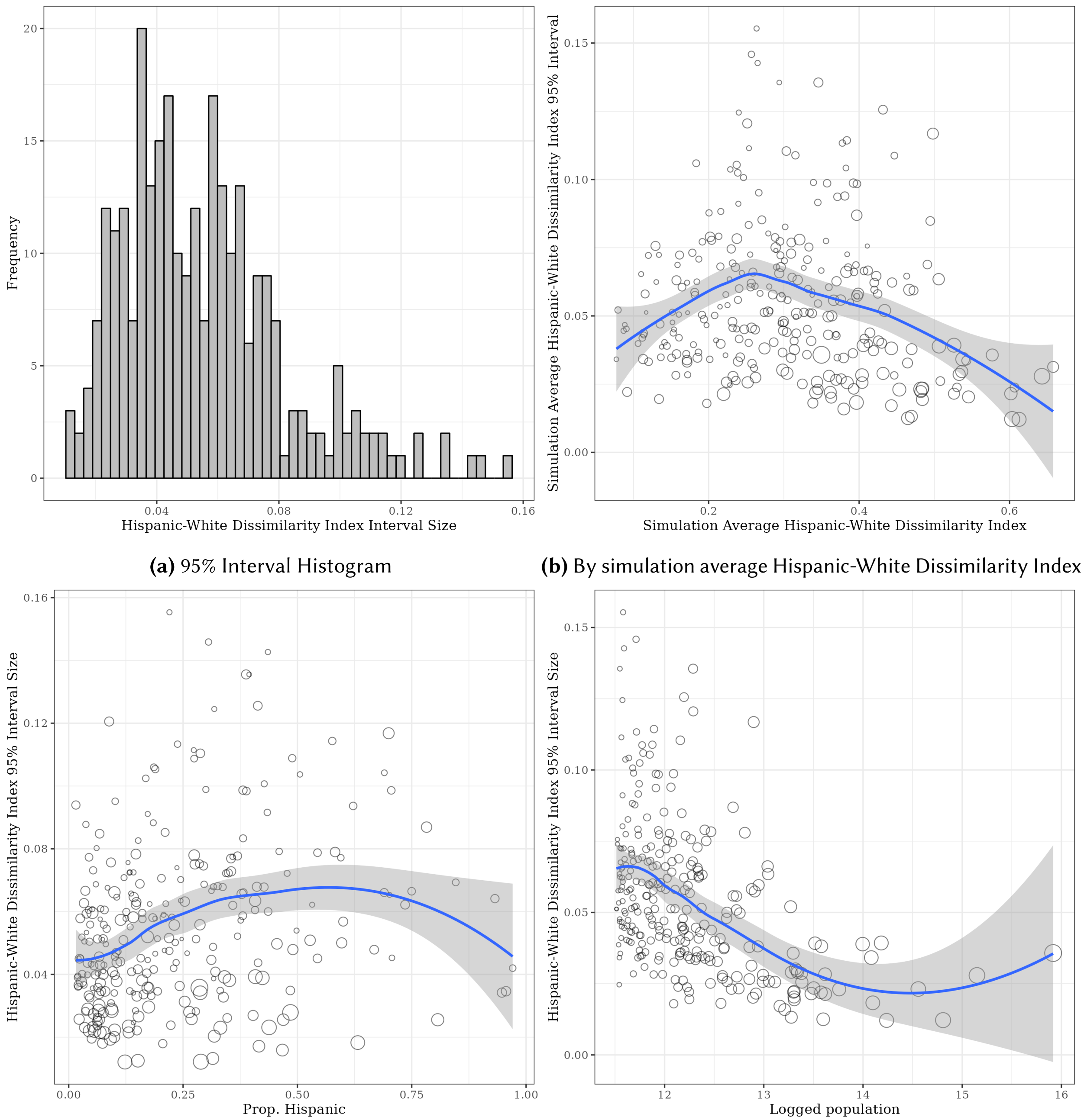


(c) By city proportion Hispanic



(d) By logged city population

Figure B18: Hispanic-White Dissimilarity Uncertainty - 2020



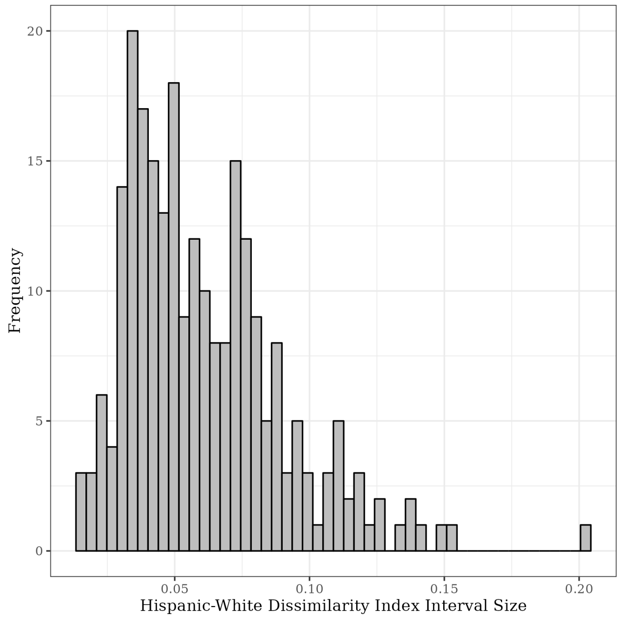
(a) 95% Interval Histogram

(b) By simulation average Hispanic-White Dissimilarity Index

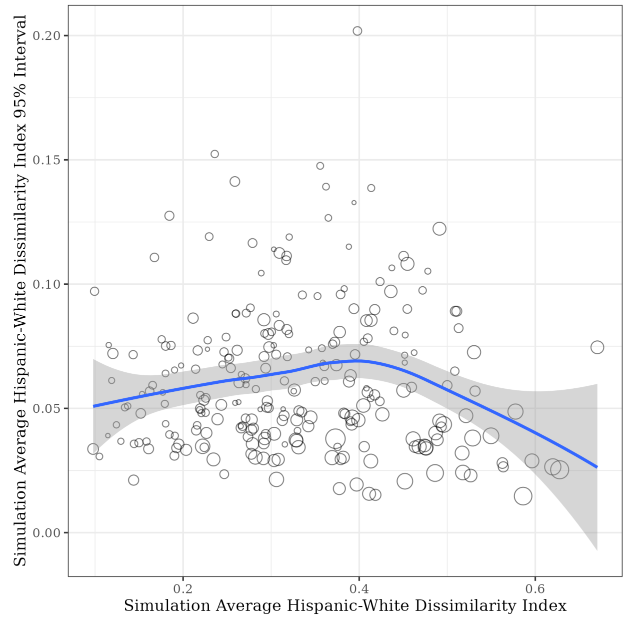
(c) By city proportion Hispanic

(d) By logged city population

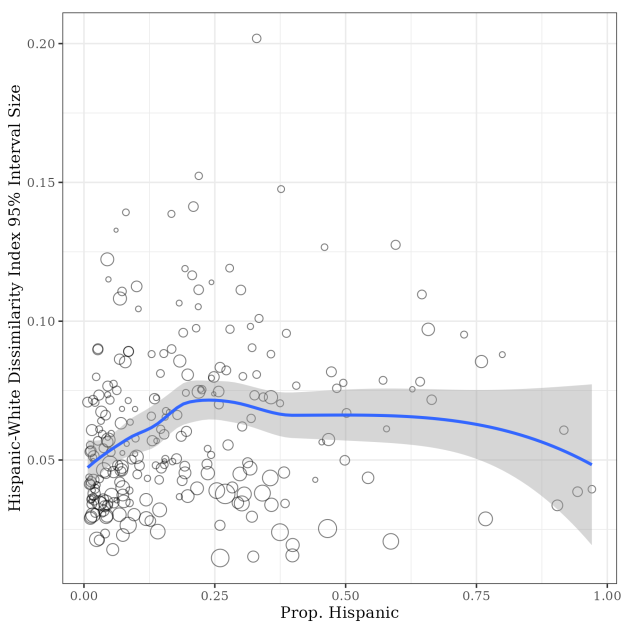
Figure B19: Hispanic-White Dissimilarity Uncertainty - 2010



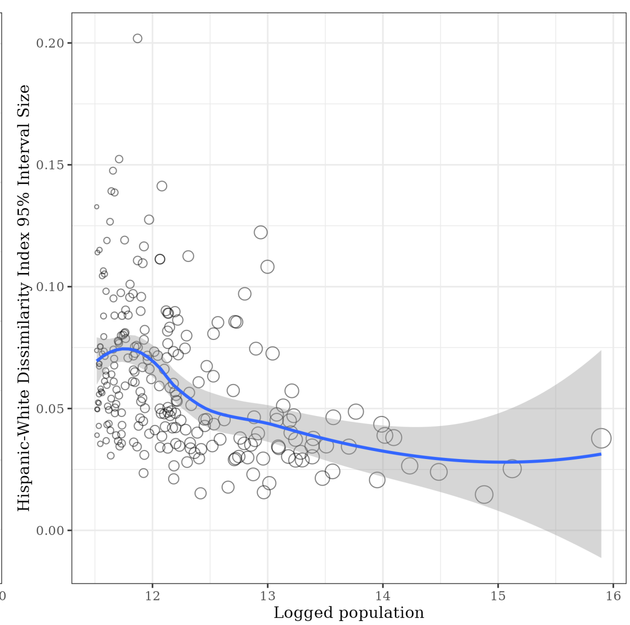
(a) 95% Interval Histogram



(b) By simulation average Hispanic-White Dissimilarity Index



(c) By city proportion Hispanic



(d) By logged city population

Figure B20: Hispanic-White Dissimilarity Uncertainty - 2000

B.5 H Index

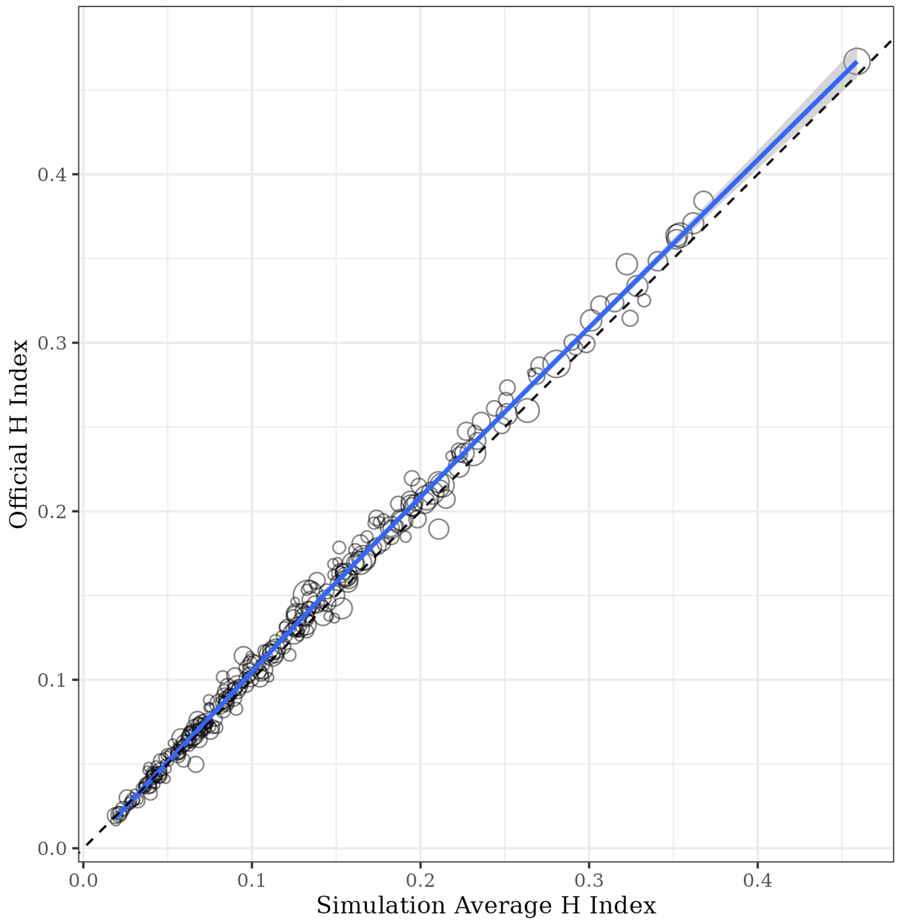


Figure B21: Official H Index versus Simulated - 2010

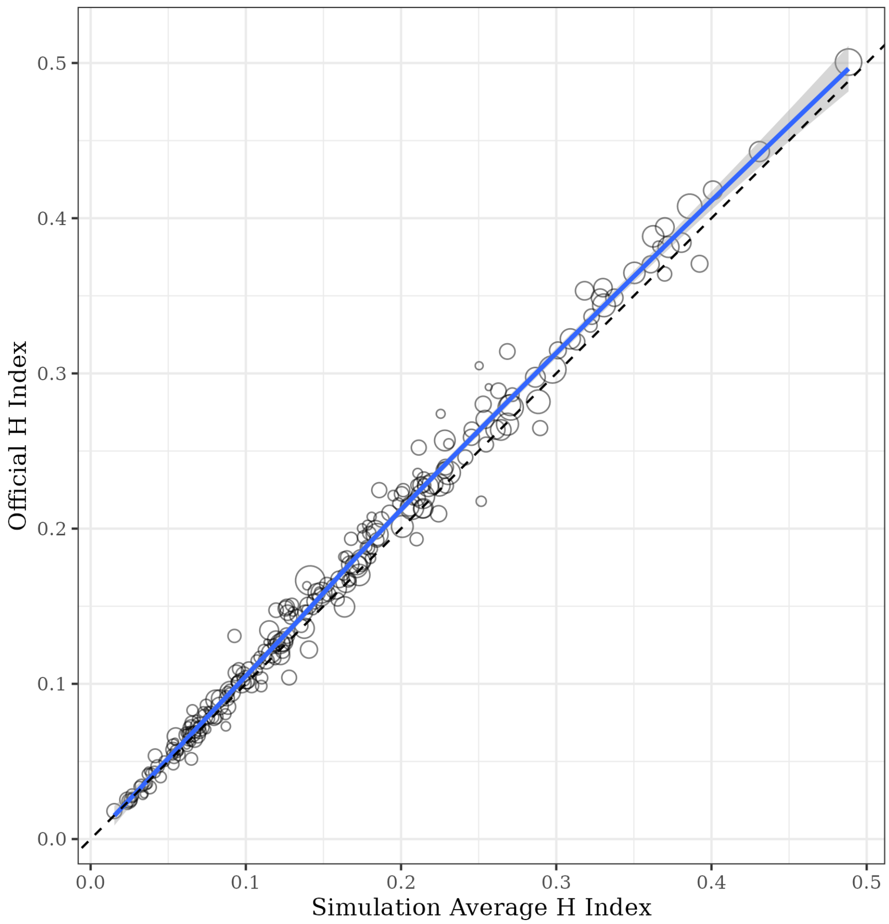
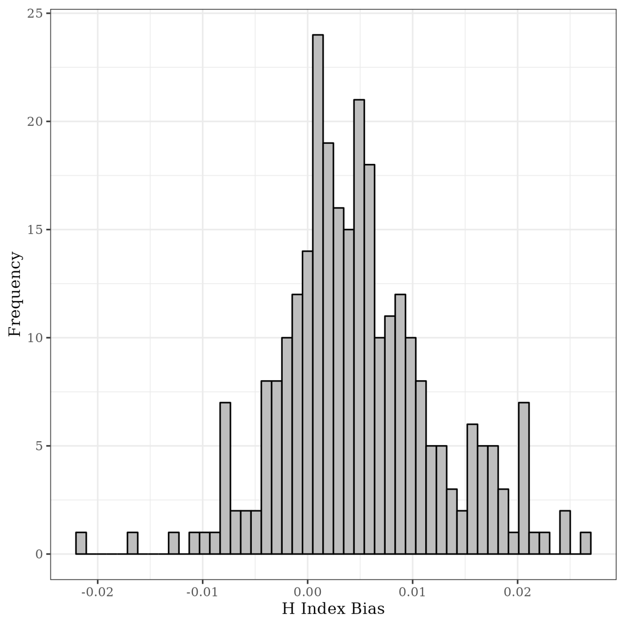
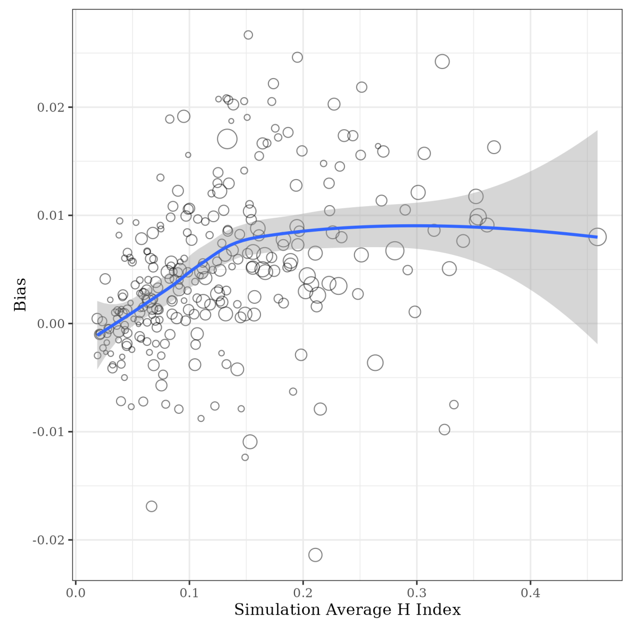


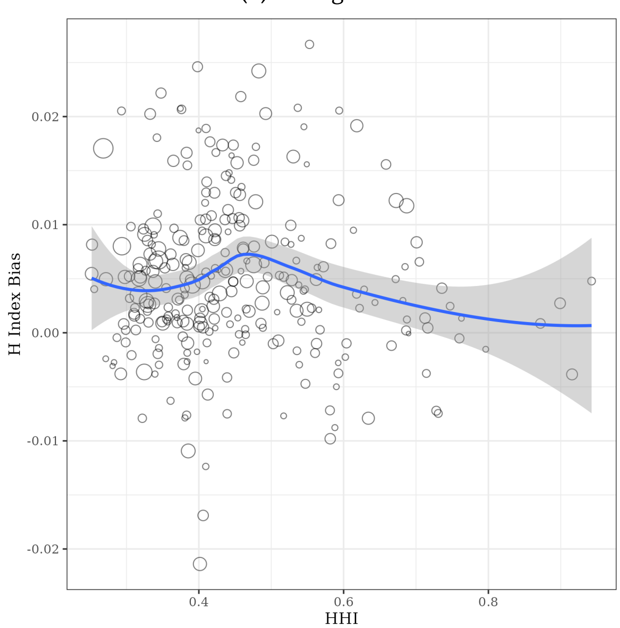
Figure B22: Official H Index versus Simulated - 2000



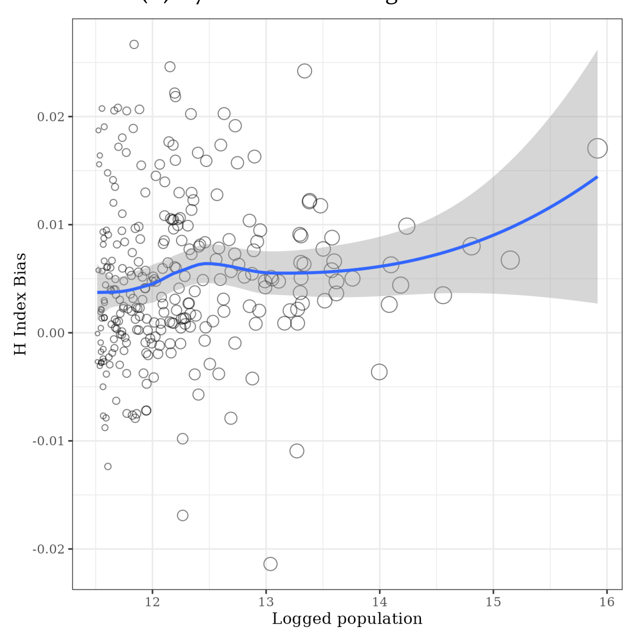
(a) Histogram



(b) By simulation average H index

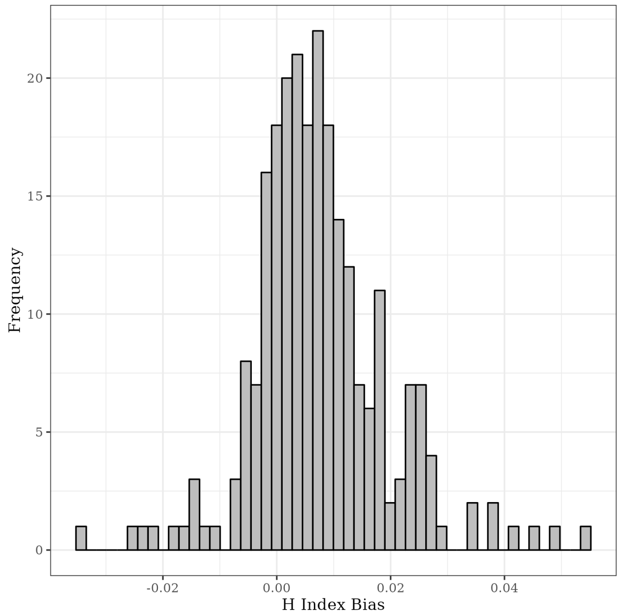


(c) By city Herfindahl Index

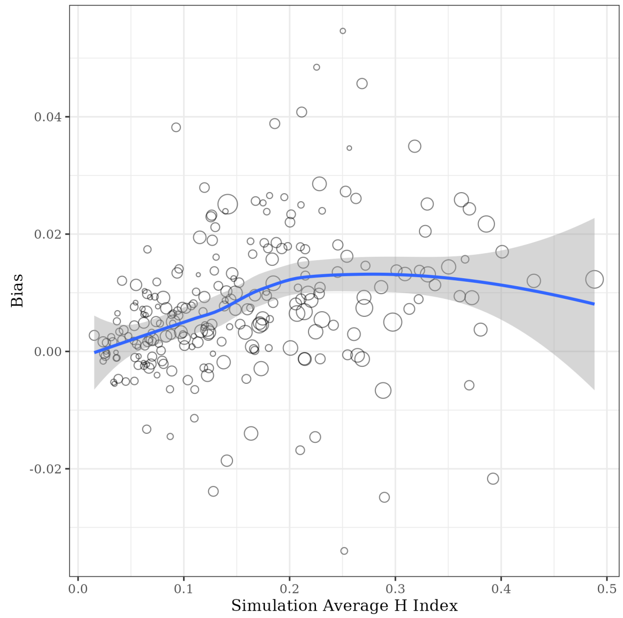


(d) By logged city population

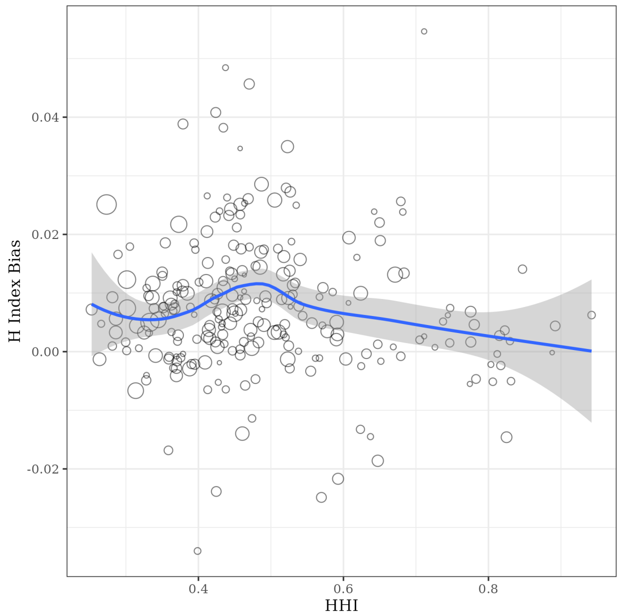
Figure B23: H Index Bias - 2010



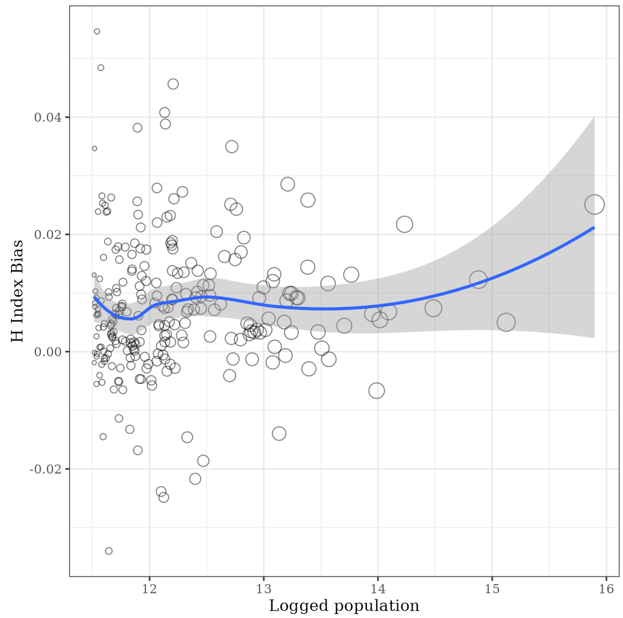
(a) Histogram



(b) By simulation average H index

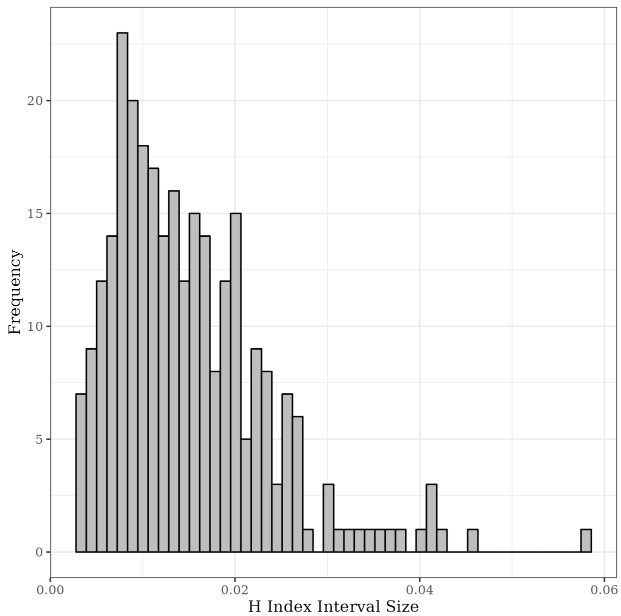


(c) By city Herfindahl Index

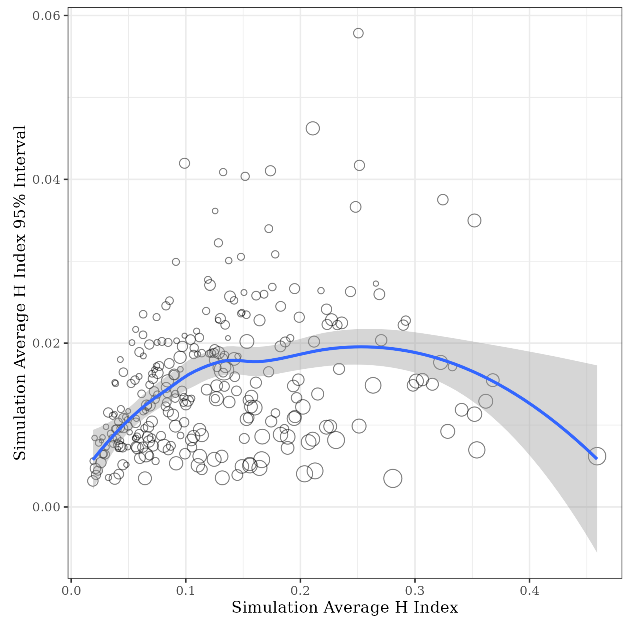


(d) By logged city population

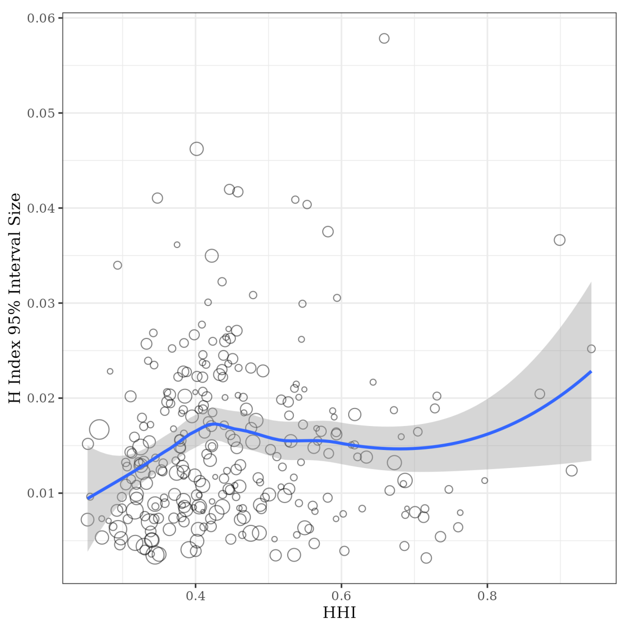
Figure B24: H Index Bias - 2000



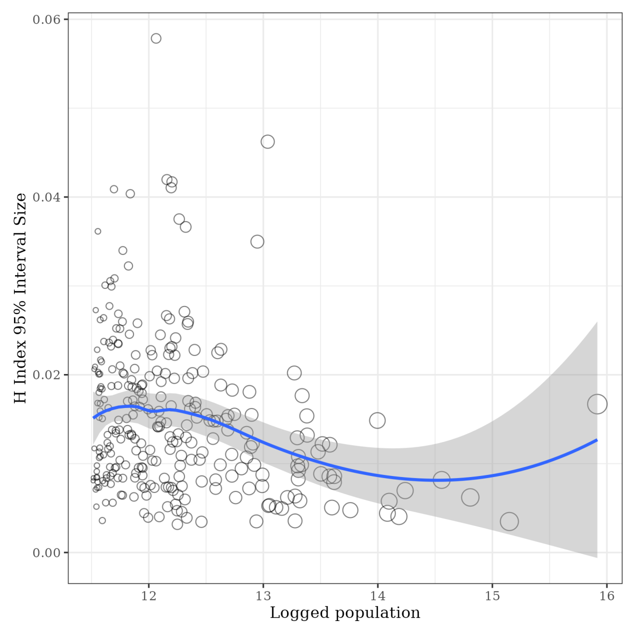
(a) 95% Interval Histogram



(b) By simulation average H index

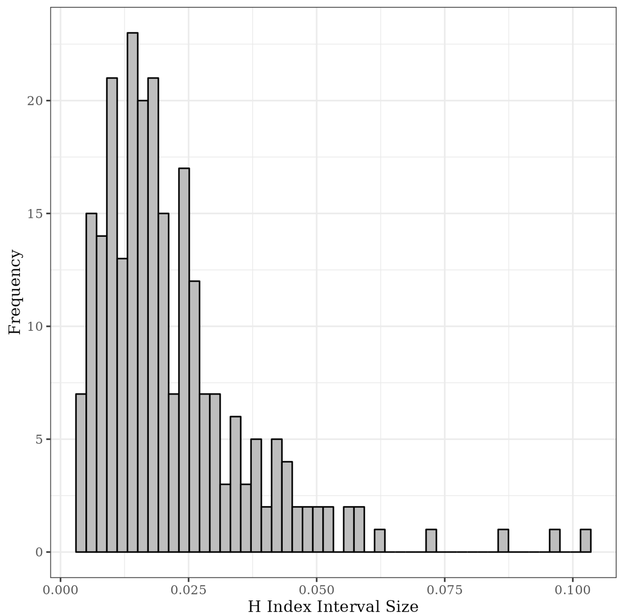


(c) By city Herfindahl Index

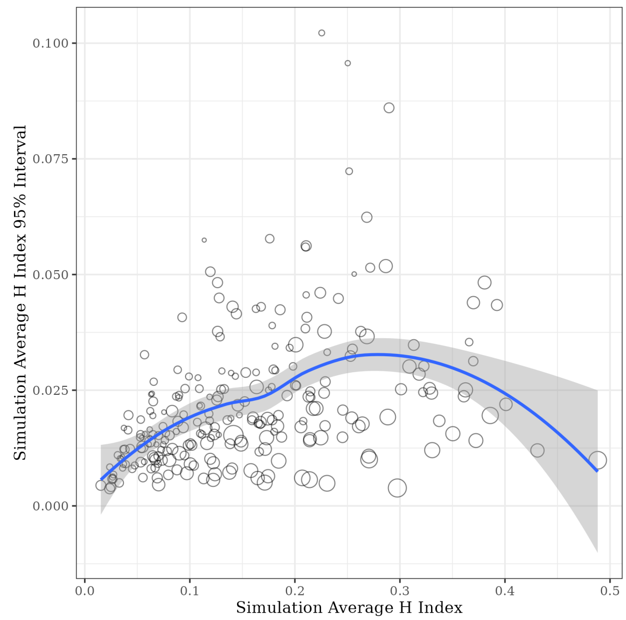


(d) By logged city population

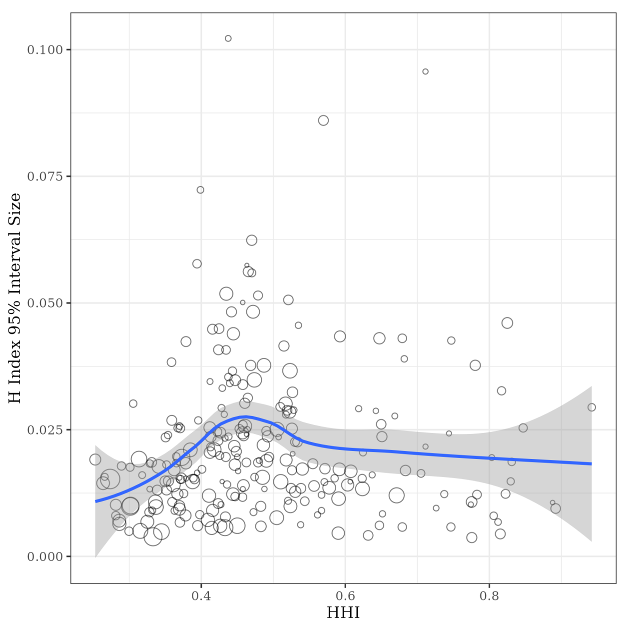
Figure B25: H Index Uncertainty - 2010



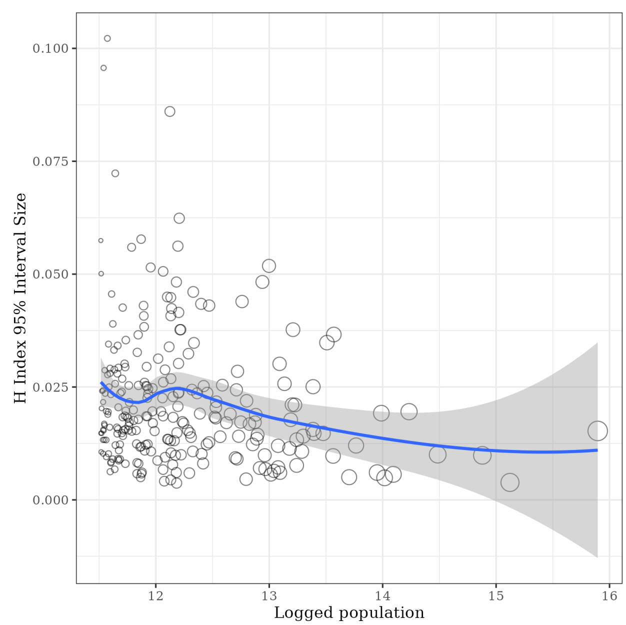
(a) 95% Interval Histogram



(b) By simulation average H index



(c) By city Herfindahl Index



(d) By logged city population

Figure B26: H Index Uncertainty - 2000

B.6 Black Isolation Index

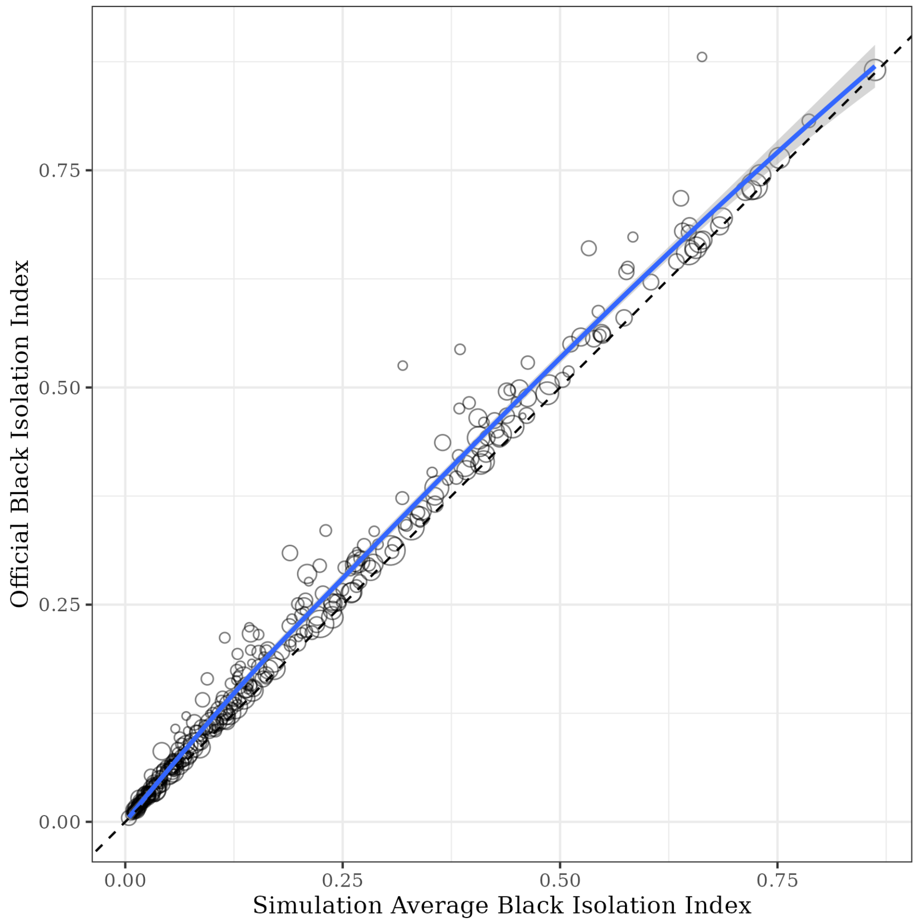


Figure B27: Official Black Isolation Index Index versus Simulated - 2020

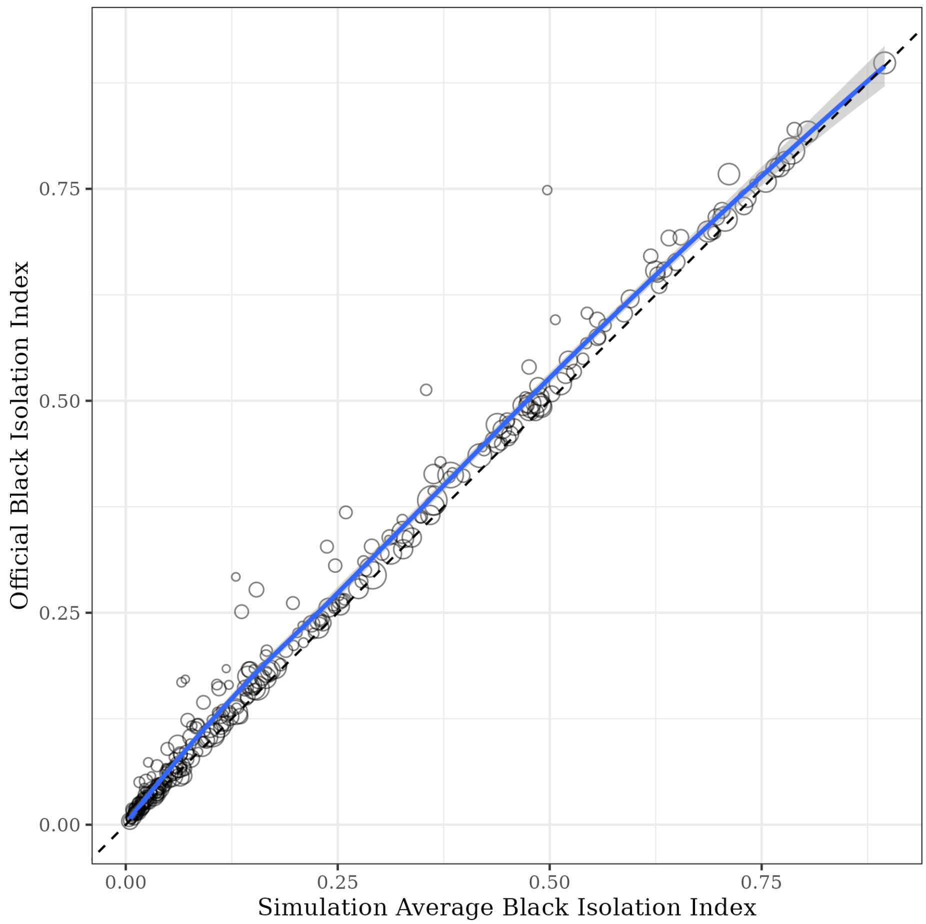


Figure B28: Official Black Isolation Index versus Simulated - 2010

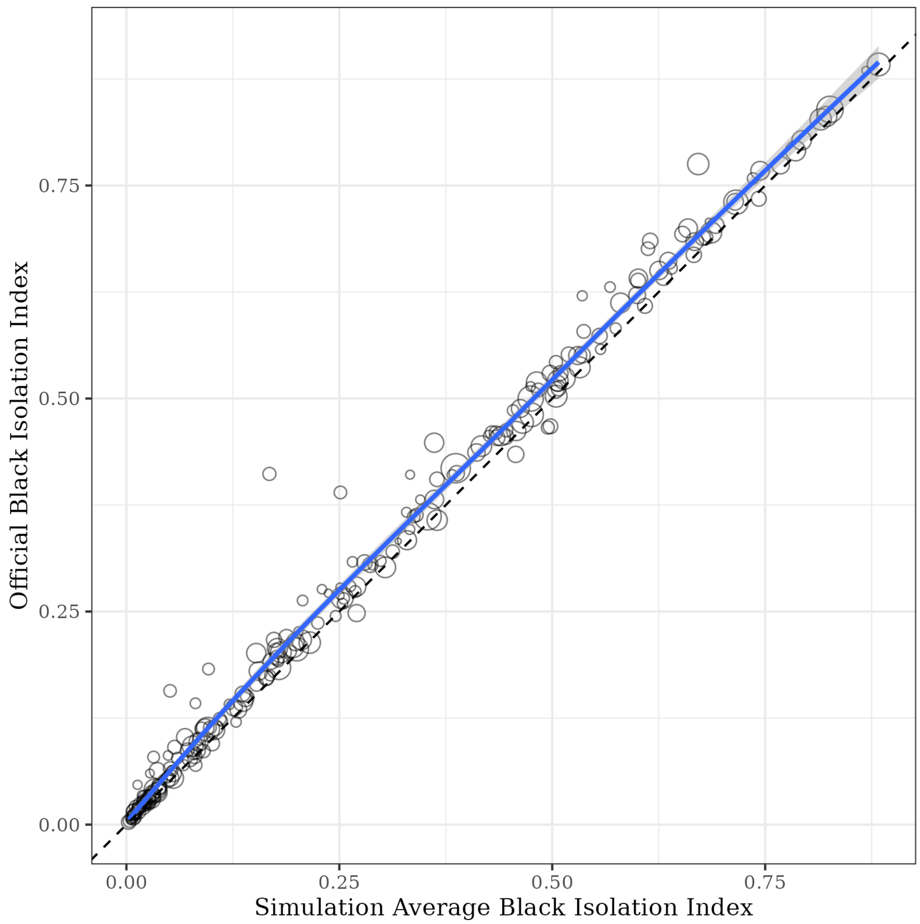
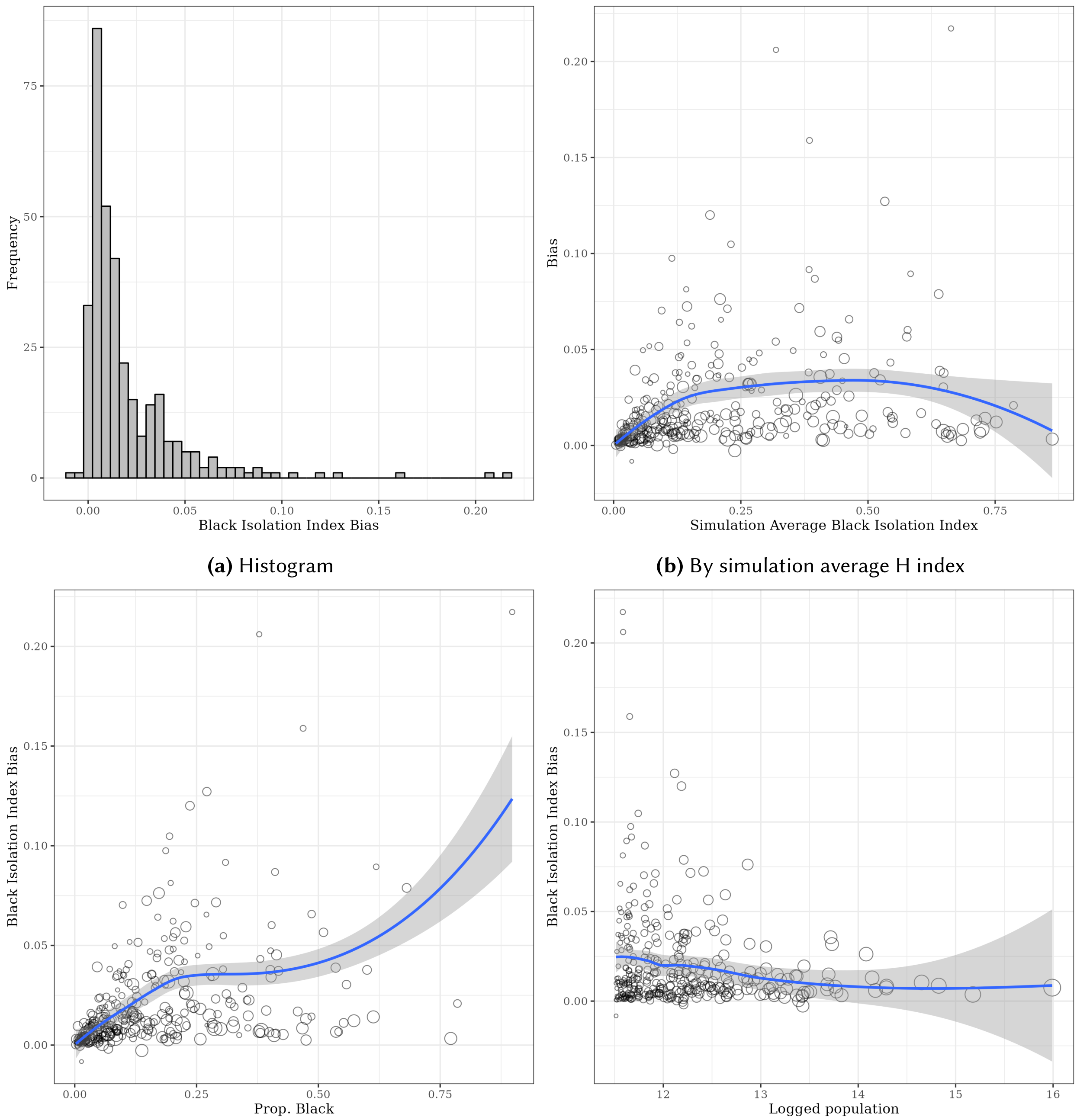


Figure B29: Official Black Isolation Index versus Simulated - 2000



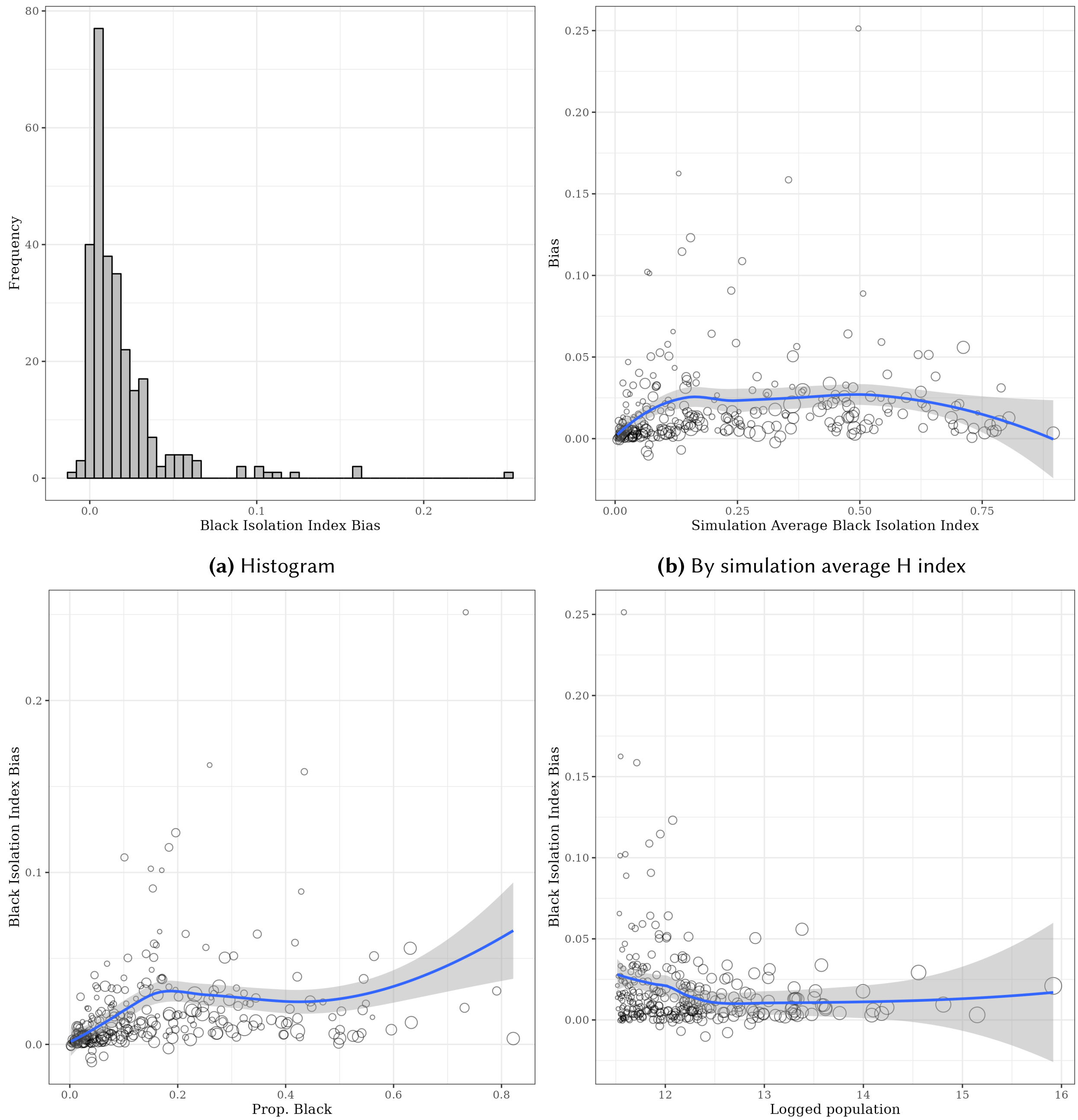
(a) Histogram

(b) By simulation average H index

(c) By city proportion Black

(d) By logged city population

Figure B30: Black Isolation Index - 2020



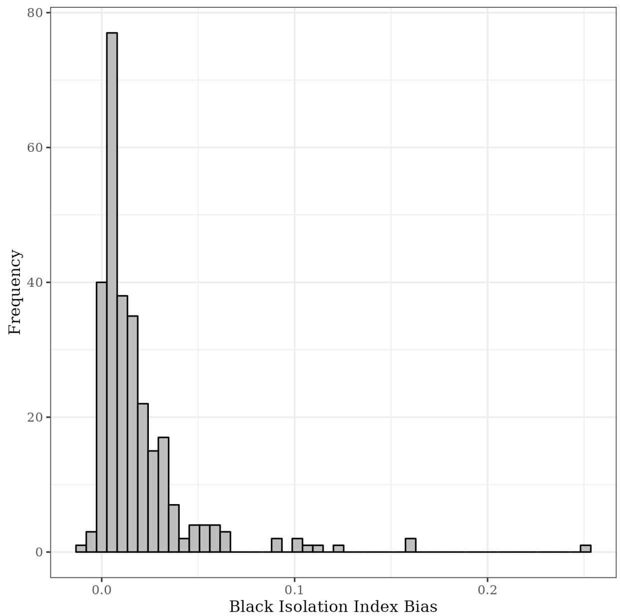
(a) Histogram

(b) By simulation average H index

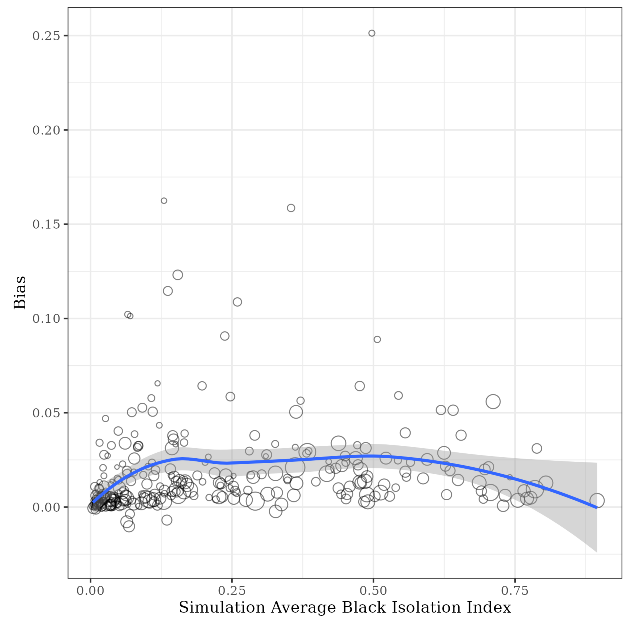
(c) By city proportion Black

(d) By logged city population

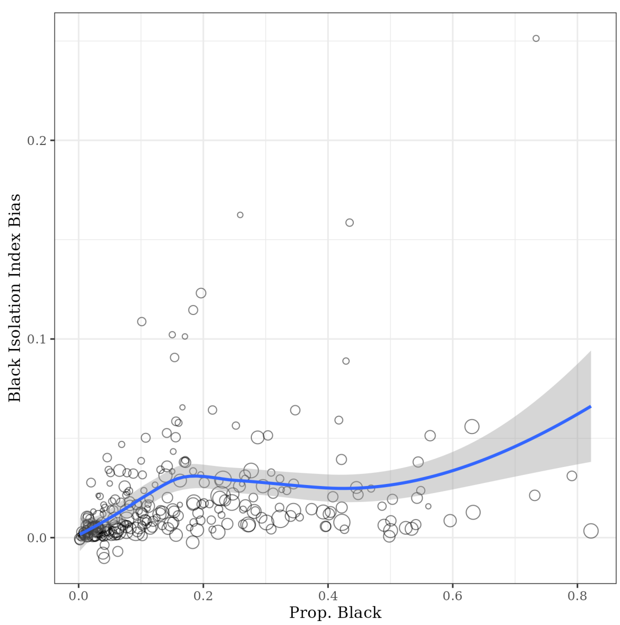
Figure B31: Black Isolation Index Bias - 2010



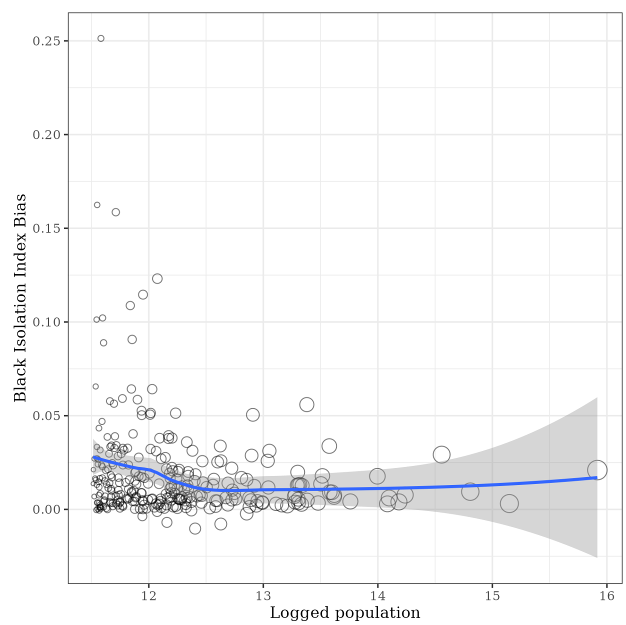
(a) Histogram



(b) By simulation average H index

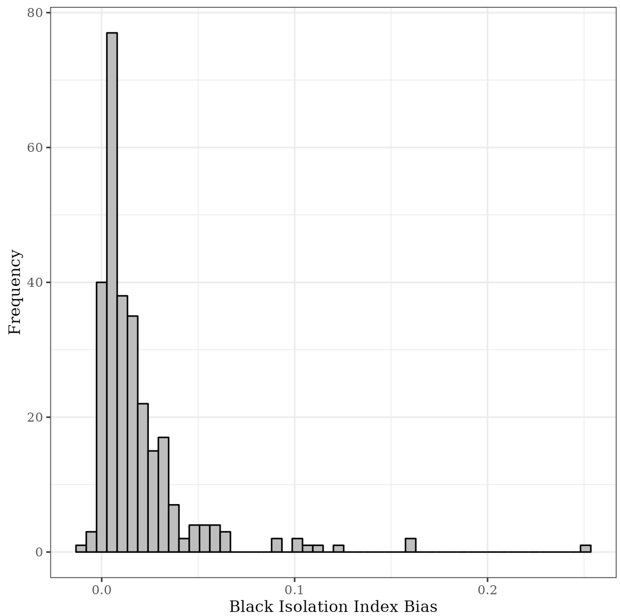


(c) By city proportion Black

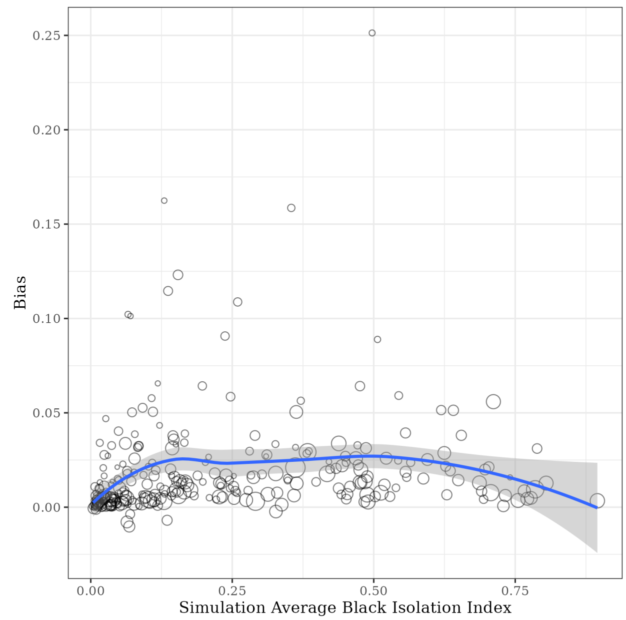


(d) By logged city population

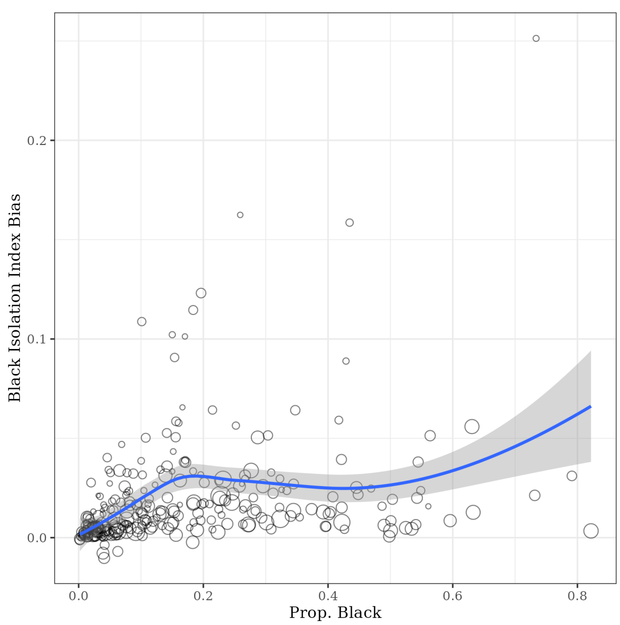
Figure B32: Black Isolation Index Bias - 2010



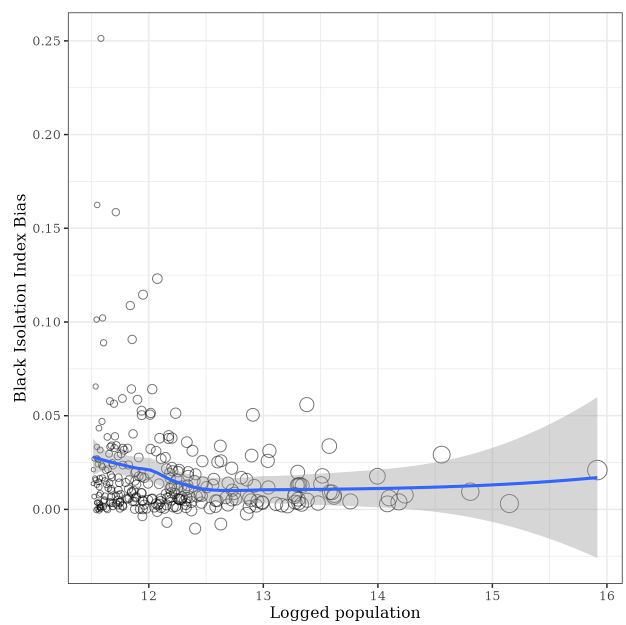
(a) Histogram



(b) By simulation average H index

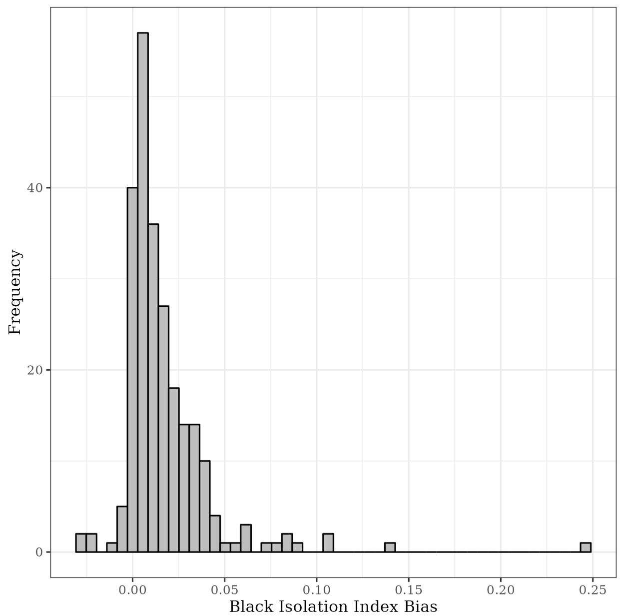


(c) By city proportion Black

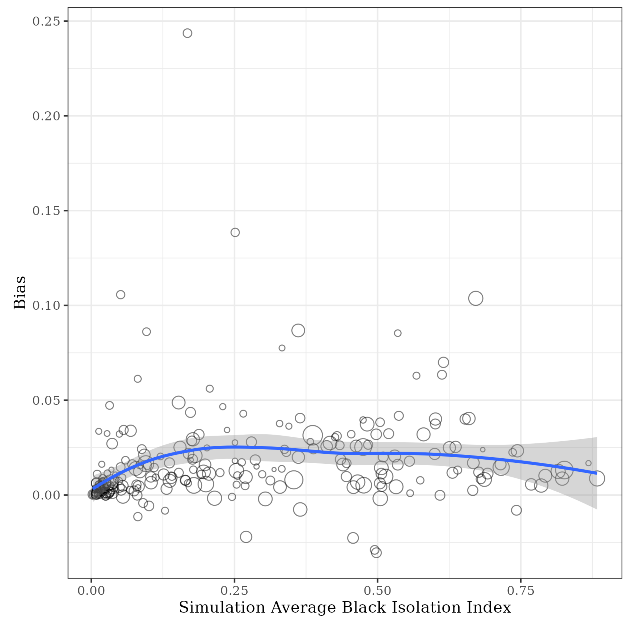


(d) By logged city population

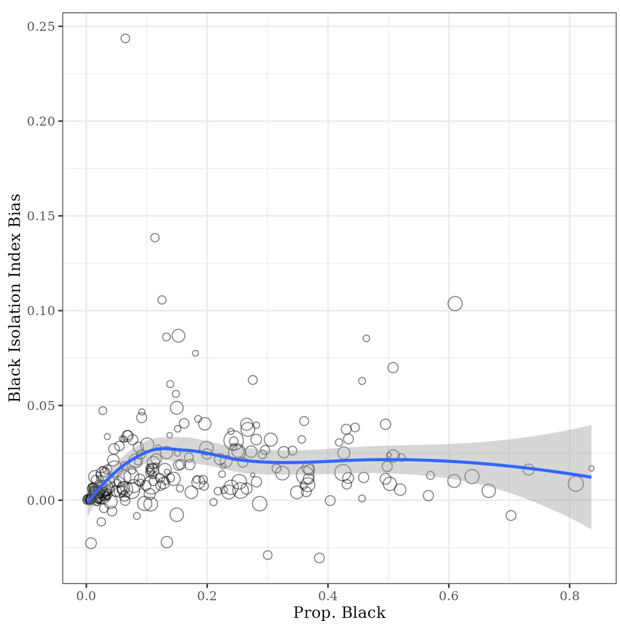
Figure B33: Black Isolation Index Bias - 2010



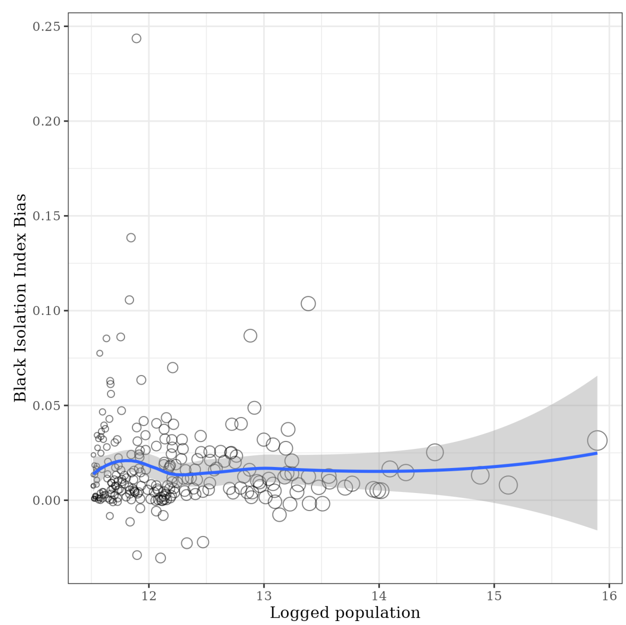
(a) Histogram



(b) By simulation average H index

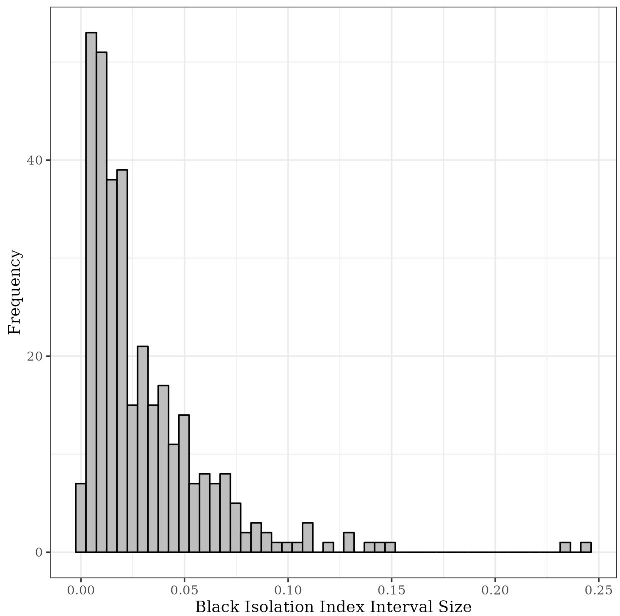


(c) By city proportion Black

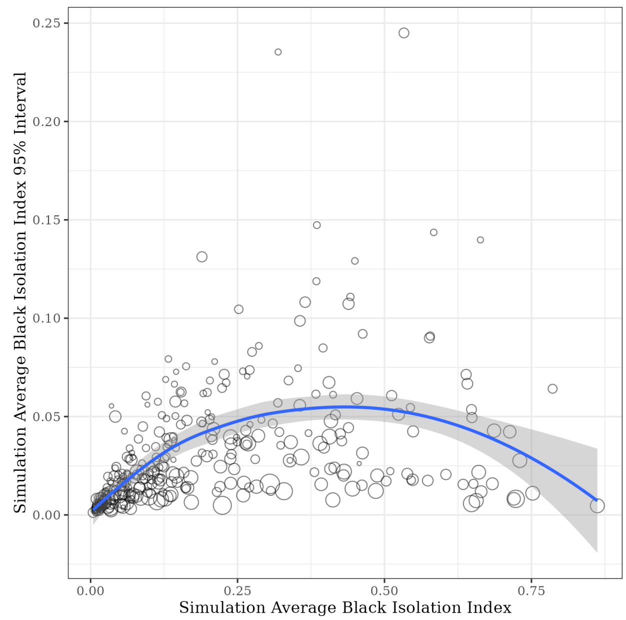


(d) By logged city population

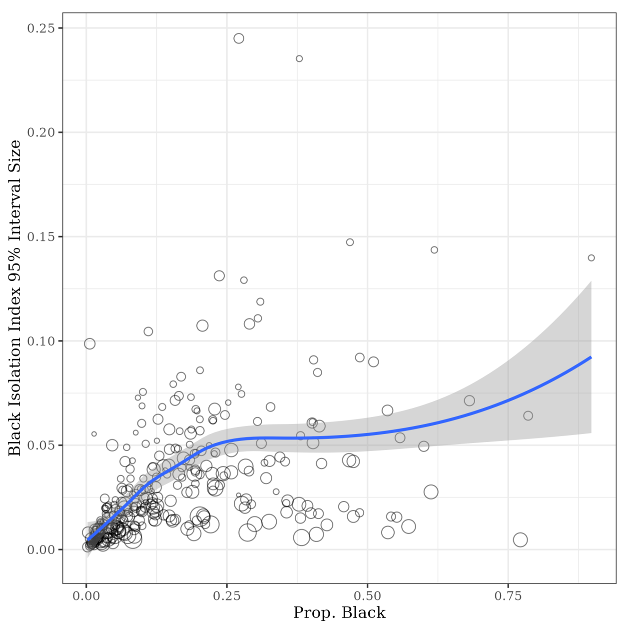
Figure B34: Black Isolation Index Bias - 2000



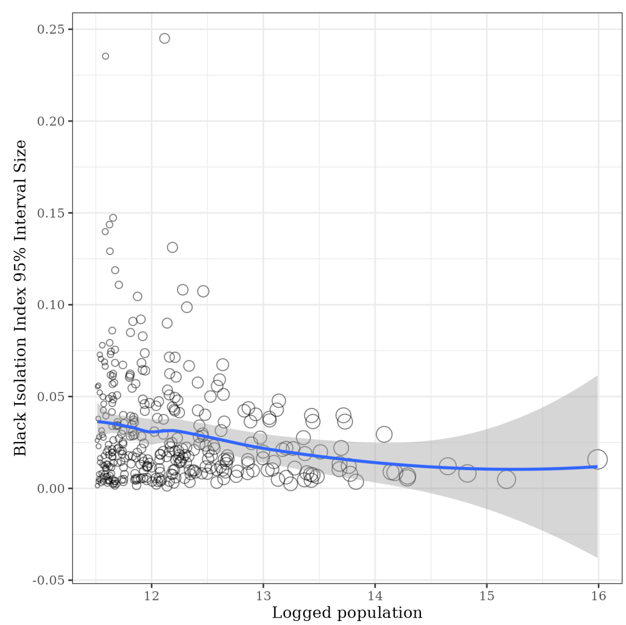
(a) 95% Interval Histogram



(b) By simulation average H index

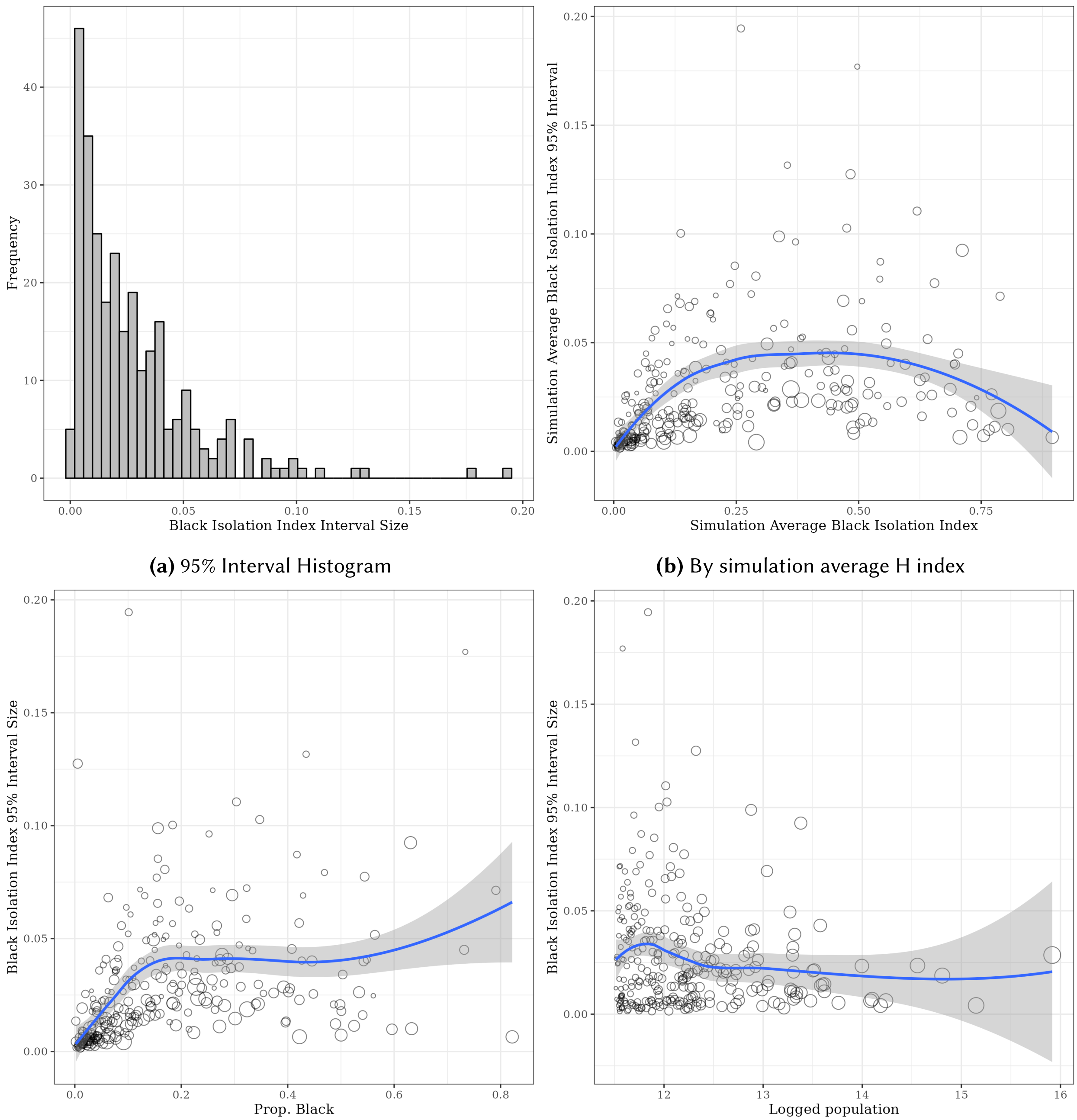


(c) By city Herfindahl Index



(d) By logged city population

Figure B35: Black Isolation Index Uncertainty - 2020



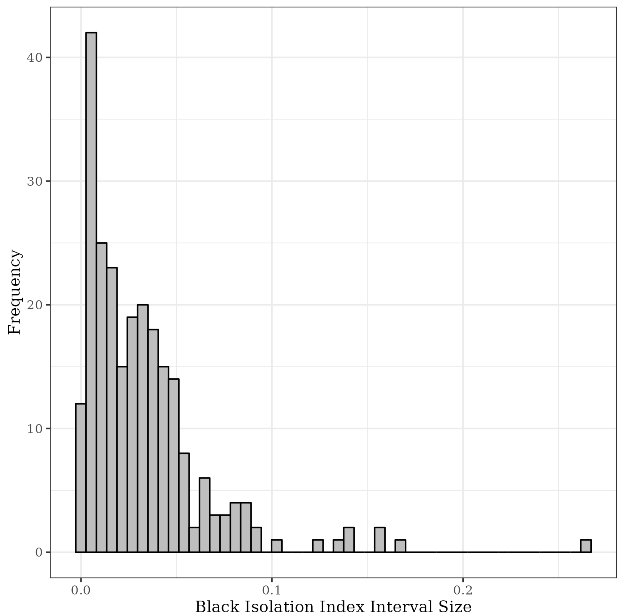
(a) 95% Interval Histogram

(b) By simulation average H index

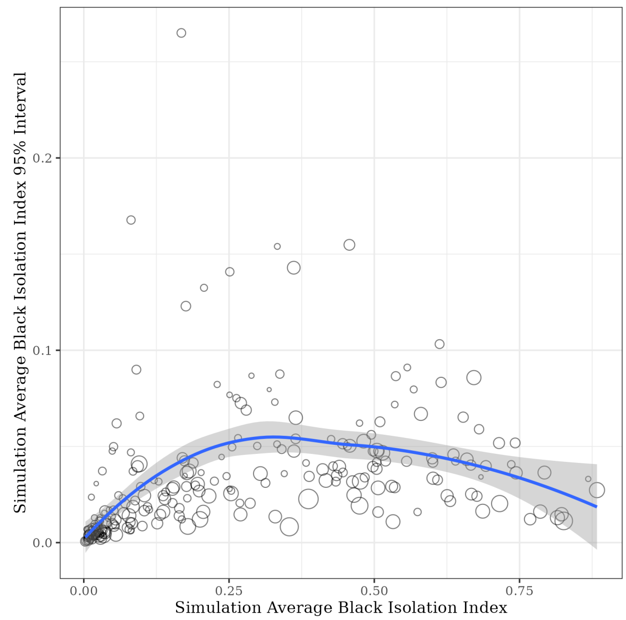
(c) By city Herfindahl Index

(d) By logged city population

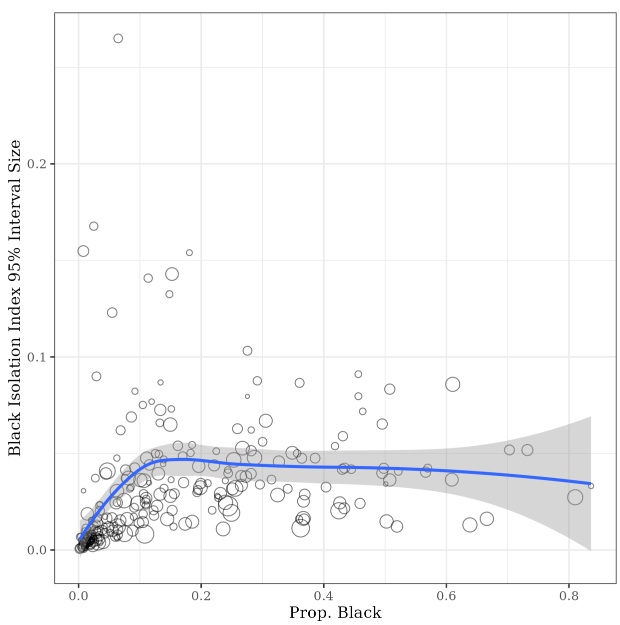
Figure B36: Black Isolation Index Uncertainty - 2010



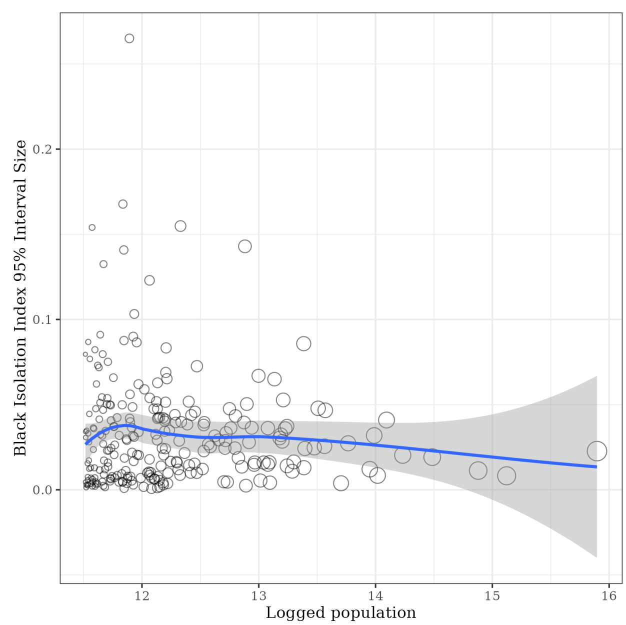
(a) 95% Interval Histogram



(b) By simulation average H index



(c) By city Herfindahl Index



(d) By logged city population

Figure B37: Black Isolation Index Uncertainty - 2000

B.7 White Isolation Index

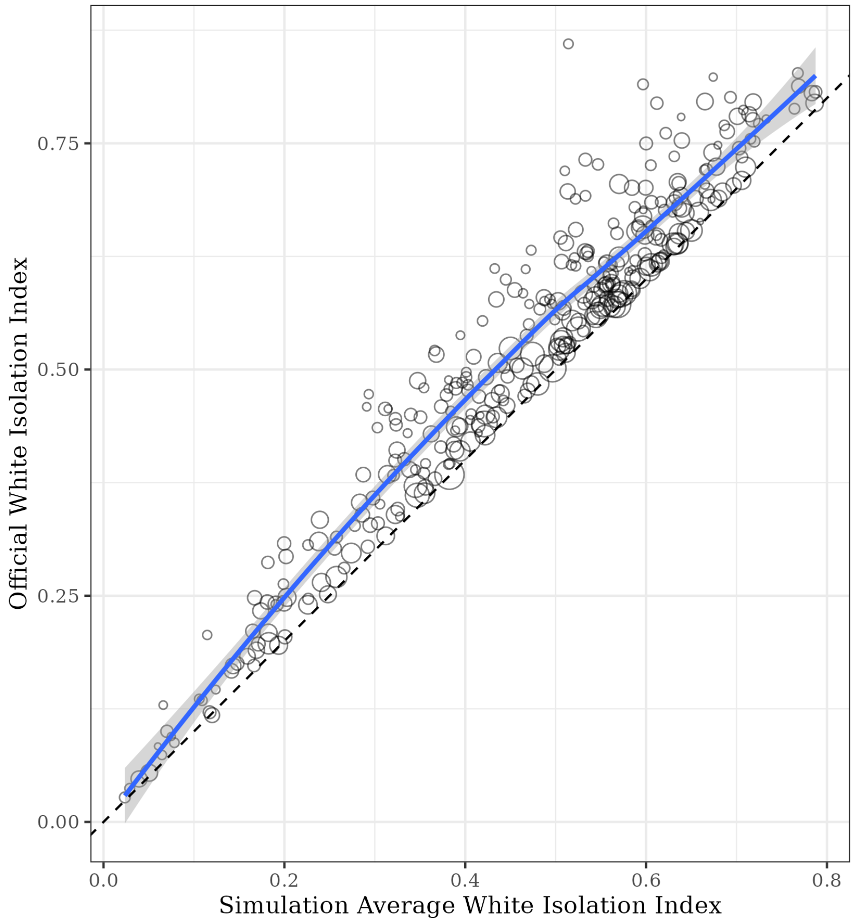


Figure B38: Official White Isolation Index Index versus Simulated - 2020

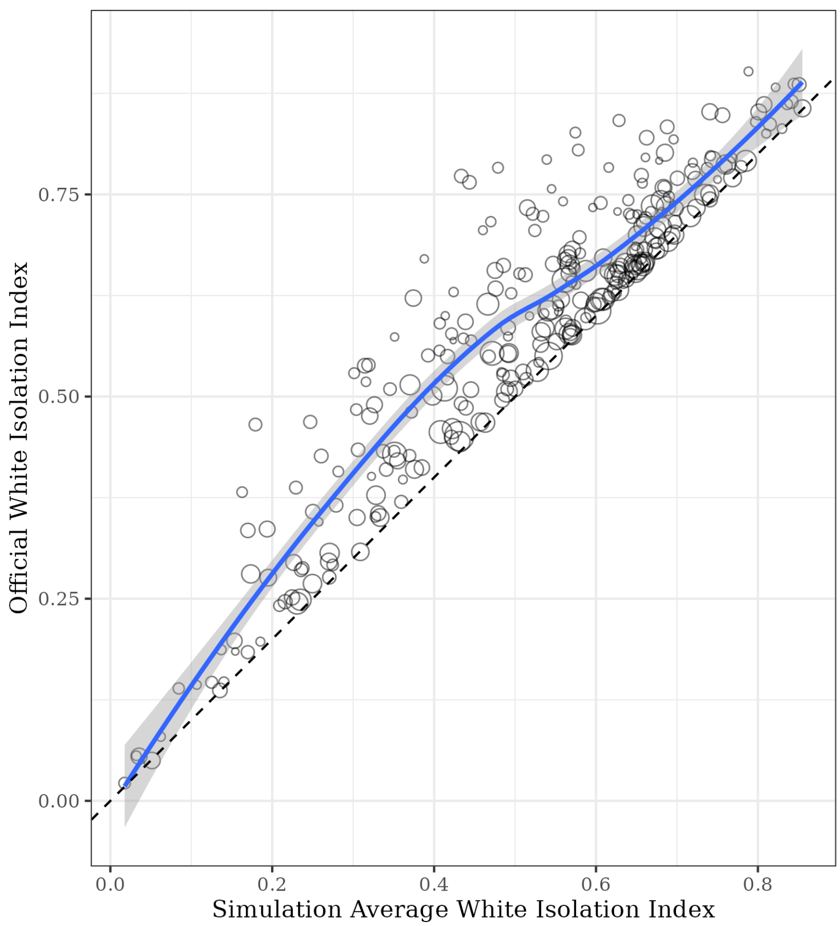


Figure B39: Official White Isolation Index versus Simulated - 2010

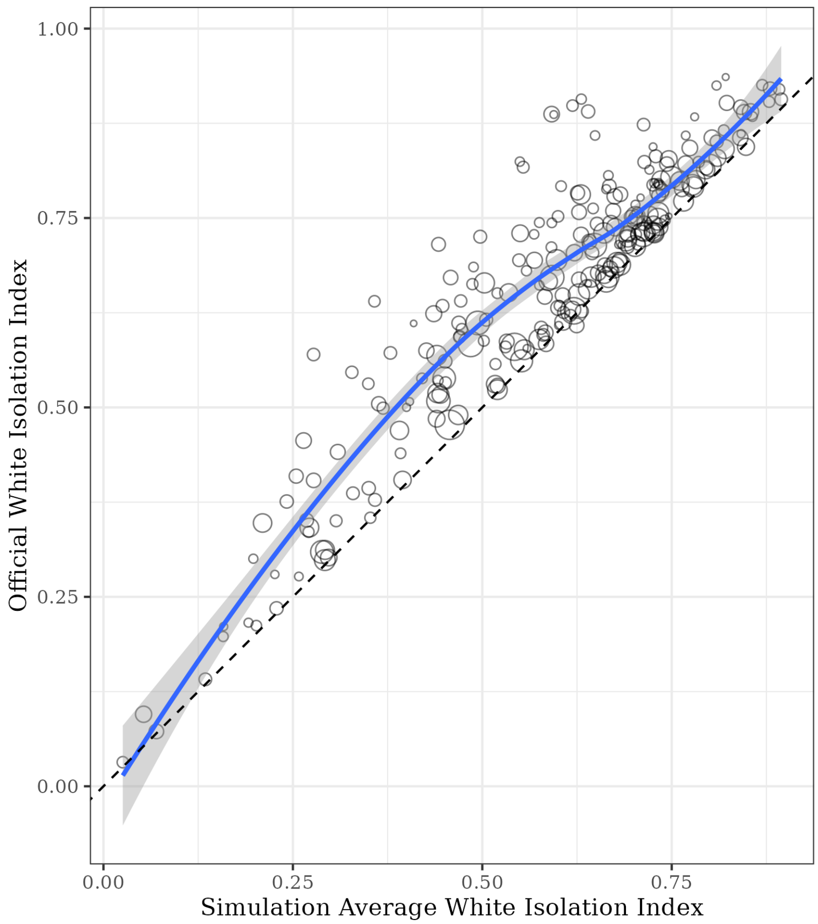
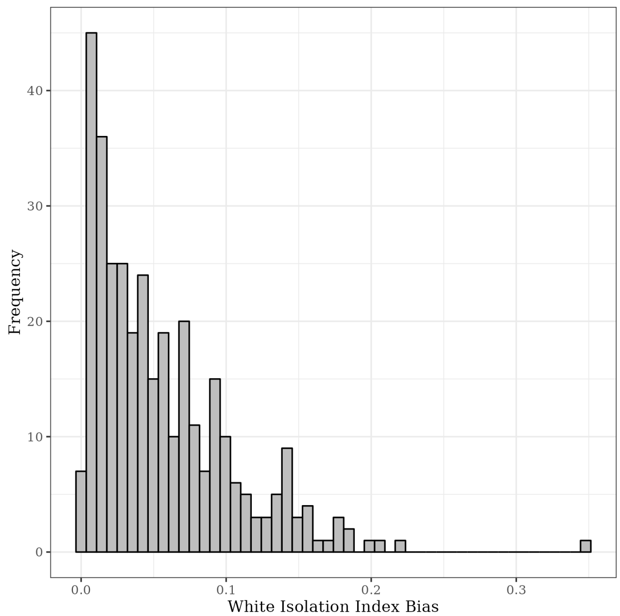
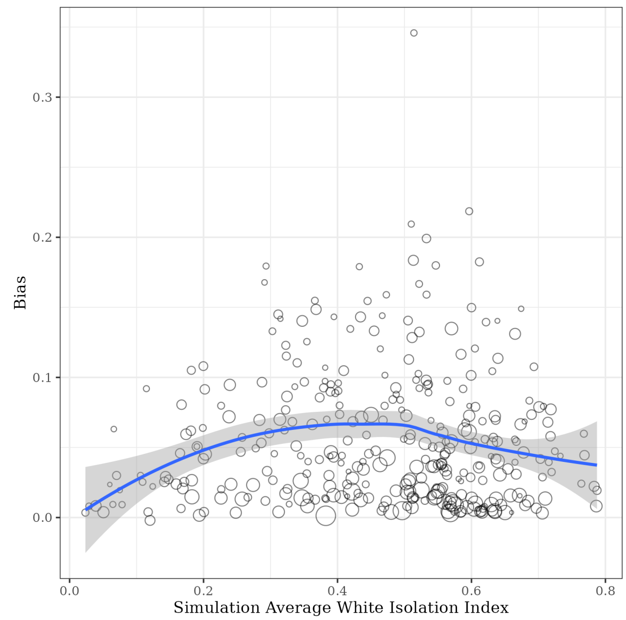


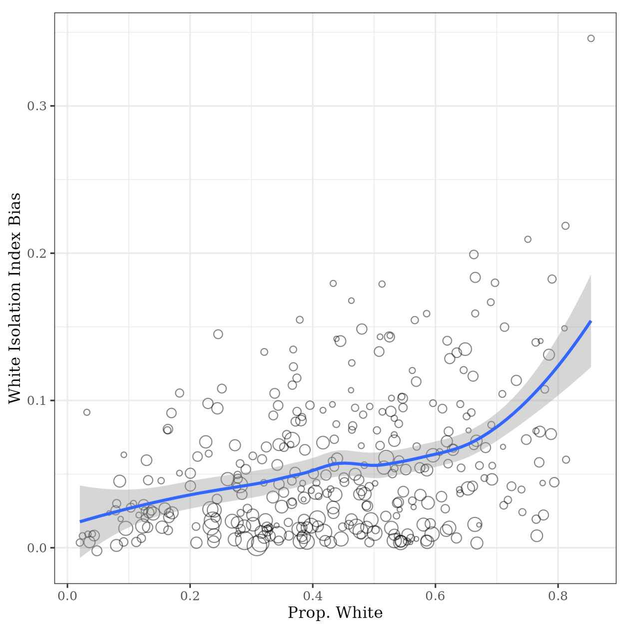
Figure B40: Official White Isolation Index versus Simulated - 2000



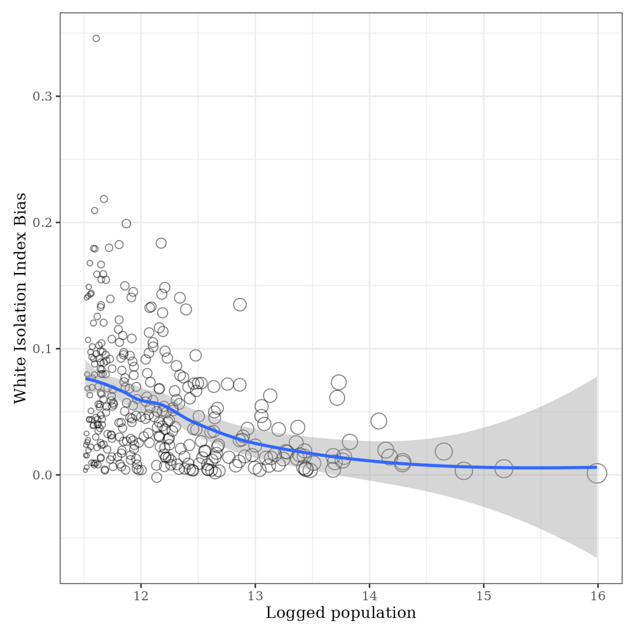
(a) Histogram



(b) By simulation average H index

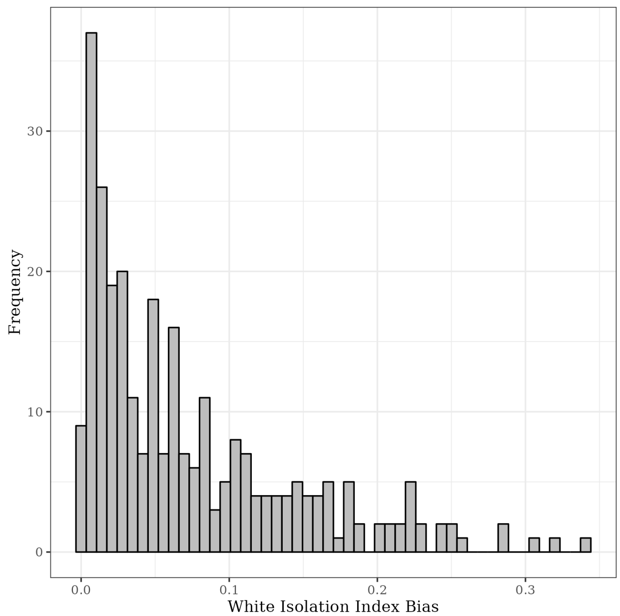


(c) By city proportion White

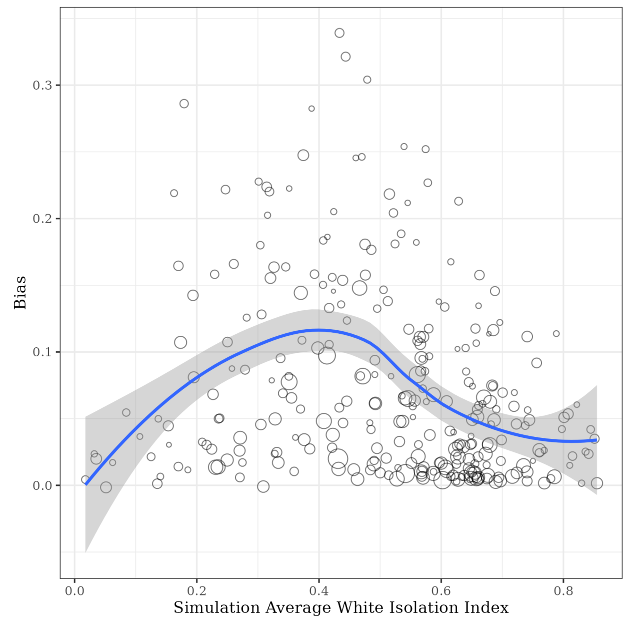


(d) By logged city population

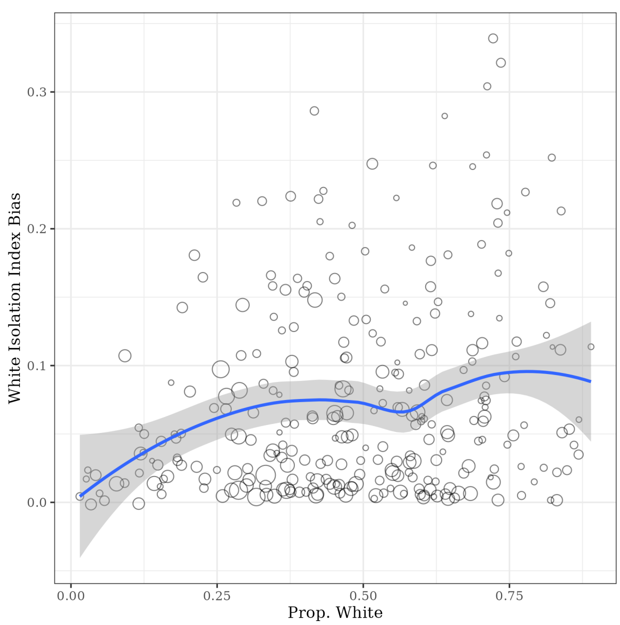
Figure B41: White Isolation Index - 2020



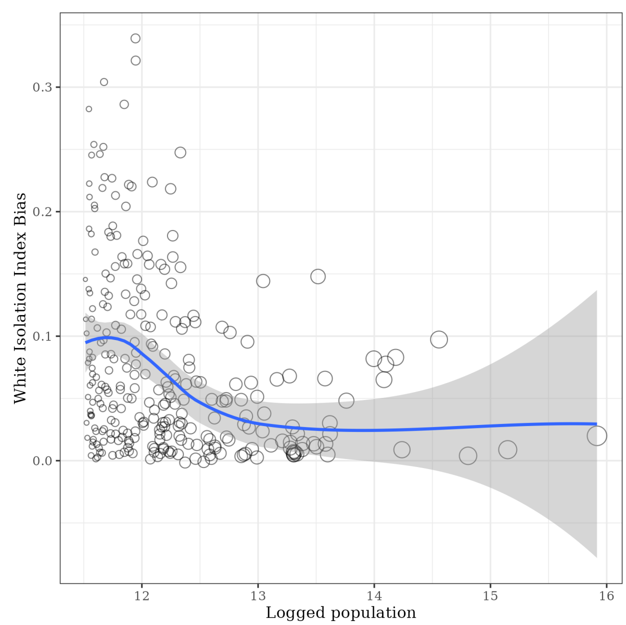
(a) Histogram



(b) By simulation average H index

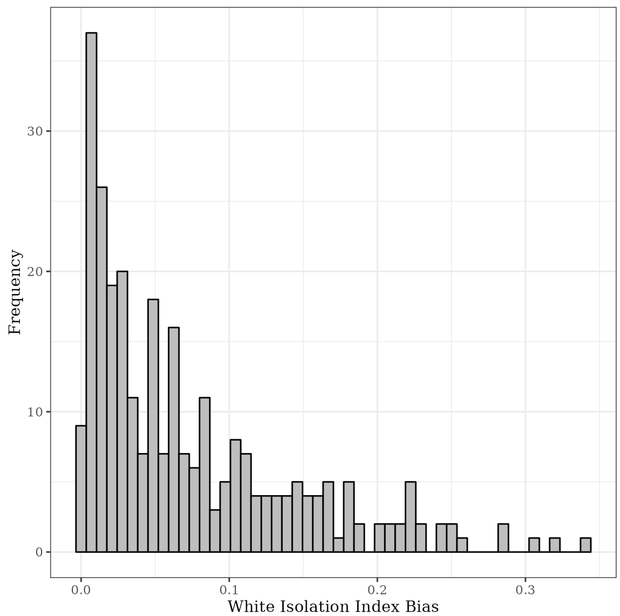


(c) By city proportion White

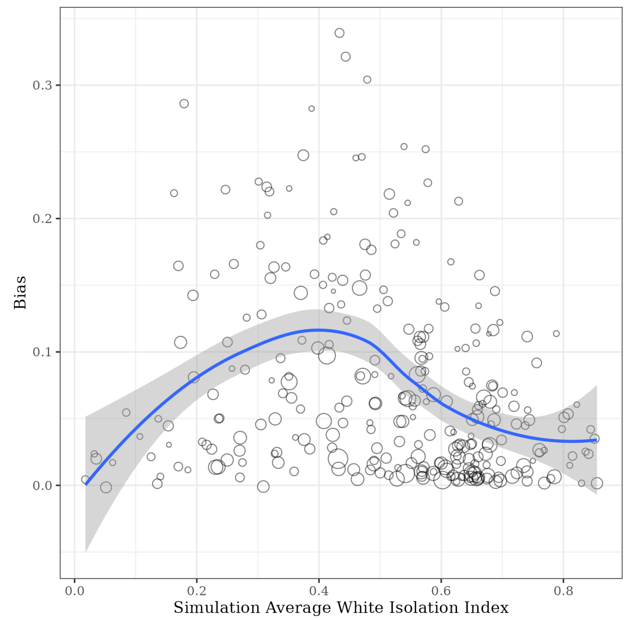


(d) By logged city population

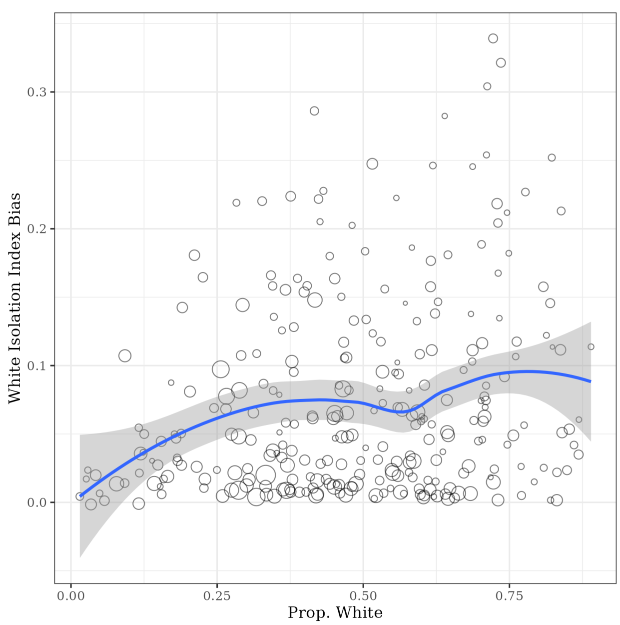
Figure B42: White Isolation Index Bias - 2010



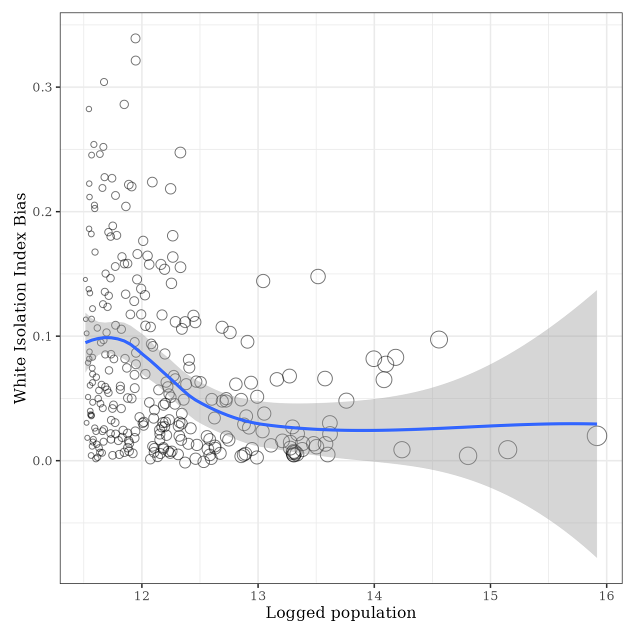
(a) Histogram



(b) By simulation average H index

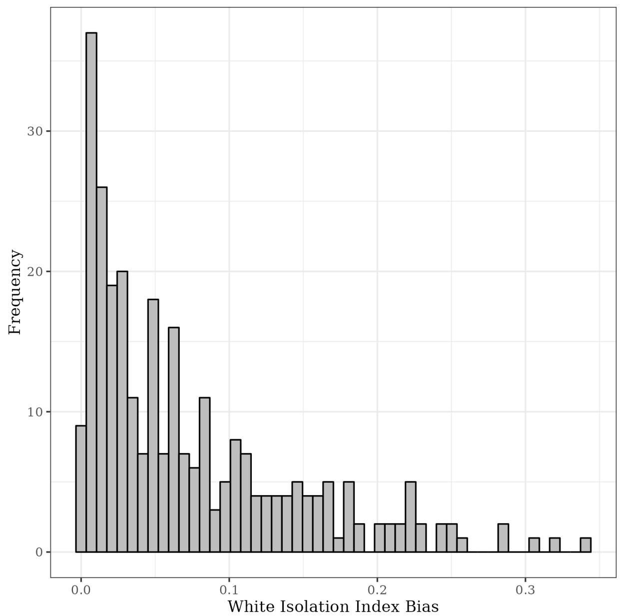


(c) By city proportion White

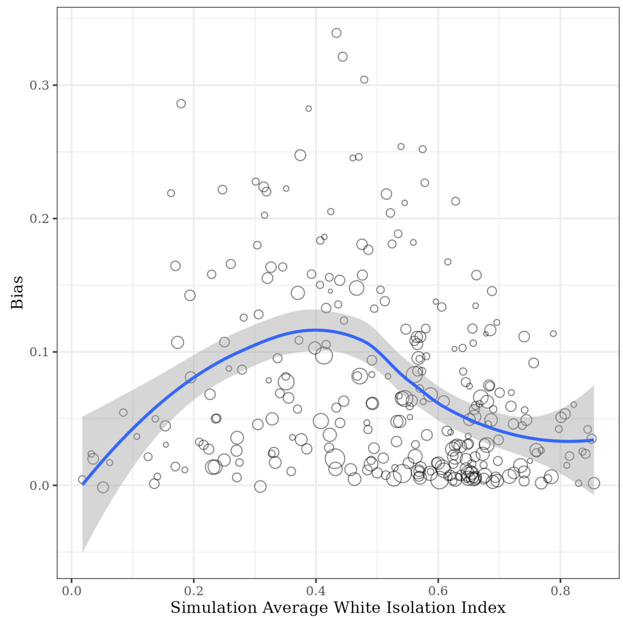


(d) By logged city population

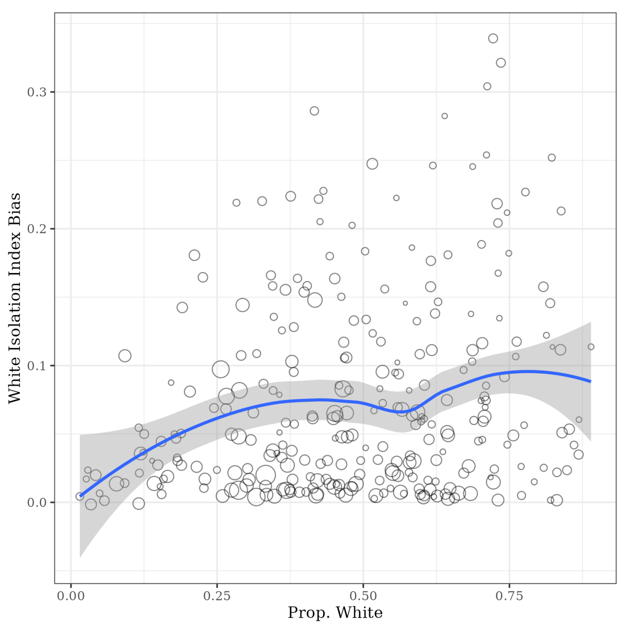
Figure B43: White Isolation Index Bias - 2010



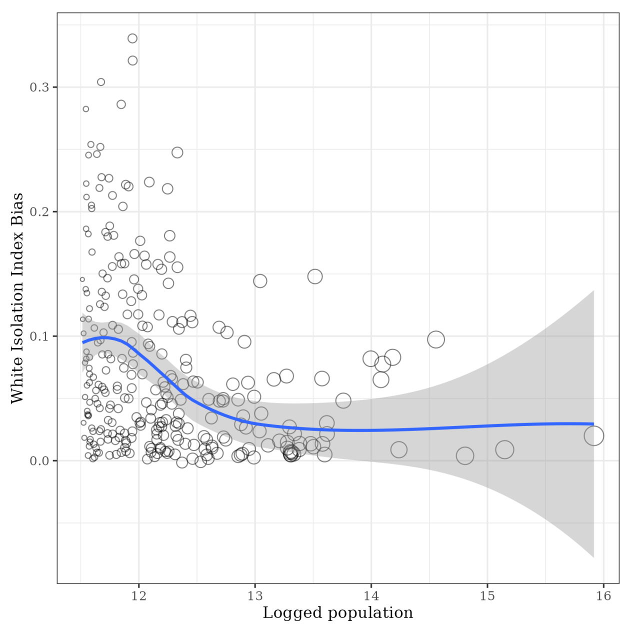
(a) Histogram



(b) By simulation average H index

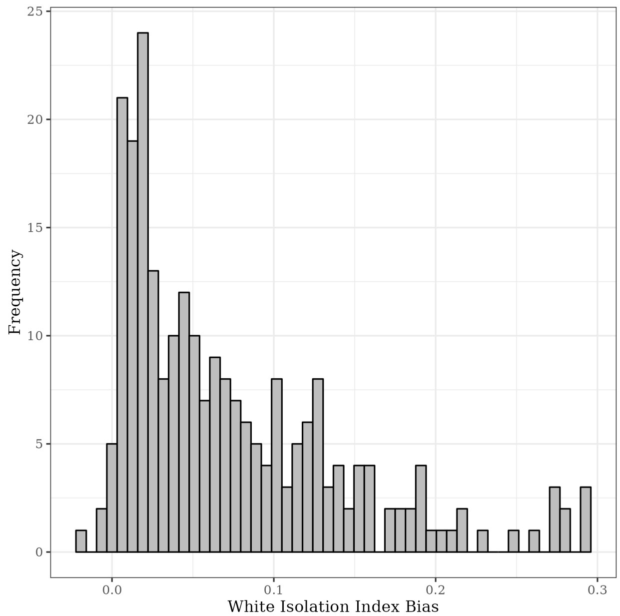


(c) By city proportion White

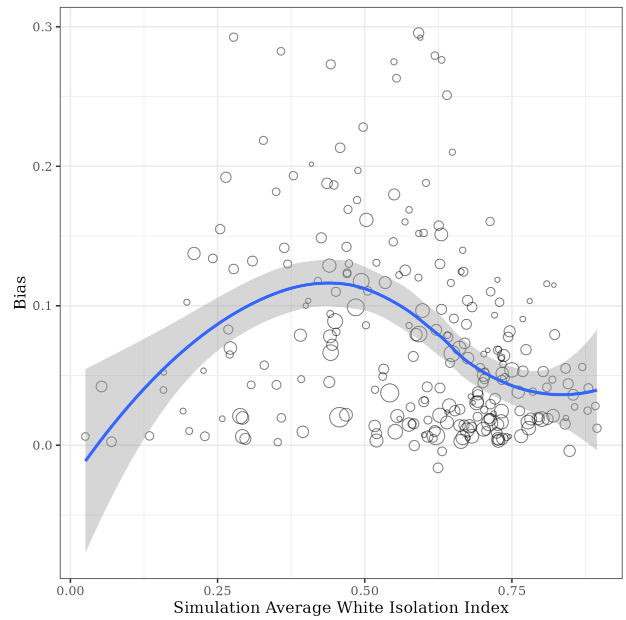


(d) By logged city population

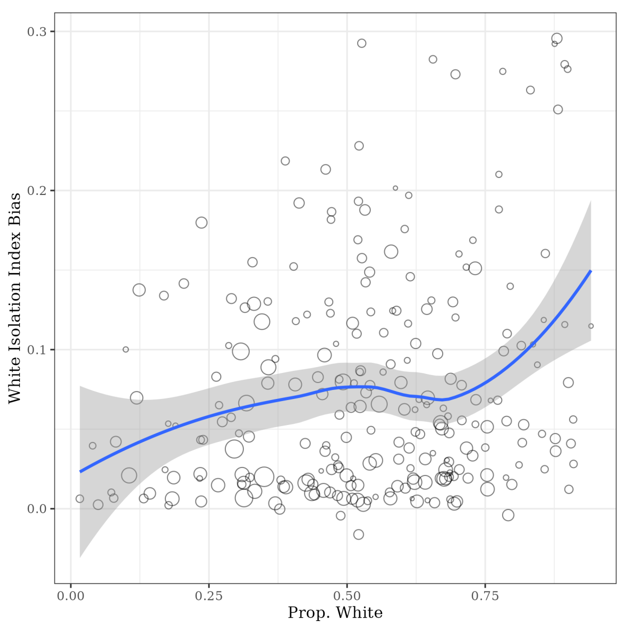
Figure B44: White Isolation Index Bias - 2010



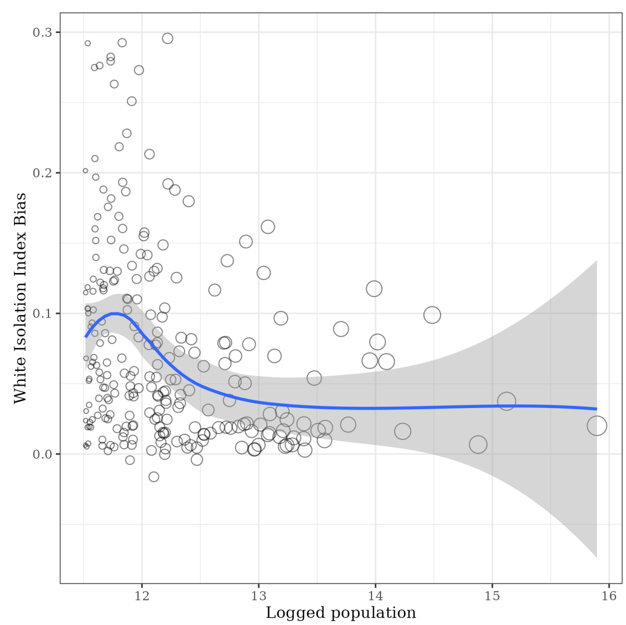
(a) Histogram



(b) By simulation average H index

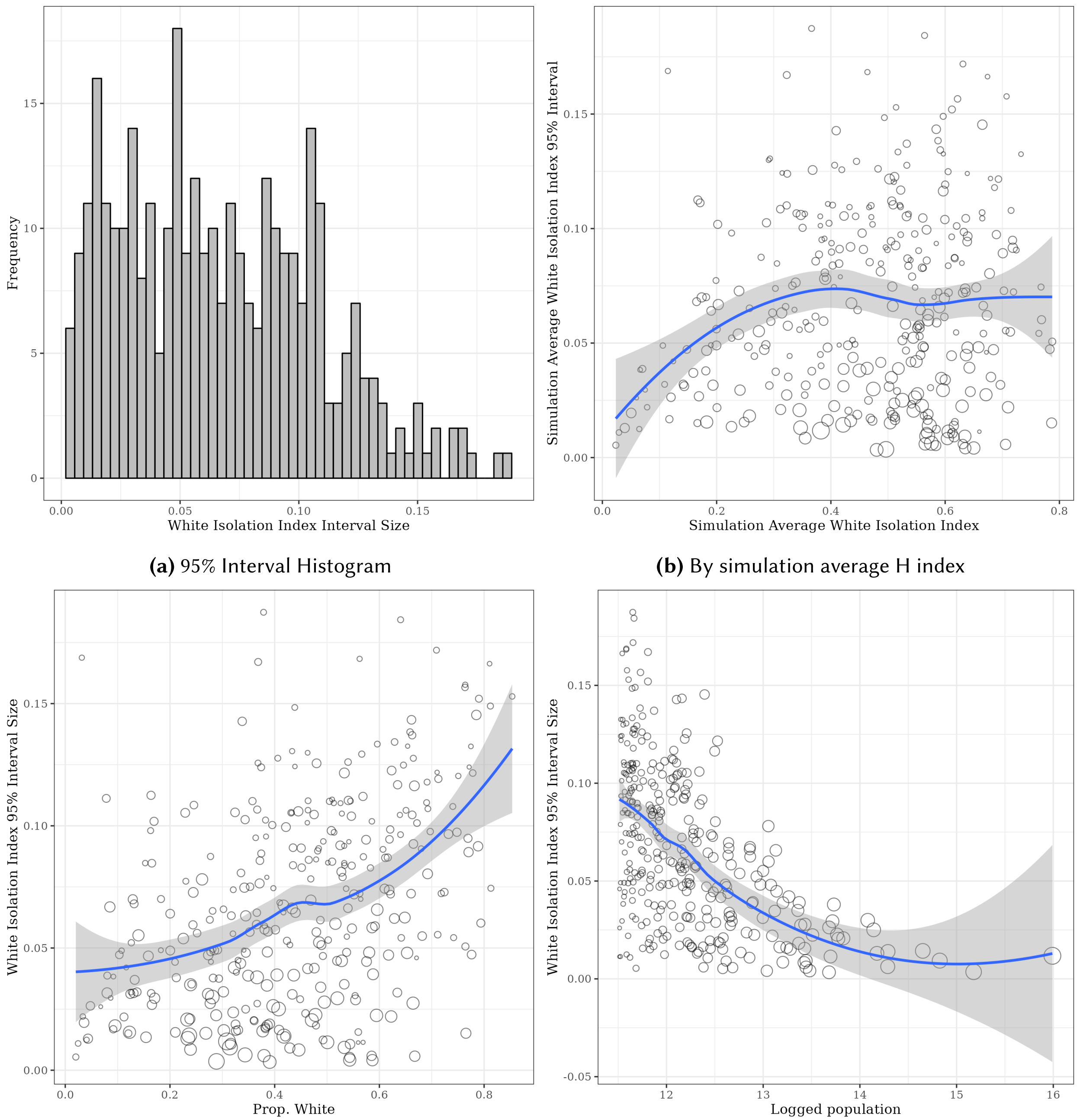


(c) By city proportion White



(d) By logged city population

Figure B45: White Isolation Index Bias - 2000



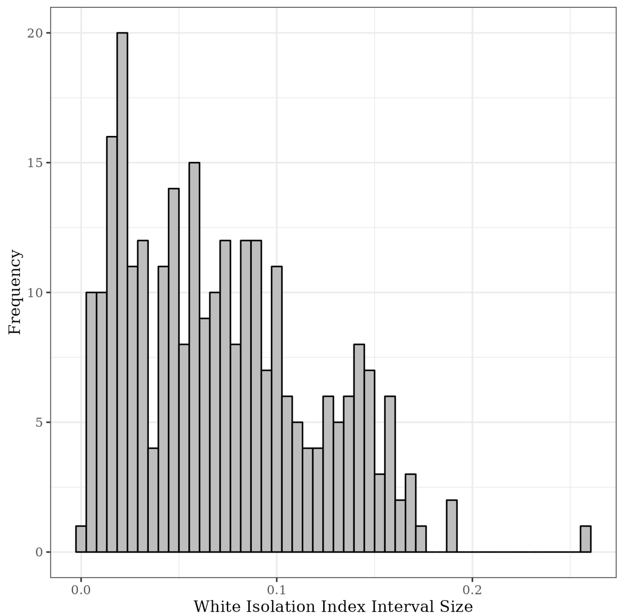
(a) 95% Interval Histogram

(b) By simulation average H index

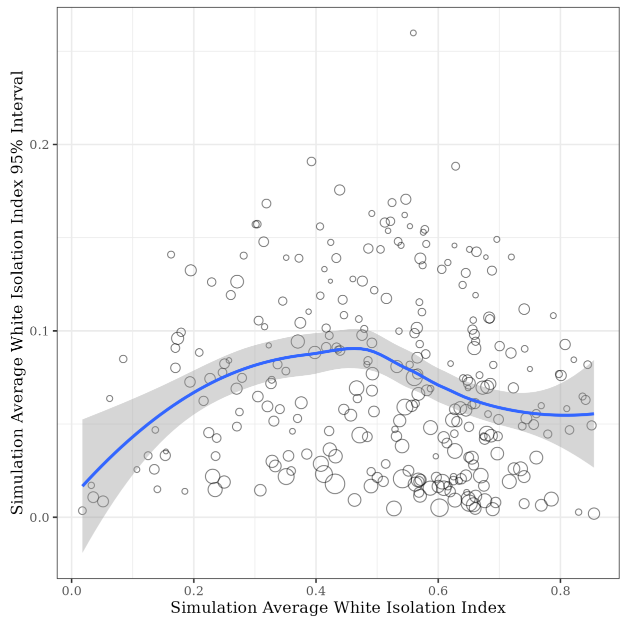
(c) By city Herfindahl Index

(d) By logged city population

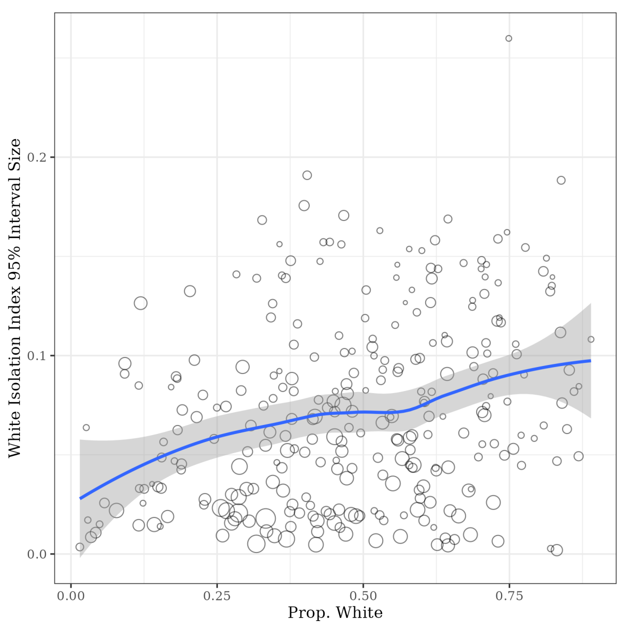
Figure B46: White Isolation Index Uncertainty - 2020



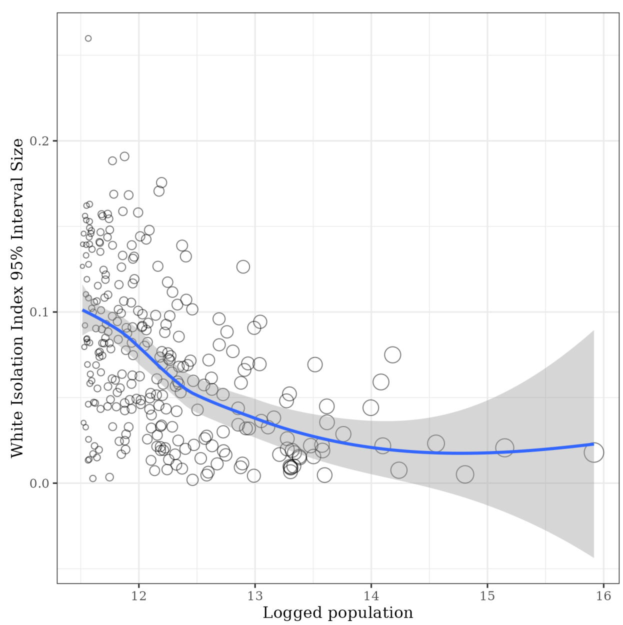
(a) 95% Interval Histogram



(b) By simulation average H index

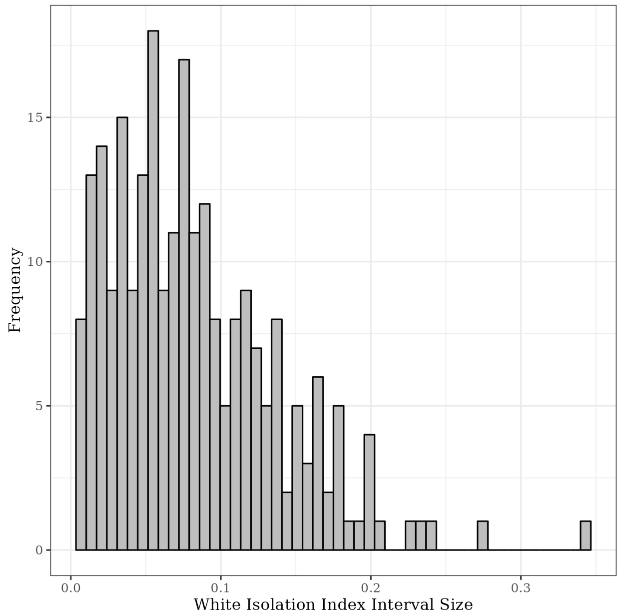


(c) By city Herfindahl Index

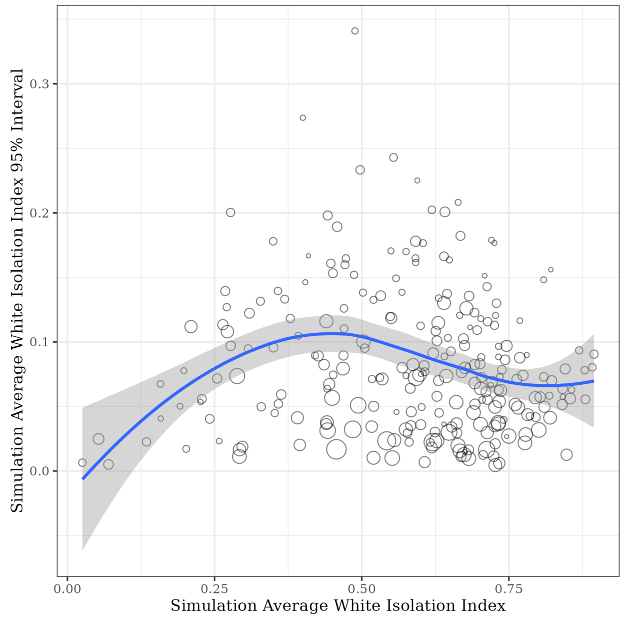


(d) By logged city population

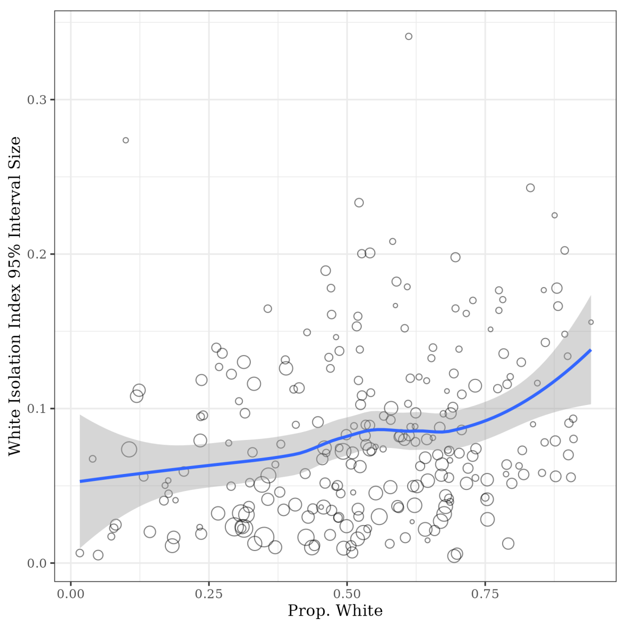
Figure B47: White Isolation Index Uncertainty - 2010



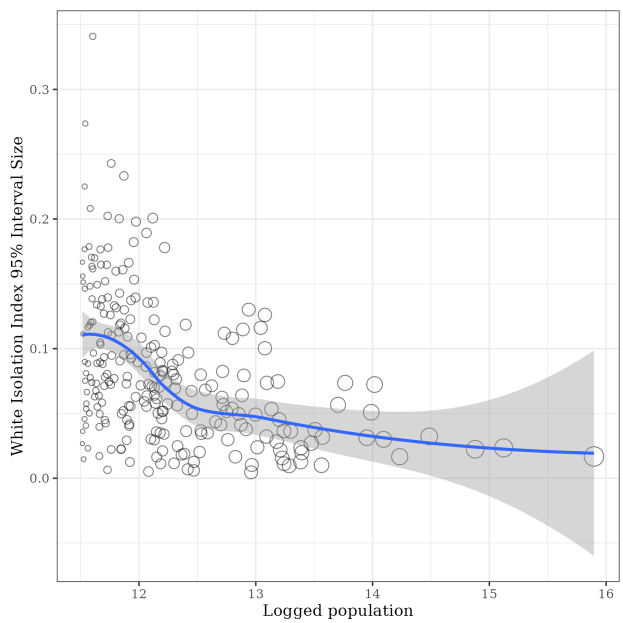
(a) 95% Interval Histogram



(b) By simulation average H index



(c) By city Herfindahl Index



(d) By logged city population

Figure B48: White Isolation Index Uncertainty - 2000

B.8 Hispanic Isolation Index

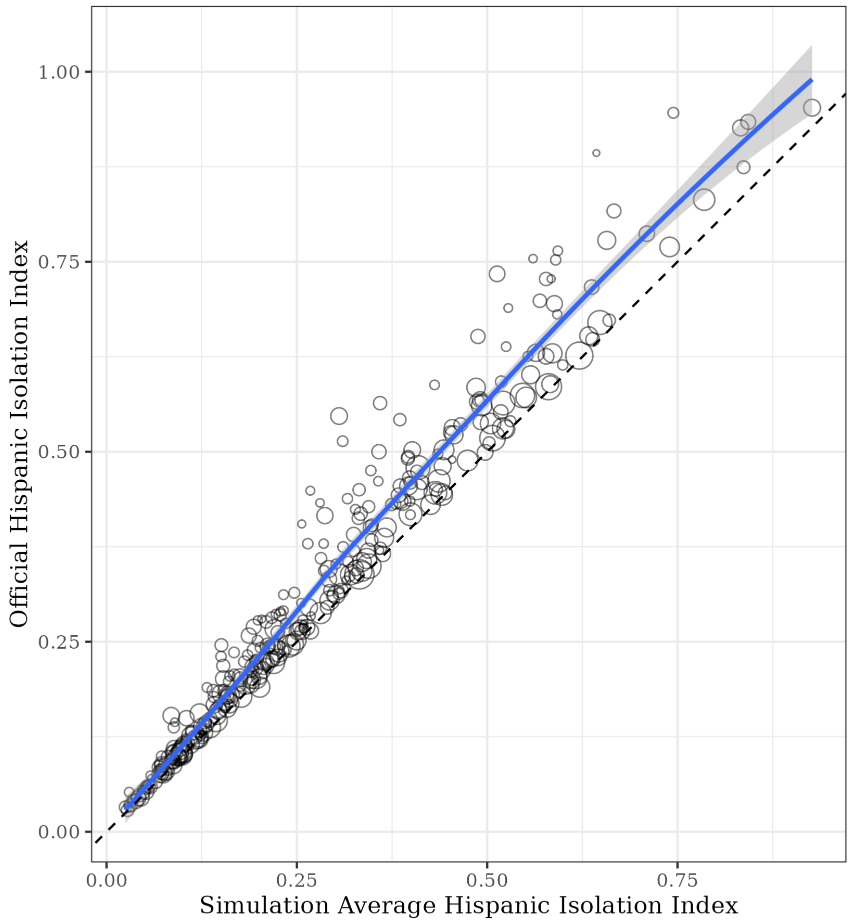


Figure B49: Official Hispanic Isolation Index Index versus Simulated - 2020

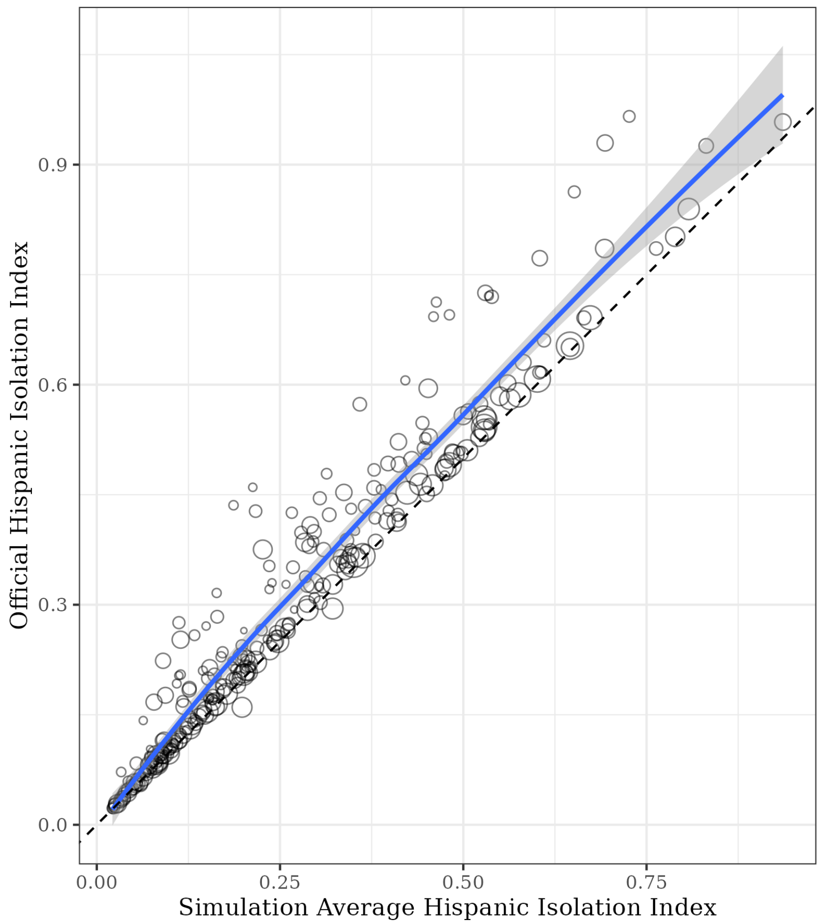


Figure B50: Official Hispanic Isolation Index versus Simulated - 2010

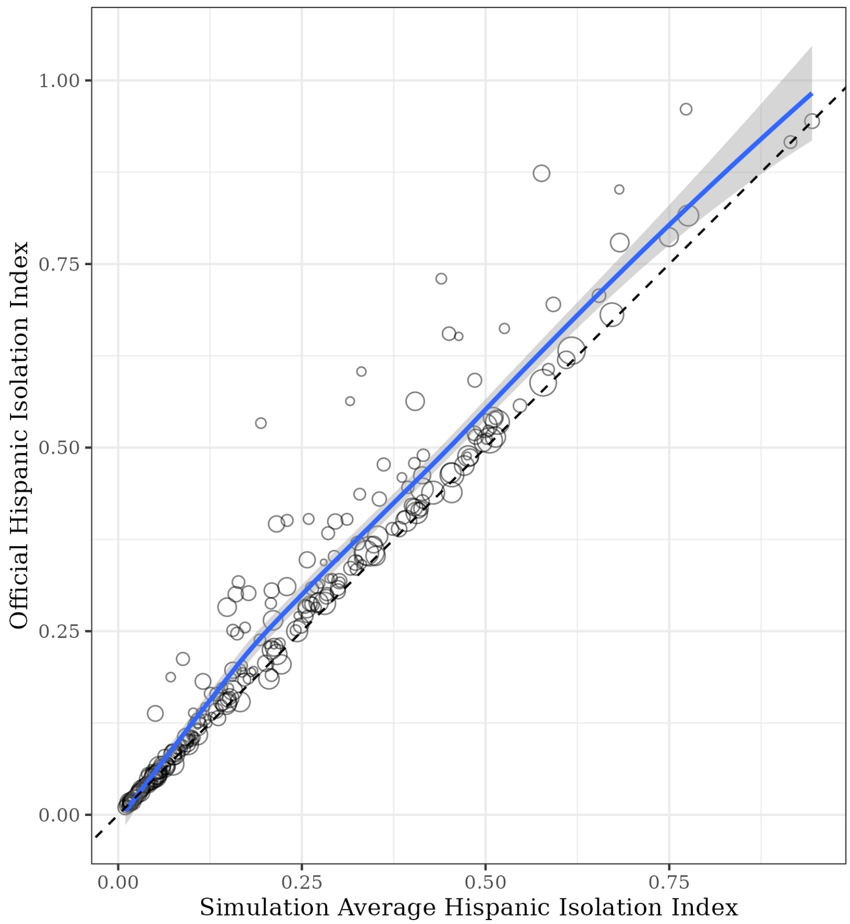
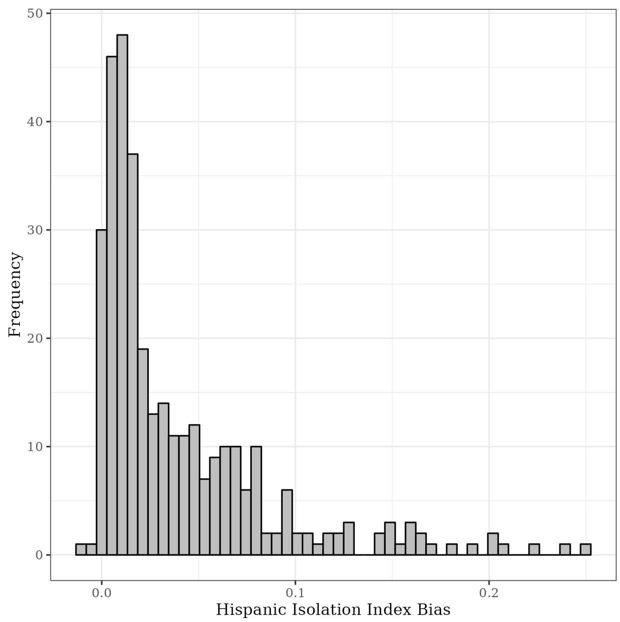
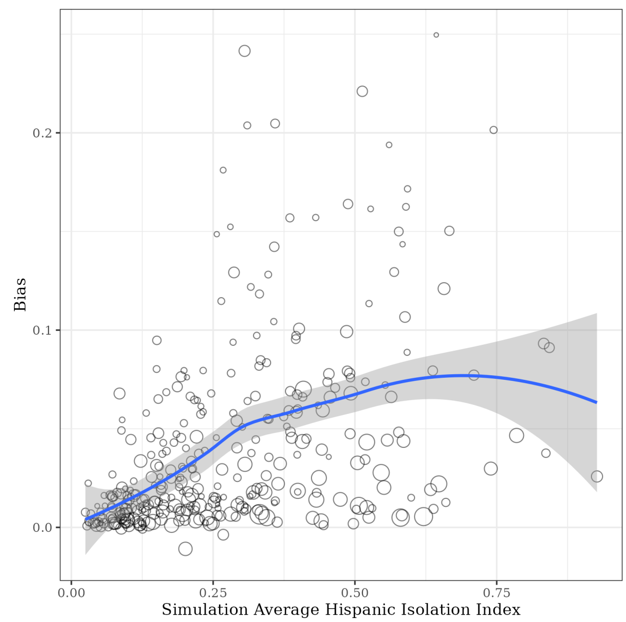


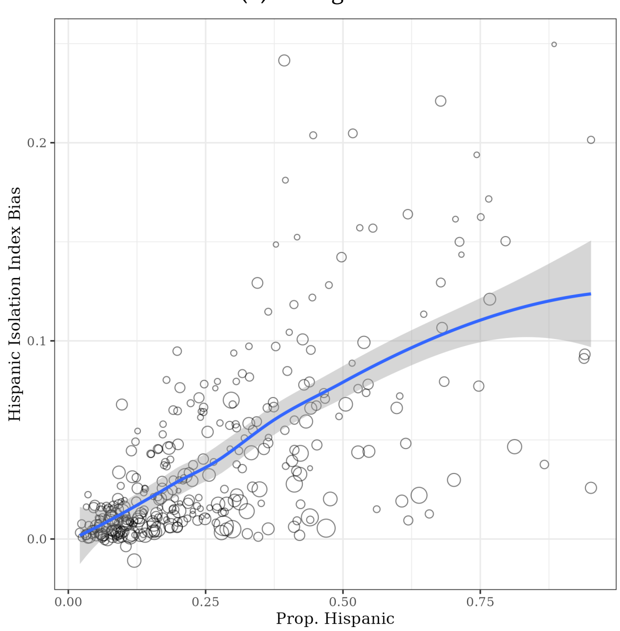
Figure B51: Official Hispanic Isolation Index versus Simulated - 2000



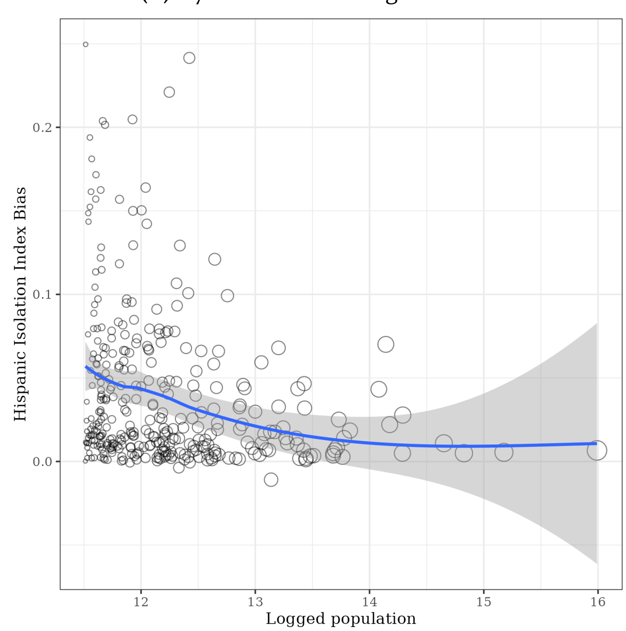
(a) Histogram



(b) By simulation average H index

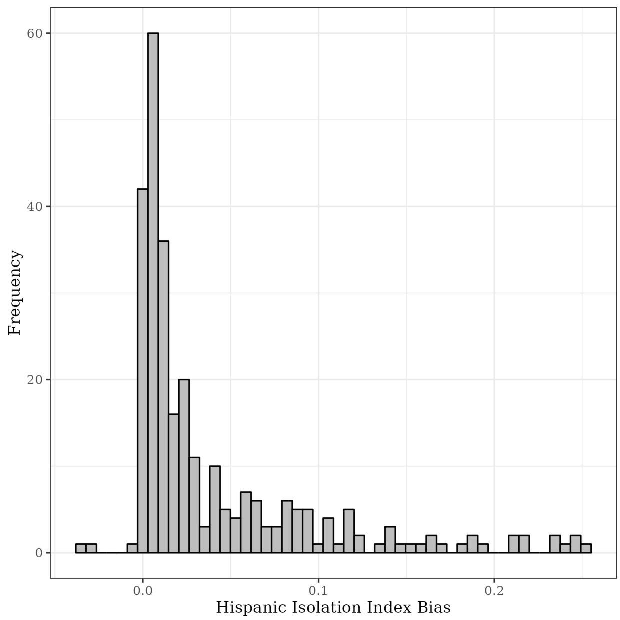


(c) By city proportion Hispanic

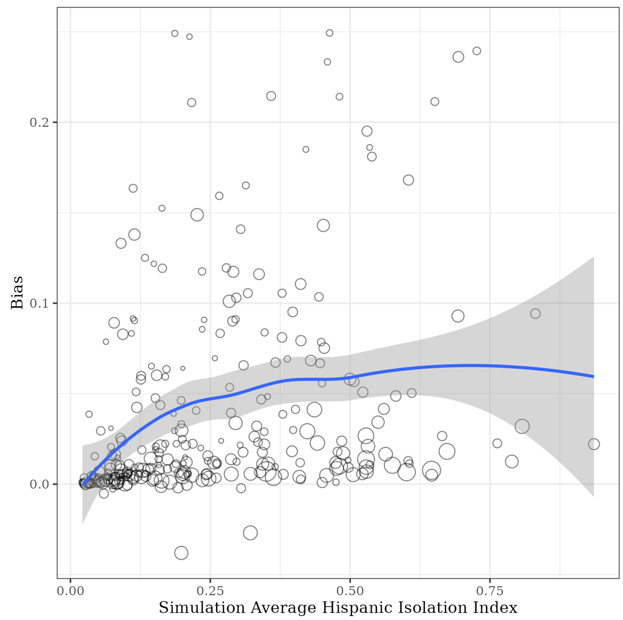


(d) By logged city population

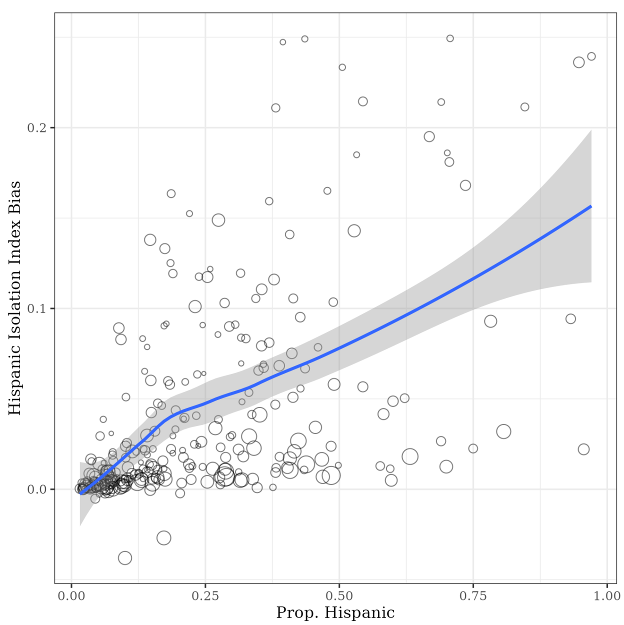
Figure B52: Hispanic Isolation Index - 2020



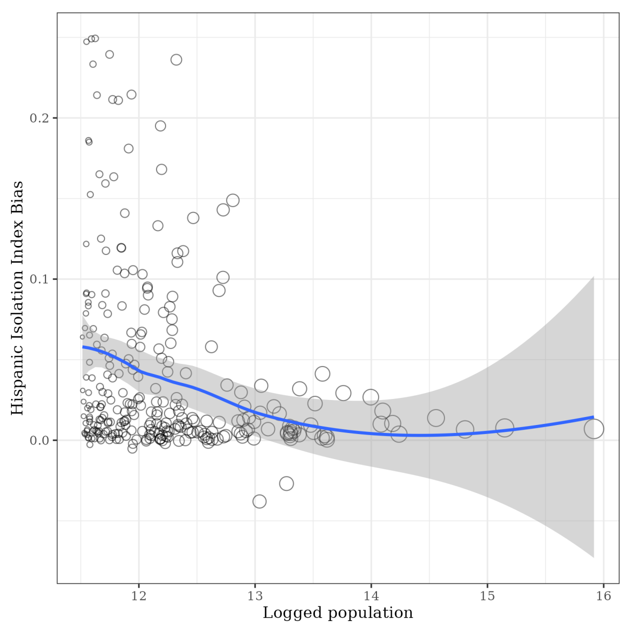
(a) Histogram



(b) By simulation average H index

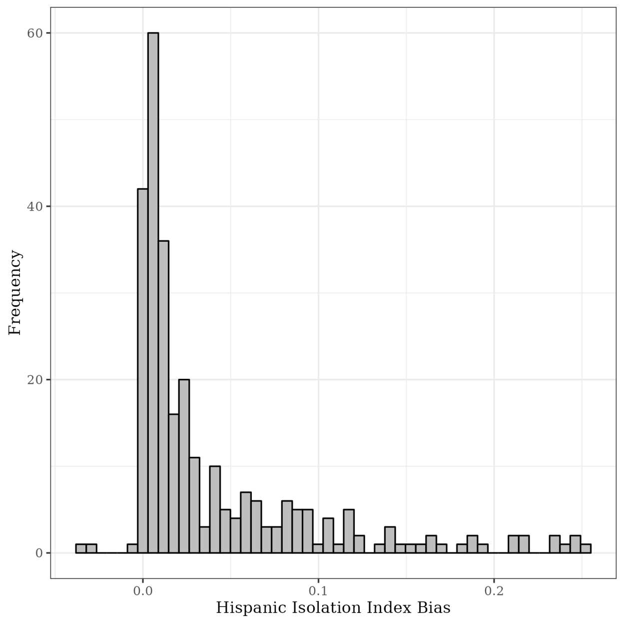


(c) By city proportion Hispanic

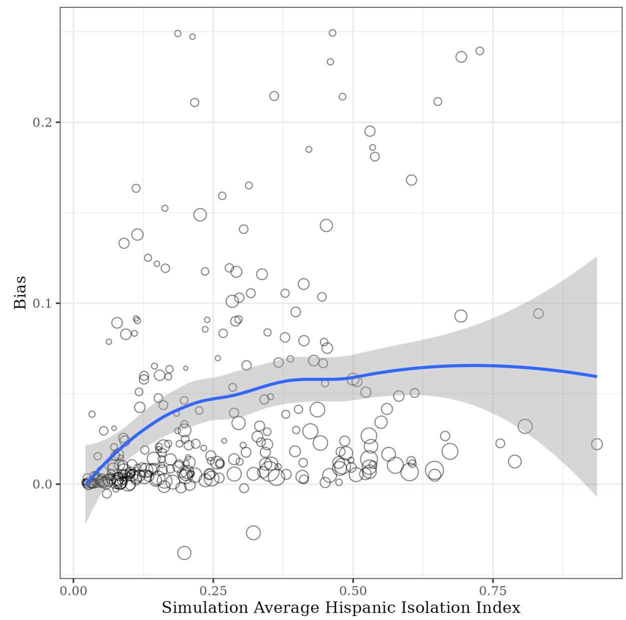


(d) By logged city population

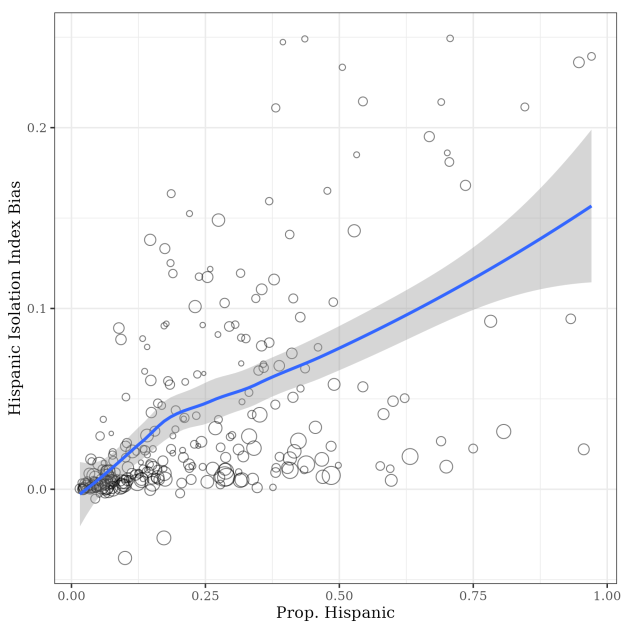
Figure B53: Hispanic Isolation Index Bias - 2010



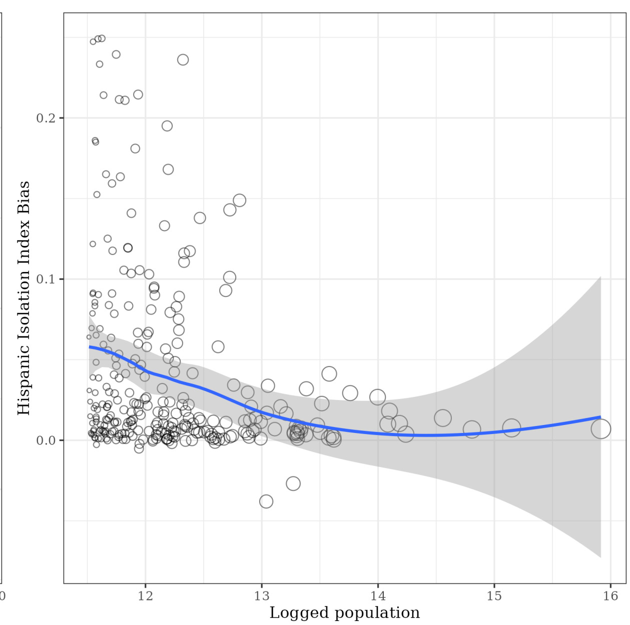
(a) Histogram



(b) By simulation average H index

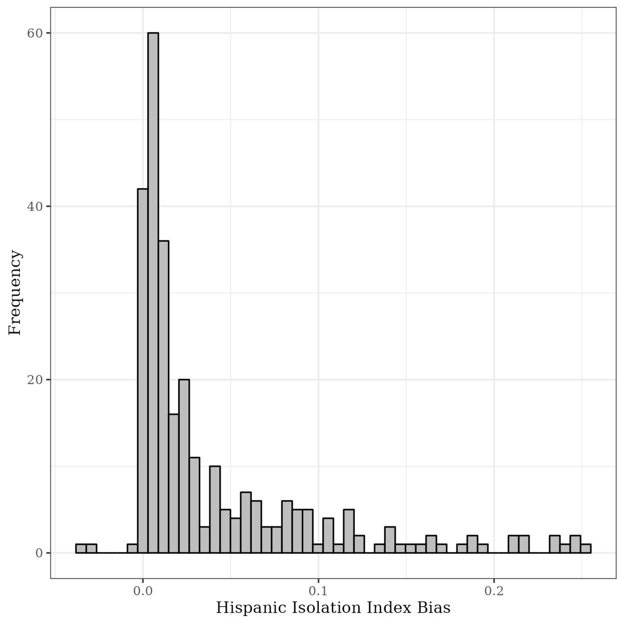


(c) By city proportion Hispanic

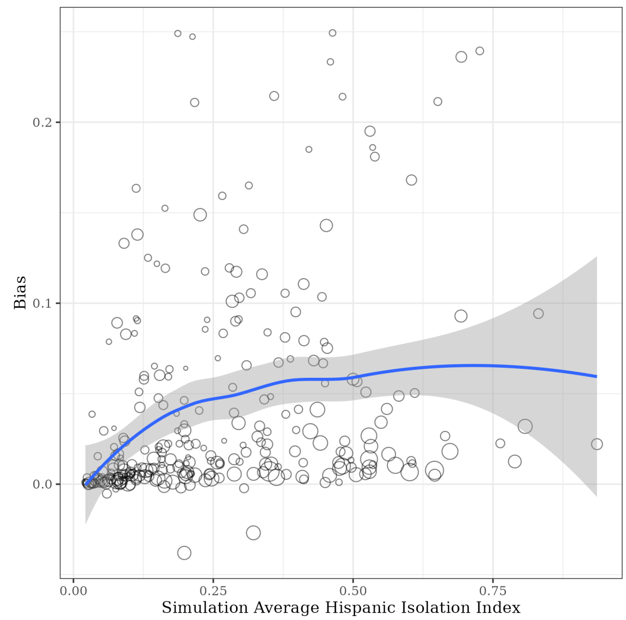


(d) By logged city population

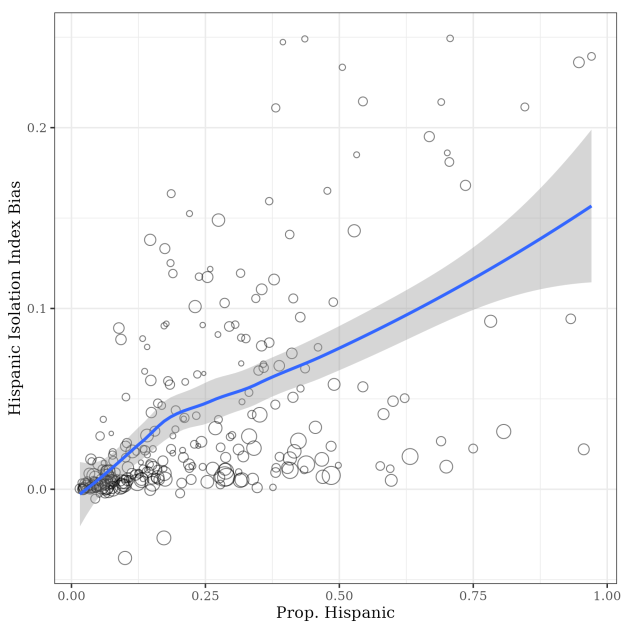
Figure B54: Hispanic Isolation Index Bias - 2010



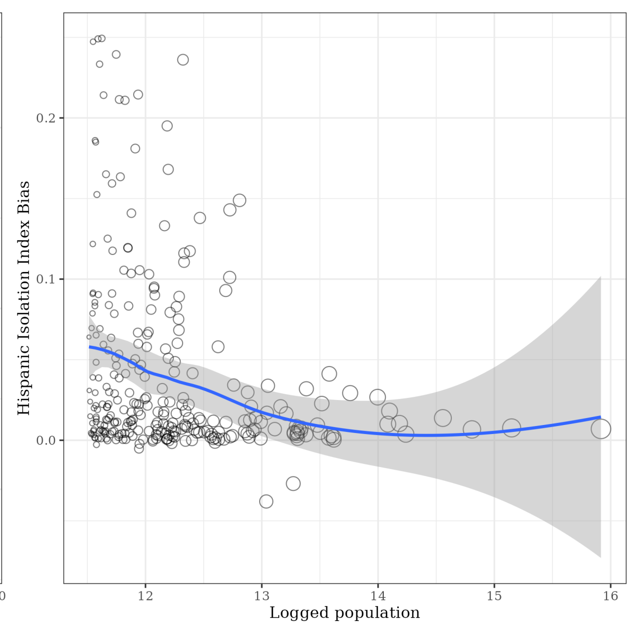
(a) Histogram



(b) By simulation average H index

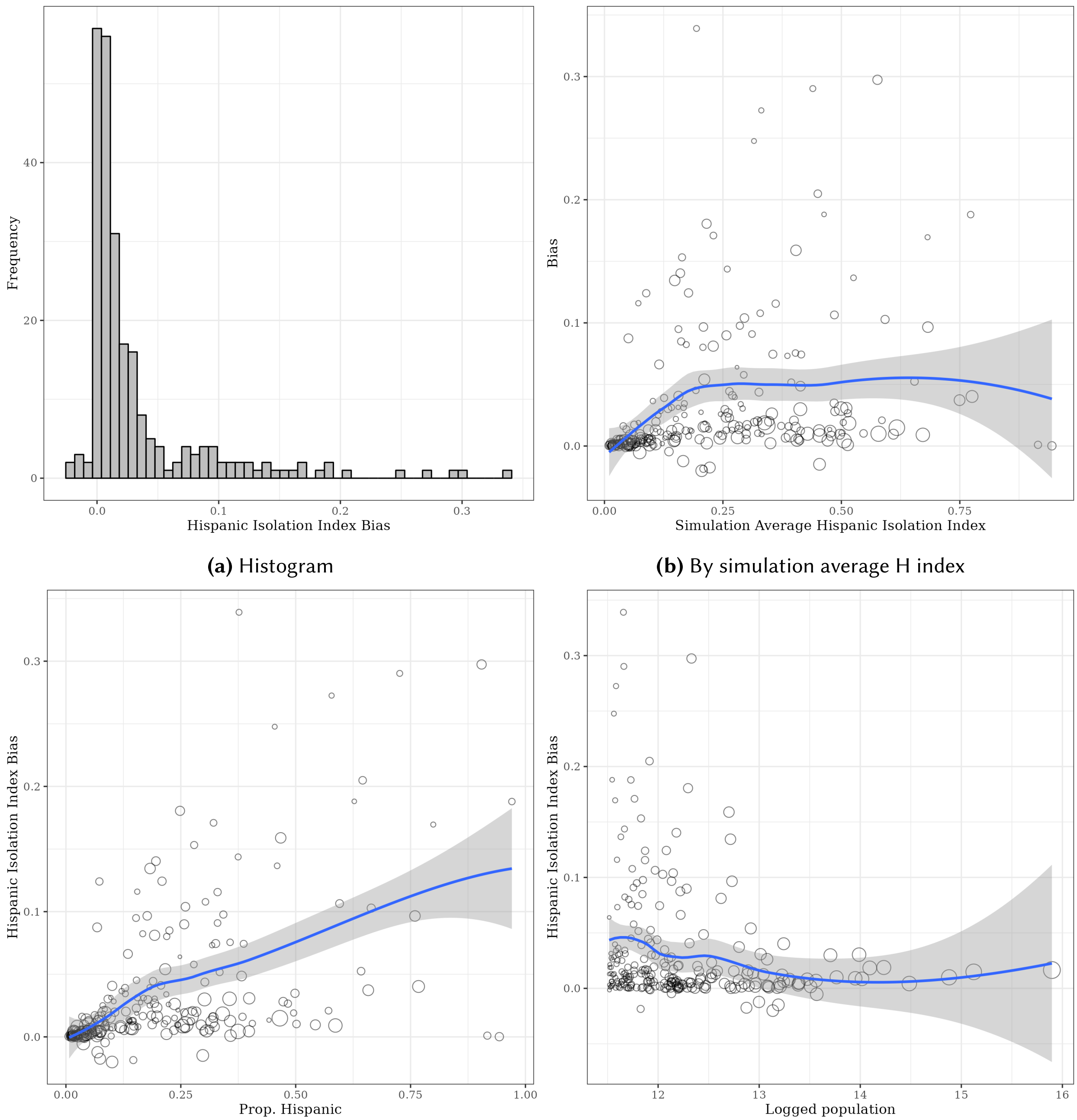


(c) By city proportion Hispanic



(d) By logged city population

Figure B55: Hispanic Isolation Index Bias - 2010



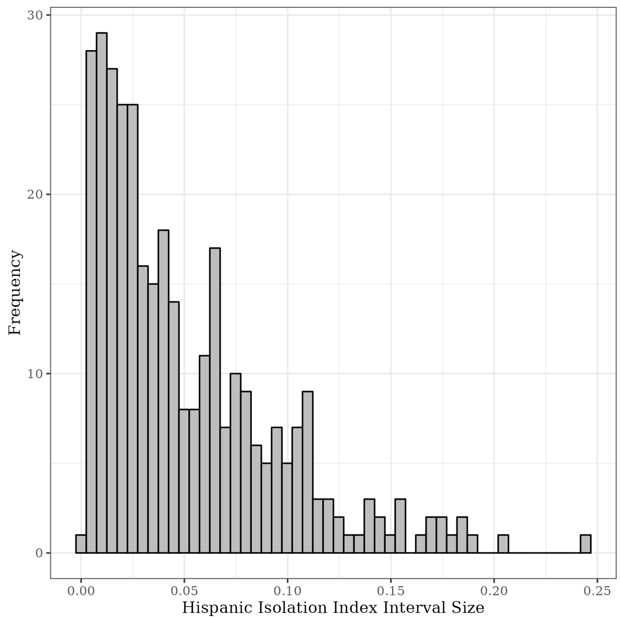
(a) Histogram

(b) By simulation average H index

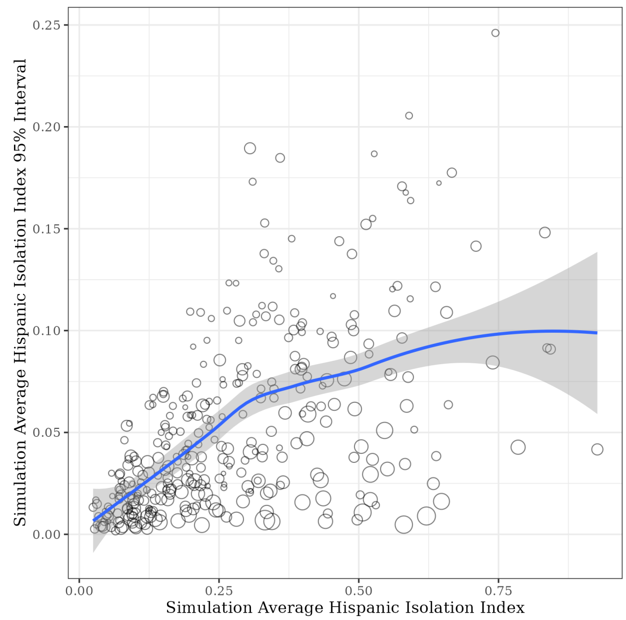
(c) By city proportion Hispanic

(d) By logged city population

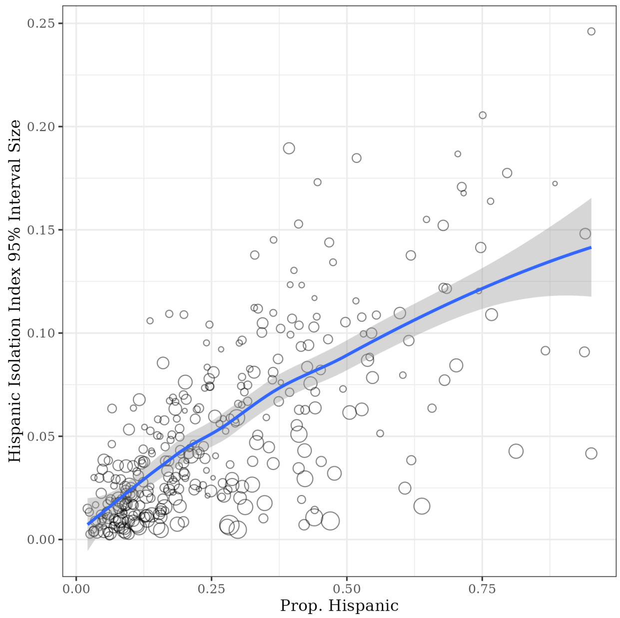
Figure B56: Hispanic Isolation Index Bias - 2000



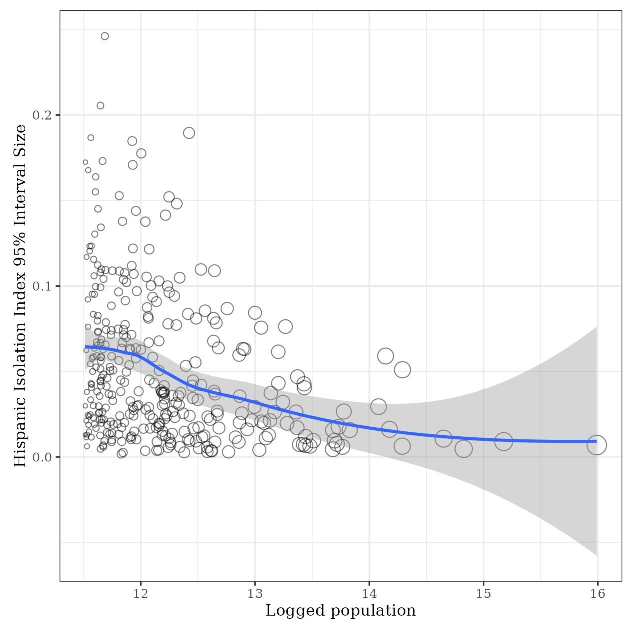
(a) 95% Interval Histogram



(b) By simulation average H index

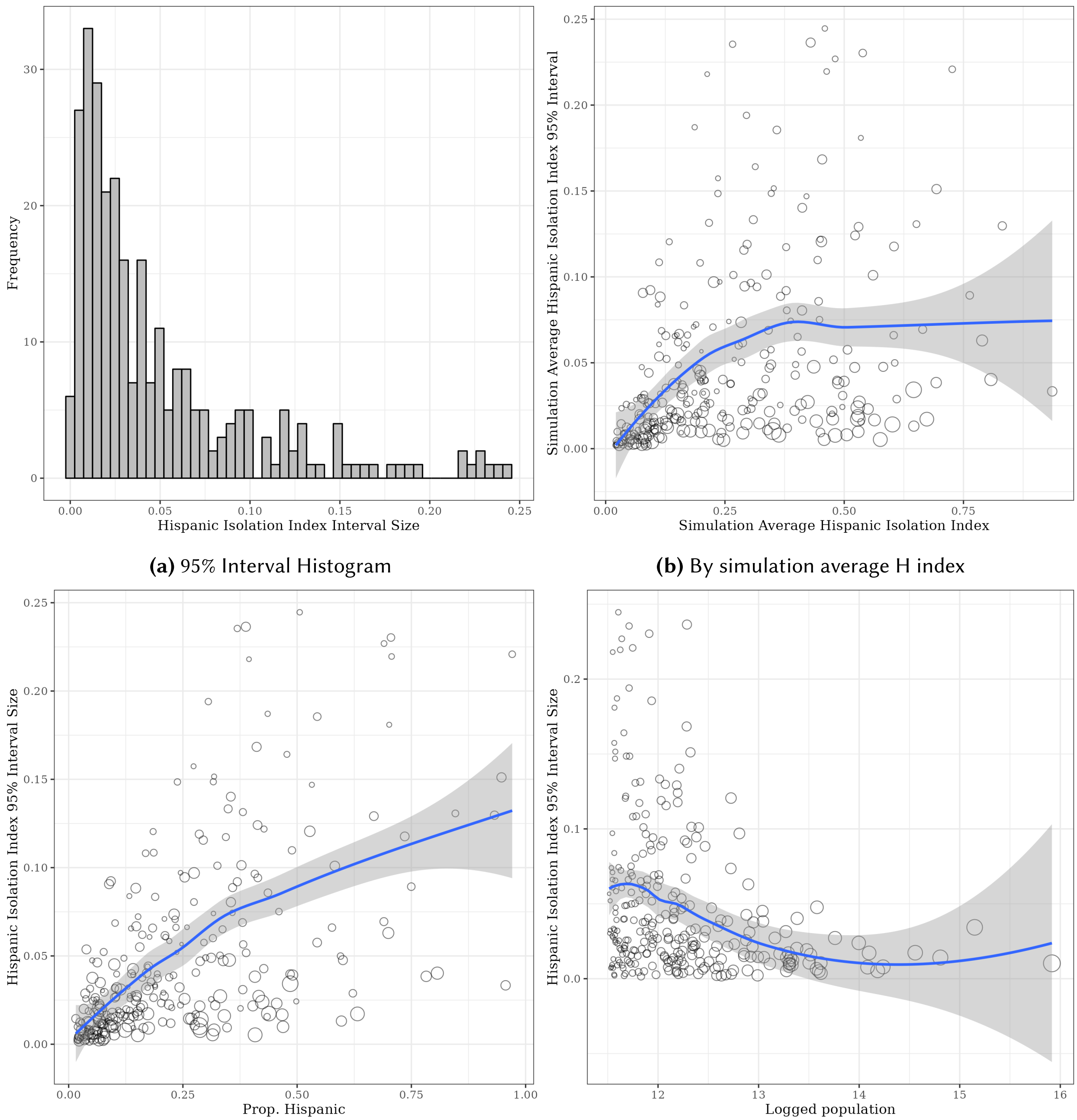


(c) By city Herfindahl Index



(d) By logged city population

Figure B57: Hispanic Isolation Index Uncertainty - 2020



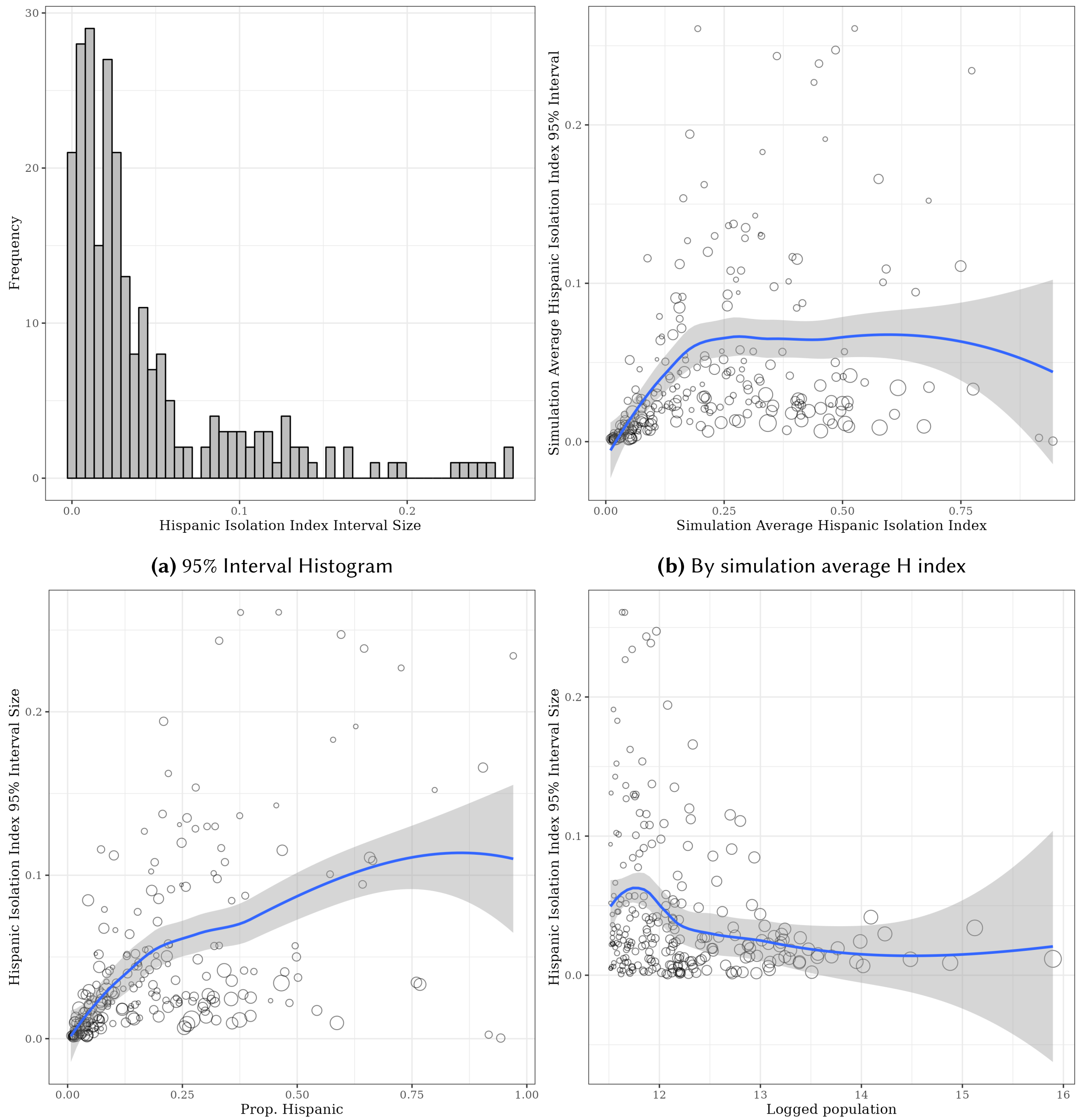
(a) 95% Interval Histogram

(b) By simulation average H index

(c) By city Herfindahl Index

(d) By logged city population

Figure B58: Hispanic Isolation Index Uncertainty - 2010



(a) 95% Interval Histogram

(b) By simulation average H index

(c) By city Herfindahl Index

(d) By logged city population

Figure B59: Hispanic Isolation Index Uncertainty - 2000