Distance from the Frontier: Local Concentration, Hub Distance, and the Black-White Inventor Gap in Rural U.S. Counties

June 2026

Sunmit Hallur

Department of Economics
Stanford University
Stanford, CA 94305
shallur@stanford.edu

Thesis Advisor: Professor Lukas Althoff
Faculty sponsor: Professor Marcelo Clerici-Arias

ABSTRACT

Patenting in the United States has long been geographically concentrated and racially unequal, yet little attention has been paid to this issue at the county level in rural America. Prior research establishes a large Black-White gap in inventorship but focuses on metro hubs where innovation clusters, so we know little about how the gap varies in rural, nonmetropolitan counties. This paper addresses that gap by constructing a county-year panel of about 2,000 nonmetropolitan US counties (USDA Rural-Urban Continuum Code of 4 or higher) from 2001 to 2023, restricted to information and communications technology patent classes. Using inventor records from PatentsView and fully Bayesian Improved Surname Geocoding (fBISG) to impute race, I construct Black-White inventorship gaps and relate them to local assignee concentration (HHI) and driving distance to the nearest pre-period innovation hub, controlling for demographics and state-by-year fixed effects via OLS. I interpret the results as descriptive evidence on where rural gaps are largest, identifying whether the gap concentrates in counties far from inventor hubs or dominated by a few assignees, to inform federal place-based innovation policy targeting underserved rural counties.

Keywords: Black-White inventor gap, rural innovation, patent geography, assignee concentration, Bayesian Improved Surname Geocoding, exposure to innovation, spatial sorting, place-based policy.

Acknowledgments: I thank Professor Marcelo Clerici-Arias for guiding this research from the first ideas I had through to this present prospectus, and Ben Davies and Brendan Moore for their feedback on the empirical design and assignment guidelines. I am especially grateful to Professor Lukas Althoff for agreeing to advise my thesis as well as other students who have supported me throughout this entire thesis writing process.

Introduction

Innovation is a key driver of economic growth, but in the United States, innovation is not equally accessible to everyone. Bell et al. (2019b) link parental income to the probability children become inventors, documenting that a child born to parents in the top 1% of the income distribution is about 10x more likely to become an inventor than a child born to parents in the bottom 50% of the income distribution. They also document a stark racial gap in who becomes an inventor: white children are more than three times as likely to become inventors as Black children, a gap that differences in early-childhood math test scores explain very little of. Cook (2014) shows us that this gap has a long history: African American patenting rose sharply postbellum and then fell just as sharply in the early twentieth century as racial violence and Jim Crow institutions raised the cost of inventive activity. A century later, this gap has narrowed but it has not closed.

The geography of U.S. patenting is similarly skewed. A small set of innovation hubs like Silicon Valley, the Greater Boston Area, and the New York Metropolitan Area account for a disproportionate share of U.S. utility patenting (Bell et al. 2019a). Diamond (2016) and Moretti (2010) document a parallel pattern in skilled labor: high-skill workers have clustered in a small number of cities since 1980, generating local multipliers that further concentrate human capital and inventive activity: innovation hubs attract inventive talent, further geographically concentrating innovation. On the other side of this concentration lies rural and nonmetropolitan America, which is home to a sizeable share of the population but accounts for a steadily declining share of U.S. patenting, even as increased Wi-Fi access and remote-work technologies have lowered the cost of distance for many other tasks.

The racial gap in inventors and the geographic concentration of innovation are not separate stories: modern innovation statistics show the effects of past policies that allocated investment and exclusion along racial lines. Aaronson, Hartley, and Mazumder (2021) discuss how neighborhoods downgraded by the 1930s HOLC redlining maps experienced long-run reductions in homeownership, credit access, and educational attainment. Derenoncourt (2022) and Collins and Wanamaker (2014) further this idea by showing that the Great Migration sorted Black families across the urban hierarchy with long-run consequences for their children and grandchildren. As a result, the racial and spatial aspects of patenting are deeply intertwined - to the extent that modern innovation hubs are themselves a product of mid-century industrial policy and human-capital investment (Moretti 2010).

Despite reading this growing literature, I did not come across a good answer to a basic question I had: how does the Black-White inventor gap vary across rural America? Most existing empirical work on inventor inequality studies national cohorts (Bell et al. 2019b) or metropolitan statistical areas (Akcigit, Grigsby, and Nicholas 2017). Even when rural counties are studied, the focus is typically on overall economic outcomes rather than on inventive activity (Diamond 2016; Fajgelbaum and Gaubert 2020). And when the racial dimensions of patenting are studied directly (Cook 2014), the matter is analyzed from a national or regional level rather than county level. As a result, I believe two questions need to be answered. First, is the Black-White inventor gap larger in rural counties that are farther from major innovation hubs, or is the gap roughly constant once we condition on local patenting rates? Second, is the gap wider in rural counties where patent ownership is concentrated in the hands of a single large assignee, or is it smaller, because that single firm might itself be a source of local opportunity?

To answer these questions, I plan to construct a county-year panel of U.S. nonmetropolitan counties from 2001 to 2023, restricted to three information and communications technology areas (software, telecommunications, and electronic media) that are economically important, geographically diffuse, and most flexible with remote work options from rural areas. Inside this panel, I plan to test two specific hypotheses. My first hypothesis is that the gap in Black-White inventor rates is widest in counties most distant from the country's top inventor commuting zones. My second hypothesis is that the gap is widest in counties where patent ownership is most concentrated across assignees. My empirical strategy will be reduced-form rather than structural: I plan not to claim that hub distance or assignee concentration cause the gap, but I do want to show if they help organize the cross-sectional and longitudinal variation in the gap once I account for state-year shocks and demographic composition.

Answering these questions is specifically relevant because it informs essential government policy. Federal geography-dependent innovation programs like the Regional Technology and Innovation Hubs program created by the CHIPS and Science Act and the Rural Innovation Initiative at the U.S. Department of Agriculture explicitly target rural and underserved counties, but they largely treat rural America as one homogeneous group. If I find that the racial gap in patenting is concentrated in particular subsets of rural America (for example, counties far from existing hubs and dominated by a single large assignee), then place-based interventions targeted at those counties will have a different expected return than uniform transfers to rural areas.

Additionally, since exposure to local inventor networks is a key channel through which inventorship is passed down across generations (Bell et al. 2019b), the geographic distribution of those networks within rural America has first-order implications for who has access to careers in invention.

The rest of this prospectus is organized as follows. I first state the research question that motivates the county-year panel, then present a literature review that examines and synthesizes prior research on racial inequality in U.S. patenting, spatial inequality in innovation, and the geography of opportunity relevant to rural counties. Then I end with a study design section that describes how I construct the panel, define the outcome variable, and estimate the baseline specification. Finally, after the references, I recount how my thesis evolved over the quarter with feedback from peers and faculty.

Research Question

Does local concentration in patent ownership and distance from innovation hubs help explain Black-White gaps in inventor patenting across rural US counties?

Literature Review

In their paper, Bell et al. (2019b) ask who becomes a patent holder in the US and in it they find that large inequalities show up across geography and childhood exposure to innovation. They merge US tax and education administrative data with patent records, follow cohorts from childhood to adulthood, and measure exposure to innovation through inventor density in the child's CZ plus inventors in family and school networks. They find that children who see more inventors are far more likely to file for a patent as an adult: a child born to parents in the top 1% of income is about 10x more likely to become an inventor than a child born to parents earning less than the median income. Additionally, they also find that exposure is not simply unobserved ability by looking at within-family comparisons (comparisons between twins or siblings). I plan to use this paper to explain the mechanics of the county-year panel, making this a seminal paper for my thesis: Bell et al. justify using hub distance and assignee concentration as stand-ins for unequal access to innovation networks rather than just as structural primitives, so I will do the

same. Their emphasis on conditional patenting propensity also motivates my aggregation to GAP(c,t) and interacting local concentration and distance with group-specific intensity when I build county-year gaps.

In another paper by the same team, Bell et al. (2019a) contrast financial incentives with exposure-based mechanisms using quasi-experimental variation, where they find that patenting moves with both prices and social environment. Akcigit et al. (2022) study taxation and innovation over the twentieth century and show that fiscal policy shapes the supply of inventors at scale. Taking both of these papers together, they serve as a warning for my interpretations because when I interpret rural or within-rural variation, I should not fold everything into one policy channel. Specifically, state-by-year fixed effects absorb part of incentive-driven variation, but even so, local exposure can still differ across counties within a state-year. Coefficients on HHI(c,t), top 3 share, or d(c) show a pattern, but they don't prove cause-and-effect. People with similar backgrounds naturally cluster in the same neighborhoods, so correlations that look like network exposure may partly reflect who chooses to live where, not only how place changes patenting propensity.

Cook (2014) constructs historical patent-inventor data and links African American patenting to political violence (like riots and lynchings) and institutional barriers. Patenting falls when violence and discrimination disrupt markets and expectations. As a result, my ICT-focused rural panel will not estimate those historical shocks and instead use Cook's paper to anchor the review: Black-White patenting gaps are economically meaningful historically, which is another motivation for studying spatial correlations today like I will be doing.

Koning, Samila, and Ferguson (2021) bring the question to the present by showing that women invent for women's health disproportionately and that there are relatively few women

inventors overall. Their framing parallels my interest in who patents and for whom, so it is very relevant even though my outcome is the Black-White gap in ICT classes rather than gendered technology fields.

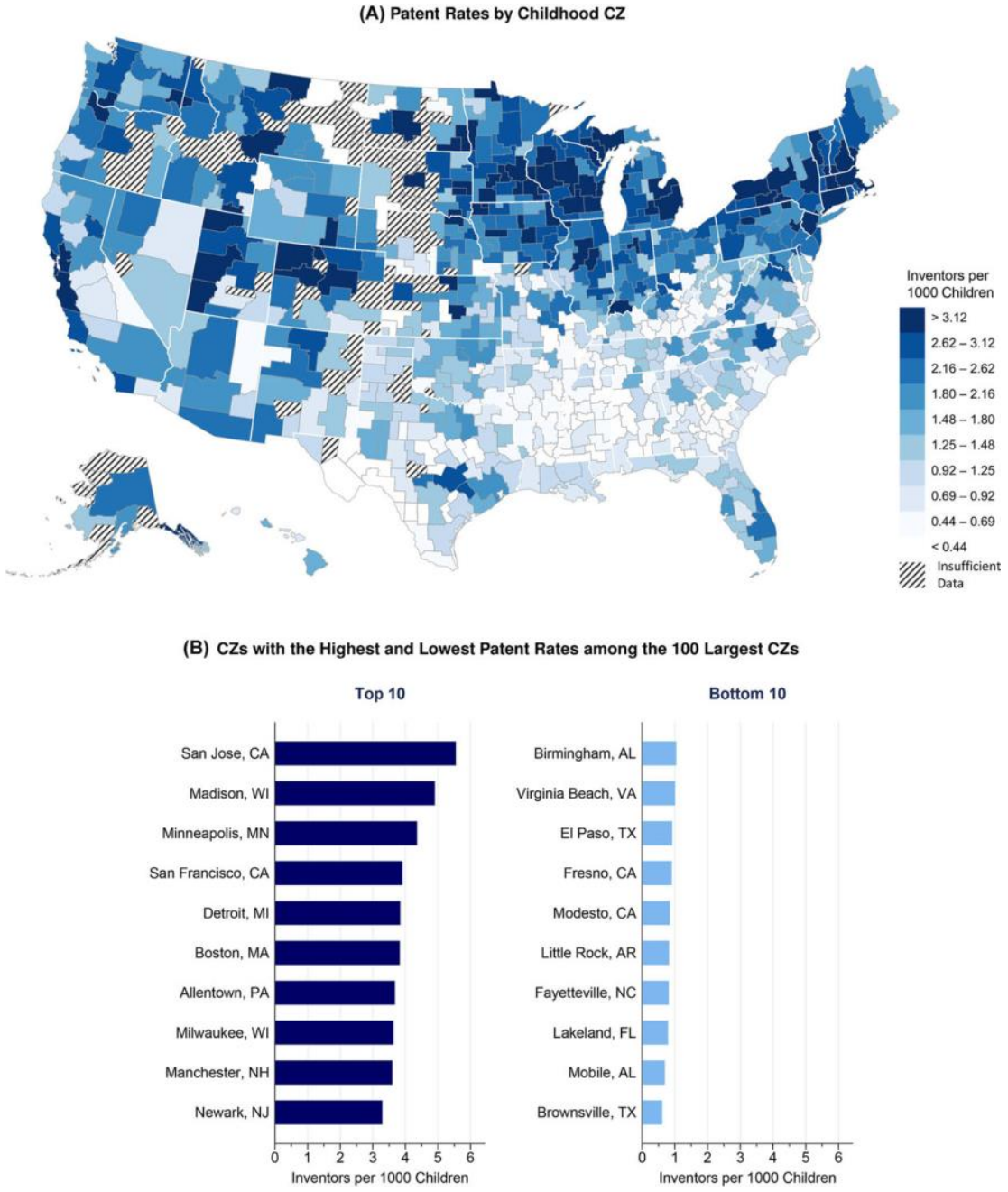


Figure 1. Reproduced from Bell et al. (2019b), Figure VIII (p. 690). Panel A maps the share of children who become inventors by childhood commuting zone for the 1980-1984 birth cohort;

darker shades indicate areas with higher inventor rates. Panel B lists the commuting zones with the highest and lowest inventor shares among the largest CZs.

Historical place-based exclusion and spatial mobility

In their paper, Aaronson, Hartley, and Mazumder (2021) estimate the long run effects of 1930s HOLC neighborhood security grades on housing, credit access, and neighborhood outcomes. This paper is especially relevant because HOLC appears in my research question as the thematic background for racialized place-based exclusion; I don't plan on using redlining as an instrument in the baseline patent panel. Cook (2014) stresses violence and legal exclusion and Aaronson et al.'s paper stresses credit markets and housing finance. Both imply that policy directly shaped where families could build wealth and invest in human capital long before my sample window.

Derenoncourt (2022) analyzes Great Migration responses to opportunity and violence and characterizes how destination labor markets absorbed migrants, and Collins and Wanamaker (2014) use linked Census data to document selection into migration and migration-related gains. Together with Derenoncourt, this paper clarifies composition margins that affect cross-county comparisons of patent intensity: county populations reflect historical sorting, so correlations between GAP(c,t) and local concentration or hub distance may partly reflect who lives where.

Boustan (2010) studies postwar suburbanization and Black migration and shows that locational choices reshaped where African American families lived relative to employment centers. This history matters to my question because rural counties today sit far from the CZs where inventor networks thickened after 1980 (analyzed in my top 50 inventor CZ).

Derenoncourt et al. (2024) quantify the U.S. racial wealth gap from 1860 to 2020 and they emphasize how policies and violence constrained wealth accumulation for Black Americans. I

read their results and plan to use it as the historical context for why exposure and place still differ by race, though I will not use it to explain causality in my baseline tables. If my data allows for it, I can also add robustness checks as suggested by this paper (see equations in Derenoncourt et al. for more detail). My baseline specification will lean on ACS controls and limit claims of causation. I do not treat these historical papers as instruments for contemporary patent gaps, but they do set expectations for persistent geographic inequality in who can access innovation-intensive labor markets, which like I said previously, was the motivation for my research question in the first place.

Spatial sorting, innovation hubs, and local concentration

Diamond (2016) documents rising spatial sorting of skill across U.S. cities and embeds the patterns in a spatial equilibrium model with wages, rents, and amenities. Because of this, I will treat d(c) partly as a sorting margin: nonmetropolitan counties may be distant from concentrated inventor CZs even after ACS controls because high-exposure networks cluster in metro areas rather than spreading evenly across similar-looking counties. Glaeser et al. (1992) show that local spillovers help explain city growth and provide an older agglomeration reference for why innovation concentrates geographically. Baum-Snow (2007) finds that highways contributed to suburbanization, so for rural counties, transportation frictions and distance remain important considerations when I interpret d(c). Diamond gives me a basis for analyzing heterogeneity in GAP(c,t) by d(c), but her paper does not examine patent class filtering or assignee concentration measures in its baseline tables. Bell et al. stresses transmission of inventor status within networks while Diamond stresses reallocation of skill across space, but in either case, they propose that rural counties far from hubs may patent less.

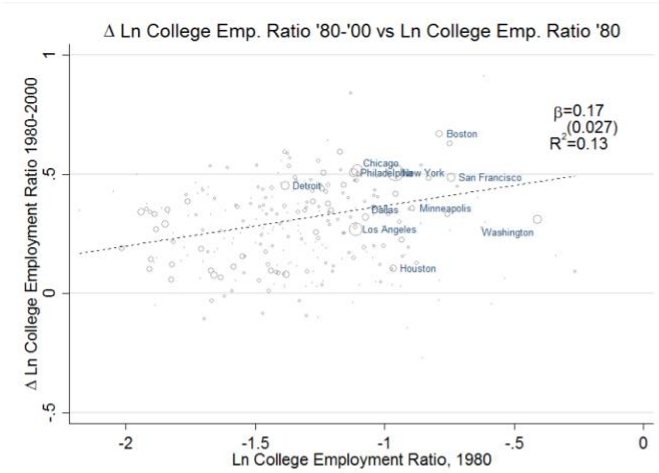


Figure 1.A

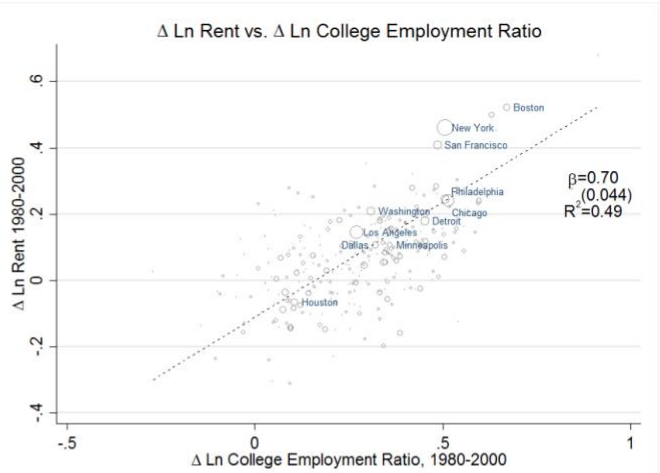
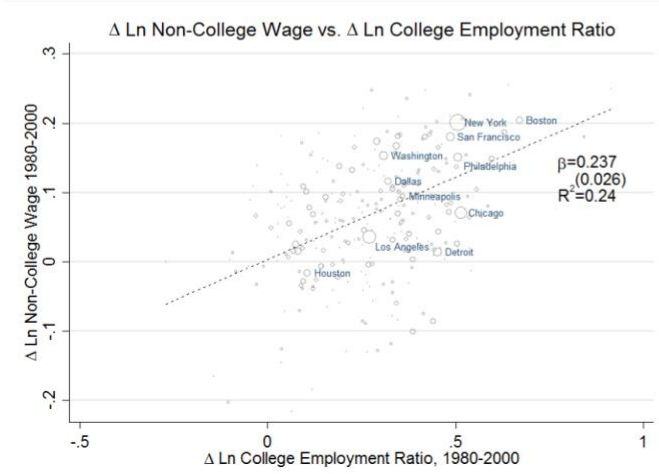


Figure 1.B



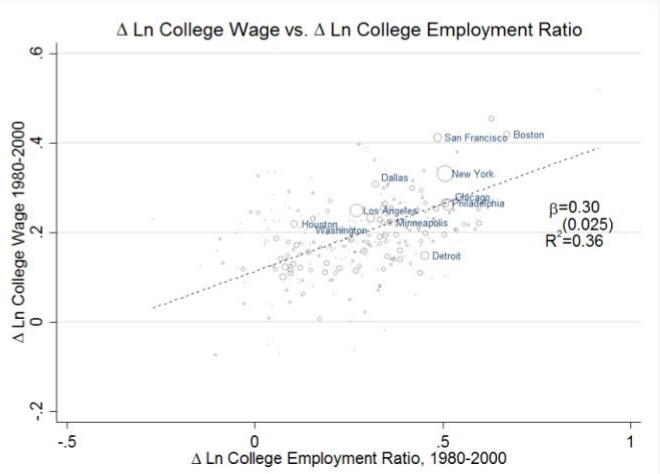


Figure 2. Reproduced from Diamond (2016), Figure 1. Panels plot changes from 1980 to 2000 in wages, rents, and college employment ratios across metropolitan areas, illustrating rising spatial sorting of skill.

Fajgelbaum and Gaubert (2020) develop a quantitative spatial model with trade, migration, and sorting and study optimal spatial policies. My baseline model stays simpler, but their paper helps me discuss why regional disparities in patent-intensive activity can persist. Moretti (2010) summarizes evidence that skilled and innovative employment raises local employment through multipliers, which helps explain why innovation hubs concentrate. I plan to read assignee

concentration (HHI(c,t) or top 3 share) as a rough proxy for how dense the local patent ecosystem is, without claiming that Moretti's multiplier numbers apply 1 to 1 to patent assignee levels. Fajgelbaum and Gaubert also do not report race-specific inventor rates for rural counties, but nevertheless, their sorting framework is the right framework for when I discuss why some nonmetro places remain thin markets for ICT patents even after controlling for education and industry mix in the ACS.

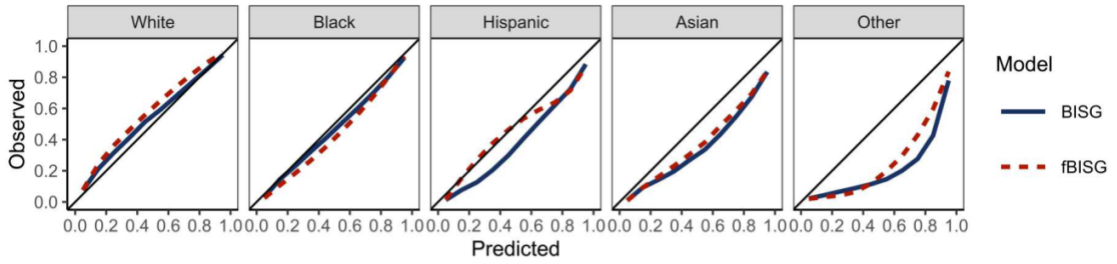
Kline et al. (2019) study rent-sharing at innovative firms and show that employees capture part of the returns to patents. Their results support treating assignee concentration as more than just a head count: a county dominated by one patent-intensive employer may concentrate both patents and wage premia locally. Alcácer, Gittelman, and Sampat (2009) analyze applicant and examiner citations in U.S. patents and document how knowledge flows through the patent system. I will cite them when I explain why citation-based networks and assignee structure are natural complements to inventor counts in county panels. None of Diamond, Fajgelbaum and Gaubert, Moretti, Kline et al., or Alcácer et al. report race-specific inventor rates for rural counties, hence the reason for this paper in my literature review.

Measuring inventor race in patent microdata

Interestingly enough, patent microdata rarely reports inventor race. Imai and Khanna (2016) develop methods for predicting individual ethnicity from voter registration records, building on the Bayesian Improved Surname Geocoding (BISG) approach that later papers apply to patents, and I plan to use this technique with the following modifications: Imai, Olivella, and Rosenman (2022) extend BISG in a fully Bayesian framework (fBISG), add name supplements, and correct Census-related processing issues that bias race probabilities. The same team published the name supplements and race priors as a reusable dataset for first, middle, and surnames in 2023. This is

key for my thesis because every county-year calculation of N(g,c,t) in my panel depends on that imputation data.

The fBISG empirical section reports priors, calibration geographies, and sensitivity analyses. Small rural counties may have large variances in their estimated gaps, so I plan on documenting this uncertainty rather than treating GAP(c,t) as exact. I also plan to aggregate posterior race probabilities to expected inventor counts in each county and year, consistent with Imai et al.'s notation, and report sensitivity to alternative priors when discussing measurement error. Pairing Imai and Khanna (2016) with Imai, Olivella, and Rosenman (2022) and Rosenman, Olivella, and Imai (2023) makes it clear that race gaps in PatentsView are model outputs, not self-reported fields. This matters for policy interpretation in my thesis conclusion because a widening GAP(c,t) in a small county could reflect true inequality, prior misspecification, or just thin cell counts, so the thesis will report counts of imputed inventors alongside rates to make interpretations of my thesis more robust.



tion curves for race predictions obtained using the standard BISG (blue) and fBISG (red) methods. The curves plot predicted probabilit
egory against the observed proportion of cases that actually fall in that category. Thus, curves closer to the 45° line indicate better calibrate
based on the 2010 Census surname dictionary. The red curve (i.e., the fBISG calibration curve) is either identical to or closer to the 45° than
calibration curve), indicating that fBISG's predictive accuracy is at least on par to that of BISG.

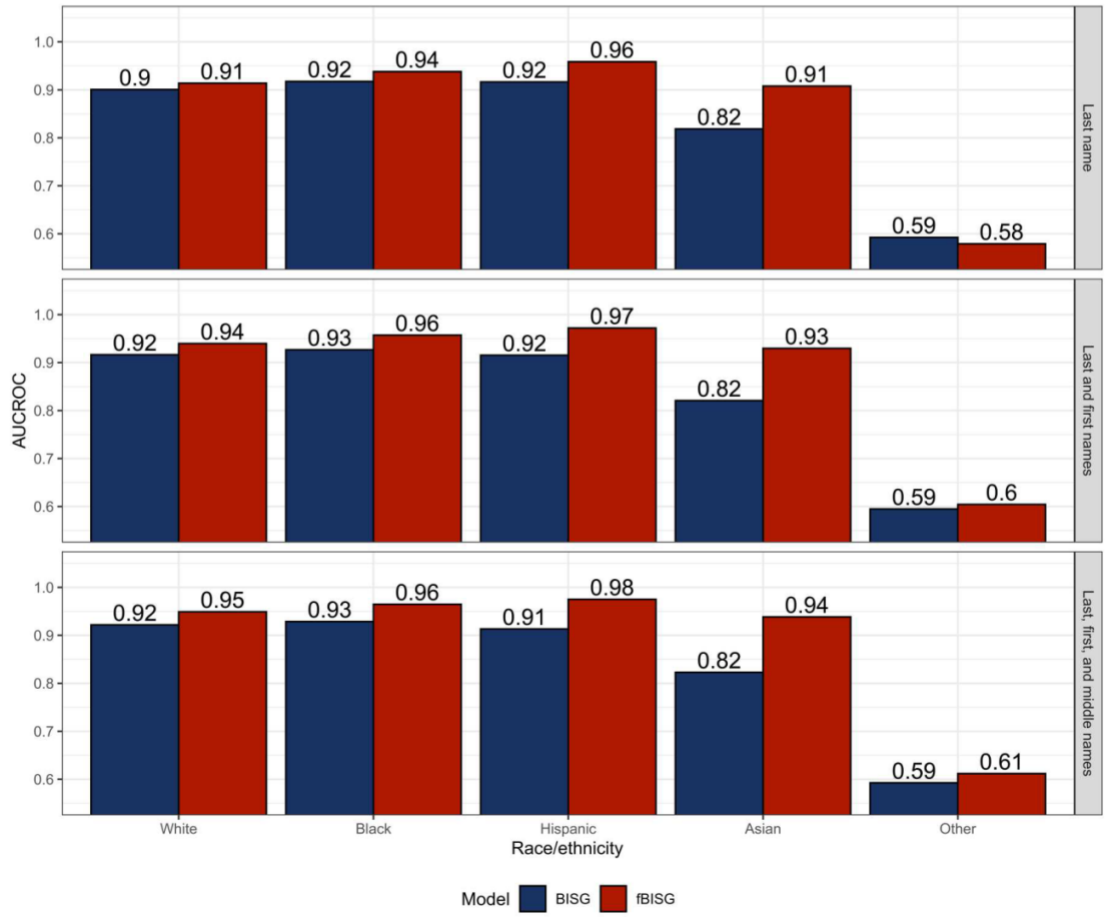


Figure 3. Reproduced from Imai, Olivella, and Rosenman (2022), Fig. 2. Calibration curves compare standard BISG and fully Bayesian BISG (fBISG) race predictions against 2010 Census truth; curves closer to the 45-degree line indicate better calibration.

How my thesis extends the literature

The papers I read on this topic cover exposure and history (Bell et al. 2019a, 2019b; Cook 2014; Koning, Samila, and Ferguson 2021; Akcigit et al. 2022), place and migration (Aaronson, Hartley, and Mazumder 2021; Derenoncourt 2022; Collins and Wanamaker 2014; Boustan 2010; Derenoncourt et al. 2024), spatial structure and hubs (Diamond 2016; Fajgelbaum and Gaubert 2020; Moretti 2010; Glaeser et al. 1992; Baum-Snow 2007; Kline et al. 2019; Alcácer, Gittelman, and Sampat 2009), and race measurement (Imai and Khanna 2016; Imai, Olivella, and Rosenman 2022; Rosenman, Olivella, and Imai 2023). Even though all of them ask surrounding questions, there still remains the question at rural county resolution regarding whether Black-White gaps in ICT patenting line up with assignee concentration and distance from pre-period innovation hubs after state-year shocks and demographics are accounted for. Related literature at the national or CZ level cannot say whether a rural county with one dominant tech assignee and a six-hour drive to the nearest top 50 inventor CZ is where the Black-White ICT gap is widest, which is the descriptive question my panel is built to answer.

More specifically, my thesis extends this prior literature by constructing a county-year panel of nonmetropolitan counties from 2001 to 2023 with race-specific ICT patenting rates and the gap GAP(c,t). Bell et al. work mainly at national cohort and CZ scale; I move to rural counties where federal place-based innovation programs operate. Diamond, Fajgelbaum and Gaubert, Moretti, Glaeser et al., Baum-Snow, Kline et al., and Alcácer et al. each stress a different spatial or firm-level force; I combine hub distance d(c) and assignee concentration HHI(c,t) in one reduced-form specification. My analysis will target where GAP(c,t) is largest in the current sample, interpreted using Bell et al. (2019a, 2019b) on exposure to innovation, with redlining and migration entering only as background context rather than direct identified shocks in my regressions' baseline specification. If the widest gaps show up in counties with both substantial

hub distance d(c) and high assignee concentration HHI(c,t), blanket rural innovation programs may channel funds poorly compared with interventions aimed at those specific places. The study design section that follows presents the data, outcome variable, empirical framework, and full econometric specification.

Study Design

1. Data Sources

My primary data source is PatentsView, the USPTO's publicly disclosed patent dataset, which provides geocoded inventor addresses, assignee identifiers, and Cooperative Patent Classification (CPC) codes for every U.S. utility patent granted between 1976 and the present. PatentsView's inventor disambiguation algorithm assigns a unique identifier to each inventor, which is essential for converting patent counts into inventor counts at the county-year level. I have already downloaded the full set of PatentsView granted patents tables (g_patent, g_inventor_disambiguated, g_assignee_disambiguated, g_location_disambiguated, and g_cpc_current) onto my project pipeline, together with a complementary set of pre-processed annual patent grant files for 2000 to 2024 that include WIPO field titles, team sizes, and inventor-level gender flags. I plan to crosscheck inventor and county assignments against the public Google Patents BigQuery dump, whose disambiguation routine is developed independently with a different geocoder, and I will use this as a secondary data point to crosscheck any inventor-county assignment errors.

My demographic and socioeconomic controls will come from the American Community Survey (ACS) five-year files from 2009 to 2023, which I plan to download next and combine with 2020 county boundaries using the Census Bureau's annual delineation files. I will then use

working-age (25 to 65 years old) population counts by singular race and Hispanic origin as the denominators for race-specific inventor rates.

I will define rural status using the USDA Economic Research Service Rural-Urban Continuum Codes (RUCC). I plan to include counties with RUCC values of 4 or higher, which by the 2023 USDA definition corresponds to 1,958 nonmetropolitan counties accounting for about 74% of U.S. land area and about 14% of the population.

Nevertheless, not all of the data is publicly available and my analysis does require some extrapolation since inventor race is not observed in patent microdata. Following Imai, Olivella, and Rosenman (2022), I will assign inventor race using fully Bayesian Improved Surname Geocoding (fBISG), which combines a surname-based prior that can be derived from the U.S. Census Bureau surname file (augmented with the Voter File Surname Supplement of Rosenman, Olivella, and Imai 2023) with a location-based prior derived from the inventor's county of residence at the time of patent application. The output is, for each inventor, a posterior probability Pr(race = r | surname, county). All downstream race-specific inventor counts in each county will be computed as expected counts using this methodology.

I plan to measure distance to innovation hubs, d(c), as driving distance from each county's population-weighted centroid to the nearest top 50 inventor commuting zone, where the top 50 are defined by cumulative inventor counts in 1995 to 2000 (before the panel window to avoid simultaneity bias). Defining the hubs in this period also captures the immediate pre-period geography of inventor concentration during the internet invention boom, which is exactly the network that 2001 to 2023 rural inventors are or are not close to, so it should provide the best reference.

2. Outcome Variable Construction

I am planning on my unit of analysis being the county-year like previous analyses (Akcigit, Grigsby, and Nicholas 2017; Diamond 2016), with the county-year being indexed by (c for county, t for year) for t = 2001, ..., 2023. I will restrict the technology basket to three WIPO fields that map cleanly over to the CPC tree: Digital Communication and Telecommunications (CPC section H04), Audio-Visual Technology (parts of CPC sections H03 and H04), and Computer Technology (CPC section G06). I am deciding to use the WIPO field labels because they are pre-computed in the annual patent files I have already downloaded and because they correspond to internationally standardized technology areas that are easier to interpret than the raw CPC tree itself. A patent is in cell (c, t) when at least one of its inventors lists a residence in county c and the patent carries at least one CPC code in the basket. I will then standardize this basket before estimation.

To find the race-specific inventor rate, let N(B,c,t) and N(W,c,t) denote the expected number of Black and non-Hispanic White inventors in county c in year t, computed by summing the BISG race posteriors, as previously described, over all inventors with at least one in-basket (see above for definition) grant in (c, t). Let P(B,c,t) and P(W,c,t) denote the corresponding working-age populations from the ACS. The race-specific inventor rate per 100,000 working-age adults is:

$$r(g,c,t) = 100{,}000 \cdot [N(g,c,t) / P(g,c,t)],\ g \in \{B, W\},$$

and the primary dependent variable is the gap [GAP(c,t) = r(B,c,t) - r(W,c,t)]. From the literature I have read, GAP(c,t) is negative for almost every county-year in expectation, and it becomes less negative when the Black-White gap narrows.

3. Empirical Framework

Based on the equation described in Bell et al. (2019b), I will model the probability that a person of race g in county c becomes an inventor as a function of three components: ability denoted by θ, exposure to local inventor networks E(g,c,t), and the return to invention R(g,c,t). Aggregating to the county level and taking the difference between racial groups, the expected gap satisfies

$$E[GAP(c,t)] = \beta_E \cdot (E(B,c,t) - E(W,c,t)) + \beta_R \cdot (R(B,c,t) - R(W,c,t)) + \beta_\theta \cdot (\theta(B,c,t) - \theta(W,c,t)),$$

where the last term denotes the difference in average ability between the two groups. For some more clarification, βE is the return to exposure to local inventor networks, βR is the return to the private payoff from inventing, and βθ is the return to innate inventive ability, with all three coefficients positive in the Bell et al. (2019b) framework. Still, to use this equation, I have to make two key assumptions. The first assumption is that talent gaps between groups are not driven by intrinsic differences in ability but instead reflect cumulative differences in early childhood investment (Aaronson and Mazumder 2011). The second assumption is that exposure to local inventor networks operates through both racial and geographic channels: in rural counties far from major hubs, Black residents are more likely than White residents to be disconnected from local inventive family members and other mentors, because Black residential and educational segregation has historically been higher (Diamond 2016; Derenoncourt 2022). As a result, I plan to use two moderating variables, the assignee HHI(c,t) and the log of hub distance log d(c) as proxies for the racial gap in exposure (E(B) - E(W)).

4. Econometric Specifications

My baseline plan is to run an OLS regression of the county-year Black-White gap GAP(c,t) on three observable moderators: the county-year assignee HHI(c,t), the share of in-basket patents accounted for by the top three assignees Top3Share(c,t), and the log of road-network distance

from county c to its nearest top 50 inventor commuting zone, log d(c). I will run this regression conditional on a vector of ACS controls like median household income, college share of the working-age population, the unemployment rate, and log population density, and I will make this regression conditional on a full set of state-by-year fixed effects, with standard errors clustered at the state level. The state-by-year fixed effects absorb anything that happens at the state level in a given year (such as state R&D tax policy or state-level IP enforcement), so that the moderator coefficients are estimated only from differences across counties within the same state-year.

My central hypothesis is that all three moderator coefficients (on HHI(c,t), Top3Share(c,t), and log d(c)) are negative, meaning that the Black-White inventor gap (already negative almost everywhere) becomes more negative in counties with concentrated patent ownership and in counties far from major innovation hubs, and a finding of significantly positive coefficients or a failure to reject zero would be evidence against the exposure mechanism documented in Bell et al. (2019b). To probe heterogeneity, I plan to re-estimate the same specification separately by Census region (Northeast, Midwest, South, West) and by individual CPC subclass within the basket (G06, H03, H04), allowing the HHI(c,t) and log d(c) to interact to test whether being both concentrated and remote is worse for the gap than the sum of the two effects in isolation. On this same train of thought, I will also re-estimate the regression separately for male and female inventors since the annual patent files I have already staged include inventor-level gender flags and this will increase the depth of my research.

My thesis will ultimately view these estimates as descriptive moderators rather than causal effects: they tell me where the gap is largest, not why it is largest there. However, three confounding variables need to be addressed before the estimates can be trusted. First, the gap depends on the BISG race imputation, which is imperfect, so I plan to report sensitivity to

alternative BISG priors (national versus state-level) and to the augmented surname dictionary as established by Rosenman, Olivella, and Imai (2023). Second, rural counties in my sample often have very small Black working-age populations, which makes the gap volatile from year to year (as n decreases, standard error increases), so I plan to pool counties into rolling three-year windows, report a complementary specification at the commuting-zone-year level, and weight each county-year by total working-age population to reduce the impact of outliers. Third, even after the state-by-year fixed effects and the four ACS controls, a county-specific event (such as the opening of a new research university or the closing of a major employer) could shift both the regressors and the gap simultaneously while biasing the moderator estimates, so I plan to report a complementary specification with county fixed effects, which identifies the assignee-concentration moderators from within-county variation over time at the cost of absorbing the cross-sectional variation in log d(c).

I am not attempting any causal identifications in this prospectus because I have not yet performed the actual data analysis. If these descriptive moderators turn out to be strong and robust to the three threats above, I plan to end my thesis with a discussion of what plausibly exogenous variation in either hub distance or assignee concentration would look like as a path to causal interpretation and accordingly give recommendations to inform future government policy.

References

Aaronson, Daniel and Mazumder, Bhashkar. 2011. "The Impact of Rosenwald Schools on Black Achievement," Journal of Political Economy, 119(5): pp. 821-888.

Aaronson, Daniel, Hartley, Daniel, and Mazumder, Bhashkar. 2021. "The Effects of the 1930s HOLC "Redlining" Maps," American Economic Journal: Economic Policy, 13(4): pp. 355-392.

Akcigit, Ufuk, Grigsby, John, and Nicholas, Tom. 2017. "The Rise of American Ingenuity: Innovation and Inventors of the Golden Age." NBER Working Paper No. 23047.

Akcigit, Ufuk, Grigsby, John, Nicholas, Tom, and Stantcheva, Stefanie. 2022. "Taxation and Innovation in the Twentieth Century," The Quarterly Journal of Economics, 137(1): pp. 329-385.

Alcácer, Juan, Gittelman, Michelle, and Sampat, Bhaven. 2009. "Applicant and Examiner Citations in U.S. Patents: An Overview and Analysis," Research Policy, 38(2): pp. 415-427.

Baum-Snow, Nathaniel. 2007. "Did Highways Cause Suburbanization?" The Quarterly Journal of Economics, 122(2): pp. 775-805.

Bell, Alex, Chetty, Raj, Jaravel, Xavier, Petkova, Neviana, and Van Reenen, John. 2019a. "Do Tax Cuts Produce More Einsteins? The Impacts of Financial Incentives versus Exposure to Innovation on the Supply of Inventors," Journal of the European Economic Association, 17(3): pp. 651-677.

Bell, Alex, Chetty, Raj, Jaravel, Xavier, Petkova, Neviana, and Van Reenen, John. 2019b. "Who Becomes an Inventor in America? The Importance of Exposure to Innovation," The Quarterly Journal of Economics, 134(2): pp. 647-713.

Boustan, Leah Platt. 2010. “Was Postwar Suburbanization “White Flight”? Evidence from the Black Migration,” The Quarterly Journal of Economics, 125(1): pp. 417-443.

Collins, William J. and Wanamaker, Marianne H. 2014. “Selection and Economic Gains in the Great Migration of African Americans: New Evidence from Linked Census Data,” American Economic Journal: Applied Economics, 6(1): pp. 220-252.

Cook, Lisa D. 2014. “Violence and Economic Activity: Evidence from African American Patents, 1870-1940,” Journal of Economic Growth, 19(2): pp. 221-257.

Derenoncourt, Ellora. 2022. “Can You Move to Opportunity? Evidence from the Great Migration,” American Economic Review, 112(2): pp. 369-408.

Derenoncourt, Ellora, Kim, Chi Hyun, Kuhn, Moritz, and Schularick, Moritz. 2024. “Wealth of Two Nations: The U.S. Racial Wealth Gap, 1860-2020,” The Quarterly Journal of Economics, 139(2): pp. 693-750.

Diamond, Rebecca. 2016. “The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000,” American Economic Review, 106(3): pp. 479-524.

Fajgelbaum, Pablo D. and Gaubert, Cecile. 2020. “Optimal Spatial Policies, Geography, and Sorting,” The Quarterly Journal of Economics, 135(2): pp. 959-1036.

Glaeser, Edward L., Kallal, Hedi D., Scheinkman, José A., and Shleifer, Andrei. 1992. “Growth in Cities,” Journal of Political Economy, 100(6): pp. 1126-1152.

Imai, Kosuke and Khanna, Kabir. 2016. “Improving Ecological Inference by Predicting Individual Ethnicity from Voter Registration Records,” Political Analysis, 24(2): pp. 263-272.

Imai, Kosuke, Olivella, Santiago, and Rosenman, Evan T. R. 2022. “Addressing Census Data Problems in Race Imputation via Fully Bayesian Improved Surname Geocoding and Name Supplements,” Science Advances, 8(49): eadc9824.

Kline, Patrick, Petkova, Neviana, Williams, Heidi, and Zidar, Owen. 2019. “Who Profits from Patents? Rent-Sharing at Innovative Firms,” The Quarterly Journal of Economics, 134(3): pp. 1343-1404.

Koning, Rembrand, Samila, Sampsa, and Ferguson, John-Paul. 2021. “Who Do We Invent for? Patents by Women Focus More on Women's Health, but Few Women Get to Invent,” Science, 372(6548): pp. 1345-1348.

Moretti, Enrico. 2010. “Local Multipliers,” American Economic Review, 100(2): pp. 373-377.

Rosenman, Evan T. R., Olivella, Santiago, and Imai, Kosuke. 2023. “Race and Ethnicity Data for First, Middle, and Surnames,” Scientific Data, 10: 299.

My Thesis Evolution

Initially, I began with the idea to study how patenting is distributed across the US to see if there were racial disparities in who gets to patent. This was in part motivated by my work at Schmeiser Olsen LLP, an IP law firm where I had worked for 1 year prior to working on this thesis, and my prior work documenting civil rights history in the US through the medium of film. What I found after some digging was that there was a plethora of literature that explored the idea of racial disparities in patenting and they found a large gap in patenting and innovation across racial lines, but their analysis ended at the city level in urban areas. This brought me to the idea of studying patenting divides in rural areas, which became the thesis prospectus above.

My early bibliography work framed this question around a Black-White gap in per-capita, inventor-based patenting in information and communications technology classes since these patent classes were the most able to be worked on remotely from rural areas. I also then decided to consider the impact of driving distance to a nearby city as a factor in how likely one is to invent, which I represented as d(c). Over the quarter, I refined this into a proper regression looking at the Black-White inventor gap GAP(c,t) and how it can be explained by hub distance d(c) and assignee concentration HHI(c,t).

The peer review on my Assignment 6 draft was especially helpful. My reviewer said the introduction already connected racial gaps in inventorship, the geographic concentration of innovation, and the historical roots of racial inequality, and that my motivation was clear, but they noted a real gap in prior work: national cohorts, metropolitan areas, and broad rural outcomes are common, but the county-level Black-White inventor gap in rural America is not. They suggested that a later draft (like this thesis prospectus) could end the introduction with an explicit contribution sentence about whether disparities concentrate where counties are remote

from hubs or dominated by few assignees. I used that advice when I merged the introduction for this prospectus and when I wrote the extension section at the end of the literature review.

Outside of feedback from classmates, I met with Ben Davies, who gave me concrete advice on how to merge the introduction, literature review, and study design into one coherent prospectus rather than treating them as separate course artifacts. I also met with Brendan Moore before refining the literature review, who gave me great suggestions on how to review and incorporate Bell et al. (2019b) as the seminal exposure paper and how to use the Bell framework without letting the tax-policy companion paper or the spatial-sorting papers talk past one another.

Faculty feedback from Professor Marcelo Clerici-Arias across the annotated bibliography and literature review assignments, and from Professor Lukas Althoff after he agreed to advise the thesis, shaped the feasibility of the PatentsView-ACS panel and kept the claims descriptive rather than causal. Overall, this feedback helped me significantly to write this prospectus.