Desegregation Paradox? A Model and Simulation of LIHTC's Effects on Economic Segregation

Peter Kirgis and Teddy Knox

January 29, 2025

Abstract

The Low Income Housing Tax Credit (LIHTC) subsidizes below-market rents for millions of low-income Americans, but research on the program's effects on economic segregation is inconclusive. We simulate "LIHTC removal" to create a counterfactual where LIHTC residents are redistributed from their status quo tracts across their home counties. Our simulation suggests that the presence of LIHTC is associated with negligible changes in CBSA-level economic segregation between extremely low-income (ELI) and non-ELI residents. However, this null finding is sufficiently caused by the universally low prevalence of LIHTC. We offer a toy model of how LIHTC siting, the composition of LIHTC developments, and the composition of the general population can lead to counterintuitive effects on economic segregation in settings with sufficiently prevalent LIHTC. When examining the subsample of only tracts that contain LIHTC developments, our simulations find that LIHTC is associated with decreased economic segregation (across the subsamples) in most CBSAs, with a median decrease of 10%. Even though poorer neighborhoods tend to have more and poorer LIHTC developments, these relationships are not strong enough to cause an increase in economic segregation.

1 Introduction

The Low Income Housing Tax Credit (LIHTC) is the nation's largest affordable housing program, enabling the construction of roughly 3.5 million rental units since 1986 [Keightley, 2023]. While proponents laud LIHTC as a model for supply-side housing reform in incentivizing private sector investments in affordable housing, others raise concerns about LIHTC's role in increasing economic segregation. Prior research has provided conflicting evidence for both the direction and magnitude of LIHTC's effect on these outcomes [Ellen et al., 2016, McClure, 2018].

Our analysis contributes to this literature by combining novel data sources and methods to estimate the association between the presence of LIHTC and economic segregation at the CBSA level. We use property-level data on LIHTC developments across the US and census tract-level data on extremely low income (ELI) households to simulate how the distribution of ELI residents would change if LIHTC residents were removed from their home tracts and redistributed across their counties. Our results support three key findings:

1. LIHTC does not have a meaningful effect on economic segregation at the CBSA level,
2. Poorer tracts tend to have more LIHTC units and poorer LIHTC developments, and
3. In areas with high concentration of LIHTC, LIHTC appears to be associated with decreased economic segregation at the CBSA level.

Though (2) and (3) may appear to conflict, we offer a theoretical model in which LIHTC can have a desegregating effect even when LIHTC units are disproportionately sited in poor areas and disproportionately occupied by poor residents, if those two relationships are not strong enough.

Section 2 reviews the mechanics of the LIHTC program and previous literature relating LIHTC to economic segregation. Section 3 introduces a theoretical model illustrating how LIHTC relates to economic segregation under different assumptions about the siting and composition of developments. Section 4 reviews our data sources and simulation method used to compare the status quo to a world without LIHTC. Section 5 presents the key findings of our main analysis. Section 6 discusses these findings in greater detail and relates them to the model presented in Section 3. Section 7 concludes by contextualizing these effects within a broader conversation about place-based policy.

2 Literature Review

LIHTC awards housing developers federal tax credits to offset construction costs on the condition that a fraction of the units prepared with the tax credits are rent-restricted to accommodate lower-income households [Keightley, 2023]. Until 2018, the tax credit was awarded when either:

1. A minimum of 20% of the units in the property are affordable to households earning 50% or below the unit's Area Median Gross Income (AMGI), with "areas" determined by states in their Qualified Allocation Plans (QAPs)
2. *OR* A minimum of 40% of units are affordable to households earning 60% or below of AMGI.

While QAPs are determined by each state, the U.S. Department of Housing and Urban Development (HUD) provides non-binding guidance. Each year, HUD ranks census tracts based on various characteristics to designate tracts that have 50% of households with incomes below 60% of the AMGI or have a poverty rate of 25% or more as "Qualified Census Tracts" (QCTs) [HUD, 2024]. LIHTC sited in the QCTs can be claimed for 130% (instead of the normal 100%) of the project's eligible basis [Keightley, 2023].

After determining which tracts qualify as QCTs, states can use that information to determine their QAPs for the geographic siting of LIHTC. Given the authority that state and local governments have to determine where LIHTC developments can be sited—both because of the design of the QCT designations and the tendency of some localities to reject developments—and given the mechanics of financing developments specifically designed to house low-income residents in a development in a specific community, it is possible that LIHTC may increase economic segregation in some places and decrease it in others.

The literature on LIHTC's effect on economic segregation is mixed, but generally suggests that while LIHTC properties tend to be sited in poorer areas, there is no meaningful relationship between LIHTC and economic segregation, possibly because of the program's flexibility in implementation.

Several existing studies document LIHTC siting patterns. Won [2022] explores the spillover effects of LIHTC units on neighborhood economic status and suggests LIHTC developments, in general, cluster households that have lower-than-average household incomes for their metropolitan statistical area (MSA). Schwartz and McClure [2022] compare communities with no LIHTC to those with, finding that 72% of all municipalities and 52% of all growing municipalities contain no LIHTC. They also find the most important predictors of LIHTC being missing from a municipality are population size, being a suburb in a large metropolitan area, and the percentage of rental and multifamily housing. This systematic bias suggests that LIHTC units are not being built in the higher-value suburbs outside of opportunity-rich metropolitan areas. Basolo et al. [2022] study California to find that LIHTC tends to be sited in neighborhoods with lower rents and a higher percentage of renters; in Los Angeles County specifically, LIHTC is sited in neighborhoods with greater economic hardship, suggesting that state and local implementation choices can impact siting.

Though LIHTC tends to be sited more in disadvantaged areas, other analyses find that it may not be adding to economic segregation. By assessing the effects of siting and tenant composition, Ellen et al. [2016] find "little evidence that the LIHTC is increasing the concentration of poverty – and [even] some evidence that it is reducing poverty rates in high-poverty neighborhoods." Freedman and McGavock [2015] find that while housing investment under LIHTC has measurable effects on the distribution of income within and across communities, there is little evidence the program contributes meaningfully to poverty concentration or residential segregation. Diamond and McQuade [2019] find that LIHTC developments in higher-income areas cause housing price declines and attract lower-income households, suggesting the potential for LIHTC to contribute to decreased economic segregation.

There is a caveat, however, to the efforts to place LIHTC developments in higher-opportunity neighborhoods: Cook et al. [2023] find that tenants of such LIHTC units are typically higher income, more educated, and far less likely to be Black, which they say is "primarily due to 'crowding out': households that only apply for assistance in higher-opportunity neighborhoods crowd out those willing to apply regardless of location."

3 Theoretical Model

Mechanically, the relationship between the presence of LIHTC and economic segregation depends on the following variables:

1. The share of LIHTC residents who are ELI, compared to the share of non-LIHTC residents who are ELI – if LIHTC developments have high concentrations of ELI residents

relative to the general population, LIHTC has a greater potential to be associated with substantial changes in segregation

2. The share of residents who are ELI and live in LIHTC units, and how this share is correlated with tract-level ELI% – the more that poorer tracts have greater shares of LIHTC×ELI individuals, the more that LIHTC is associated with increased economic segregation

3. The share of all residents that live in LIHTC developments – if this is small, the status quo vs. the counterfactual with LIHTC removal will be negligibly different

To illustrate how these variables fit together, we can set up a toy model with two tracts, one rich and one poor, and examine how dissimilarity differs between the world with LIHTC and the counterfactual with LIHTC removed. For this exercise, we hold all variables constant except for the tract population share that is ELI and living in LIHTC. Table 1 illustrates the mechanics of how variations in this value (X) can lead to changes in the dissimilarity metric.

Table 1: Toy Model Specifications

Tract	**Rich**	**Poor**
ELI%	10%	25%
ELI% in LIHTC properties	50%	50%
% ELI and LIHTC	5%	X%
% of residents in LIHTC	5% / 50% = 10%	$X\%/50\% = 2X\%$
ELI% with LIHTC removed	$\frac{10-5}{100-10} \approx 6\%$	$\frac{25-X}{100-2X}$

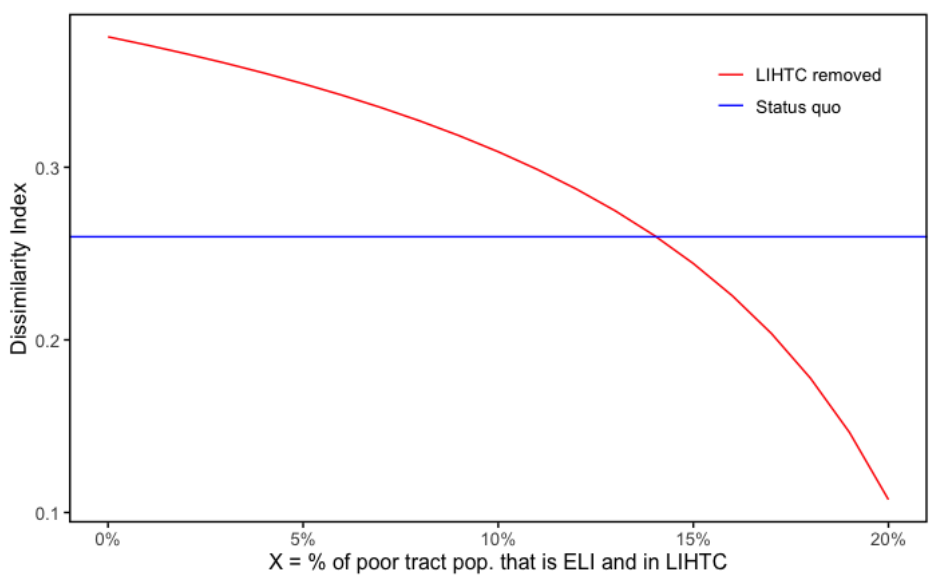

Figure 1: Dissimilarity of Two Tracts by ELI x LIHTC % in Poor Tract.

Figure 1 visualizes the outputs of the toy model under these assumptions. In the status quo (which includes LIHTC), this two-tract CBSA has a dissimilarity of 0.26. Under simulated LIHTC removal, dissimilarity varies based on X, the assumed share of the poor tract's population that is ELI and in LIHTC. When $X < 14\%$, the presence of LIHTC is associated with a decrease in economic segregation (status quo dissimilarity is lower than counterfactual); when $X > 14\%$, LIHTC is associated with increased segregation. Note the window, $5\% < X < 14\%$, where the poor tract has more ELI×LIHTC residents, but the presence of LIHTC is still associated with a decrease in segregation.

4 Data and Methods

We use a decomposition approach to compare current economic segregation against a simulated counterfactual without LIHTC. Our approach requires information on (1) the total population and (2) a low income population, within and outside of LIHTC developments, at a low level of US geography with full geographic coverage. We use census tracts as the low-level unit and study tract-level segregation across CBSAs. We operationalize poverty using the metric of extremely low income (ELI), defined as a household earning less than 30% of the area median household income (AMHI).

Data on LIHTC siting and composition comes from three sources: the National Housing Preservation Database (NHPD), HUD's LIHTC Property Database, and HUD's LIHTC Tenant Database; these yield a sample of 37,875 LIHTC developments comprising 2.9 million units. We match this property file to HUD's LIHTC Tenant Data file which has property-level household income breakdowns, including the ELI%. We use an iterative matching process, first looking for matches in 2022, then 2019, and then 2016 (the last three years tenant-level data are available), yielding 16,655 developments with ELI data, 44% of our total sample. (See Appendix B for details). We then use a random forest model to impute missing ELI proportion values for the other $\sim 56\%$ of our sample. Data on total and ELI households by tract come from the CHAS, which consolidates multiple ACS 2017–2021 5-year surveys.

We validate the full data processing methodology by comparing our aggregated and imputed sample against state-level tenant data released by HUD. Thirteen states either do not report or do not have sufficient data to be included in HUD's report, but of the other 37 states in our sample, the mean absolute error (MAE) is 4.5% (See Appendix B). The outliers suggest that there are some states where our post-processing is biased, but for most states the two sources of data are very closely aligned.

Given data on ELI and non-ELI households living in LIHTC developments and the census tracts where they are sited, we can perform our simulation. First, we simulate "LIHTC removal" by subtracting all ELI and non-ELI households living in LIHTC properties from the totals for each census tract. Next, we aggregate the total ELI and non-ELI households removed at the county level and redistribute them to new census tracts proportionally based on the distribution of ELI and non-ELI households across each county. Finally, we calculate the economic segregation in the simulated counterfactual. We define economic segregation at the CBSA level using the dissimilarity index with the following specification:

$$D = \frac{1}{2} \sum_{i=1}^{n} \left| \frac{E_i}{E} - \frac{N_i}{N} \right|,$$

where E_i is the number of ELI households in tract i, E is the total number of ELI households in the CBSA, N_i is the number of non-ELI households in tract i, and N is the total number of non-ELI households in the CBSA.

The dissimilarity index ranges from 0 to 1 and can be interpreted as the fraction of ELI households that would need to be swapped with non-ELI households in other census tracts to achieve an equal distribution of ELI households across the CBSA. Our target variable of interest is the difference in the dissimilarity index for each CBSA for both our real world and counterfactual without LIHTC, which we call the "dissimilarity difference":

$$\Delta D = D_{\text{With LIHTC}} - D_{\text{Without LIHTC}}.$$

A negative value for the change in the dissimilarity index should be interpreted as saying that for a given CBSA, the presence of the LIHTC program is associated with a decrease in economic segregation.

5 Results

5.1 Full CBSA Analysis

At the CBSA level, the presence of LIHTC is not associated with a change in economic segregation. We simulate the effects of LIHTC removal on economic segregation between extremely low income ("ELI") and non-ELI populations. We find that the average dissimilarity difference is -0.003, with a standard deviation of 0.015, suggesting that the average change in economic segregation associated with the presence of LIHTC is essentially zero, and that the majority of CBSAs have very small changes in economic segregation under the presence of LIHTC (see Figure 2 for the full distribution). Only 12% of CBSAs have an absolute dissimilarity difference greater than .02, and as the CBSAs with larger changes tend to be smaller in population, these 12% represent only 2% of households in our sample. Among the top five CBSAs (by number of LIHTC units), the greatest absolute dissimilarity difference is -0.007 (see Table 2).

Critically, the dissimilarity difference within a CBSA is constrained by the share of its residents that live in LIHTC developments. Across the CBSAs in our sample, only 1.5% of units are in LIHTC developments, and only 4% of CBSAs have greater than 3% of their households in LIHTC. Furthermore, we find that CBSAs with a greater LIHTC share tend to see slightly greater (negative) dissimilarity differences from the presence of LIHTC (see Figure 5 and Table 3). Thus, the null finding here does not indicate that LIHTC could not have more substantial effects on economic segregation if the program existed on a greater scale.

In analyzing the composition of LIHTC developments, we observe two trends that relate to our theoretical model. First, we observe that tracts with LIHTC developments have higher ELI%s on average. LIHTC developments tend to be sited in tracts with a greater

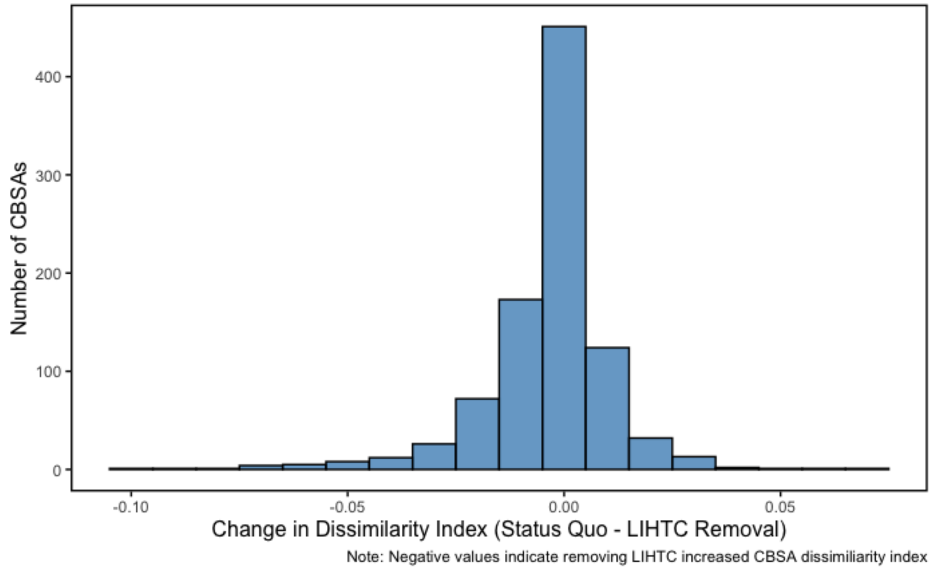

Figure 2: Change in CBSA Dissimilarity: Status Quo vs. LIHTC Removal.

share of ELI residents. Among all tracts in the US without any LIHTC developments, the mean (median) ELI% is 13% (10%); among all tracts with at least one LIHTC development, it is 20% (17%), a 7-point (7-point) difference. Although the two distributions overlap, LIHTC tracts have a substantially thicker right rail; they are almost 3 times as likely (27% vs. 10%) to have more than a quarter of their population be ELI. See Figure 10 for the full distributions.

Second, we observe that LIHTC developments have very large within-property ELI%s on average. LIHTC developments are home to significantly larger concentrations of ELI residents than the US at large. 13% of all US households are ELI, whereas 55% of LIHTC residents are ELI and the median LIHTC property is 60% ELI. Furthermore, ELI% is fairly tightly distributed across LIHTC units; the standard deviation of ELI% among LIHTC developments is 21 percentage points, and 47% of developments have ELI%s between 50% and 70%. See Figure 7 for the full distributions.

5.2 LIHTC Tract Analysis

In most CBSAs, we find no association between the presence of LIHTC and a change in economic segregation; however, because most CBSAs have a very small share of residents in LIHTC, this observation cannot rule out that LIHTC may have an effect on economic segregation in areas with a larger concentration of LIHTC developments.

The trends we observe in LIHTC composition and siting, however, suggest that in areas with a large concentration of LIHTC developments, LIHTC may have a larger relationship with economic segregation. To explore that dynamic, it is helpful to examine some CBSA subsets that have greater concentrations of LIHTC residents. Thus, we study the set of residents living in LIHTC tracts within each CBSA.

As before, we simulate the removal of LIHTC residents from their tracts of residence and their redistribution across their counties. When considering economic segregation among only tracts that contain LIHTC developments, most CBSAs find that in the real world (with LIHTC), dissimilarity is substantially *higher* than in the simulation (without LIHTC). In other words, if the set of LIHTC tracts within a CBSA were considered to be their own community, within that community the LIHTC is typically associated with *lower* economic segregation.

Across CBSAs, the average dissimilarity difference is -0.031, with a standard deviation of 0.045 (see Figure 3). Considering that the average dissimilarity index across CBSAs is 0.24, differences of this magnitude are substantial; the median CBSA sees a 10% decrease in dissimilarity associated with LIHTC; the top five CBSAs are no exceptions (see Table 4).

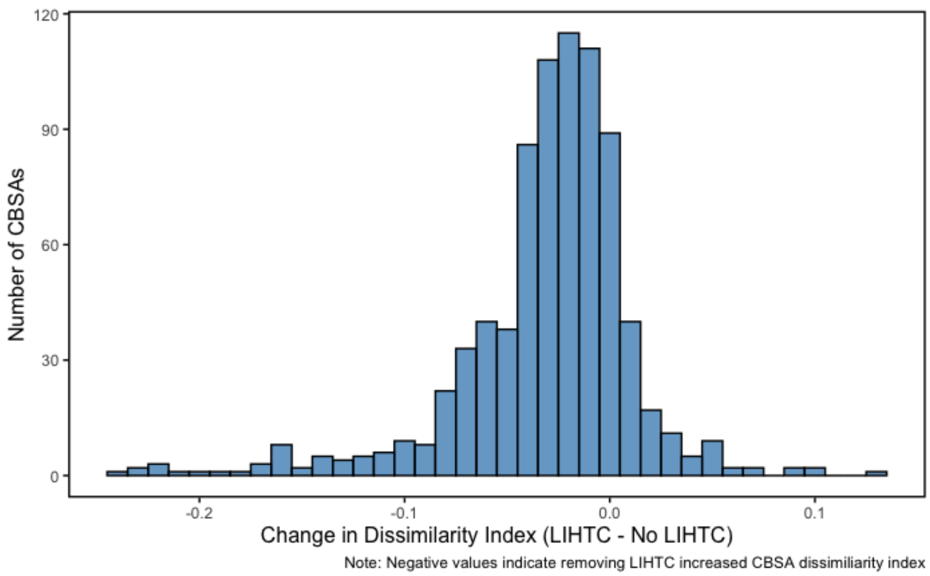

Figure 3: Change in CBSA Dissimilarity (LIHTC Tracts Only).

Of course, these subsamples are not meaningful communities in the same sense that CBSAs are, so these findings do not directly imply that a CBSA with a large LIHTC presence would have substantially decreased economic segregation. However, this analysis can help illustrate how LIHTC could decrease economic segregation in a hypothetical community with a large LIHTC%.

We find that on average, among tracts with LIHTC developments, there is a positive relationship between the tract's overall ELI % (removing LIHTC residents) and its share of residents that are ELI and LIHTC. The association is statistically significant but substantively very small: a 1pp increase in tract ELI% is associated with a 0.09pp increase in the share of the tract that is ELI and LIHTC. Recall that this relationship is critical to the dissimilarity difference outcome in the toy model. The relationship picks up superlinearly among tracts that are greater than 75% ELI, but there are very few of these (see Figure 4).

This phenomenon can be decomposed into two contributing characteristics of poorer tracts: they have slightly more LIHTC units (see Figure 8) and their LIHTC units tend to have slightly greater concentrations of ELI residents (see Figure 9).

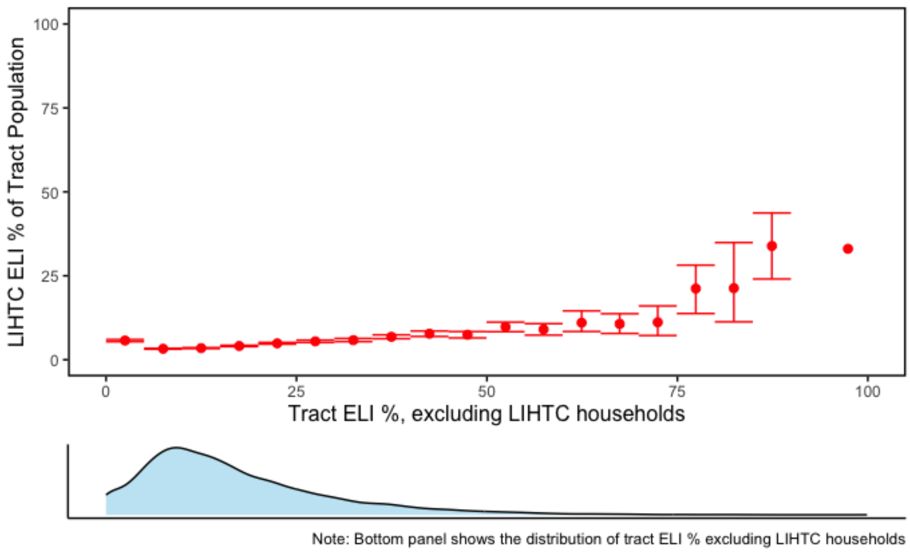

Figure 4: LIHTC & ELI % of Tract Population by Tract ELI %.

6 Discussion

For the vast majority of CBSAs, LIHTC is associated with negligible change in economic segregation. This result is sufficiently but not necessarily explained by the very low prevalence of LIHTC residency in almost all cities. However, our analysis suggests that it LIHTC likely to have a significant relationship with economic segregation in areas where it is more highly concentrated: LIHTC developments tend to be sited in tracts that have larger ELI populations, and they tend to contain much greater concentrations of ELI residents than the general population does. Recall from our toy model that these conditions, combined with a sufficiently large concentration of LIHTC, can lead to a substantial relationship between LIHTC and economic segregation.

In most CBSAs, when examining only LIHTC tracts, the presence of LIHTC is associated with substantial decreases in economic segregation, with a median decrease of 10% (vs. the LIHTC removal counterfactual). This investigation provides an empirical illustration of the toy model's implications: higher-ELI tracts tend to have greater shares of ELI$\times$LIHTC residents, suggesting that most CBSAs are in the region that corresponds to $X > 5\%$ in the toy model, but likely only marginally right of 5%, still well within the region to the left of $X = 14\%$, where LIHTC is associated with decreased segregation relative to LIHTC removal.

There may be some concern that examining a subsample that only includes LIHTC tracts ignores the effects of LIHTC siting. But this is partially addressed by the fact that there is great variation in LIHTC% among LIHTC tracts; in other words, tracts with only a few small LIHTC developments are not so different from tracts with no LIHTC developments, and in that sense we do capture a great deal of siting variance within the sample.

Neither the full-sample analysis nor the LIHTC tracts-only analysis represent causal estimates of the impact of the LIHTC program. There are many factors external to our analysis that would likely make real-world LIHTC removal look different from our simulations. Rather, the full-sample analysis provides evidence that LIHTC does not exist at a large enough scale to meaningfully impact economic segregation at the CBSA level, and the LIHTC tracts-only analysis illustrates how patterns of LIHTC siting and composition might produce desegregating effects in areas where LIHTC is highly prevalent.

This analysis is also limited in scope by its inability to account for spillovers and other second-order effects of LIHTC siting on neighborhoods. Our county-level redistribution allows us to assume some degree of crowd-out, but a more local model of migration would be superior. Similarly, we unrealistically assume there are no implications of LIHTC removal on immediately neighboring census tracts without LIHTC. Nevertheless, our analysis represents a novel contribution to the literature on first-order effects, both in completing a full nationwide decomposition of a low income ELI population in LIHTC and proposing a theoretical model for LIHTC's desegregating effect.

7 Conclusion

One of the most profound dilemmas in social policy is the conflict between investments in people and investments in places. In many cases, a policymaker must accept at least one of two drawbacks: either a policy disincentivizes spatial mobility by only investing in "poor places" or it promotes a cycle of disinvestment by only investing in select poor people. Over the years, changes to LIHTC's program design reflect these tensions. The rise of the state-run QAP process and the declining weight of the nationwide QCT system is indicative of a general anxiety over the potential for the program to inadvertently economically segregate American cities. Our paper supports previous work in finding LIHTC to, at the margin, be associated with reductions, not increases, in economic segregation as measured by the dissimilarity index [Diamond and McQuade, 2019, Ellen et al., 2016]. With the help of a basic model, we propose that this effect is driven by the high number of ELI renters in LIHTC developments and the weak correlation between LIHTC ELI composition and census tract ELI composition. Our findings, in conjunction with this previous work, support the view that LIHTC may be threading the needle in improving affordability and reducing economic segregation.

References

Victoria Basolo, Edith Huarita, and Jongho Won. A neighborhood-level analysis of low-income housing tax credit developments in the state of california and los angeles county. *Urban Science*, 6(2):39, 2022. doi: 10.3390/urbansci6020039. URL https://doi.org/10.3390/urbansci6020039.

Cody Cook, Pearl Z Li, and Ariel Binder. Where to build affordable housing?: Evaluating the tradeoffs of location. 2023. URL https://economics.yale.edu/sites/default/files/2024-01/Cook%20Where%20to%20Build%20Affordable%20Housing.pdf. Working Paper.

Rebecca Diamond and Tim McQuade. Who wants affordable housing in their backyard? an equilibrium analysis of low-income property development. *Journal of Political Economy*, 127(3):1063–1117, 2019. doi: 10.1086/701354. URL https://doi.org/10.1086/701354.

Ingrid G. Ellen, Keren M. Horn, and Katherine M. O'Regan. Poverty concentration and the low income housing tax credit: Effects of siting and tenant composition. *Journal of Housing Economics*, 34:49–59, 2016. doi: 10.1016/j.jhe.2016.08.001. URL https://doi.org/10.1016/j.jhe.2016.08.001.

Matthew Freedman and Tamara McGavock. Low-income housing development, poverty concentration, and neighborhood inequality. *Journal of Policy Analysis and Management*, 34(4):805–834, 2015. doi: 10.1002/pam.21856. URL https://doi.org/10.1002/pam.21856.

HUD. Qualified census tracts and difficult development areas — hud user, 2024. URL https://www.huduser.gov/portal/datasets/qct.html. Accessed December 9, 2024.

Mark P. Keightley. An introduction to the low-income housing tax credit. April 26 2023. URL https://crsreports.congress.gov/product/pdf/RS/RS22389.

Kirk McClure. What should be the future of the low-income housing tax credit program? *Housing Policy Debate*, 29(1):65–81, 2018. doi: 10.1080/10511482.2018.1431794. URL https://doi.org/10.1080/10511482.2018.1431794.

Alex Schwartz and Kirk McClure. The geography of absence: Cities, towns, and suburbs with no lihtc housing. *Housing Policy Debate*, 33(4):963–982, 2022. doi: 10.1080/10511482.2023.2171740. URL https://doi.org/10.1080/10511482.2023.2171740.

Jongho Won. Exploring spatial clustering over time and spillover effects of the low-income housing tax credit on neighborhood-level income segregation. *Urban Affairs Review*, 58(3):799–831, 2022. doi: 10.1177/1078087420973436. URL https://doi.org/10.1177/1078087420973436.

A Appendix A: Analytical Figures and Tables

Table 2: Dissimilarity difference under LIHTC removal simulation for top 5 CBSAs

CBSA	With LIHTC	Without LIHTC
New York-Newark-Jersey City	0.348	0.355
Los Angeles-Long Beach-Anaheim	0.278	0.276
Chicago-Naperville-Elgin	0.338	0.338
Dallas-Fort Worth-Arlington	0.356	0.358
Houston-Pasadena-The Woodlands	0.354	0.357

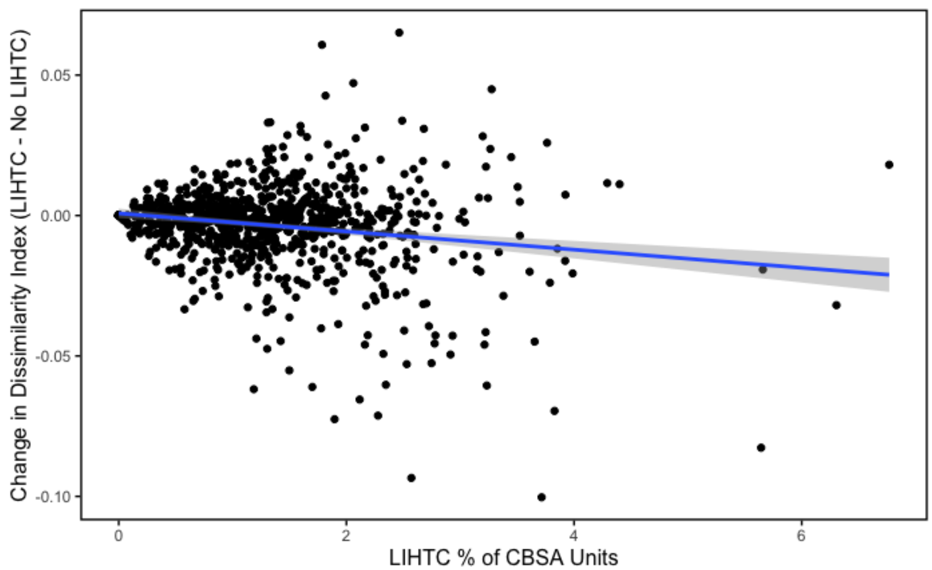

Figure 5: Change in CBSA Dissimilarity Index by LIHTC Coverage.

Table 3: Change in CBSA Dissimilarity Index by LIHTC Coverage: Regression Results

Term	Coefficient	Std. Error	Test Statistic	P-value
(Intercept)	0.001	0.001	0.898	0.369
LIHTC %	-0.003	0.001	-5.791	0.000***

Note: * p<0.05, ** p<0.01, *** p<0.001

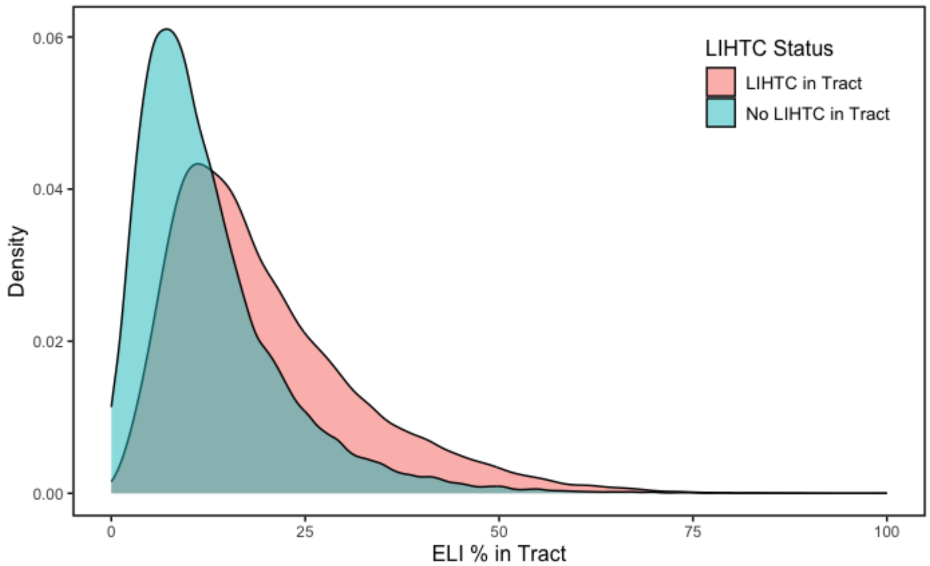

Figure 6: Density of ELI Units by Presence of LIHTC

Table 4: Dissimilarity difference under LIHTC removal simulation for top 5 CBSAs (LIHTC Tracts Only

CBSA	**With LIHTC**	**Without LIHTC**
New York-Newark-Jersey City	0.332	0.374
Los Angeles-Long Beach-Anaheim	0.229	0.252
Washington-Arlington-Alexandria	0.285	0.354
Chicago-Naperville-Elgin	0.320	0.357
Philadelphia-Camden-Wilmington	0.344	0.370

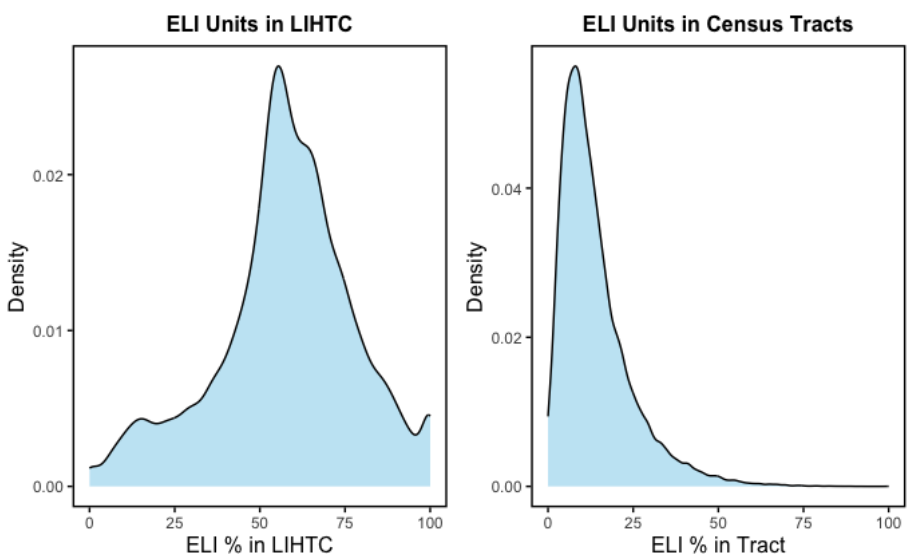

Figure 7: Comparison of ELI Unit Density (LIHTC vs. Tracts)

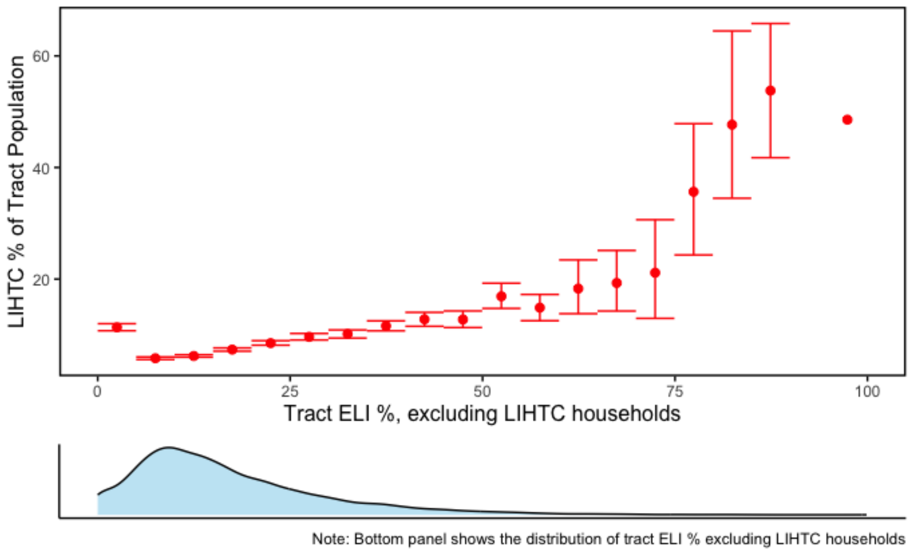

Figure 8: LIHTC % of Tract Population by Tract ELI %

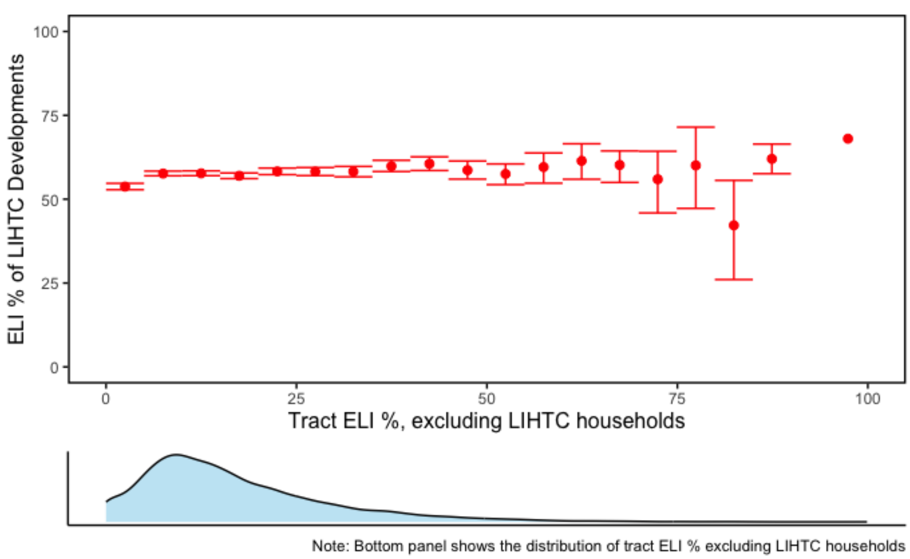

Figure 9: ELI % of LIHTC Population by Tract ELI %

B Appendix B: LIHTC ELI Methodology

Appendix B provides additional information and statistics on our pipeline for obtaining complete data on the number of extremely low income (ELI) households living in developments with active LIHTC subsidies as of 2022.

B.1 Raw ELI LIHTC Tenant Data Processing

Our data on ELI households comes from Table 9: Total Annual Household Income Relative to Derived Area Median Gross Income (AMGI). This data was released in 2022, 2019, and 2016. This is not mandatory information reported to HUD, so many values in each year are blank, and the data is also considered PII, so values less than 11 are suppressed. To obtain the largest sample possible, we look for a value for ELI % in any of those years, prioritizing the most recent value if multiple are present. The data includes other relative income brackets, with the total % summing to 100. We are able to counteract some of the bias from the suppression of smaller ELI % values by "backing" out small cells and redistributing residual percentages evenly across suppressed cells when more than half of the income brackets have nonzero values.

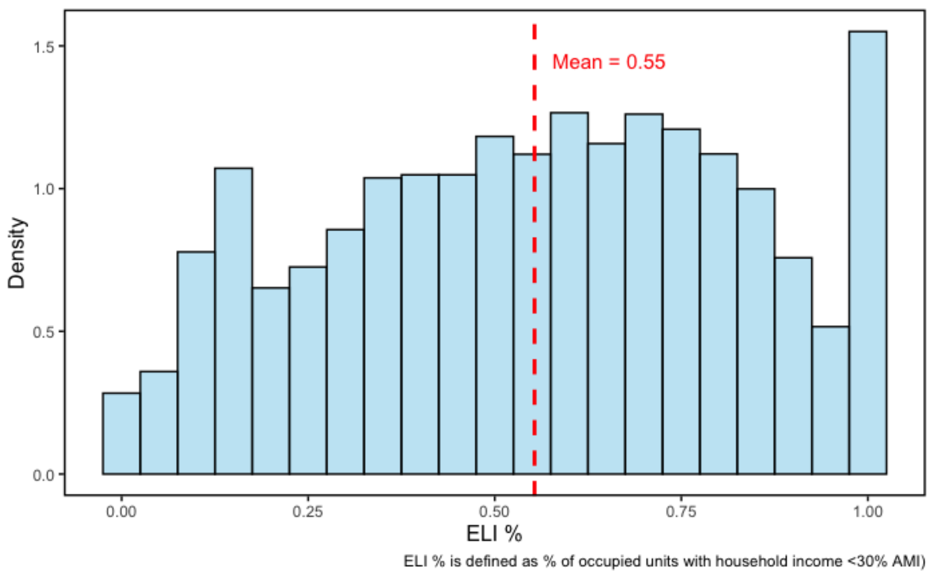

Figure 10: Distribution of ELI % within LIHTC Developments

B.2 ELI LIHTC Tenant Imputation

Once we have our baseline data from the LIHTC database, we use a multi-stage imputation process to estimate ELI % for the rest of the developments. Because ELI data is missing for entire states, such as Texas and Alabama, we can be confident that our data does not meet the missing completely at random (MCAR) assumption. To meet the missing at random assumption necessary for imputation, we use characteristics of the LIHTC developments from the NHPD and LIHTC data sources and census tract ACS profiles to build a predictive model of missingness. Using logistic regression, we are able to achieve 78% accuracy. We use the predicted probabilities from this model to re-weight our sample using inverse probability weighting (IPW). Our imputation process uses a random forest algorithm with the same development and census tract features as our missingness model.

We do not have full development profiles for all of the data points, so we are not able to fully impute the ELI % using this method. We use mean imputation to fill in the rest of the sample, iterating from census tracts, to CBSAs, to global means based on data availability. This process yields the following sources to our development ELI %s:

Table 5: Breakdown of ELI Data Sources

	Properties (%)	**ELI Units (%)**
LIHTC Tenant Data	44.0	43.5
Model-Based Imputation	47.8	46.0
Tract Mean Imputation	0.5	0.8
CBSA Mean Imputation	7.1	9.0
Global Mean Imputation	0.6	0.5

This process does meaningfully change the distribution of ELI % compared to what is observed in the original LIHTC tenant data. The mean remains unchanged, but the influence of the two clusters of data points at 0 and 1 is significantly reduced.

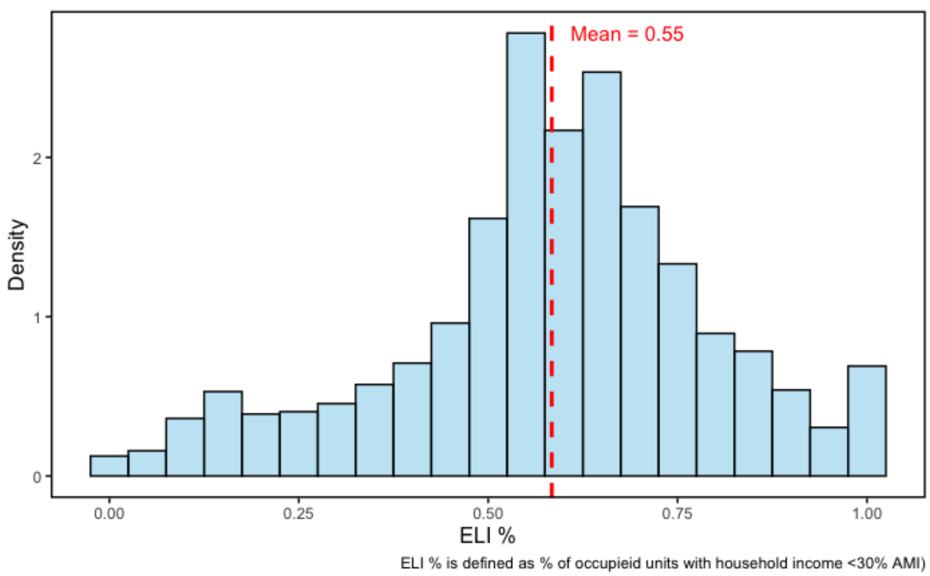

Figure 11: Distribution of ELI % within LIHTC Developments after Imputation

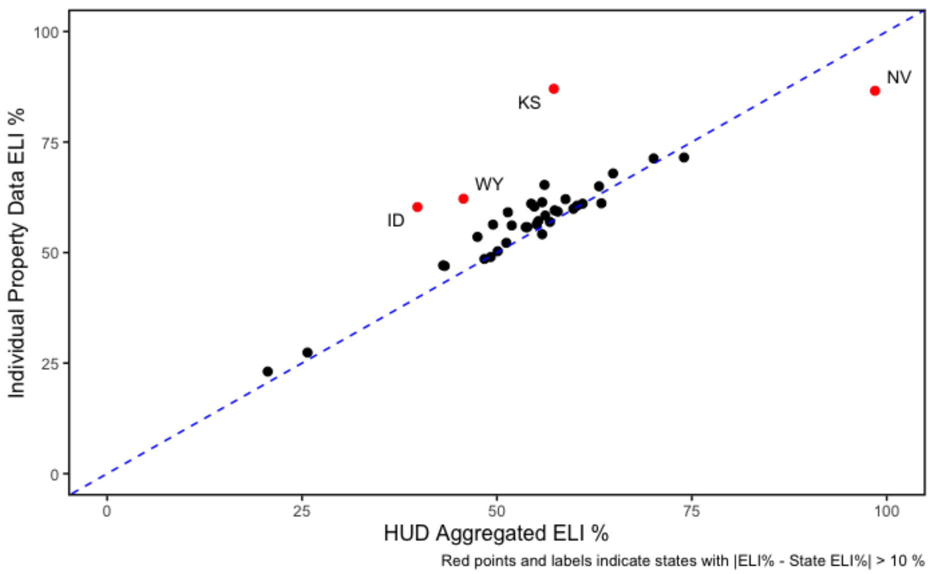

Figure 12: LIHTC State ELI Rates from HUD Property and Aggregated Data