Student Debt and High-Skill Worker Location Choice*

Zachary Sauers †

June 14, 2023

Working Paper

ELECTRONIC LINK TO LATEST VERSION

Abstract

Over the past 40 years, the United States has experienced significant skill-based geographic sorting, with high-skill workers increasingly concentrating in large cities. During the same period, the growth of student debt has far exceeded the rate of inflation, and the average debt load is now roughly $30,000. In this paper, I document a link between these two facts. First, I employ the recently-developed University of California Consumer Credit Panel to estimate the effect of student debt on post-schooling location choices by exploiting an expansion of federal student loan limits in 2008-2009. Using a difference-in-differences framework, I find that an increase of $10,000 in debt makes individuals 6.5 percent more likely to locate in large metropolitan counties. Next, I provide descriptive evidence that student debt makes borrowers more elastic to nominal wages than local prices in migration decisions. Embedding this mechanism in a spatial equilibrium model, I find that the rise in student debt from 1980 to 2019 can account for 5-19 percent of the increase in skill-based sorting over this period. Counterfactual simulation of three policy proposals – debt forgiveness, tuition-free college, and income-driven repayment – show that only income-driven repayment can eliminate distortions to location choices while improving welfare.

*I am extremely grateful to my advisors Christian Dippel, Paola Giuliano, Melanie Wasserman, and Romain Wacziarg for their guidance and support. This paper also benefitted immensely from discussions with Gregor Schubert, Giovanni Peri, Joan Monras, Enrico Moretti, Sandra Black, Rebecca Diamond, and many others. I would also like to thank the participants of the 2022 North American Urban Economics Association Meeting, Global Migration Center Summer School, 2021 All California Labor Economics Conference, and UCLA Anderson Global Economics and Management Workshop for their many helpful comments and suggestions. Additionally, I would like to thank Evan White, Eduard Tomany, Gary Gremaux, April Chang, and the rest of the California Policy Lab staff for their immense help and patience in my work with the UC-CCP dataset. Taara Cason and Bethan Consic from the Institute of Education Studies and Harry Morall, Freddy Lopez, Kevin Kurata, and Karla Zapeda from UCLA were all incredibly helpful in my work with the Education Longitudinal Study of 2002. I am also grateful for the help of Linda Eggenberger and other members of the Institute of Social Research at the University of Michigan in my work with the Panel Study of Income Dynamics. I gratefully acknowledge receiving financial support from the UCLA Ziman Center for Real Estate, the California Policy Lab, and the UCLA Dissertation Year Fellowship while working on this research.

†University of California, Los Angeles. Email: zsauers@ad.ucla.edu. Website: www.zacharysauers.com.

1 Introduction

Over the past 40 years, there has been a significant amount of skill-based geographic sorting in the United States, with high-skill workers increasingly concentrating in large, dense locations (Diamond, 2016; Moretti, 2013; Berry and Glaeser, 2005). As a result, by 2020, the average high-skill worker share in top labor markets was 57 percent while it was just 37 percent in rural areas – double the difference in 1980. This sorting has been of interest to researchers and policy-makers alike due to the implications at the national and local level for productivity, inequality, housing markets, and voting patterns (Diamond, 2016; Glaeser, 2008; Scala and Johnson, 2017). While skill-based spatial sorting is now a well-known fact, I document the thus-far overlooked fact that migration decisions in early adulthood drive this divergence. As shown in Panel (a) of Figure 1, high-skill workers are significantly more likely to move to higher-density locations in early adulthood relative to low-skill workers. This results in differential location patterns that stabilize by age 30, as shown in Panel (b). Together, these patterns suggest that factors weighing heavily on individuals in early adulthood could contribute to aggregate sorting patterns.

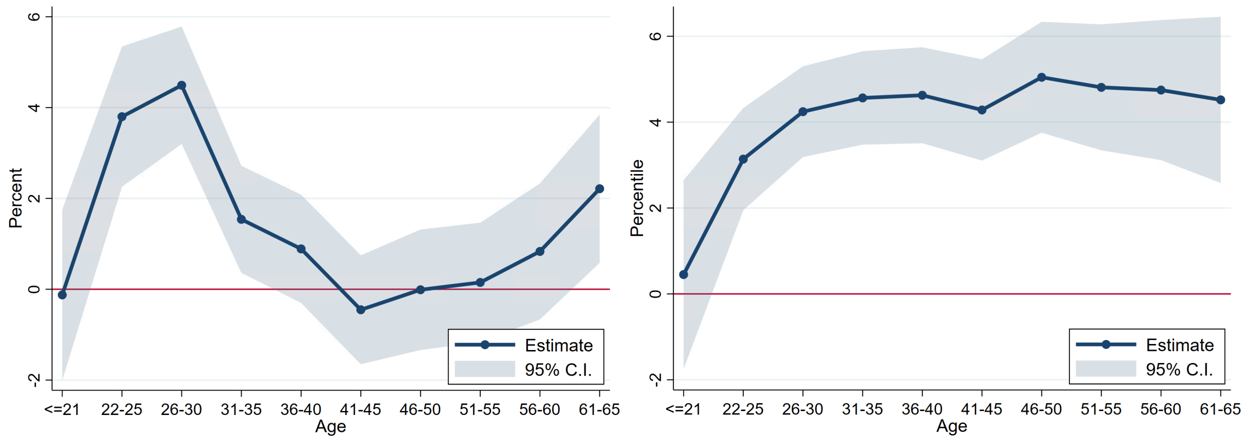
In this paper, I investigate one such factor and show that it distorts location choices of high-skill workers: student debt. The cost of postsecondary education in the U.S. has increased fourfold since 1980 and students have increasingly turned to loans as a way to finance their education. I show that increased student debt pushes high-skill workers to large, densely-populated areas where it is widely known that wages and costs tend to be higher (Glaeser, 2008). The proposed mechanism by which student debt affects post-college location choice is that it makes borrowers more elastic to nominal wages than local prices. This stems from the structure of student debt repayment plans which, for the majority of borrowers, are fixed and independent of local prices and income.

To build intuition for this mechanism, consider individuals with and without student debt (D) that choose location based only on nominal wages (W) and local prices (P), i.e., they maximize $ln\left(\frac{W-D}{P}\right)$ – a simplified version of the standard Rosen (1979)-Roback (1982) framework. For individuals with no debt, location choice is equally elastic to nominal wages and local prices. When individuals carry student debt, the elasticity with respect to wages ($\frac{1}{1-\frac{D}{W}}$) is strictly greater than the elasticity with respect to prices (-1), and it increases with the level of debt. This is because debt is repaid from the dollar gap between nominal wages and housing costs, which is typically higher in urban areas even if the real standard of living is lower. Individuals with student debt then maximize utility by choosing to locate in areas with relatively higher nominal wages.

In the first part of the paper, I estimate the causal effect of student debt on post-college location choice. Since eligibility for federal student loans is nearly universal, uptake could be endogenous to personal, social, and economic characteristics. As a result, isolating the causal effect of student debt is challenging. To solve this problem, I exploit an increase of federal student loan limits in 2008 and 2009 that created cohort variation in the maximum amount individuals could borrow. I employ a difference-in-differences approach that compares post-schooling location choices of students who enrolled in college before and after the loan limit increases across two groups: (1) students unconstrained by limits, and (2) students likely to be constrained by limits. The key

Figure 1: Association Between Being High-Skill and Migration Choices Over Life Cycle

(a) Probability of Moving to Higher Density County (b) Population Density Percentile of Residence



Panel (a) shows the coefficients from a linear probability model of an indicator for moving to a county with a higher population density relative to previous county of residence on the interaction between being a high-skill worker and age categories. Panel (b) shows the coefficients from the same regression when using the percentile of residence county population density as a dependent variable. Both estimated using a PSID sample of individuals born between 1940 and 1979 and observed biennially 1968 to 2019. Excludes waves when individuals enrolled in postsecondary institutions. Regressions include wave, cohort, and demographic controls. See Appendix A for a complete discussion.

identifying assumption is that, in the absence of loan limit increases, differences in post-college location choices of constrained and unconstrained students would be similar across cohorts.[1]

I implement this strategy using newly available data in the University of California Consumer Credit Panel (UCCCP). The UCCCP covers the complete population of individuals with credit reports who lived in California between 2004 and 2019 – containing all credit reports for these individuals throughout this period, even when they reside outside of California. Precise location is observed throughout while detailed account information allows me to observe educational borrowing from 2002 to 2019. I first employ the difference-in-differences analysis to confirm that increased access to educational loans resulted in additional borrowing. I find that increased loan limits resulted in an average of $2,600 in additional loans for constrained students relative to unconstrained students through year 4 of borrowing – the last year of postsecondary education for the majority of students.[2]

I then turn to the location choices of student borrowers in post-college years. I find that $10,000 in additional debt increases the population density of a borrower's initial post-college county by 8.5 percent. Focusing on migration to top urban areas, I find that $10,000 in additional borrowing increases the probability of locating in counties classified as being in the 95th percentile by population density by 6.2 percentage points or 11.4 percent from the mean. Indicative of lasting effects on migration choices, I show that this effect is persistent through year 9 from first education loan, the last year with available data. These results are robust to alternative classifications of counties provided by the Department of Agriculture's Rural-Urban Continuum Codes.

In the second part of the paper, I provide descriptive evidence consistent with high-skill workers

[1]This empirical approach is similar to Black et al. (2020), who examine financial well-being outcomes.

[2]This result is similar to Black et al. (2020), who find roughly $1,800 of additional borrowing in a national sample.

differentially valuing the higher nominal wages of cities. For this analysis, I use data from the Education Longitudinal Study of 2002 (ELS), a longitudinal survey of 2002 high school seniors that, unlike the UCCCP, provides a rich set of controls to perform heterogeneity analysis. This analysis stems from a testable implication of the proposed mechanism: the association between student debt and urban post-college location choice should be higher for individuals who face steeper urban wage premiums. Using American Community Survey microdata containing nominal wages, location, and degree fields for high-skill workers, I compute urban wage premiums by degree field and map them onto the universe of academic majors in the ELS. As predicted by the proposed mechanism, the association between student debt and urban post-college location choice is driven by individuals who major in fields with above-median urban wage premiums even after accounting for spatial variation in the distribution of industries. For individuals who major in degree fields with below-median urban wage premiums, there is no significant relationship between student debt and urban post-college location choice.

One alternative explanation for the identified effect of student debt is that preference for urban locations may be correlated with willingness to accept student debt. This would require individuals make college, student debt, and post-college location choices at least partially as a joint decision. I present two descriptive facts from the ELS that suggest this mechanism does not play a large role. First, I show that the association between student debt and urban post-college location choice is robust to controlling for high school and college county urbanicity as proxies for urban preference. Second, as a large portion of college graduates live and work near their postsecondary institution (Conzelmann et al., 2022), this implies that individuals making a joint decision would likely have the same college and post-college location. I show that the documented association between student debt and urban post-college location choices is robust to analysis of only the subpopulation of individuals who left their college county – a group which may be making post-college location choices more independently of college and debt decisions.

How important is student debt for aggregate sorting patterns? What effect may policy proposals such as student loan forgiveness, a free college option, or widespread adoption of income-driven repayment plans have? To study these counterfactual questions, in the third part of the paper I develop and calibrate a spatial equilibrium model that enables me to quantify the role of student debt in skill-based sorting and simulate outcomes under various policy proposals. Workers in the model choose skill and debt levels endogenously and have heterogeneous preferences over locations. Locations differ by skill-specific productivity and amenity levels. Housing markets differ across locations due to heterogeneity in their elasticity of housing supply. To conduct my analysis with complete coverage of the US, I estimate new 1990 Commuting Zone-level elasticities following the methods of Saiz (2010) and Howard et al. (2018).

After calibrating the model using parameters from the literature, I present simulated outcomes under various counterfactual scenarios. I present results in full equilibrium, where prices and wages respond to population changes, and partial equilibrium, where prices and wages remain fixed and changes in outcomes reflect only a labor supply response via migration. By remaining agnostic about wage and price responses, the partial equilibrium results offer short-term predictions while full equilibrium results, which account for spillover effects on low-skill workers through price and wage adjustments, may more accurately depict long-term outcomes. In both, location- and skill-specific amenities remain fixed.

I begin the counterfactual analysis by estimating that the growth of student debt accounts for 3.5-4.5 percent of the increase in skill-based sorting from 1980 to 2019 (11.5-18.7 percent in partial equilibrium). I then turn to assessing the impact of various policy proposals. Debt forgiveness is estimated to decrease the level of skill-based sorting by 2.4-3.4 percent (8.1-14.0 in partial equilibrium). This is accompanied by a small decline in welfare with high-skill workers gaining from the reduction in debt and the elimination of the location distortions while low-skill workers are worse off due to the introduction of a uniform tax to fund the policy. Widespread adoption of income-driven repayment plans, which do not distort location choices, results in similar reductions to skill-based geographic sorting; however, all workers now experience an increase in welfare as location choices adjust to fully reflect preferences without imposing a tax on low-skill workers. The introduction of tuition-free postsecondary education has a negligible net effect on skill-based sorting as two opposing forces interact: the rising national share of high-skill workers are still drawn to cities by higher productivity, but their location choices are no longer distorted by debt repayment. Welfare increases for high-skill workers as they no longer bear the full cost of education while it declines for low-skill workers as they assume some of the cost. For debt forgiveness and income-driven repayment, aggregate output declines by a fraction of a percent as high-skill workers express preferences to locate in lower productivity regions. As tuition-free college increases the share of high-skill workers, it increases aggregate production by 1.5-2 percent.

These findings have important implications for policy-makers at both the national and local level. In 2022, the Biden Administration announced the nation's largest student loan forgiveness program, covering up to $20,000, and is pushing for increased adoption of income-driven repayment plans. Implementing a free college option is also part of some political platforms. As these policies are novel in the United States, the findings in this paper provide valuable insights about implications for internal migration and labor markets. Additionally, these findings provide a new avenue to address the so-called 'brain drain' from rural areas to urban ones – the phenomenon where top-performing students leave rural areas rather than enhancing the local stock of college-educated adults. While leaders have frequently employed place-based policies to address this problem with mixed success (Neumark and Simpson, 2015) and unclear implications for aggregate welfare (Kline and Moretti, 2014; Glaeser and Gottlieb, 2008), reducing outflows of highly productive workers by switching to income-driven repayment plans for student debt offers a promising new strategy to boost economic activity in rural areas via the supply-side.

Related Literature. My work builds on a number of papers exploring the causes of skill-based geographic sorting. There is robust evidence that the primary driver is changes in local productivity (Berry and Glaeser, 2005; Moretti, 2013) resulting in part from skill-biased agglomeration economies (Baum-Snow et al., 2018; Giannone, 2017). Amenity differences have also been documented as an important contributing factor (Shapiro, 2006; Albouy et al., 2016) that amplifies productivity-driven sorting (Diamond, 2016). I add to this literature by identifying student debt as new driver of skill-based spatial sorting that, unlike the previously-documented factors, is the result of a policy decision governing the structure of debt repayment rather than underlying economic forces.

This paper also adds to a growing literature examining the implications of student debt for post-graduation outcomes. While student debt is considered a good investment for most people (Barrow and Malamud, 2015; Oreopoulos and Petronijevic, 2013) even in the face of growing debt (Avery and Turner, 2012), conclusions about its effects on post-schooling life are less clear. There is increasing evidence that student borrowers do not behave as a standard life-cycle model would suggest, i.e., student debt should have a minimal effect on post-college decisions because the ratio of debt to the present discounted value of lifetime earnings is small (less than 1 percent). Instead, student debt has been found to affect many key decisions in adulthood: household formation via homeownership (Mezza et al., 2020; Bleemer et al., 2021) and co-residence with parents (Dettling and Hsu, 2018); family formation via marriage and fertility (Gicheva, 2016); job choice (Di Maggio et al., 2020; Rothstein and Rouse, 2011); and income (Gervais and Ziebarth, 2019; Luo and Mongey, 2019; Bettinger et al., 2019).[3,4] While nearly all find significant effects of student debt on the respective outcome, the magnitude and direction varies by sample and empirical strategy. I contribute to this literature by identifying a missing dimension, across which many of these outcomes vary significantly, that could account for the mixed results: geography.

Methodologically, the empirical work in this paper builds on those in the student debt literature that use experimental or quasi-experimental variation. The most closely related is Black et al. (2020), who use the same policy change to reflect on educational attainment and post-college financial well-being. Other empirical work exploits variation from grant aid, tuition, and bankruptcy regulations. Additionally, this paper augments a small but growing number of studies in the student debt literature employing large consumer credit panels (Chakrabarti et al., 2020; Black et al., 2020). The structural component of this paper builds on the original Rosen (1979)-Roback (1982) framework as well as more recent spatial equilibrium models like Hsieh and Moretti (2019) and Diamond (2016). The proposed mechanism is similar to that in Albert and Monras (2017), who find that immigrants concentrate in expensive U.S. cities because remittences to origin countries reduce sensitivity to local price levels. In my setting, debt repayments at a national price, i.e. fixed across locations, make student borrowers relatively less sensitive to local price levels.

In the following section, I discuss relevant institutional details on student borrowing and the policy change that generates the identifying variation. Section 3 outlines the empirical strategy for identifying the effect of student debt on post-college location choice and presents the results. In Section 4, I provide supporting evidence for the proposed mechanism. Section 5 embeds the proposed mechanism in a spatial equilibrium model, which is calibrated in Section 6. Section 7 simulates outcomes under various policy proposals. Section 8 concludes.

[3]See as well: household formation (Black et al., 2020; Akers and Chingos, 2014; Houle and Berger, 2015; Gicheva and Thompson, 2015; Chakrabarti et al., 2020; Bleemer et al., 2014), family formation (Shao, 2014), job choice (Krishnan and Wang, 2019), income (Di Maggio et al., 2020; Chapman, 2015; Minicozzi, 2005; Weidner, 2016).

[4]Contradictions of the life-cycle model are attributed to debt aversion (Burdman, 2005; Callender and Jackson, 2005; Field, 2009) or credit constraints after college (Rothstein and Rouse, 2011; Gicheva and Thompson, 2015).

2 Background on Student Borrowing

The cost of postsecondary education in the United States has increased rapidly over the past 40 years. As shown in Figure 2, inflation-adjusted tuition and fees are now four times the 1980 level. Students have absorbed these rising prices by taking out additional debt - with loans per student increasing nearly fivefold in real terms over the same period. In 2019, the typical bachelor's degree-holder leaves school with roughly $30,000 in debt. Federal lending dominates the landscape of student borrowing for post-secondary education, comprising 88-93% of all educational loans over the past decade (Baum et al., 2019).[5] Under the umbrella of federal lending, Stafford Loans account for roughly two thirds of borrowing (Baum et al., 2019). Stafford Loans have historically been provided by one of two federal lending programs: the Federal Family Education Loan (FFEL) Program, authorized as part of the Higher Education Act of 1965 (HEA), and the Federal Direct Loan Program, created as a 1992 amendment to the HEA. Though the source of the funds has varied over time, the function of Stafford Loans has remained consistent from the student's perspective.[6]

Undergraduate students have essentially uniform access and terms for Stafford Loans. To qualify for Stafford Loans or any other type of federal student aid, students must complete the Free Application for Federal Student Aid (FAFSA), an annual form which collects demographic and financial information. This information, which includes assets and income, pertain to the student and their household for dependent students. The Department of Education uses the information provided in the FAFSA to determine a student's eligibility across the two types of Stafford Loans: subsidized and unsubsidized. Subsidized Stafford Loans are need-based and do not accrue interest while the student is enrolled. Unsubsidized Stafford Loans are not based on financial need and do accrue interest while the student is pursuing their degree. Although the package of offered Stafford Loans varies by student, this has limited implications for cumulative borrowing over the duration of an individual's enrollment.[7] The more meaningful constraint and determinant of cumulative borrowing is the federal student loan limit, which governs the maximum allowable loans over both types of Stafford Loans for a given academic level (i.e., freshmen, sophomore, or upper level).

The majority of college graduates are awarded their degree in 4 years or less and repay educational debt using plans with fixed payments across space. As shown in Figure 3, over 60% of students who attained a Bachelor's degree in 2017 did so in 4 years or less. This increases to 85% by year 5. Repayment typically begins after graduation and a grace period of 6 months. Repayment periods are typically up to 10 years for single loans and between 10 and 30 years for consolidation loans. Income-driven repayment plans were first introduced in 2009; however, as of 2020, roughly

[5] This has been remarkably stable over time. In 1998-99, the first period with nonfederal borrowing data, federal loans accounted for 91% of all education borrowing (Baum et al., 2019).

[6] The FFEL was eliminated in 2010 as part of the Health Care and Education Reconciliation Act of 2020. After this date, nearly all federal lending is through the Direct Loan Program. Stafford Loans under the Direct Loan Program were issued directly from the Department of Education, while loans under the FFEL Program were issued by private sector institutions and guaranteed by the federal government. There is no practical difference from the student's perspective.

[7] The value of the in-school interest subsidy varies by entry year and duration of schooling. Subsidized Stafford Loans had a slightly lower interest rates from 2008 to 2013. Black et al. (2020) estimate that this subsidy ranges from $34 to $82 for a $1,000 loan when repayment starts one year after origination.

70% of borrowers are still on traditional repayment plans with amounts set independent of income or location (Figure 4). Durante et al. (2017) find an average monthly payment of $393 among those actively making payments in 2016.

2.1 Federal Revisions of Stafford Loan Limits

The borrowing limits for Stafford Loans can only be adjusted via federal legislation – something that has only been happened twice in the 21st Century. The first was with the Higher Education Reconciliation Act of 2005, which went into effect in the 2007-08 academic year. The second adjustment occurred as part of the Ensuring Continued Access to Student Loans Act of 2008, which took effect starting in the 2008-09 academic year. Both changes increased the loan limits. The experienced borrowing limits are reported by academic year and level in Table 1. The first wave of increases only impacted individuals in their freshmen and sophomore years, while the second adjustment increased limits for all academic levels. Although aggregate cumulative limits were adjusted as well (last Column in Table 1), these limits would never constrain a borrower who attains their degree in 4 years further than academic year-x-level limits.[8]

Table 2 shows how these changes impacted students by entry cohort. While the first wave of loan limit increases took effect in the 2007-08 academic year, individuals in earlier cohorts can still be affected if they are in school when adjustments took effect. This combination of staggered introduction over time and uneven increases across academic levels generates the identifying variation used in the following section. For example, students who entered in the 2005-06 academic year may have experienced a $2,000 increase in their year 4 borrowing limit if they were still enrolled through year 4. The increased loan availability phases in over the 2005-06 through 2008-09 cohorts and peaks at an additional $9,875 in borrowing ability (last column of Table 2).

3 Identifying the Effect of Student Debt on Location Choice

In this section, I estimate the effect of student debt on post-schooling location choice by exploiting variation in student borrowing driven by a policy change that increased the maximum amount students are able to borrow for postsecondary education from Federal sources. I estimate this effect for the full population of California student borrowers in the recently developed University of California Consumer Credit Panel (UCCCP) assembled by the California Policy Lab.

3.1 Data

The UCCCP is an individual-level longitudinal dataset following roughly 60 million consumers with credit reports on a quarterly basis since 2004. While not the first credit panel used in the literature (Black et al., 2020; Chakrabarti et al., 2020), it is significantly larger than the Federal Reserve

[8] Considering cohorts 2002-3 through 2012-3 (the sample used for identification in Section 3) and individuals who enroll in postsecondary education for up to 6 years, aggregate limits would only constrain borrowers further than academic year-x-level limits for individuals in the 2002-03 cohort who are enrolled and borrow the maximum amount each year through year 6.

Bank of New York's Consumer Credit Panel (approx. 13 million) and the Consumer Financial Protection Bureau's Consumer Credit Panel (approx. 5 million). The underlying credit histories are sourced from Experian, one of the three nationwide credit bureaus. The UCCCP is composed of two samples: a nationally representative 2 percent sample of U.S. adult consumer with credit records, and a full 'sample' consisting of 100 percent of Californians with credit histories. The California sample includes all consumers with credit reports who lived in California between 2004 and 2019.[9] This includes those who originated in the state, moved to California for college, or resided there after schooling. For individuals that meet this inclusion criteria, the UCCCP includes all available reports regardless of location in any given quarter. Although both samples provide the same coverage of variables, this analysis uses the California sample due to its size.

The UCCCP includes information on tradeline-level account information, credit scores, location, and demographics of consumers. Tradelines include student loans, auto loans, credit cards, mortgages, and other forms of credit. Data on tradelines include account opening date, account type, account condition (open, closed, in deferment, in repayment, etc.), principal amount (for loans), borrowing limits (for credit cards), and latest balance amount among others. Geographic information consists of 5-digit ZIP codes sourced from tradeline mailing addresses, which I then map to county-level characteristics for analysis.[10] Demographic information includes gender, month and year of birth, and education codes.[11]

Since the UCCCP does not include enrollment information, I use information on student loans contained in credit histories to infer entry cohorts and build a dataset of borrower-x-year since entry observations. To do so, I assume that the first academic year an individual is observed opening a student loan is the first year that they enter school, i.e. their entry cohort.[12,13] Since each quarterly archive of the UCCCP provides a snapshot of an individual's full credit report at a given point in time, I am able to construct first-year borrowing information for cohorts back to 2002 using 2004 archives.[14] Loan transfers and consolidations in addition to changing tradeline identifiers make linking and tracking the exact evolution of borrowing beyond first year unreliable, particularly for cohorts before 2004. To solve this problem, I measure a borrower's cumulative borrowing at any given point as the sum of all active loans. While eliminating the need to trace all transfers and consolidations, this limits observations to first-year borrowing and cumulative borrowing through years 4, 5, and 6 from first education loan for all individuals in cohorts 2002 to 2013.

[9] Though the UCCCP contains continually adds archives up to the present, I limit the sample to 2019 to avoid confounding factors arising from the COVID-19 pandemic.

[10] ZIP codes mapped to counties using a crosswalk from the U.S. Department of Housing and Urban Development.

[11] Demographic information is often limited in credit reports because federal law prohibits discrimination in credit transactions on the basis of race, ethnicity, religion, sex, age, marital status, or receipt of public assistance. Education codes are modeled/estimated by the credit bureau using sample surveys.

[12] Academic year defined as July through June and denoted as the calendar year it ends. For example, the 2003-2004 academic year, denoted just by 2004, includes loans opened from July 1st, 2003 through June 30th, 2004.

[13] Most individuals that ever borrow to finance their college education do so in their first year. Black et al. (2020) estimate that 73 percent of all dependent undergraduates in the 2016 National Postsecondary Student Aid Study who ever took out student loans and graduated in 2016 borrowed in their first year.

[14] As I am using the 2004 archive, this approach conveniently avoids the problem that credit histories prior to 2004 often suffer from incomplete reporting of student loans.

I restrict the population in two ways to target borrowers financing first-time undergraduate education. First, I only include borrowers who open their first education loan between the ages of 16 and 20. Second, I exclude all borrowers whose first-year loans exceed the Federal student loan borrowing limits for first year undergraduate students in a given academic year. The intention of both restrictions is to reduce the inclusion of individuals who first borrow for a graduate degree (and thus face a different labor market), individuals who first borrow in upper academic levels of undergraduate education, and those that are independent students.

To create a balanced panel, I only include individuals from the resulting sample that are observed at a minimum through year 6 from first loan and for up to 9 years. As shown in Figure 3, a vast majority of borrowers who attain an undergraduate degree do so in 4 or 5 years. I ensure all individuals are observed through year 6 because I will consider this the start of when location information in credit archives reflect their post-college location. The resulting dataset contains roughly 940,000 student borrowers who entered in 2002 through 2013 cohorts and are observed for up to 9 years after entry.

3.2 Empirical Strategy

The empirical strategy for identifying the causal effect of student debt on post-college location choices compares outcomes for students likely to be constrained and unlikely to be constrained (unconstrained) by original borrowing limits in years before and after the federal loan limit increases. The identifying assumption is that, in the absence of loan limit increases, differences in post-college location choices of constrained and unconstrained students would be similar across cohorts. I begin the empirical analysis with an event study framework which allows me to analyze dynamics over the cohorts, such as differences in baseline characteristics, cumulative borrowing ('first stage'), and location choices ('reduced form'). I then move on to the difference-in-differences analysis to get the main estimate of interest: the effect of additional borrowing on post-college location choices. I am also able to examine persistence through year 9 from initial borrowing in this specification.

The event study framework is given by:

$$\begin{aligned} Y_{isc} =& \alpha + \beta_1 \text{Constrained}_i + \sum_{c \neq 2005} \beta_2^c \left[\mathbb{1}[\text{Cohort}_i = c] \text{ x Constrained}_i \right] \\ &+ \mathbf{X_i'} \boldsymbol{\beta_x} + \delta_c + \delta_s + \epsilon_{isc}, \end{aligned} \tag{1}$$

where Y_{isc} is the outcome of interest (baseline characteristic, cumulative borrowing through a certain year, or post-college location choice) for individual i who attended college in state s and first borrowed as part of cohort c. When considering post-college location choices, urbanicity of a location is defined in three ways: (1) a continuous measure of a county's population density (in logs); (2) an indicator function for if the county is in the 95$^{\text{th}}$ percentile by population density; and (3) an indicator function for if the county is classified in the top category (metropolitan area with a population of at least 1 million) using the Rural-Urban Continuum Codes (RUCC). Whether individuals are likely or unlikely to be constrained by original borrowing limits is captured by the binary variable Constrained$_i$. The coefficients of interest, β_2^c, capture the interaction between the indicator for being constrained and an indicator being in cohort c. The limited demographic variables available in credit reports are included in X_i. These include sex, age when individual was first issued an education loan, and characteristics of the borrower's credit history prior to entry.[15] I also include cohort (δ_c) and college state fixed effects (δ_s) to capture common trends and unobserved characteristics that may vary by state of college attendance.[16] Errors are clustered by college state.[17]

In the event study specification and the difference-in-differences framework to follow, the final cohort designated as unaffected by the loan limit increases contains those who entered in 2005. This reflects the variation in possible cumulative borrowing through year 4 as shown in Table 2. This choice, consistent with Black et al. (2020), was based on evidence that the majority of students who attain a Bachelor degree do so in 4 years or less, as illustrated in Figure 3. To the extent that the share of individuals staying more than four years experience additional loan limits prior to 2005, the event study estimates would show pre-trends and the difference-in-differences estimates would underestimate the full effect of additional borrowing.[18]

The main advantage of the event study framework is to provide insight into the identifying assumption of parallel trends in the difference-in-differences specification. The parallel trends assumption in this application is that in the absence of loan limit increases, differences in post-college location choices of constrained and unconstrained students would be similar across cohorts. Although this assumption is untestable, it suggests that the difference in outcomes between groups should be similar across the untreated cohorts, i.e., pre-2005 cohorts. As shown in Figure 7-8 and Figure 9-11 for borrowing and location outcomes, respectively, there is little evidence of pre-trends. The only variable displaying differences between constrained and unconstrained borrowers that vary significantly from the 2005 cohort level are for borrowing outcomes ('first stage'), all of which are small in magnitude and do not suggest a clear pattern. There are no pre-trends in location outcomes. While not conclusive, this provides evidence that is consistent with the parallel trends assumption being valid in this setting.

The main estimates for the effect of student debt on post-college location outcomes comes from

[15] This includes an indicator for the existence of a prior credit report, an indicator for a credit score, the credit score, and indicators for the most common accounts in early adulthood, auto loans and credit cards.

[16] College states other than California exist in the data as the UCCCP tracks Californians with a credit report prior even when they leave the state. The UCCCP sample also includes individuals who moved to California after college (with historical credit reports from time outside the state). College state FEs are included to capture selection bias involved with the two, but results are also robust to restricting the sample to only individuals who are in California during postsecondary education.

[17] College state is captured by location in year 3 from first educational loan. This closely follows the framework of Black et al. (2020), although they focus on post-college financial outcomes.

[18] The UCCCP does not include information on enrollment or degree-level. As a result, I am not able to distinguish increases in post-year 4 borrowing as additional borrowing for an undergraduate degree, new borrowing for a graduate or professional degree, or restructuring of existing loans.

the following difference-in-differences specification aggregating cohorts based on treatment status:

$$\begin{aligned} Y_{isc} = &\alpha + \beta_1 \text{Constrained}_i + \beta_2 \left[\mathbb{1}[\text{Cohort}_i > 2005] \text{ x Constrained}_i\right] \\ &+ \mathbf{X_i'}\boldsymbol{\beta_x} + \delta_c + \delta_s + \epsilon_{isc} \end{aligned} \tag{2}$$

where the coefficient of interest, β_2, captures the effect of being constrained across all cohorts. The rest of the specification, including covariates, fixed effects, and error clustering is the same as in the event study framework. To avoid distortion from year-specific fluctuations that affect constrained and unconstrained borrower outcomes equally, I continue to include flexible cohort year fixed effects rather than a simple indictor for the post-policy change period. This specification provides an estimate that is a weighted average of the effects for each treated cohort.

3.3 Identifying Constrained Borrowers

The empirical strategy identifying the effect of student debt on post-schooling location choice relies on the ability to classify students as likely or unlikely to be or have been constrained by *pre-policy-change* borrowing limits. I do so using observed loans in a student's first year of borrowing.[19] For students who entered in years unaffected by loan limit increases for first-year borrowing (pre-2008), I classify students who borrow exactly at the loan limit of \$2,625 to be likely constrained. For students who experienced increased borrowing limits in their first year, they are classified as likely to have been constrained if they borrow at or above the original limit of \$2,625. In all years, students who borrow below the original first-year limit are classified as unlikely to be constrained. This classification system is depicted in Figure 5.

This strategy assumes that students who borrow exactly at the loan limit in years before the increases would have borrowed more if given the possibility. Although it is impossible to verify this assumption using the data in the UCCCP, the distribution of first year borrowing across cohorts indicates that it is likely to be true for the majority of borrowers at the limit. As shown in Figure 6, the borrowing distribution for individuals who first took out education loans in 2002-2007 academic years has a large mass exactly at the limit that year (\$2,625). For individuals who first borrowed in the 2008 academic year, when the loan limit increased to \$3,500, the largest mass in the distribution of first year borrowing shifts to the new limit. This shift to the new limit is also observed for borrowers in 2009-2013 entry cohorts for whom the first year loan limit was \$5,500. This bunching and the quick shifts of the distribution to new limits suggest students are constrained at limits and would borrow more when given the possibility.[20]

Characteristics of constrained and unconstrained borrowers in the UCCCP are reported in Table 3. Federal laws prohibit creditors from discriminating against applicants on the basis of

[19] This is the same classification strategy used in Black et al. (2020). They estimate that 73 percent of all dependent undergraduate students who ever borrowed and graduated in 2016 borrowed in their first year.

[20] Another possibility is that students may be likely to accept financial aid as it is packaged by schools and nearly all four-year institutions include the maximum available Stafford Loans in aid packages (Marx and Turner, 2019). As argued by Black et al. (2020), this channel still induces additional borrowing and the empirical strategy still produces a causal estimate of student debt on student outcomes.

many personal characteristics, including sex, race, color, religion, and marital status. As a result, credit histories, and thus the UCCCP, contain little demographic information; however, I am able to observe age at first education loan and sex along with variables to characterize borrower's credit profiles. As shown in Row 1 of Table 3, the sample contains borrowers that were a little older than 18 at the time of their first education loan. Consistent with national statistics showing women make up a majority of recent postsecondary degree recipients, women make up slightly over half of the sample. As for credit characteristics of the UCCCP sample, about 30-40 percent had a credit report prior to opening an education loan, with this share increasing in later years. The share of individuals with a credit score, which is only calculated after 3 to 6 months of credit activity, also increases over the sample period from about 10 percent to 30 percent. For borrowers with a credit score, the average is in the low 600s, which falls in the 'fair' category. About 20 percent of individuals have a credit card account and very few have an auto loan.

The difference-in-differences framework allows me to compare the baseline characteristics of borrowers presented in Table 3 across treatment groups and the sample period. To do so, I modify Equation 2 only by omitting any baseline characteristics. The results of this analysis are reported in Table 4. As shown, individuals who are constrained in the post-period tend to be slightly older (5 days) and a slightly higher percentage are women (1.4 percentage points). There is no significant variation in the presence of a credit report, though constrained borrowers in the post-period are less likely to have a credit score and have a slightly lower credit score when present – both indicative that constrained borrowers in the post period may have less financial experience than in the pre-period; however, a larger share have credit cards. Although differences are small relative to variable means, I control for all of these characteristics in both the event study and difference-in-differences framework to reduce any bias introduced by changes in the sample population.

The estimates provided by the event study and difference-in-differences specifications reflect the effect of increased access to educational loans, but there are limits to the interpretation. Due to the setup and strategy for identifying the constrained status of borrowers, estimates reflect effects for students *already enrolled and borrowing*. One concern is that increased borrowing limits enabled students to enroll in postsecondary education that were previously too credit constrained to enroll; however, Marx and Turner (2019) find only minor effects of increased borrowing ability on enrollment. Another concern is the implications for school choice. While UCCCP data does not allow me to identify enrollment institutions, Black et al. (2020) find no evidence that increased access to borrowing led to more transfers from community college to four-year institutions in a sample of Texas students. Lastly, it is possible that students who are defined as unconstrained by first-year borrowing become constrained in subsequent years. To the extent that this occurs, results will underestimate the true effect.

3.4 Results

3.4.1 The Effect of Increased Loan Limits on Borrowing

Increased access to educational loans resulted in higher debt for students who were students likely to have been constrained by original loan limits. The results of the event study specification in Equation 1 when examining first-year borrowing are reported in Figure 7. As expected since first-year loan limits remain at the original level, the effect of increased borrowing limits on first-year borrowing is flat through the 2007 cohort. The estimated effect on first-year borrowing then increases sharply to around $2,000 by the 2010 cohort. This is similar to the magnitude increase of the expansion in first-year loan limits shown in Table 2, suggesting constrained students take near-full advantage of additional credit in the first year.

Increased loan access also resulted in additional cumulative borrowing through year 4 from entry – the last year of postsecondary enrollment for the majority of students. Estimated coefficients from the event study specification are illustrated in Figure 8. Cumulative borrowing of constrained relative to unconstrained students increases starting with the 2008 cohort and levels out around $4,000 by the 2010 cohort. Columns 1 and 2 in Table 5 present the corresponding difference-in-differences results when examining first-year and cumulative borrowing through year 4. The effect of increased loan access on first year and cumulative year 4 borrowing are $1,214 and $2,600, respectively.[21] This is roughly a quarter of the increase one might expect if all constrained students fully take advantage of higher loan limits. This could reflect misclassification of some individual's constrained status or changes in this status over the enrollment period. These explanations would both bias results down when examining borrowing and location choices, but do not pose a threat to identification.

3.4.2 The Effect of Increased Loan Limits on Post-College Migration

The additional debt for constrained borrowers caused them to choose initial post-college destinations with higher population densities. Figure 9 shows the event study coefficients when considering the effect of increased loan limits on county population density in year 6 from entry. As shown, there is a significant increase in year 6 county population density for all post-period cohorts except 2007 and 2008. Column 3 of Table 5 shows the aggregate estimate from the difference-in-differences specification. Additional loan access increased the year 6 county population density by 2.2 percentage points. Scaling by the additional debt estimated in the previous section, this suggests that $1,000 in additional student loans increased the population density of a borrower's year 6 county by 5.7 percent.

The additional debt for constrained borrowers also resulted in increased probability of locating in *top* urban areas after school. Figure 10 shows the event study estimates when considering the probability of locating in counties classified as being in the 95$^{\text{th}}$ percentile by population density. Figure 11 shows estimates from the same regression when considering the probability of locating in

[21] Black et al. (2020) estimate that an increase of $1,800 through year 4 in a national credit panel.

counties classified as a top metropolitan area with a population of more than 1 million, as specified by the Rural-Urban Classification Codes. A similar pattern to that when examining population density emerges with higher precision when considering these outcomes. Difference-in-differences estimation (Column 4 of Table 5) finds that a $1,000 increase in debt caused constrained borrowers to be 4.2 percentage points more likely to locate in a county in the 95$^{\text{th}}$ percentile by population density – a roughly 8 percent increase from the mean. Similarly, a $1,000 increase in student debt increased the probability of locating in a RUCC-defined top metropolitan area by 4.4 percentage points – a 6 percent increase from the mean. Columns 6 through 8 show that these results remain consistent when limiting the sample to individuals who reported a college address in California (year 3 from entry).

The Great Recession, officially starting in December of 2007 and ending in June of 2009, likely affected students differently depending on their academic level. Cohorts 2004 through 2009 all experienced the crisis at different points during their postsecondary education (assuming a 4-year degree). Event study results present some unexpected cohort heterogeneity that may reflect this fact. First, estimated effects on location outcomes for 2007 and 2008 cohorts are consistently lower despite additional borrowing for constrained individuals in the 2008 cohort. One possible explanation is that individuals entering in these years, when the effects of the recession were most acute, may have adjusted their career goals. Liu et al. (2019) find that individuals who began postsecondary education in recession years were less likely to major in business and finance, both of which are fields that have relatively high urban wage premiums (as explored further in Section 4). If this phenomenon is widespread in 2007 and 2008 cohorts, it is possible that urban areas, which would normally attract student debtholders because of higher nominal wages, did not offer higher wages for these individuals and this is why the effect is dampened for these cohorts. Unfortunately, I cannot control for or explore this potential channel as college major is not observed in the UCCCP.

Another puzzling feature of the event study estimates is that constrained borrowers in the 2006 cohort showed little increase in student debt through year 4, but subsequently exhibited increased urban post-college location choice. One possible explanation is a change in enrollment behavior. Long (2014) find a decline in full-time enrollment accompanied by an increase in part-time enrollment during the Great Recession.[22] As the 2006 cohort was only exposed to increased borrowing limits in year 4, it is possible that many had to shift to part-time enrollment, drop out, or extend the time to graduation. In the case of the latter, they may have benefited from additional borrowing capacity when they resumed full-time enrollment. Unfortunately, I cannot observe enrollment intensity or academic level so it is impossible for me to unpack this channel in this setting.[23]

While data limitations prevent me from digging into these dynamics further in the current setup, subpopulation analysis allows me to minimize any confounding factors introduced by the

[22] Van Horn et al. (2012) find that most individuals who were no longer enrolled as full-time students report the inability to afford the cost of college as the main reason.

[23] The main issue is that additional borrowing in year 5, 6, etc. could be used to finish an undergraduate degree or for a graduate/profession degree. To some extent, Column 8 of Table 4 suggests little effect of additional borrowing on education attainment. Nevertheless, all specification include controls for education level reported in credit histories.

economic downturn. Specifically, I drop all cohorts affected by the Great Recession (2004-2009) from the difference-in-differences estimation. Results of this restricted analysis are presented in Table 6. As shown, the results are very similar albeit larger given the omission of cohorts with only partial increases in loan access.

In Table 7, I explore the persistence of the effect of additional borrowing on the location choice of constrained borrowers. For all three outcomes measures characterizing destination counties, the effects are persistent through year 9 after entry (the last year with available data). While I cannot determine graduation dates at the individual level, this would typically be 5 years after exiting school for most individuals. This provides some initial evidence for how student debt may impact location choices in the long run as well.

4 Mechanisms

While results in the previous section show that additional student debt causes individuals to locate in urban areas at higher rates, data limitations prevent me from examining potential mechanisms within that empirical framework. To that end, there are two primary potential explanations for the identified effect of student debt on post-college location choice. The first is the proposed mechanism that student debt makes individuals more sensitive to nominal wages than local price levels. The second is that individuals who have a preference for urban areas are willing to borrow more for postsecondary education. This could be the case if, for example, individuals who aspire to live and work in New York City after school would also like to attend college in a large city, where tuition and related costs tend to be higher and may require additional borrowing. A necessary component of this alternative mechanism is that individuals are making debt, college, and post-college location choices at least partially as a joint decision. In this section, I first show that the association between student debt and urban post-college location choice is robust to controlling for the second potential mechanism – suggesting it plays at most a minor role. Second, I present descriptive evidence supporting the proposed mechanism that student debt makes borrowers more sensitive to nominal wage differences in location choices.

To better understand the proposed mechanism and produce a testable implication of it, consider individuals with and without student debt (D) that choose a location j from choice set J based only on nominal wages (W_j) and local prices (P_j):[24]

$$\max_{j \in J} \; ln\left(\frac{W_j - D}{P_j}\right) \tag{3}$$

$$\epsilon_w = \frac{1}{1 - \frac{D}{W_j}} \geq 1 \;\; , \;\; \epsilon_p = -1$$

When individuals consider location options, one step is the consideration of the partial derivative with respect to wages and prices. For individuals with no student debt, their location choices are

[24]This framework is a simplified version of the standard Rosen (1979) and Roback (1982) frameworks.

equally elastic to nominal wages and local price levels. In this case, the standard result holds: individuals locate in places with the highest real wages. When individuals carry student debt, the elasticity with respect to wages is strictly greater than the elasticity with respect to prices. As a result, individuals with student debt choose to locate in higher nominal wage locations. Additionally, this framework shows that, among individuals with the same level of student debt, urban locations should only be relatively more attractive if they provide significantly higher nominal wages for that individual. This is the testable implication I will examine further.

4.1 Data

To explore these potential mechanisms, I exploit rich data from the Education Longitudinal Study of 2002 (ELS). The main advantage of using the ELS instead of the UCCCP is that it provides a rich set of individual-level characteristics to perform heterogeneity analysis on the association between student debt and urban post-college location choice. It contains a nationally representative sample of over 16,000 10$^{\text{th}}$ graders in 2002 and 12$^{\text{th}}$ graders in 2004. The ELS includes a survey of students, their parents, and school officials with the stated goal of understanding student trajectories from high school through postsecondary education and into the workforce. Individuals are surveyed 4 times: 2 times during high school years (2002 and 2004); 1 time two years after prospective high school graduation during typical college-going years (2006); and 1 time 8 years after prospective high school graduation (2012), which would typically be 4 years after postsecondary education is completed. At each point the survey documented their location at the county level. The ELS also collects information on total amount borrowed for postsecondary education. Demographic information includes parental education/occupation, ability (via high school grade point average), college major, sex, and race as additional factors that may influence migration choices.

To reduce unobserved heterogeneity, I limit the ELS sample in two ways. First, I examine only outcomes of those with a postsecondary degree – omitting both individuals who end the observation period with higher or lower levels of educational attainment. As labor market opportunities vary significantly by education level, this sample restriction aims to eliminate this confounding factor. Second, I limit the sample to individuals who borrowed a positive amount to fund postsecondary education, i.e. the intensive margin. This was done to reduce unobserved heterogeneity as there is likely a larger amount of unobserved factors that enter into the borrowing decision on the extensive margin than on the intensive margin. Finally, both restrictions offer the added benefit of constructing a sample that is similar to the UCCCP sample used in Section 3. The resulting sample includes roughly 2,200 individuals.

4.2 Urban Preference Mechanism

I begin by showing a robust association between student debt and urban post-college location choice in the ELS sample. I do so by estimating the following cross-sectional regression:

$$\begin{aligned}\text{Post-College Cty. Urban}_i = \alpha + \beta_1 \text{Debt}_i + \mathbf{X_i'}\boldsymbol{\beta_x} + \mathbb{1}[\text{Origin Cty. Urban}_i] \\ + \mathbb{1}[\text{College Cty. Urban}_i] + \gamma_{\text{Degree Field}(i)} + \epsilon_i \quad ,\end{aligned} \tag{4}$$

where the outcome of interest is the urban classification of the post-college location of an individual. The urbanicity of a location is defined in three ways: (1) a continuous measure of a county's population density (in logs); (2) an indicator function for if the county is in the 80$^{\text{th}}$ percentile by population density; and (3) an indicator function for if the county is classified in the top category (metropolitan area with a population of at least 1 million) using the Rural-Urban Continuum Codes (RUCC). The coefficient of interest, β_1, captures the association between urban post-college location choice and student debt (in logs).

To partially account for urban preferences, I control for the urban classifications of high school and college locations based on population density or RUCC. To account for geospatial variation in industry concentration and the resulting variation in available employment opportunities for individuals with different academic majors, I include fixed effects for postsecondary degree field as reported in the ELS (γ). I also account for a rich set of individual characteristics ($X_i'\beta_x$), including father's education, father's income, race, sex, and ability (via high school GPA). The regression is weighted based on the ELS sample design with errors clustered at the high school level.

To provide a baseline, I first estimate the coefficients from Equation 4 when not including the proxies for urban preference. These results are presented in Columns 1-3 of Table 8. As Column 1 shows, student debt has a positive and significant relationship with post-college county population density. A 10 percent increase in student debt is associated with a 1.4 percent increase in post-college county population density. The positive relationship between student debt and urban post-college location choice is also present when we consider locating in top urban areas. As shown in Column 2, a 10 percent increase in student debt is associated with a 0.3 percent increase in the probability of locating in a county in the 80$^{\text{th}}$ percentile of population density. Results in Column 3, where urbanicity is defined as the top category in the RUCC, are similar in magnitude but estimated with lower precision.

Controlling for urban preference reduces the magnitude of these estimates, but only accounts for roughly one third of the association between student debt and urban post-college location. Estimated coefficients from Equation 4 when including the proxies for urban preference are presented in Columns 4-6 of Table 8. As shown, a 10 percent increase in student debt is still associated with a 1.1 percent increase in post-college county population density. The association is also still present when we consider locating in top urban areas. As shown in Column 5, a 10 percent increase in student debt is associated with a 0.2 percent increase in the probability of locating in a county in the 80$^{\text{th}}$ percentile of population density.[25] Results in Column 6, where urbanicity is defined as the top category in the RUCC, are qualitatively similar but now lack precision.

These results provide two indications that the urban preference mechanism is unlikely to be the primary driver of the results in Section 3. First, as expected, both origin and college county urban indicators have a strong association with post-college urban location; however, a robust relationship between student debt and post-college location choice is still found despite including these proxies for urban preference. Second, as noted above, the urban preference mechanism requires that individuals make college, student debt, and post-college location choices at least partially as a joint decision. Since Conzelmann et al. (2022) have shown that a large portion of college graduates live and work near their postsecondary institution, this implies that individuals making a joint decision that includes post-college location would likely have the same college and post-college location. This suggests one possible way to address this concern would be to examine individuals who *did not* locate in the same place for postsecondary education and post-college life. These individuals may be making their post-college location choices more independently of college and debt decisions. Columns 7 through 9 in Table 8 conduct the same analysis on the subpopulation of individuals who left their college county. As shown, the results are qualitatively similar and slightly larger although not significantly so.[26]

4.3 Association Driven by Individuals Facing High Urban Wage Premiums

The proposed mechanism for why additional student debt would cause individuals to locate in urban areas at higher rates is that student debt makes individuals more sensitive to nominal wage differences across locations than differences in local price levels. As noted above, an implication of this mechanism is that the association between student debt and urban post-college location choice should be higher for individuals that face steeper urban wage premiums. To test this, I estimate the wage premium associated with increases in population density for each degree field reported in the ACS 2010 sample.[27] By mapping ELS-reported postsecondary academic major to ACS degree field, I can then determine the urban wage premium that each student faces upon graduation.

Empirically, I expand on Equation 4 by interacting student debt with an indicator function for whether an individual entered a field with an above- or below-median urban wage premium:

$$\begin{aligned}\text{Post-College Cty. Urban}_i = \alpha &+ \beta_1^0 \text{Debt}_i \times \mathbb{1}[\text{Below Median Urban Wage Premium}_i] \\ &+ \beta_1^1 \text{Debt}_i \times \mathbb{1}[\text{Above Median Urban Wage Premium}_i] \\ &+ \mathbf{X_i'}\boldsymbol{\beta_x} + \mathbb{1}[\text{Origin Cty. Urban}_i] \\ &+ \mathbb{1}[\text{College Cty. Urban}_i] + \gamma_{\text{Degree Field}(i)} + \epsilon_i \quad ,\end{aligned} \tag{5}$$

[25] As the 80$^{\text{th}}$ percentile threshold is arbitrary, I adjust the cutoff in Figure A10.

[26] The same result holds when considering the subpopulation of individuals that left their college commuting zone, which perhaps is a more accurate measure of the entire local labor market where the college is located.

[27] I derive wage premium as the coefficient on PUMA population density interacted with postsecondary degree field in a regression of nominal wages on sex, age, race fixed effects, and degree field fixed effects. Sample includes employed prime working-age (22-54) individuals making a positive income. ACS degree field only recorded for Bachelor's degree holders.

where β_1^0 and β_1^1 capture the association between student debt and urban post-college location choice for individuals facing below- and above-median urban wage premiums by field, respectively. The rest of the specification remains the same as in Equation 4 with the exception of the degree field fixed effects which now capture ACS rather than ELS postsecondary degree fields. Though the divisions change slightly, these fixed effects still capture spatial variation in employment opportunities by degree field.

The results of this analysis are reported in Table 9. Columns 1, 4, and 7 repeat the results in Table 8, while columns 2, 5, and 8 conduct this analysis on the sample that have majors mapped to ACS degree fields. Results are nearly identical, suggesting there is no issue of sample selection in matching rates by ELS major. Columns 3, 6, and 9 report the results from the estimation of Equation 5 on the three measures of post-college location urbanicity. As shown, the positive relationship between student debt and urban post-college location is driven by individuals facing above-median urban wage premiums. For individuals who enter fields with below-median urban wage premiums, additional student debt is not associated with any increased likelihood of locating in urban areas. This is consistent with the proposed mechanism that student debt makes individuals more sensitive to spatial variation in nominal wages.

5 Spatial Equilibrium Model

In this section, I introduce a spatial equilibrium model that delivers the empirical regularities presented so far and additionally affords a framework to simulate counterfactual outcomes under various policy interventions. The setup is similar to the canonical Rosen (1979) and Roback (1982) framework, but I allow for heterogeneity in workers' productivity and location preferences as well as cities' productivity and housing supply. To focus on the role of student debt, I minimize departures from 'standard' spatial equilibrium models in the literature, and follow recent versions from Diamond (2016), Hsieh and Moretti (2019) and Albert and Monras (2017).

The model has J regions that differ by low- and high-skill productivity, amenity level, and housing supply. Firms in all locations produce an identical good that is freely traded between regions at no cost. Individuals consume this tradable good as well as a non-tradable local good, which for simplicity I will call housing. The housing sector generates the congestion force in the model, as housing prices and population have a positive relationship.

In the rest of the section, I imbed the basic mechanism outlined in Equation 3 in a comprehensive discrete location choice model with two periods. In period 1, individuals choose a level of education/skill and take out student debt if they choose to become high-skill workers. In period 2, individuals choose a location, work, consume local and tradable goods, and pay off student debt. In period 1, the model only has one component: a combined education and debt choice by individuals. In period 2, the model has three components: labor supply (location choice), labor demand, and the housing market. A banking sector spans both periods to facilitate borrowing for education. To allow for policy counterfactuals involving debt relief and tuition-free postsecondary education, a 'federal' government institutes a tax to cover the associated cost. Because decisions in period 1 are made recursively based on utility in period 2, I present period 2 setup first followed by period 1.

5.1 Labor Supply (Period 2)

Individuals enter period 2 as low or high-skill workers ($g \in \{l, h\}$). If they are part of the high-skill group, they also carry a positive student debt (D_g), which is constant for all individuals in the group. Student debt is zero for all low-skill individuals. Workers choose location j, consumption of a tradable good C_T (numeraire good), and consumption of a non-tradable good, C_{NT} (at price p_j). The utility of individual i from group g locating in j is given by:

$$\ln U_{igj} = (1-\beta)\ln C_T + \beta \ln C_{NT} + \ln A_{gj} + \ln \epsilon_{ij} \ , \tag{6}$$

where $(1-\beta)$ denotes the expenditure share devoted to tradable goods. A_{gj} is the utility derived from local amenities in location j, and is group-specific. Finally, individuals have an idiosyncratic location preference ϵ_{ij} that has a Frechét distribution with inverse shape parameter $\alpha \geq 0$, which governs the variance of the idiosyncratic taste shocks. C_{NT} represents the consumption of housing and other non-tradable goods which need to be consumed in location j. For simplicity, I will call this housing.

Individuals maximize utility subject to the following budget constraint:

$$C_T + p_j C_{NT} + D_g \leq W_{gj}, \tag{7}$$

where wages, W_{gj}, vary by group (skill level) and location. The demand for each good is given by:

$$C_T = (1-\beta)(W_{gj} - D_g) \quad , \quad C_{NT} = \beta \left(\frac{W_{gj} - D_g}{p_j} \right) \tag{8}$$

Plugging the optimal demand functions into the utility function, the indirect utility of living in location j is:

$$\ln V_{igj} = \ln V_{gj} + \ln \epsilon_{ij} = \kappa + \ln(W_{gj} - D_g) - \beta \ln p_j + \ln A_{gj} + \ln \epsilon_{ij}, \tag{9}$$

where $\kappa = \beta \ln \left[\beta (1-\beta)^{\frac{1-\beta}{\beta}} \right]$. Indirect utility can be broken into a group-specific valuation of location j, V_{gj}, and the individual idiosyncratic preference for locating in j, ϵ_{ij}.

Given the distribution of the idiosyncratic taste parameter, the share of workers in group g locating in j is equal to:

$$\begin{gathered} \pi_{gj} = \frac{(V_{gj})^{\alpha}}{\sum\limits_{k} (V_{gk})^{\alpha}} = \left(\frac{V_{gj}}{V_{gJ}} \right)^{\alpha} \\ \text{where } V_{gJ} = \left[\sum_{j' \in J} (V_{gj'})^{\alpha} \right]^{\frac{1}{\alpha}} , \end{gathered} \tag{10}$$

where V_{gJ} represents the expected value, or welfare, of being in this economy for a worker in group

g. The shape parameter α on the idiosyncratic preferences governs the elasticity of migration with respect to changes in indirect utility of locations.[28] Assuming each worker inelastically supplies one unit of labor, the overall supply of low- and high-skill labor in location j is given by:

$$L_j = \pi_{lj} N_{lJ} \tag{11}$$

$$H_j = \pi_{hj} N_{hJ}, \tag{12}$$

where N_{lJ} and N_{hJ} are total low- and high-skill workers in the economy, respectively.

I, the econometrician, observe wages (W_{gj}), low- and high-skill population (L_j and H_j), price levels (p_j), and high-skill debt levels (D_g). Exogenous amenities (A_{gj}) and workers' idiosyncratic taste for each location (ϵ_{ij}) are unobserved. Parameters to be calibrated are the worker expenditure share devoted to non-tradable goods, β, and the shape parameter for the idiosyncratic location preference, α.

5.2 Labor Demand (Period 2)

Each location j has a single firm that produces the tradable goods with a production function that combines low- and high-skill labor as the only inputs. The output in location j is given by:

$$\begin{gathered} Y_j = \left(\theta_{lj} L_j^{\rho} + \theta_{hj} H_j^{\rho}\right)^{\frac{1}{\rho}} \\ \theta_{gj} = exp(\epsilon_{gj}) \end{gathered} \tag{13}$$

where firms combine low- and high-skill labor as imperfect substitutes in production with a constant elasticity of substitution, $\frac{1}{1-\rho}$.[29] Skill-specific productivity, θ_{lj} and θ_{hj}, differ across locations and are determined exogenously. Labor markets are assumed to be perfectly competitive such that wages equal the marginal product of labor. Profit maximization leads to the following demand for low- and high-skill labor in location j:

$$W_{lj} = \theta_{lj} L_j^{\rho-1} (\theta_{lj} L_j^{\rho} + \theta_{hj} H_j^{\rho})^{\frac{1-\rho}{\rho}} \tag{14}$$

$$W_{hj} = \theta_{hj} H_j^{\rho-1} (\theta_{lj} L_j^{\rho} + \theta_{hj} H_j^{\rho})^{\frac{1-\rho}{\rho}} \tag{15}$$

I observe wages (W_{gj}), low, and high-skill population (L_j and H_j). Local exogenous productivity, ϵ_{lj} and ϵ_{hj}, are unobserved. The parameter governing the elasticity of substitution between high- and low-skill labor, ρ, needs to be calibrated.

[28] Previous work, including Diamond (2016) and Bound and Holzer (2000), has found differences between high- and low-skill workers in migration elasticity with respect to real wages. In the counterfactual analysis, I simulate outcomes with both uniform and skill-specific shape parameters, α_g.

[29] This production function is prominent in the literature examining wage inequality and its relation to supply of high- and low-skill labor (Katz and Murphy, 1992; Katz and Autor, 1999; Acemoglu, 2002; Diamond, 2016).

5.3 Housing Market (Period 2)

In each location, the supply of housing is produced using land for homes (T_j), which is a fixed factor, and the tradable good (Y_j^T) according to the following production function:[30]

$$Y_j^{NT} = \zeta_j^{-\zeta_j}(Y_j^T)^{\zeta_j}(T_j)^{1-\zeta_j} \quad , \tag{16}$$

where $1-\zeta_j$ is the weight of land in the production of housing. I assume land is owned by absentee landlords.[31] This results in the following housing supply equation:

$$Y_j^{NT} = p_j^{\gamma_j} T_j, \tag{17}$$

where $\gamma_j = \frac{\zeta_j}{1-\zeta_j}$ is the housing supply elasticity. Note that γ_j differs across locations, which could capture a combination of factors such as limits on the amount of land or land use regulations (Saiz, 2010). Cities with limited land or cumbersome land use regulations have a lower γ_j (lower elasticity), while cities with few limits on available land or land use have a higher γ_j (higher elasticity).

Total demand for housing is given by the sum of the local demands of individuals in each location. Local housing prices are implicitly defined by market clearing in each location:

$$p_j^{\gamma_j} T_j = L_j \beta \frac{W_{lj}}{p_j} + H_j \beta \frac{W_{hj} - D_h}{p_j} \tag{18}$$

This equation captures one difference between my model and standard spatial equilibrium models (such as Hsieh and Moretti (2019)): the demand for housing in each location depends on the size *and* skill composition of the population.[32]

As the econometrician, I observe price levels (p_j), low- and high-skill population (L_j and H_j), and high-skill debt levels (D_h). Location-specific housing supply elasticities (γ_j) need to be estimated and land available for homes (T_j) is unobserved.

5.4 Government (Period 2)

In the baseline model, there is no government; however, certain policy counterfactuals involving student debt forgiveness and the elimination of tuition require a government to facilitate them. In each case, there is an aggregate cost, T, that will funded by uniform taxation of nominal wages at rate τ.[33] The government budget constraint is given by:

$$T = \sum_j (L_j W_{lj} + H_j W_{hj}) * \tau \tag{19}$$

[30] Housing production function also used in Albert and Monras (2017). Resulting housing supply identical to reduced-form version used in Hsieh and Moretti (2019).

[31] A common assumption in the literature. See Albert and Monras (2017); Diamond (2016); Hsieh and Moretti (2019); Eeckhout et al. (2014).

[32] This is similar to the model in Albert and Monras (2017), where demand for housing depends on location-specific size and *immigrant* composition of the population.

[33] In the case of debt forgiveness, $T = D_h * N_{hJ}$. When considering a free college option, $T = D_h * N_{hJ}^{CF}$ for the counterfactual high-skill population.

It is important to note that, unlike debt repayment, the tax is proportional to wages and, therefore, does not distort location choices. This can be seen by considering the group-specific indirect utility of living in location j (Equation 9) under a policy eliminating debt and instituting a tax:

$$\ln V_{gj} = \kappa + \ln(W_{gj} * (1-\tau)) - \beta \ln p_j + \ln A_{gj} \tag{20}$$

As shown, τ is separable from wages and thus cancels out in Equation 10 determining worker shares in each location.

5.5 Education and Debt Choice (Period 1)

In period 1, individuals choose their skill level for period 2 employment. There is no cost to becoming a low-skill worker. To become a high-skill worker, individuals must pay a monetary cost, D, that is paid by borrowing in period 1 and repayment in period 2. There is also an idiosyncratic utility cost, z_i, which follows a Frechét distribution with inverse shape parameter $\mu \geq 0$. This utility cost captures two traits: (1) an individual's ability to become a high-skill worker (e.g., aptitude); and (2) an individual's ability (including willingness) to take on student debt. The joint education-debt decision is made based on the period 2 expected value of being in this economy as a worker of each type ($g \in \{l, h\}$):

$$U_i^g = V_{gJ} + z_i * \mathbb{1}[g = h], \tag{21}$$

where V_{gJ} is the same as in Equation 10 and represents expected value of being in the period 2 economy as a worker in group g. An individual chooses to become a high-skill worker if $U_i^h > U_i^l$.

Given the distribution of the idiosyncratic preferences, the share of individuals that become high-skill workers can be represented as:

$$s_h = \frac{(e^{U_i^h})^\mu}{(e^{U_i^l} + e^{U_i^h})^\mu} \tag{22}$$

As shown, μ, governs the elasticity of worker type choice with respect to changes in expected utility of being a high-skill worker. This includes changes induced by adjustments to student debt – an important detail for the counterfactual exercises.

As the econometrician, I do not observe U_i^H, U_i^L, μ, or z_i; however, as discussed further in the calibration and counterfactual analysis sections, the relevant dimension is $\frac{\partial s_h}{\partial D}$.

Banking Sector (Periods 1 and 2). For simplicity, the banking sector offers credit to all individuals at 0 interest rate.

5.6 Equilibrium

Definition I. *The spatial equilibrium is defined as follows:*

1. *Workers decide where to live and how much to consume of each good.*

2. *Firms decide how many workers to hire to maximize profits.*

3. *Landlords decide how much housing to supply.*

4. *Tradable goods, labor, and housing markets clear.*

5. *Government budget constraint satisfied.*

6 Calibration

The location choice set is defined as the 722 1990-defined Commuting Zones (CZ) in the contiguous United States.[34] Related prior literature has used metropolitan statistical areas, but this does not allow for full coverage of the U.S. and would omit migration across the dimension of interest: the urban-rural divide. There are two main sources of data for the structural analysis: the 2015-2019 American Community Survey (ACS)[35] and the National Center for Education Statistics (NCES). ACS provides the baseline levels of wages and population for low- and high-skill workers. Local prices are derived from housing costs in ACS, following the approach of Moretti (2013) for 1990 Commuting Zones. Baseline levels of student debt among high-skill workers was estimated from the NCES data on cumulative borrowing for bachelor's degree-holders.[36]

6.1 Parameters in Labor Supply

There are two parameters that need to be calibrated in labor supply: worker expenditure share devoted to non-tradable goods, β, and the shape parameter governing idiosyncratic location preference and migration elasticity, α. The expenditure share devoted to non-tradable goods, β, will be taken from the literature and set to 0.6.[37] Migration elasticity, α, is taken from the literature as well, but there is far less consensus. Most estimates center around 4, but some are as low as 0.4 and others as high as 12.8.[38] The breadth of these estimates likely reflect the paper-specific variation and estimation strategy in each. Higher values may also reflect elasticity over a longer time horizon to some extent. Additionally, the literature has found that migration elasticity may vary by skill group, with low-skill workers being less mobile than high-skill workers.[39] As this parameter is a crucial component to generating counterfactual population distributions, the simulations below will report results under 2 values of migration elasticity. First, I consider homogeneous migration elasticity across skill groups and set it around the median in the literature, 4. I then consider

[34] Alaska and Hawaii excluded due to the unique nature of their labor markets and the discontinuous costs associated with moving to either location.

[35] Downloaded from IPUMS (Ruggles et al., 2022). See Section Appendix B for more information.

[36] Payments based on typical repayment period (10 years) and interest rate (4.99%) for federal student loans.

[37] Hsieh and Moretti (2019) find β of 0.6, which is used by Monras and Albert (2022). Moretti (2013) finds local good expenditure of 0.59. Diamond (2016) uses 0.62, supported by analysis of the Consumer Expenditure Survey.

[38] Hsieh and Moretti (2019) and Hornbeck and Moretti (2018) estimate 3.3. Suárez et al. (2016) estimate 0.75-4.2; Caliendo et al. (2019) estimate 2; Albert and Monras (2017) estimate 12.8; and Monras (2018) estimate 0.4.

[39] Hornbeck and Moretti (2018) estimates $\alpha_l = 2.6$ and $\alpha_h = 6.7$; Diamond (2016) estimates $\alpha_l = 2.1$ and $\alpha_h = 4$, but that low-skill migration elasticity is only significant at the 10-percent level. Bound and Holzer (2000) find that college workers' migration is elastic to local demand, but that low-skill workers are inelastic.

heterogeneity in this parameter by setting low- and high-skill migration elasticity to the values estimated in Hornbeck and Moretti (2018).

Exogenous amenities (A_{gj}) are estimated from Equation 10. Specifically, by setting one location (k) as the reference location, we can consider the share of individuals in skill group g locating in j relative to k :

$$\ln\left(\frac{\pi_{gj}}{\pi_{gk}}\right) = \alpha \ln\left(\frac{\frac{W_{gj}-D_g}{P_j^\beta}}{\frac{W_{gk}-D_g}{P_k^\beta}}\right) + \alpha \ln\left(\frac{A_{gj}}{A_{gk}}\right) \tag{23}$$

Population shares, wages, and debt are all observed, allowing me to derive amenities as the residual. This allows for a perfect fit of the location choice data. In counterfactual analysis, amenities will remain fixed.

6.2 Parameters in Labor Demand

The parameter governing the elasticity of labor substitution in production, ρ, will be taken from the literature. Diamond (2016) and Card (2009) provide estimates using the same production framework, finding estimates of $\rho = 0.4$ and $\rho \in (0.3, 0.6)$.[40] For the counterfactual exercises, I will use ρ equal to 0.4, which implies an elasticity of labor substitution of 1.7. With ρ, I estimate skill-specific productivities for each location j using labor demand (Equation 14) and baseline data on wages and population. Exogenous productivity will remain fixed in counterfactual simulations.

6.3 Parameter in Housing Market

There are two elements of the housing market that need to be calibrated or estimated: elasticity of housing supply in each location (γ_j) and the land available for housing (T_j). Location-specific elasticities of housing supply, γ_j, are calculated by commuting zone similar to Saiz (2010) and Howard et al. (2018). The land available for housing, T_j, is calculated using the housing market clearing condition (Equation 18) and observed low- and high-skill population, wages, debt, and price levels.

6.4 Parameters in Education/Debt Choice

For the counterfactual analysis, the parameter of interest is the elasticity of high-skill worker population share with respect to required student debt ($\frac{\partial s_h}{\partial D}$). Although estimates of this elasticity exist in the literature, the counterfactual analysis below only requires an estimate for the change in high-skill worker share associated with one discrete change in debt: a free college option. As most of the literature estimates this elasticity from marginal changes in the cost of attendance, a complete elimination of cost is likely to have an effect that is different from estimates extrapolated from this literature. For this reason, I calibrate the change in high-skill worker share associated with the introduction of a free college option from the literature examining this exact scenario.

[40] Katz and Autor (1999) also provides similar estimates using earlier data.

Kane (2003) exploits discontinuities in the California Grant program eligibility, which is a program that provides enough aid to cover the cost of attending University of California or California State University institutions. They estimate that a free college option increases enrollment by 3-4 percentage points. In a similar analysis examining expansion of the Georgia HOPE Scholarship Program, which waives tuition and fees for eligible students, Dynarski (2008) finds that a free college option increases college completion by 3 percentage points.

7 Counterfactual Analysis

With the model fully calibrated, I now use it to estimate counterfactual outcomes under four scenarios. First, I simulate outcomes if the average debt required for postsecondary education had stayed constant at the 1980 level. The purpose of doing so is to estimate the contribution of the growth in student debt to post-1980 skill-based geographic sorting. In the remaining counterfactual simulations, I estimate the effect of 3 policy proposals that address student debt on sorting, consumer welfare, and aggregate output: (1) debt forgiveness for *existing* borrowers[41]; (2) income-driven repayment plans for *existing* borrowers; and (3) the institution of tuition-free postsecondary education (accessible to all individuals). When considering debt forgiveness or the elimination of tuition, I pay for the policy with a uniform tax on all individuals.

To analyze effects on skill-based geographic sorting for each counterfactual, I use the model to simulate the population distribution then discuss the effect on the *level* of skill-based geographic sorting. I do so by calculating the correlation between the counterfactual high-skill worker shares and the population density in each commuting zone.[42] This approach, consistent with the literature's method of documenting skill-based sorting, provides a concise and transparent comparison to observed sorting. The observed levels of spatial sorting are presented in Figure 12. In the case of setting debt to the 1980 level, I also consider the effect on the level *relative to 1980*.

The model also enables me to reflect on aggregate outcomes of output and consumer welfare when considering the three policy proposals. The change in aggregate output can be calculated by summing across location-specific production (Equation 13) under baseline and counterfactual population estimates. The change in aggregate consumer welfare is measured using the expected value of being in the baseline and counterfactual economy for a worker in each group (V_{gJ} in Equation 10):

$$\widehat{W} = \frac{N_l}{N}\widehat{V}_{lJ} + \frac{N_h}{N}\widehat{V}_{hJ} \tag{24}$$

The results of the counterfactual analysis, reported in Table 10-12, are broadly split into two panels for partial and full equilibrium simulations separately as both offer unique advantages and together provide a more complete picture. Partial equilibrium results allow for only a labor supply response via migration, but keep wages and local prices fixed. By remaining agnostic about wage and price responses, both of which have been scarcely estimated in the literature for rural areas, the

[41] The debt forgiveness program proposed by the Biden administration forgives up to $20,000 for student debtholders, which is paid for by taxpayers. In my counterfactual analysis, I consider complete debt relief.

[42] Results consistent across alternative measures of urbanicity, including local wages and prices.

partial equilibrium results offer the most straightforward analysis. Additionally, partial equilibrium results provide an understanding of short-term outcomes. Full equilibrium results, in contrast, allow for wage and local price responses. In doing so, they account for spillover effects on low-skill workers and perhaps more accurately depict long-term outcomes. In both, location- and skill-specific amenities remain fixed. Individuals can adjust their joint education-debt decisions in both considerations of tuition-free postsecondary education. Within each, I simulate the model under various values of the most important parameter: migration elasticity to real wages, as discussed in Section 6.

Counterfactual simulations that hold wages and prices fixed (partial equilibrium) may overestimate the effect of resorting for two possible reasons. First, to the extent that prices in initially low-population areas respond to a rising population, this would discourage large growth in rural areas depending on location-specific housing supply elasticity. As a result, partial equilibrium solutions would overestimate the counterfactual resorting. Second, the shape of skill-specific demand curves. In the case of traditional downward-sloping demand for skill workers, wages for high-skill workers would increase in areas where the high-skill population declines – in turn, reducing the outflow. This would mean that partial equilibrium counterfactuals are an overestimate of resorting.[43]

7.1 Counterfactual 1: Debt Reversed to 1980 Level

I begin the counterfactual analysis by examining the role of rising student debt in post-1980 skill-based geographic sorting. The partial equilibrium results for this analysis are reported in Panel A of Table 10, Columns 1 and 2. Column 1 shows the percentage reduction in post-1980 sorting due to a decrease in debt to the 1980 level. In other words, these estimates reflect the share of post-1980 sorting that can be attributed to the growth of student debt over this period. The estimates range from 11.5 percent to 18.7 percent in the partial equilibrium solutions. Since wages and prices remain fixed, there are no spillovers to low-skill workers and variation in partial equilibrium solutions reflect only adjustments to migration elasticity for high-skill workers: higher elasticity equates to a larger exodus from urban areas when debt is decreased. Column 2 reports the estimated impact on the *level* of 2019 geographic sorting. The implications for geographic sorting are directionally the same but smaller in magnitude when considering full equilibrium solutions (see Panel B of Table 10, Column 1 and 2). The growth of student debt explains 3.5-4.5% of the post-1980 skill-based geographic sorting in this case – a smaller adjustment that reflects the dampening effects of rising rural prices and urban high-skill wages over the long-term. These results are also depicted in Figure 13 and 14 under the heterogeneous migration elasticities provided by Hornbeck and Moretti (2018). As shown, the observed growth of student debt resulted in the largest increases in high-skill worker

[43] Alternatively, demand for high-skill workers could be upward-sloping depending on the strength of agglomeration economies in production. For simplicity, the model presented here does not incorporate agglomeration economies, but Diamond (2016) offers a framework to do so. If high-skill worker demand exhibits strong enough agglomeration economies, wages for high-skill workers would decline in urban areas as high-skill workers relocate. As a result, partial equilibrium results would *underestimate* the reduction in sorting.

shares in dense urban commuting zones, while rural areas experience a decrease in their high-skill workers shares relative to the counterfactual where student debt remains at the 1980 level in real terms.

7.2 Counterfactual 2: Debt Forgiveness

Next, I examine the model-predicted outcomes for the first of three counterfactual policies: debt forgiveness for *existing borrowers*. To facilitate debt forgiveness, the government levies a uniform tax on all individuals that is a fixed percentage of wages. As the tax is a constant share of income, it does not distort the spatial allocation of low or high-skill workers. Predictions for the effect on the *level* of skill-based geographic sorting in 2019 are reported in Column 3 of Table 10. In partial equilibrium, debt forgiveness significantly reduces sorting by 8.1-14.0 percent depending on the assumed migration elasticities. In full equilibrium simulations, this percentage drops to a still significant 2.4-3.4 reduction. Debt forgiveness and the resulting migration decisions increase welfare for high-skill workers, but comes at the cost of low-skill workers who now pay a tax without the benefit of debt relief (Column 1 in Table 11). There is also a reduction in aggregate output as high-skill workers relocate to lower-productivity areas, as shown in Table 12, but these results should be interpreted with caution as production functions outside of urban areas have been poorly estimated in the literature.

7.3 Counterfactual 3: Income-Driven Repayment

The second counterfactual policy proposal I consider is switching all high-skill workers to income-driven repayment plans. To do so without reducing the aggregate debt level, I calibrate the repayment rate such that aggregate debt repayments remain fixed at the baseline level.[44] The resulting effect on skill-based geographic sorting is reported in Column 4 of Table 10. In partial equilibrium, the results are the same as in the counterfactual of debt forgiveness – as expected given that both entail a shift to debt repayment methods that do not distort location choices. In the full equilibrium solutions, the picture is a bit different: the reduction in spatial sorting is lower than under debt forgiveness. When considering debt forgiveness, the costs of education are paid uniformly by all; however, with income-driven repayment, the costs remain with high-skill workers. This does not distort high-skill worker location choices, but it does affect prices in the full equilibrium solution. It lowers prices in areas with larger high-skill worker shares (see Equation 18), which in turn reduces the exodus of high-skill workers. This dampening effect is stronger when high-skill workers are more mobile than low-skill workers. The maps provided in Figure 15 and 16 provide insight on how resorting affects commuting zone high-skill worker shares across the country. As shown, the largest declines are in commuting zones containing large metropolitan areas like New York City, Los Angeles, and Austin, Texas. The largest increases are seen in the Midwest.

[44] In real life, income-driven repayment plans are capped at the standard repayment amount, often extend repayment periods, and involve some measure of debt relief for remaining balances at the end of the repayment period. To fully capture these details, heterogeneity in debt as well as a multi-period model is needed.

The effects on welfare are also different under this counterfactual policy as shown in Column 2 of Table 11. The most notable difference is that they are an order of magnitude smaller than when considering debt forgiveness – a fact that reflects the absence of a shift of the education costs from high to low-skill workers. The remaining changes in welfare are driven by migration choices that reflect preferences. In both partial and general equilibrium, welfare for high-skill workers increases as the distortion of debt repayments is removed. In partial equilibrium, there is no effect on low-skill workers as prices and wages remain the same. In full equilibrium, low-skill workers actually see an increase in welfare as they are less 'crowded out' from urban areas. There is also a reduction in aggregate output as high-skill workers relocate to lower-productivity areas, as shown in Table 12.

7.4 Counterfactual 4: Tuition-Free Postsecondary Education

In the final counterfactual simulation, I consider the policy proposal of eliminating tuition for postsecondary education. The cost will instead be borne by all via a uniform tax rate. Individuals will be able to adjust their education decisions, i.e. additional individuals can choose to become high-skill workers that would not otherwise have done so when debt is required. The resulting effect on skill-based geographic sorting is reported in Column 5 of Table 10. In partial equilibrium, the reduction in sorting ranges from 5.7-11.6% while there is virtually no change in full equilibrium. This reflects how two opposing forces interact: a rising national share of high-skill workers are still drawn to cities by higher productivity, but their location choices are no longer distorted by debt repayment. Welfare increases for high-skill workers as they no longer bear the full cost of education while it decreases for low-skill workers as they assume some of the cost. Output increases as the aggregate supply of high-skill labor, which is more productive, increases.

8 Conclusion

In this paper, I have shown that student debt distorts the location choices of high-skill workers, increasing the probability that they locate in dense, urban areas. I estimated this effect using a policy change that increased federal student loan limits, a recently developed large consumer credit panel, and a difference-in-differences framework. I then presented descriptive evidence supporting the proposed mechanism that student debt makes individuals more sensitive to nominal wages in location decisions. More specifically, I showed that the association between student debt and urban post-college residency is driven by those who attain degrees with higher urban wage premiums, i.e., those for whom urban areas offer relatively large nominal wage gains. I also provided evidence that correlation between urban preference and student debt is unlikely to be the primary driver of the identified effect.

To quantify the impact on aggregate spatial sorting, I incorporated student debt into a standard spatial equilibrium model and ran counterfactual simulations to show that the growth of student debt since 1980 accounted for 3.5-4.5 percent of skill-based geographic sorting over the same period.

Estimates based on partial equilibrium analysis, which remains agnostic about price and wage responses, suggest this share could be 11.5-18.7 percent. The model also enabled me to reflect on the implications of various policy proposals under consideration by policy-makers, including debt forgiveness, income-driven repayment plans, and a free college option. While all policies eliminated the distortion of high-skill location preferences caused by traditional student debt repayment, only income-driven repayment did so while strictly increasing welfare.

The findings in this paper represent the first examination of the impact of student debt on individual location choices and the aggregate spatial distribution of high-skill workers. Regarding the magnitude of aggregate counterfactual estimates, I would like to highlight three caveats to the model predictions. First, the model assumes homogeneous high-skill workers in both productivity and debt. While it is clear that student debt is not evenly distributed (Looney and Yannelis, 2018), there is limited knowledge about the full distribution and potential correlation with worker productivity. If individuals who borrow the most are differentially productive in urban areas and earn higher wages as result, the model would overestimate resorting.

Second, the model assumes low- and high-skill productivity are independent of the skill-mix in a location despite robust evidence in the literature to the contrary (Diamond, 2016; Moretti, 2011). While the presence of agglomeration economies among high-skill workers in cities would mean the model underestimates out-migration of high-skill workers from urban areas in counterfactual simulations, little is known about how productivity responds to a changing population outside of major metropolitan areas. Due to the uncertainty around production in rural areas, partial equilibrium results may present the most straightforward predictions as they do not impose estimates of city labor demand functions on rural areas. Finally, the model maintains a fixed migration elasticity within skill groups despite knowledge of varying mobility over demographic characteristics like age. While I did not incorporate endogenous migration elasticity into the model for simplicity, recent evidence suggests that migration choices of debt-holders are not entirely path-dependent and borrowers do indeed respond to debt relief (Di Maggio et al., 2020).

The focus of this paper is on empirically documenting the effect of student debt on location choices of high-skill workers and providing a qualitative understanding of how it impacts aggregate sorting and welfare. In future research, I plan to extend this work to better understand how student debt interacts with optimal sorting patterns by developing a structural framework that more accurately captures production responses in urban and rural environments. Additionally, future work should examine the implications of the COVID-19 pandemic, which resulted in a decoupling of wages and location, for student debt-driven spatial sorting.

References

Acemoglu, D. (2002). Technical change, inequality, and the labor market. *Journal of economic literature 40*(1), 7–72.

Akers, B. and M. M. Chingos (2014). *Is a student loan crisis on the horizon?* Brown Center on Education Policy at the Brookings Institution Washington, DC.

Albert, C. and J. Monras (2017). Immigrants' residential choices and their consequences.

Albouy, D., W. Graf, R. Kellogg, and H. Wolff (2016). Climate amenities, climate change, and american quality of life. *Journal of the Association of Environmental and Resource Economists 3*(1), 205–246.

Avery, C. and S. Turner (2012). Student loans: Do college students borrow too much–or not enough? *Journal of Economic Perspectives 26*(1), 165–92.

Barrow, L. and O. Malamud (2015). Is college a worthwhile investment? *Annu. Rev. Econ. 7*(1), 519–555.

Baum, S., J. Ma, M. Pender, and C. Libassi (2019). Trends in student aid 2019. *New York, NY: College Board.*

Baum-Snow, N., M. Freedman, and R. Pavan (2018). Why has urban inequality increased? *American Economic Journal: Applied Economics 10*(4), 1–42.

Berry, C. R. and E. L. Glaeser (2005). The divergence of human capital levels across cities. *Papers in regional science 84*(3), 407–444.

Bettinger, E., O. Gurantz, L. Kawano, B. Sacerdote, and M. Stevens (2019). The long-run impacts of financial aid: Evidence from california's cal grant. *American Economic Journal: Economic Policy 11*(1), 64–94.

Black, S. E., J. T. Denning, L. J. Dettling, S. Goodman, and L. J. Turner (2020). Taking it to the limit: Effects of increased student loan availability on attainment, earnings, and financial well-being. Technical report, National Bureau of Economic Research.

Bleemer, Z., M. Brown, D. Lee, K. Strair, and W. Van der Klaauw (2021). Echoes of rising tuition in students' borrowing, educational attainment, and homeownership in post-recession america. *Journal of Urban Economics 122*, 103298.

Bleemer, Z., M. Brown, D. Lee, W. Van der Klaauw, et al. (2014). Tuition, jobs, or housing: What's keeping millennials at home.

Bound, J. and H. J. Holzer (2000). Demand shifts, population adjustments, and labor market outcomes during the 1980s. *Journal of labor Economics 18*(1), 20–54.

Burdman, P. (2005). The student debt dilemma: Debt aversion as a barrier to college access.

Caliendo, L., M. Dvorkin, and F. Parro (2019). Trade and labor market dynamics: General equilibrium analysis of the china trade shock. *Econometrica 87*(3), 741–835.

Callender, C. and J. Jackson (2005). Does the fear of debt deter students from higher education? *Journal of social policy 34*(4), 509–540.

Card, D. (2009). Immigration and inequality. *American Economic Review 99*(2), 1–21.

Chakrabarti, R., N. Gorton, and M. F. Lovenheim (2020). State investment in higher education: Effects on human capital formation, student debt, and long-term financial outcomes of students. Technical report, National Bureau of Economic Research.

Chapman, S. (2015). Student loans and the labor market: Evidence from merit aid programs. *Northwestern University Department of Economics, Chicago, IL.*

Conzelmann, J. G., S. W. Hemelt, B. Hershbein, S. M. Martin, A. Simon, and K. M. Stange (2022). Grads on the go: Measuring college-specific labor markets for graduates. Technical report, National Bureau of Economic Research.

Dettling, L. J. and J. W. Hsu (2018). Returning to the nest: Debt and parental co-residence among young adults. *Labour Economics 54*, 225–236.

Di Maggio, M., A. Kalda, and V. Yao (2020). Second chance: Life without student debt.

Diamond, R. (2016). The determinants and welfare implications of us workers' diverging location choices by skill: 1980-2000. *American Economic Review 106*(3), 479–524.

Durante, A., J. Larrimore, C. Park, A. Tranfaglia, et al. (2017). Report on the economic well-being of us households in 2016. Technical report, Board of Governors of the Federal Reserve System (US).

Dynarski, S. (2008). Building the stock of college-educated labor. *Journal of human resources 43*(3), 576–610.

Eeckhout, J., R. Pinheiro, and K. Schmidheiny (2014). Spatial sorting. *Journal of Political Economy 122*(3), 554–620.

Field, E. (2009). Educational debt burden and career choice: Evidence from a financial aid experiment at nyu law school. *American Economic Journal: Applied Economics 1*(1), 1–21.

Gervais, M. and N. L. Ziebarth (2019). Life after debt: Postgraduation consequences of federal student loans. *Economic Inquiry 57*(3), 1342–1366.

Giannone, E. (2017). Skilled-biased technical change and regional convergence. *University of Chicago. Unpublished manuscript.*

Gicheva, D. (2016). Student loans or marriage? a look at the highly educated. *Economics of Education Review 53*, 207–216.

Gicheva, D. and J. Thompson (2015). The effects of student loans on long-term household financial stability. *Student loans and the dynamics of debt*, 287–316.

Glaeser, E. L. (2008). *Cities, agglomeration, and spatial equilibrium.* OUP Oxford.

Glaeser, E. L. and J. D. Gottlieb (2008). The economics of place-making policies. Technical report, National Bureau of Economic Research.

Greenwood, M. J. (1975). Research on internal migration in the united states: A survey. *Journal of economic literature*, 397–433.

Hornbeck, R. and E. Moretti (2018). Who benefits from productivity growth? direct and indirect effects of local tfp growth on wages, rents, and inequality.

Houle, J. N. and L. Berger (2015). Is student loan debt discouraging homeownership among young adults? *Social Service Review 89*(4), 589–621.

Howard, G., J. Liebersohn, et al. (2018). The geography channel of house price appreciation. *Fisher College of Business Working Paper No.*

Hsieh, C.-T. and E. Moretti (2019). Housing constraints and spatial misallocation. *American Economic Journal: Macroeconomics 11*(2), 1–39.

Kane, T. J. (2003). A quasi-experimental estimate of the impact of financial aid on college-going.

Katz, L. F. and D. H. Autor (1999). Changes in the wage structure and earnings inequality. In *Handbook of Labor Economics*, Volume 3A, pp. 1463–1555. Elsevier.

Katz, L. F. and K. M. Murphy (1992). Changes in relative wages, 1963–1987: supply and demand factors. *The quarterly journal of economics 107*(1), 35–78.

Kline, P. and E. Moretti (2014). People, places, and public policy: Some simple welfare economics of local economic development programs. *Annu. Rev. Econ. 6*(1), 629–662.

Krishnan, K. and P. Wang (2019). The cost of financing education: can student debt hinder entrepreneurship? *Management Science 65*(10), 4522–4554.

Ladinsky, J. (1967). The geographic mobility of professional and technical manpower. *Journal of Human Resources*, 475–494.

Liu, S., W. Sun, and J. V. Winters (2019). Up in stem, down in business: changing college major decisions with the great recession. *Contemporary Economic Policy 37*(3), 476–491.

Long, B. T. (2014). The financial crisis and college enrollment: How have students and their families responded? In *How the financial crisis and Great Recession affected higher education*, pp. 209–233. University of Chicago Press.

Looney, A. and C. Yannelis (2018). Borrowers with large balances: Rising student debt and falling repayment rates. *The Brookings Institution: Washington, DC*.

Luo, M. and S. Mongey (2019). Assets and job choice: Student debt, wages and amenities. Technical report, National Bureau of Economic Research.

Malamud, O. and A. Wozniak (2012). The impact of college on migration evidence from the vietnam generation. *Journal of Human resources 47*(4), 913–950.

Manson, S., J. Schroeder, D. Van Riper, T. Kugler, and S. Ruggles (2022). Ipums national historical geographic information system: Version 17.0 [dataset]. ipums.

Marx, B. M. and L. J. Turner (2019). Student loan nudges: Experimental evidence on borrowing and educational attainment. *American Economic Journal: Economic Policy 11*(2), 108–41.

Mezza, A., D. Ringo, S. Sherlund, and K. Sommer (2020). Student loans and homeownership. *Journal of Labor Economics 38*(1), 215–260.

Minicozzi, A. (2005). The short term effect of educational debt on job decisions. *Economics of Education Review 24*(4), 417–430.

Molloy, R., C. L. Smith, and A. Wozniak (2011). Internal migration in the united states. *Journal of Economic perspectives 25*(3), 173–96.

Monras, J. (2018). Economic shocks and internal migration.

Monras, J. and C. Albert (2022). Immigration and spatial equilibrium: The role of expenditures in the country of origin. *The Center for Growth and Opportunity.*

Moretti, E. (2011). Local labor markets. In *Handbook of labor economics*, Volume 4, pp. 1237–1313. Elsevier.

Moretti, E. (2013). Real wage inequality. *American Economic Journal: Applied Economics 5*(1), 65–103.

Neumark, D. and H. Simpson (2015). Chapter 18 - place-based policies. In G. Duranton, J. V. Henderson, and W. C. Strange (Eds.), *Handbook of Regional and Urban Economics*, Volume 5 of *Handbook of Regional and Urban Economics*, pp. 1197–1287. Elsevier.

Oreopoulos, P. and U. Petronijevic (2013). Making college worth it: A review of research on the returns to higher education.

Roback, J. (1982). Wages, rents, and the quality of life. *Journal of political Economy 90*(6), 1257–1278.

Rosen, S. (1979). Wage-based indexes of urban quality of life. *Current issues in urban economics*, 74–104.

Rothstein, J. and C. E. Rouse (2011). Constrained after college: Student loans and early-career occupational choices. *Journal of Public Economics 95*(1-2), 149–163.

Ruggles, S., S. Flood, R. Goeken, M. Schouweiler, and M. Sobek (2022). Ipums usa: Version 12.0 [dataset]. ipums.

Saiz, A. (2010). The geographic determinants of housing supply. *The Quarterly Journal of Economics 125*(3), 1253–1296.

Scala, D. J. and K. M. Johnson (2017). Political polarization along the rural-urban continuum? the geography of the presidential vote, 2000–2016. *The ANNALS of the American Academy of Political and Social Science 672*(1), 162–184.

Shao, L. (2014). Debt, marriage, and children: The impact of student loans on marriage and fertility. *PhD diss., Department of Economics, University of Chicago, San Diego.*

Shapiro, J. M. (2006). Smart cities: quality of life, productivity, and the growth effects of human capital. *The review of economics and statistics 88*(2), 324–335.

Suárez, S., J. Carlos, and O. Zidar (2016). Who benefits from state corporate tax cuts? a local labor markets approach with heterogeneous firms. *American Economic Review 106*(9), 2582–2624.

Van Horn, C., C. Zukin, M. Szeltner, and C. Stone (2012). Left out. forgotten? recent high school graduates and the great recession. *John J. Heldrich Center for Workforce Development, Rutgers University.*

Weidner, J. (2016). Does student debt reduce earnings? *Princeton University*.

Wozniak, A. (2010). Are college graduates more responsive to distant labor market opportunities? *Journal of Human Resources 45*(4), 944–970.

Table 1: Federal Borrowing Limits by Academic Year and Level (Class)

Academic Year	Freshmen	Sophomore	Upper Level	Aggregate Limit
2006-07 and earlier	$2,625	$3,500	$5,500	$23,000
2007-08	$3,500	$4,500	$5,500	$23,000
2008-09 and later	$5,500	$6,500	$7,500	$31,000

NOTE: Federal Stafford Loan limits for dependent undergraduate students. Limits apply to the sum of both subsidized and unsubsidized Stafford Loans. Independent students can borrow an additional $4,000 in Freshmen and Sophomore years, and an additional $5,000 in upper levels. Aggregate limits for independent students were double the dependent student limit through 2007-08 academic year, and has been $57,500 since.

Table 2: Federal Borrowing Limits by Entry Cohort

	Individual Level Limits				4-Year Sum	
Entry Cohort	Freshmen	Sophomore	Junior	Senior+	Limit	Relative 2004-05
2004-05 and earlier	$2,625	$3,500	$5,500	$5,500	$17,125	-
2005-06	$2,625	$3,500	$5,500	$7,500	$19,125	$2,000
2006-07	$2,625	$4,500	$7,500	$7,500	$22,125	$5,000
2007-08	$3,500	$6,500	$7,500	$7,500	$25,000	$7,875
2008-09 and later	$5,500	$6,500	$7,500	$7,500	$27,000	$9,875

NOTE: Federal Stafford Loan limits for dependent undergraduate students. Limits apply to the sum of both subsidized and unsubsidized Stafford Loans. Independent students can borrow an additional $4,000 in Freshmen and Sophomore years, and an additional $5,000 in upper levels.

Table 3: Student Characteristics by Treatment Status and Cohort

	Unconstrained		Constrained	
Variable	2002-2005	2006-2013	2002-2005	2006-2013
Age at Entry	18.6	18.6	18.3	18.4
(s.d.)	(0.9)	(0.9)	(0.7)	(0.7)
$\mathbb{1}$[Female]	0.57	0.57	0.53	0.55
Before First Education Loan:				
$\mathbb{1}$[Credit Report]	0.42	0.48	0.32	0.36
$\mathbb{1}$[Credit Score]	0.12	0.35	0.11	0.25
Credit Score	624	628	654	651
(s.d.)	(78)	(85)	(66)	(82)
$\mathbb{1}$[Credit Card]	0.27	0.24	0.20	0.18
$\mathbb{1}$[Auto Loan]	0.002	0.002	0.002	0.001
Number of Students	49,242	93,067	136,383	661,102

NOTE: Sample includes individuals in UCCCP California sample that first took out education loans between July 1, 2001 and June 30, 2013 (i.e., 2002-2013 cohorts) and were less than 21 years old at that time. Sample was also restricted to those who had borrowed a cumulative amount in their first year that was at or below the Federal Stafford Loan limit for first year borrowers in their respective cohort (see Table 2). Finally, individuals had to maintain a credit report through year 6 from entry.

Table 4: Student Characteristics Comparison: Difference-in-Differences Specification

			Before First Education Loan (Entry)	
	(1)	(2)	(3)	(4)
Variable	Age at Entry	$\mathbb{1}$[Female]	$\mathbb{1}$[Credit Report]	$\mathbb{1}$[Credit Score]
$\mathbb{1}$[Constrained] $\times$	0.015^{*}	0.014^{***}	-0.003	-0.050^{***}
$\mathbb{1}$[Cohort>2005]	(0.007)	(0.004)	(0.005)	(0.004)
Dep. Variable Mean	18.385	0.549	0.367	0.233
Number of Students	939,794	939,794	939,794	939,794

*** $p< 0.01$, ** $p< 0.05$, * $p< 0.1$.

NOTE: Sample includes individuals in UCCCP California sample that first took out education loans between July 1, 2001 and June 30, 2013 (i.e., 2002-2013 cohorts), were less than 21 years old at that time, borrowed a cumulative amount in their first year that was at or below the Federal Stafford Loan limit for first year borrowers in their respective cohort (see Table 2), and maintained a credit report through year 6 from entry. Coefficients from DID specification in Equation 2 on respective outcome variable with only year 3 state FEs as controls. Standard errors (clustered by year 3 state) are reported in parentheses.

Table 4: Student Characteristics Comparison: Difference-in-Differences Specification *(continued)*

	Before First Education Loan (Entry)			Last Obs.†
	(5)	(6)	(7)	(8)
Variable	Credit Score	$\mathbb{1}$[Credit Card]	$\mathbb{1}$[Auto Loan]	Education
$\mathbb{1}$[Constrained] $\times$	-5.0^{**}	0.015^{***}	$-2.9\text{E}{-5}$	0.003
$\mathbb{1}$[Cohort>2005]	(2.0)	(0.002)	$1.2\text{E}{-4}$	(0.004)
Dep. Variable Mean	647.0	0.195	$1.5\text{E}{-3}$	2.055
Number of Students	939,794	939,794	939,794	939,794

*** $p< 0.01$, ** $p< 0.05$, * $p< 0.1$.

† All individuals are observed through at least year 6 and 76% are observed through year 9 from first education loan.

NOTE: Sample includes individuals in UCCCP California sample that first took out education loans between July 1, 2001 and June 30, 2013 (i.e., 2002-2013 cohorts), were less than 21 years old at that time, borrowed a cumulative amount in their first year that was at or below the Federal Stafford Loan limit for first year borrowers in their respective cohort (see Table 2), and maintained a credit report through year 6 from entry. Coefficients from DID specification in Equation 2 on respective outcome variable with only year 3 state FEs as controls. Standard errors (clustered by year 3 state) are reported in parentheses.

Table 5: Difference-in-Differences Main Results

Variable	First Stage: Borrowing (1) Year 1	(2) Year 4 (Cumulative)
$\mathbb{1}$[Constrained] × $\mathbb{1}$[Cohort>2005]	$1,214.32^{***}$	$2,600.25^{***}$
	(10.46)	(149.47)
$\mathbb{1}$[Constrained]	$1,000.40^{***}$	$2,829.30^{***}$
	(22.70)	(239.77)
Dep. Variable Mean	3,548.18	13,940.23
Dep. Variable S.D.	(1,392.74)	(10,530.49)
Number of Students	939,794	939,794
Cohort FEs	✓	✓
Demographic Controls	✓	✓
Credit History Controls	✓	✓

*** p< 0.01, ** p< 0.05, * p< 0.1.

NOTE: Sample includes individuals in UCCCP California sample that first took out education loans between July 1, 2001 and June 30, 2013 (i.e., 2002-2013 cohorts), were less than 21 years old at that time, borrowed a cumulative amount in their first year that was at or below the Federal Stafford Loan limit for first year borrowers in their respective cohort (see Table 2), and maintained a credit report through year 6 from entry. Coefficients from DID specification in Equation 2 on respective outcome variable. Standard errors (clustered by year 3 state) are reported in parentheses.

Table 5: Difference-in-Differences Main Results *(continued)*

	Second Stage: Year 6 Location					
	Full Sample			Year 3 California Sample		
Variable	(3) Pop. Density (logs)	(4) $\mathbb{1}$[Top 5% Pop. Density]	(5) $\mathbb{1}$[Top RUCC Metro]	(6) Pop. Density (logs)	(7) $\mathbb{1}$[Top 5% Pop. Density]	(8) $\mathbb{1}$[Top RUCC Metro]
$\mathbb{1}$[Constrained] × $\mathbb{1}$[Cohort>2005]	0.022***	0.016***	0.017***	0.026**	0.017***	0.020***
	(0.008)	(0.002)	(0.003)	(0.011)	(0.004)	(0.003)
$\mathbb{1}$[Constrained]	0.090***	0.016***	0.018***	0.099***	0.017***	0.018***
	(0.008)	(0.002)	(0.002)	(0.010)	(0.003)	(0.003)
Dep. Variable Mean	6.762	0.545	0.748	6.793	0.562	0.785
Dep. Variable S.D.	1.493	-	-	1.424	-	-
Number of Students	939,794	939,794	939,794	624,847	624,847	624,847
Cohort FEs	✓	✓	✓	✓	✓	✓
Demographic	✓	✓	✓	✓	✓	✓
Credit History	✓	✓	✓	✓	✓	✓

*** $p< 0.01$, ** $p< 0.05$, * $p< 0.1$.

NOTE: Sample includes individuals in UCCCP California sample that first took out education loans between July 1, 2001 and June 30, 2013 (i.e., 2002-2013 cohorts), were less than 21 years old at that time, borrowed a cumulative amount in their first year that was at or below the Federal Stafford Loan limit for first year borrowers in their respective cohort (see Table 2), and maintained a credit report through year 6 from entry. Coefficients from DID specification in Equation 2 on respective outcome variable. Standard errors (clustered by year 3 state) are reported in parentheses.

Table 6: Difference-in-Differences: Great Recession Robustness

	First Stage: Borrowing	
	(1)	(2)
	Year 1	Year 4
Variable		(Cumulative)
$\mathbb{1}$[Constrained] × $\mathbb{1}$[Cohort>2005]	$1,989.48^{***}$	$4,070.99^{***}$
	(45.74)	(163.00)
$\mathbb{1}$[Constrained]	955.97^{***}	$2,952.70^{***}$
	(25.82)	(246.23)
Dep. Variable Mean	4,091.53	14,445.09
Dep. Variable S.D.	1,417.86	10,229.86
Number of Students	500,020	500,020
Cohort FEs	✓	✓
Demographic	✓	✓
Credit History	✓	✓

*** $p< 0.01$, ** $p< 0.05$, * $p< 0.1$.

NOTE: Sample includes individuals in UCCCP California sample that first took out education loans between July 1, 2001 and June 30, 2003 (i.e., 2002-2003 cohorts) or between July 1, 2009 and June 30, 2013 (i.e., 2009-2013 cohorts), were less than 21 years old at that time, borrowed a cumulative amount in their first year that was at or below the Federal Stafford Loan limit for first year borrowers in their respective cohort (see Table 2), and maintained a credit report through year 6 from entry. Coefficients from DID specification in Equation 2 on respective outcome variable. Standard errors (clustered by year 3 state) are reported in parentheses.

Table 6: Difference-in-Differences: Great Recession Robustness *(continued)*

	Second Stage: Year 6 Location					
	Full Sample			Year 3 California Sample		
Variable	(3) Pop. Density (logs)	(4) $\mathbb{1}$[80th Percentile Pop. Density]	(5) $\mathbb{1}$[Top RUCC Metro]	(6) Pop. Density (logs)	(7) $\mathbb{1}$[80th Percentile Pop. Density]	(8) $\mathbb{1}$[Top RUCC Metro]
$\mathbb{1}$[Constrained] × $\mathbb{1}$[Cohort>2005]	0.038***	0.024***	0.027***	0.037**	0.023***	0.029***
	(0.009)	(0.003)	(0.005)	(0.018)	(0.006)	(0.005)
$\mathbb{1}$[Constrained]	0.088***	0.013***	0.011**	0.093***	0.013**	0.010**
	(0.010)	(0.003)	(0.005)	(0.016)	(0.006)	(0.005)
Dep. Variable Mean	6.776	0.550	0.758	6.797	0.564	0.787
Dep. Variable S.D.	1.478	-	-	1.420	-	-
Number of Students	500,020	500,020	500,020	352,638	352,638	352,638
Cohort FEs	✓	✓	✓	✓	✓	✓
Demographic	✓	✓	✓	✓	✓	✓
Credit History	✓	✓	✓	✓	✓	✓

*** $p< 0.01$, ** $p< 0.05$, * $p< 0.1$.

NOTE: Sample includes individuals in UCCCP California sample that first took out education loans between July 1, 2001 and June 30, 2003 (i.e., 2002-2003 cohorts) or between July 1, 2009 and June 30, 2013 (i.e., 2009-2013 cohorts), were less than 21 years old at that time, borrowed a cumulative amount in their first year that was at or below the Federal Stafford Loan limit for first year borrowers in their respective cohort (see Table 2), and maintained a credit report through year 6 from entry. Coefficients from DID specification in Equation 2 on respective outcome variable. Standard errors (clustered by year 3 state) are reported in parentheses.

Table 7: Persistence of Location Choices

Variable	(1) Pop. Density (logs)	(2) $\mathbb{1}$[80$^{\text{th}}$ Percentile Pop. Density]	(3) $\mathbb{1}$[Top RUCC Metro]
Year 6 (Full Sample, N=939,794)	0.022***	0.016***	0.017***
	(0.008)	(0.002)	(0.003)
Year 6 (N=715,067)	0.018*	0.015***	0.014***
	(0.010)	(0.004)	(0.004)
Year 7 (N=715,067)	0.014	0.015***	0.015***
	(0.011)	(0.003)	(0.005)
Year 8 (N=715,067)	0.021*	0.017***	0.017***
	(0.011)	(0.003)	(0.005)
Year 9 (N=715,067)	0.021**	0.018***	0.019***
	(0.010)	(0.002)	(0.004)
Cohort FEs	✓	✓	✓
Demographic Controls	✓	✓	✓
Credit History Controls	✓	✓	✓

*** p< 0.01, ** p< 0.05, * p< 0.1.

NOTE: Sample includes individuals in UCCCP California sample that first took out education loans between July 1, 2001 and June 30, 2013 (i.e., 2002-2013 cohorts), were less than 21 years old at that time, borrowed a cumulative amount in their first year that was at or below the Federal Stafford Loan limit for first-year borrowers in their respective cohort (see Table 2), and maintained a credit report through year 6 from entry. In all but the first row, the sample is limited to those observed through year 10. Coefficients from DID specification in Equation 2 on the respective outcome variable. Standard errors (clustered by year 3 state) are reported in parentheses.

Table 8: Student Debt and Post-College Location Choice

	Post-College County		
	Full Sample		
	(1) Pop. Density (logs)	(2) $\mathbb{1}$[80th Percentile Pop. Density]	(3) $\mathbb{1}$[Top RUCC Metro]
Student Debt ($thds, log)	0.1407***	0.0312***	0.0270*
	[0.0478]	[0.0109]	[0.0144]
Origin County			
$\mathbb{1}$[80th Percentile Pop. Density]			
$\mathbb{1}$[Top RUCC Metro]			
College County			
$\mathbb{1}$[80th Percentile Pop. Density]			
$\mathbb{1}$[Top RUCC Metro]			
Observations	2,210	2,210	2,210
R-squared	0.1767	0.1466	0.1232
Dep. Mean	6.6460	0.8269	0.5954
Student Loan Mean	3.1660	3.1660	3.1660
Student Loan S.D.	0.9567	0.9567	0.9567
Demographic Controls	✓	✓	✓
ELS Major FEs	✓	✓	✓

*** $p< 0.01$, ** $p< 0.05$, * $p< 0.1$.

NOTE: Dependent variable pertains to post-college county and classifies them based on varying definitions of urban. Columns (1), (4), and (7) use a continuous measure of population density (in logs); Columns (2), (5), and (8) use an indicator for whether the county is in the top 20 percentile of counties by population density; and Columns (3), (6), and (9) use an indicator for whether the county is classified as a top metropolitan area with a population of 1 million+ using the Rural-Urban Continuum Codes. Origin and college county location controls are similar indicator functions based on the stated urban criteria. Demographic controls include sex, race, father's education, father's occupation, and high school GPA. ELS major fixed effects are reported academic major in postsecondary education. Columns (1)-(6) use the full ELS sample of individuals who attained a postsecondary degree. Columns (7)-(9) restrict this sample to individuals who left their college county after graduation. Number of observations rounded to nearest 10 per IES restricted-use guidelines. Robust standard errors clustered by sample design (survey base-year high school) in parenthesis.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Base Year", 2002, "First Follow-up", 2004, "High School Transcripts", 2005, "Second Follow-up", 2006, and "Third Follow-up", 2012.

Table 8: Student Debt and Post-College Location Choice *(continued)*

	Post-College County					
	Full Sample			"Leavers" (College Cty. $\neq$ Post-College Cty.)		
	(4) Pop. Density (logs)	(5) $\mathbb{1}$[80th Percentile Pop. Density]	(6) $\mathbb{1}$[Top RUCC Metro]	(7) Pop. Density (logs)	(8) $\mathbb{1}$[80th Percentile Pop. Density]	(9) $\mathbb{1}$[Top RUCC Metro]
Student Debt ($thds, log)	0.1054**	0.0207**	0.0072	0.1359**	0.0337***	0.0185
	[0.0415]	[0.0092]	[0.0100]	[0.0634]	[0.0126]	[0.0148]
Origin County						
$\mathbb{1}$[80th Percentile Pop. Density]	1.2865***	0.3750***		1.5182***	0.4556***	
	[0.1040]	[0.0295]		[0.1295]	[0.0347]	
$\mathbb{1}$[Top RUCC Metro]			0.4601***			0.5399***
			[0.0280]			[0.0310]
College County						
$\mathbb{1}$[80th Percentile Pop. Density]	0.8474***	0.2542***		0.4461***	0.0810**	
	[0.1066]	[0.0316]		[0.1295]	[0.0321]	
$\mathbb{1}$[Top RUCC Metro]			0.3421***			0.1306***
			[0.0259]			[0.0282]
Observations	2,210	2,210	2,210	1,470	1,470	1,470
R-squared	0.3548	0.4201	0.5378	0.3827	0.4069	0.4837
Dep. Mean	6.6460	0.8269	0.5954	6.5087	0.7952	0.5959
Student Loan Mean	3.1660	3.1660	3.1660	3.1902	3.1902	3.1902
Student Loan S.D.	0.9567	0.9567	0.9567	0.9280	0.9280	0.9280
Demographic Controls	✓	✓	✓	✓	✓	✓
ELS Major FEs	✓	✓	✓	✓	✓	✓

*** p< 0.01, ** p< 0.05, * p< 0.1.

NOTE: See table notes on the previous page. Number of observations rounded to nearest 10 per IES restricted-use guidelines. Robust standard errors clustered by sample design (survey base-year high school) in parenthesis.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Base Year", 2002, "First Follow-up", 2004, "High School Transcripts", 2005, "Second Follow-up", 2006, and "Third Follow-up", 2012.

Table 9: Student Debt and Post-College Location Choice: Heterogeneity by Urban Wage Premium

	Post-College County		
	Pop. Density (logs)		
	(1) Full Sample	(2) ACS-Matched Sample	(3) ACS-Matched Sample
Student Debt ($thds, log)	0.1054**	0.1062**	
	[0.0415]	[0.0533]	
$\times$ $\mathbb{1}$[**Below Median** Urban Wage Prem.]			0.0292
			[0.0697]
$\times$ $\mathbb{1}$[**Above Median** Urban Wage Prem.]			0.1434**
			[0.0717]
Origin County			
$\mathbb{1}$[80th Percentile Pop. Density]	1.2865***	1.2715***	1.2732***
	[0.1040]	[0.1052]	[0.1054]
$\mathbb{1}$[Top RUCC Metro]			
College County			
$\mathbb{1}$[80th Percentile Pop. Density]	0.8474***	0.7887***	0.7989***
	[0.1066]	[0.1302]	[0.1286]
$\mathbb{1}$[Top RUCC Metro]			
Observations	2,210	1,700	1,700
R-squared	0.3548	0.2942	0.2951
Dependent Variable Mean	6.646	6.7556	6.7556
Student Loan Mean	3.1660	3.2344	3.2344
Student Loan S.D.	0.9567	0.9270	0.9270
Demographic Controls	✓	✓	✓
ELS Major FEs	✓		
ACS Degree Field FEs		✓	✓

*** p< 0.01, ** p< 0.05, * p< 0.1.
NOTE: Dependent variable pertains to post-college county and classifies them based on varying definitions of urban. Columns (1), (4), and (7) use a continuous measure of population density (in logs); Columns (2), (5), and (8) use an indicator for whether the county is in the top 20 percentile of counties by population density; and Columns (3), (6), and (9) use an indicator for whether the county is classified as a top metropolitan area with a population of 1 million+ using the Rural-Urban Continuum Codes. Origin and college county location controls are similar indicator functions based on the stated urban criteria. Demographic controls include sex, race, father's education, father's occupation, and high school GPA. ELS major fixed effects are reported academic major in postsecondary education. Columns (1)-(6) use the full ELS sample of individuals who attained a postsecondary degree. Columns (7)-(9) restrict this sample to individuals who left their college county after graduation. Number of observations rounded to nearest 10 per IES restricted-use guidelines. Robust standard errors clustered by sample design (survey base-year high school) in parentheses.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Base Year", 2002, "First Follow-up", 2004, "High School Transcripts", 2005, "Second Follow-up", 2006, and "Third Follow-up", 2012.

Table 9: Student Debt and Post-College Location Choice: Heterogeneity by Urban Wage Premium *(continued)*

	Post-College County					
	𝟙[80$^{\text{th}}$ Percentile Pop. Density]			𝟙[Top RUCC Metro]		
	(4) Full Sample	(5) ACS-Matched Sample	(6) ACS-Matched Sample	(7) Full Sample	(8) ACS-Matched Sample	(9) ACS-Matched Sample
Student Debt ($thds, log)	0.0229**		0.0072	0.0040		
	[0.0092]	[0.0108]		[0.0100]	[0.0119]	
× 𝟙[**Below Median** Urban Wage Premium]			0.0099			−0.0207
			[0.0194]			[0.0216]
× 𝟙[**Above Median** Urban Wage Premium]			0.0292**			0.0159
			[0.0130]			[0.0144]
Origin County						
𝟙[80$^{\text{th}}$ Percentile Pop. Density]	1.2865***	0.3482***	0.3485***			
	[0.0295]	[0.0326]	[0.0326]			
𝟙[Top RUCC Metro]				0.4601***	0.4652***	0.4649***
				[0.0280]	[0.0337]	[0.0337]
College County						
𝟙[80$^{\text{th}}$ Percentile Pop. Density]	0.8474***	0.2342***	0.2359***			
	[0.0316]	[0.0377]	[0.0377]			
𝟙[Top RUCC Metro]				0.3421***	0.3165***	0.3179***
				[0.0259]	[0.0306]	[0.0307]
Observations	2,210	1,700	1,700	2,210	1,700	1,700
R-squared	0.4201	0.3576	0.3580	0.5378	0.4938	0.4948
Dependent Variable Mean	0.8269	0.8444	0.8444	0.5954	0.6235	0.6235
Student Loan Mean	3.1660	3.2344	3.2344	3.1660	3.2344	3.2344
Student Loan S.D.	0.9567	0.9270	0.9270	0.9567	0.9270	0.9270
Demographic Controls	✓	✓	✓	✓	✓	✓
ELS Major FEs	✓			✓		
ACS Degree Field FEs		✓	✓		✓	✓

*** $p< 0.01$, ** $p< 0.05$, * $p< 0.1$.

NOTE: See table notes on the previous page. Number of observations rounded to nearest 10 per IES restricted-use guidelines. Robust standard errors clustered by sample design (survey base-year high school) in parenthesis.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), “Base Year”, 2002, “First Follow-up”, 2004, “High School Transcripts”, 2005, “Second Follow-up”, 2006, and “Third Follow-up”, 2012.

Table 10: Model Results: Skill-Based Geographic Sorting

	Counterfactual				
	1980 Debt Level		Debt Forgiveness	Income-Driven Repayment	Free College Option
	(1)	(2)	(3)	(4)	(5)
Model	Δ Sorting (Rel. to 1980)	Δ Sorting (2019 Level)	Δ Sorting (2019 Level)	Δ Sorting (2019 Level)	Δ Sorting (2019 Level)
Panel A: Partial Equilibrium					
Homogeneous Migration Elasticity	-11.5%***	-6.7%***	-8.1%***	-8.1%***	-5.7%***
Heterogeneous Migration Elasticity†	-18.7%***	-11.2%***	-14.0%***	-14.0%***	-11.6%***
Panel B: Full Equilibrium					
Homogeneous Migration Elasticity	-3.5%***	-2.0%***	-2.4%***	-2.4%***	0.0%
Heterogeneous Migration Elasticity†	-4.5%***	-2.8%***	-3.4%***	-2.2%***	0.2%***

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
† Elasticity estimates from Hornbeck and Moretti (2018).
NOTE: Reported values reflect the change in the coefficient on commuting zone population density high-skill worker share in a regression weighted by commuting zone population when comparing counterfactual to observed data.

Table 11: Model Results: Consumer Welfare

	Counterfactual								
	(1) Debt Forgiveness			(2) Income-Driven Repayment			(3) Tuition-Free College		
Model	Δ Aggregate	Δ HS	Δ LS	Δ Aggregate	Δ HS	Δ LS	Δ Aggregate	Δ HS††	Δ LS††
Panel A: Partial Equilibrium									
Homogeneous Migration Elasticity	-0.64%	2.06%	-2.40%	0.05%	0.13%	0.00%	-0.67%	1.93%	-2.53%
Heterogeneous Migration Elasticity†	-0.64%	2.07%	-2.41%	0.05%	0.12%	0.00%	-0.72%	1.93%	-2.54%
Panel B: Full Equilibrium									
Homogeneous Migration Elasticity	-1.08%	1.52%	-2.78%	0.12%	0.18%	0.09%	-0.66%	1.93%	-2.53%
Heterogeneous Migration Elasticity†	-1.10%	1.52%	-2.79%	0.11%	0.17%	0.08%	-0.71%	1.94%	-2.53%

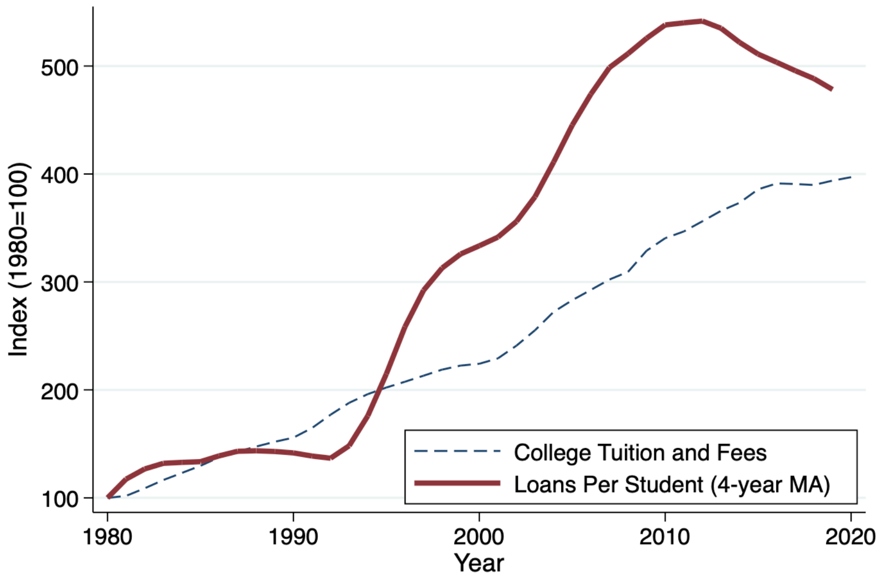
† Elasticity estimates from Hornbeck and Moretti (2018).
†† For tuition-free college counterfactual, reported value reflects individual worker changes, i.e. not accounting for changing skill-mix of the general population.
NOTE: Reported values reflect the change in aggregate consumer welfare, the change in welfare for high-skill workers, and the change in welfare for low-skill workers.

Table 12: Model Results: Aggregate Output

	Counterfactual		
	(1) Debt Forgiveness	(2) Income-Driven Repayment	(3) Tuition-Free College
Model	Δ Output (2019 Level)	Δ Output (2019 Level)	Δ Output (2019 Level)
Panel A: Partial Equilibrium			
Homogeneous Migration Elasticity	-0.30%	-0.30%	1.72%
Heterogeneous Migration Elasticity†	-0.52%	-0.52%	1.48%
Panel B: Full Equilibrium			
Homogeneous Migration Elasticity	-0.09%	-0.10%	1.99%
Heterogeneous Migration Elasticity†	-0.10%	-0.11%	1.97%

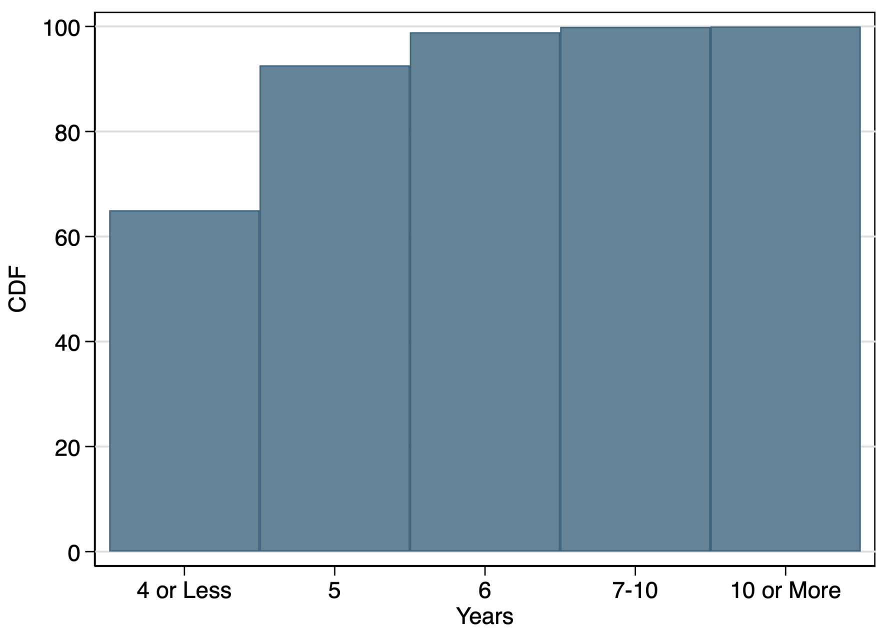
† Elasticity estimates from Hornbeck and Moretti (2018).

Figure 2: Inflation-Adjusted Cost of Post-Secondary Education



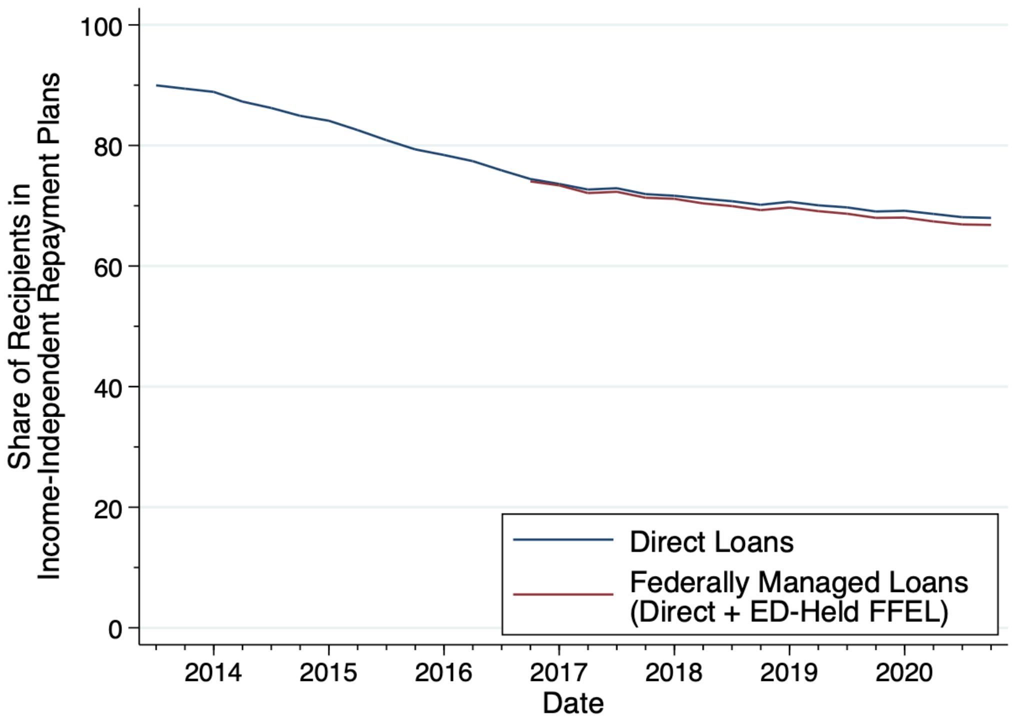
SOURCE: Loan data from College Board; includes undergraduate average federal and nonfederal loans per full-time equivalency student. College tuition and fees index from U.S. Bureau of Labor Statistics.

Figure 3: Time to Degree Completion



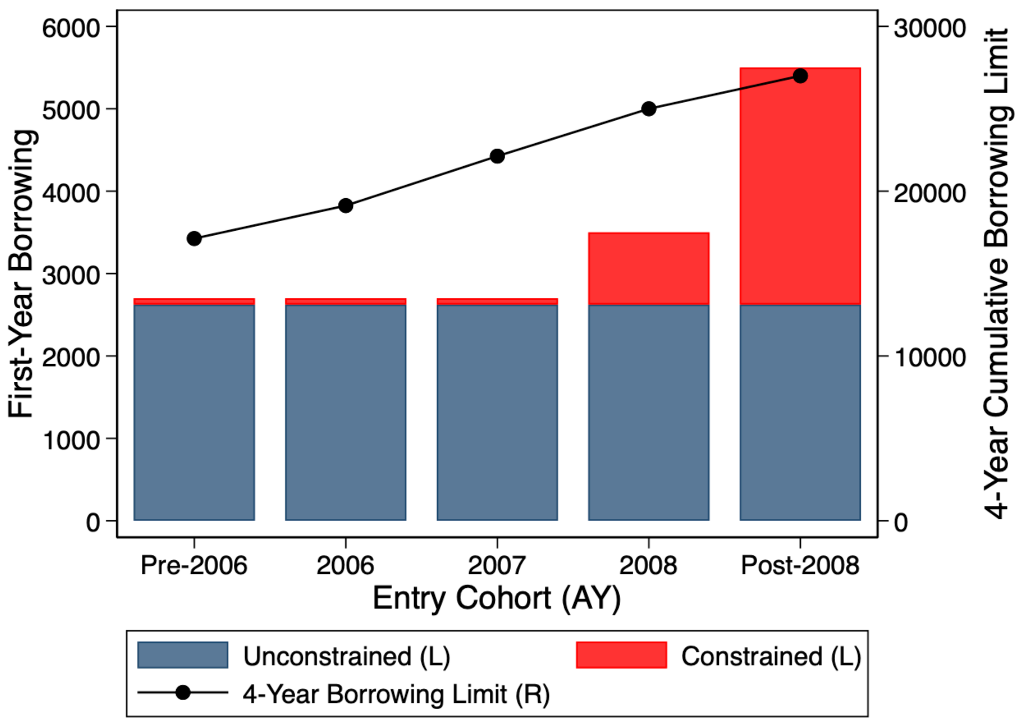
NOTE: Time to Bachelor degree completion among individuals who attain a degree.
SOURCE: U.S. Department of Education, National Center for Education Statistics (2019). Baccalaureate and Beyond (B&B:16/17): A First Look at the Employment and Educational Experiences of College Graduates, 1 Year Later (NCES 2019-106), Table 2.

Figure 4: Share of Loan Recipients on Income-Independent Repayment Plans



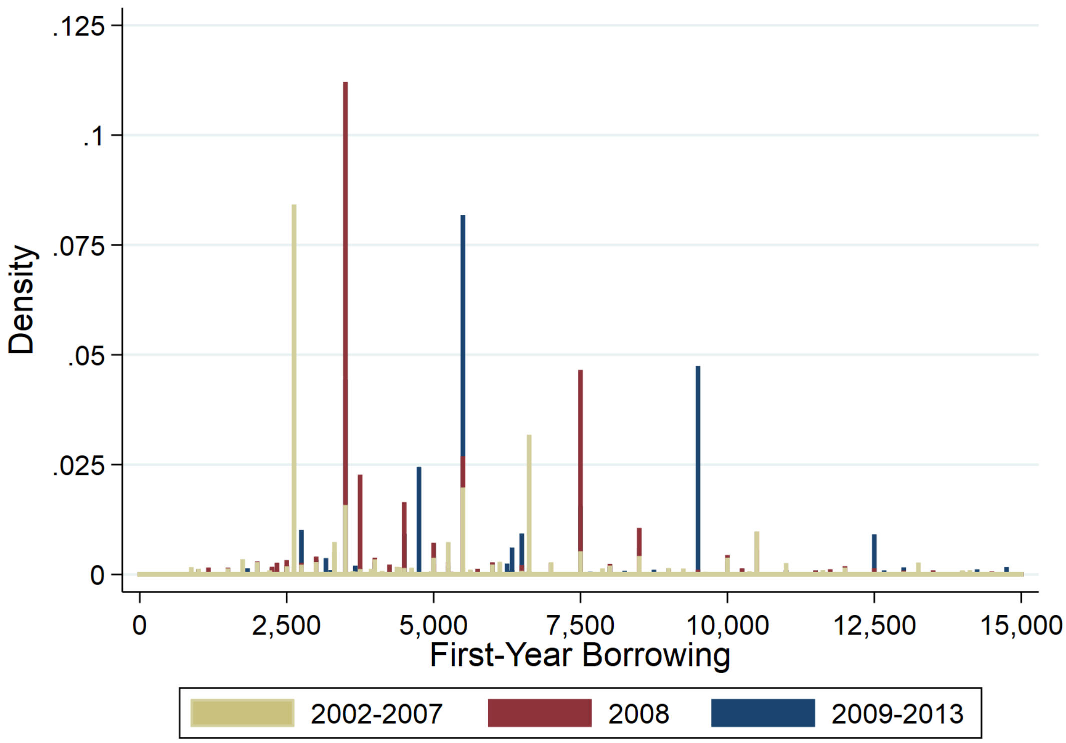
NOTE: Income-independent plans are all plans not classified as income-driven repayment plans (income-contingent, income-sensitive, income-based, Pay As You Earn (PAYE), and Revised Pay As You Earn (RePAYE)). All data excludes borrowers in default, in-school, or in grace periods before repayment begins. Recipient counts are based at loan level. As a result, recipients may be counted multiple times across varying loan statuses.
SOURCE: Data on Federal Student Loan Portfolio provided by Federal Student Aid; original source is National Student Loan Data System.

Figure 5: Defining Constrained Borrowers



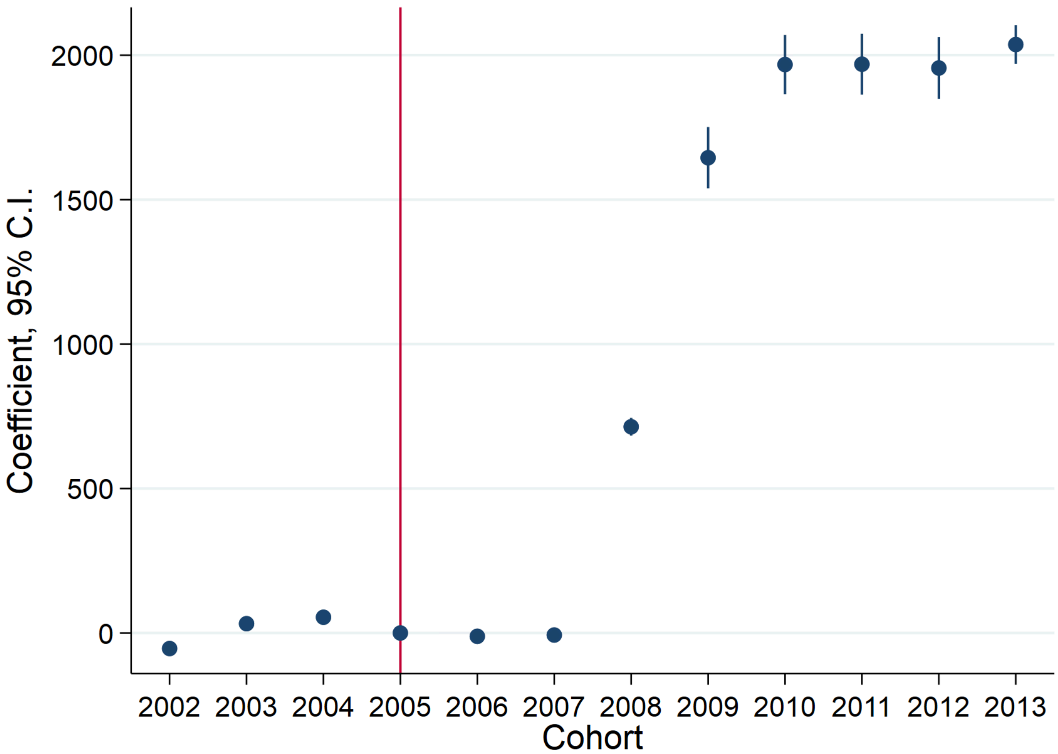
NOTE: Students are classified as likely to be constrained if: (1) they borrow exactly at the original first-year borrowing limit ($2,625) in years up to the 2007 academic year; or, (2) if they borrow between the original limit ($2,625) and the new limit in their given entry year. The bar for constrained borrowers in years prior to 2008 appears larger than the single point ($2,625) in this graph for illustrative purposes.

Figure 6: First-Year Borrowing



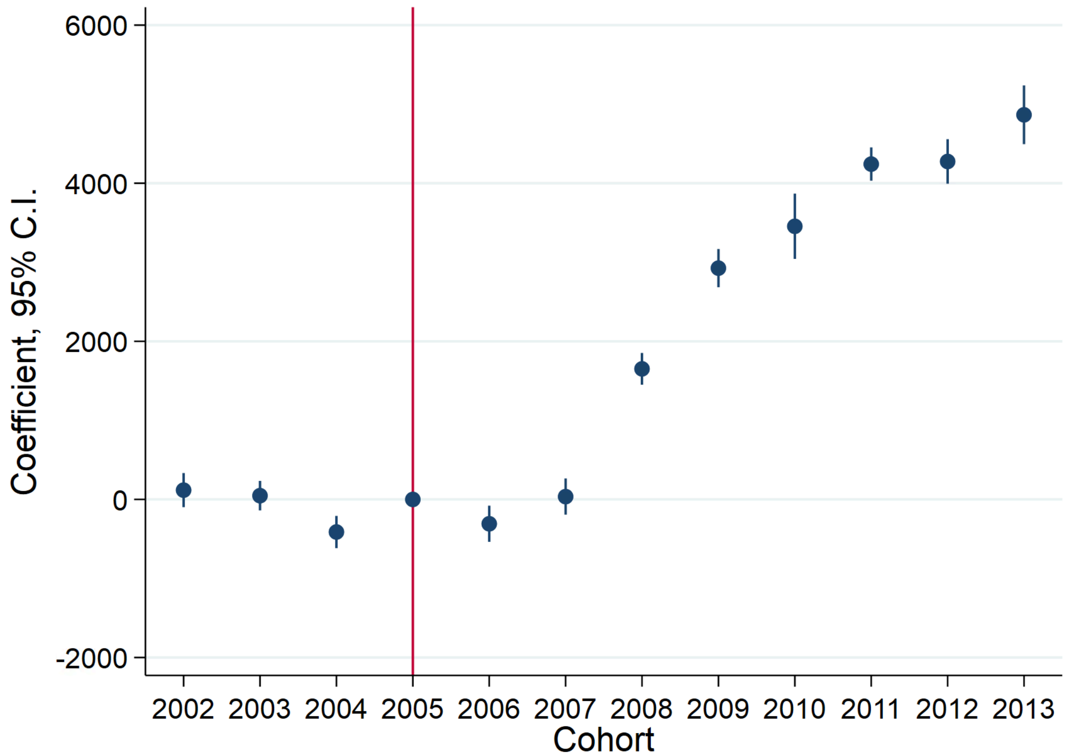
NOTE: This distribution reflects the cumulative sum of loans in the first year of borrowing for each individual.
SOURCE: University of California Consumer Credit Panel, California Sample.

Figure 7: Effect of Increased Loan Access on First-Year Borrowing



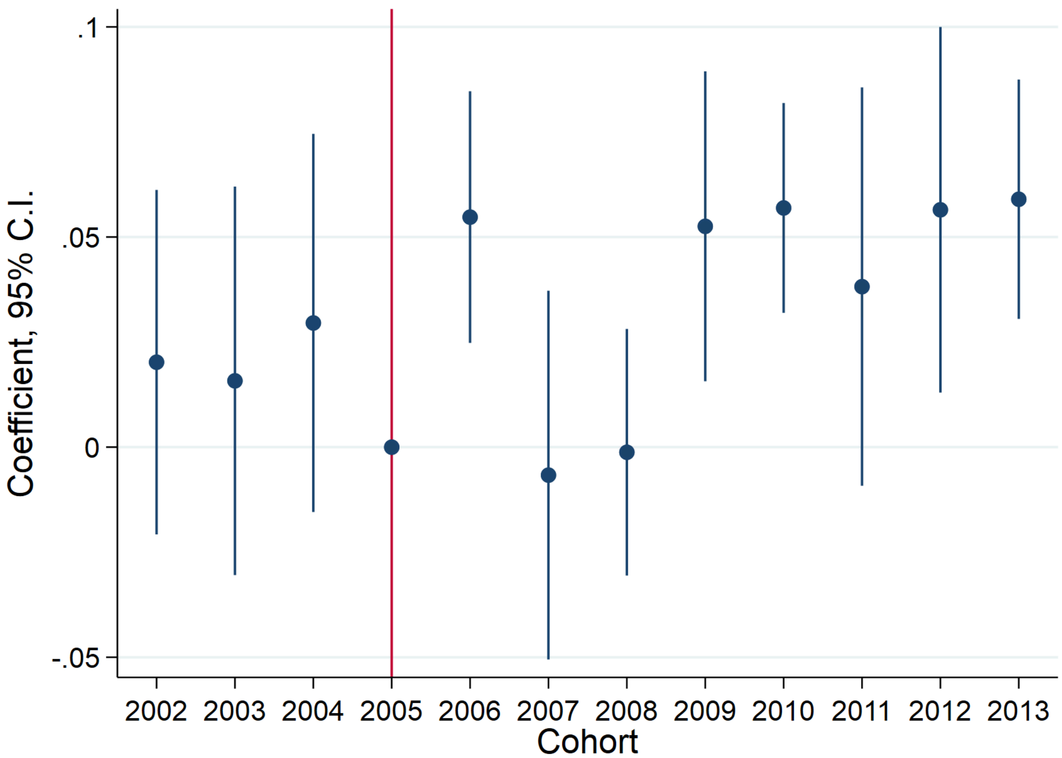
NOTE: Coefficients from the estimation of Equation 1 with the dependent variable being cumulative education loans in year 1 of borrowing. Confidence intervals reflect standard errors clustered by year 3 state.
SOURCE: University of California Consumer Credit Panel, California Sample.

Figure 8: Effect of Increased Loan Access on
Cumulative Borrowing Through Year 4



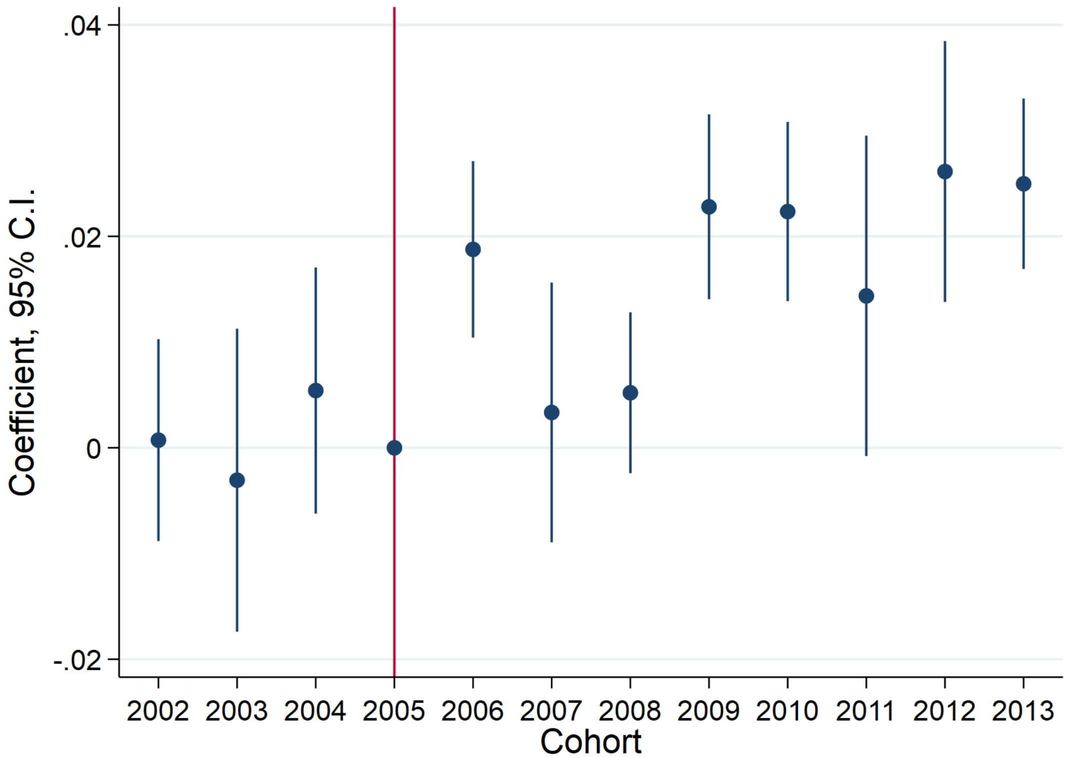
NOTE: Coefficients from the estimation of Equation 1 with the dependent variable being cumulative education loans through year 4 of borrowing. Confidence intervals reflect standard errors clustered by year 3 state.
SOURCE: University of California Consumer Credit Panel, California Sample.

Figure 9: Effect of Increased Loan Access on
Year 6 County Population Density (logs)



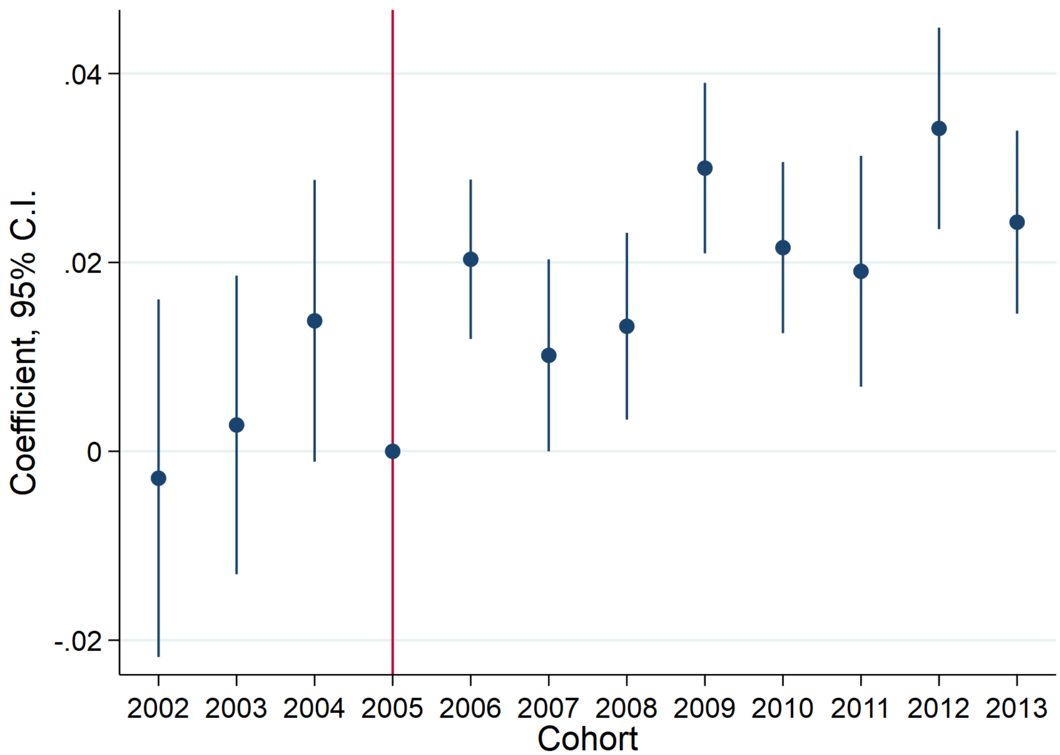
NOTE: Coefficients from the estimation of Equation 1 with the dependent variable being (log) population density for the county of residence in year 6 from entry. Confidence intervals reflect standard errors clustered by year 3 state.
SOURCE: University of California Consumer Credit Panel, California Sample.

Figure 10: Effect of Increased Loan Access on
Probability of Year 6 County Being 80th Percentile by Population Density



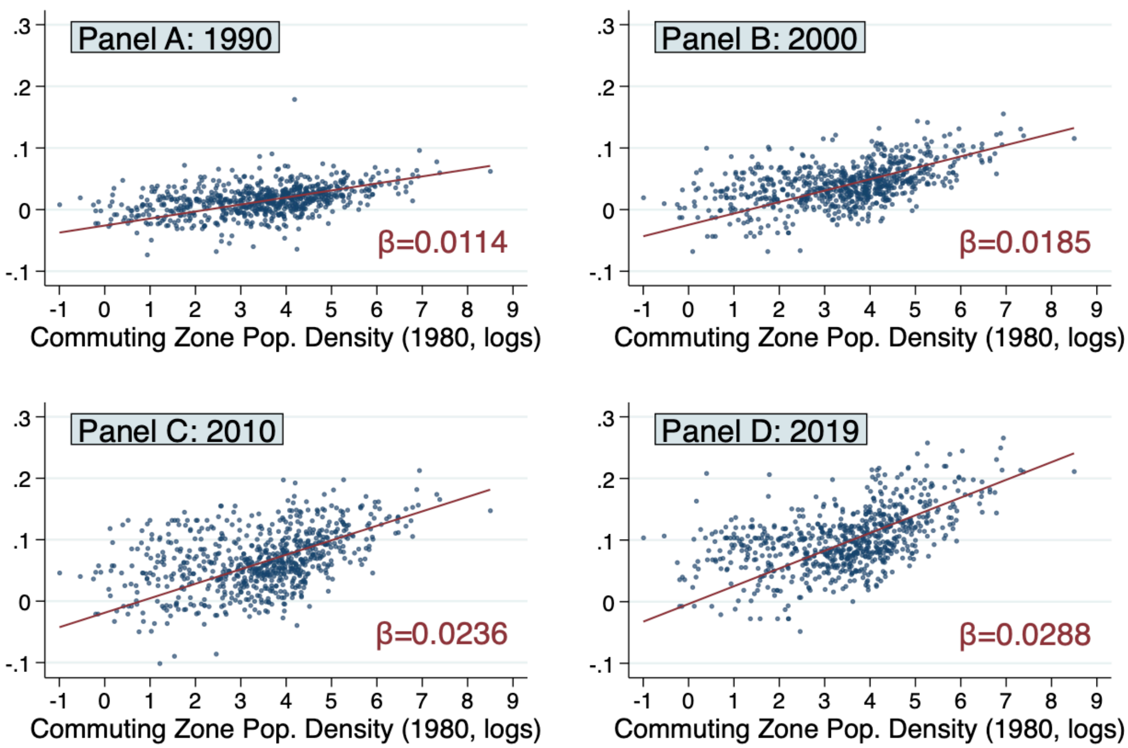
NOTE: Coefficients from the estimation of Equation 1 with the dependent variable being an indicator function for county of residence in year 6 from entry being in the 80th percentile by county population density. Confidence intervals reflect standard errors clustered by year 3 state.
SOURCE: University of California Consumer Credit Panel, California Sample.

Figure 11: Effect of Increased Loan Access on
Probability of Year 6 County Being Top Metro RUCC



NOTE: Coefficients from the estimation of Equation 1 with the dependent variable being an indicator function for county of residence in year 6 from entry being in the top metropolitan category (1m+) of the RUCC. Confidence intervals reflect standard errors clustered by year 3 state.
SOURCE: University of California Consumer Credit Panel, California Sample.

Figure 12: Skill-Based Geographic Sorting:
Change in High Skill Share Since 1980



NOTE: X-axis population reflects total population; skill shares are among employed, working-age (25-54). Points are 1990-defined commuting zones. Best fit line weighted by commuting zone total population in 1980.
SOURCE: U.S. Census 1980 5% state, 1990 5% state, 2000 5%, 2010 ACS, and 2019 5-year ACS.

Figure 13: Role of Student Debt in Post-1980 Geographic Sorting: Partial Equilibrium

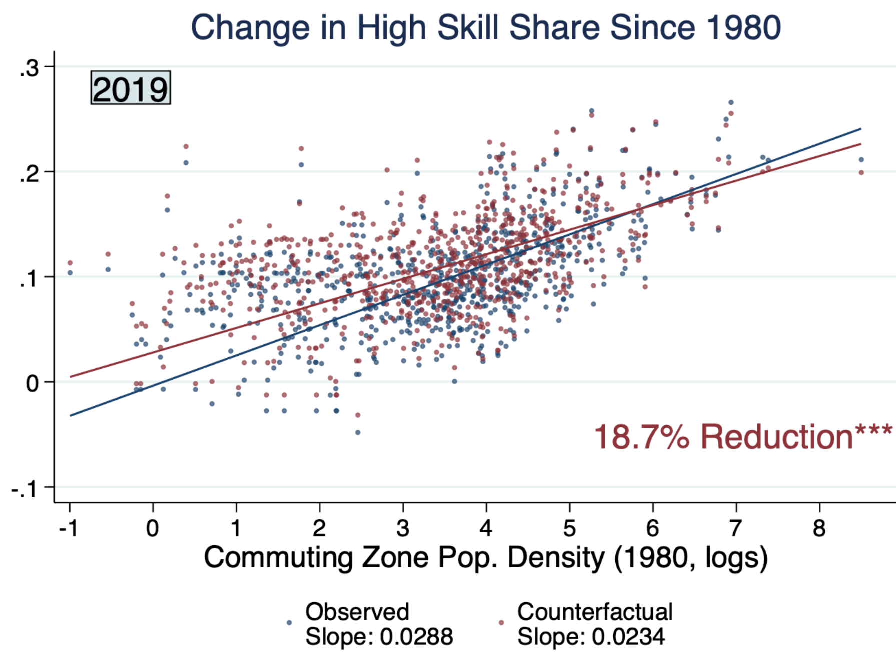


Figure 14: Role of Student Debt in Post-1980 Geographic Sorting: Full Equilibrium

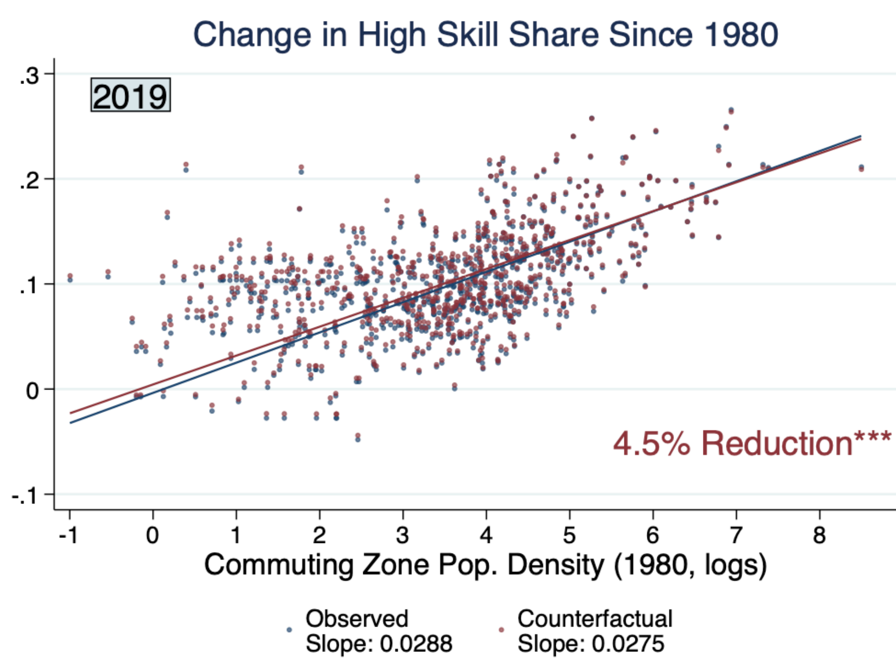


Figure 15: Income-Driven Repayment Counterfactual Map:
Partial Equilibrium

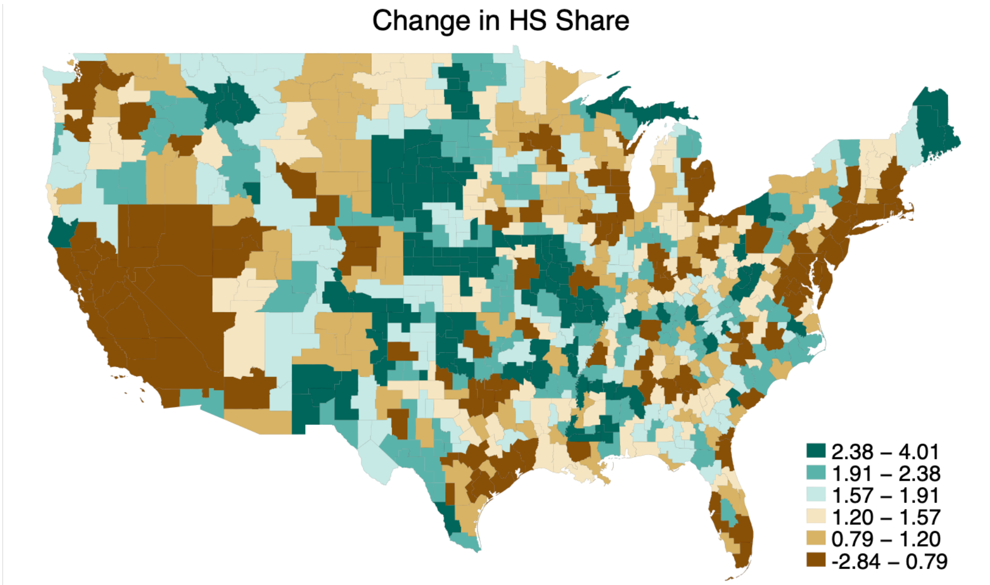
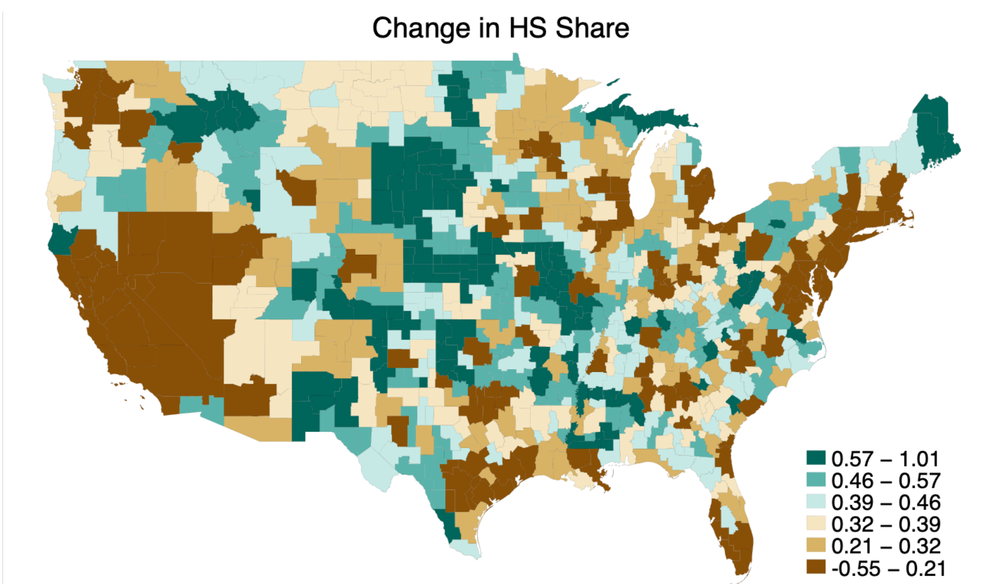


Figure 16: Income-Driven Repayment Counterfactual Map:
Full Equilibrium



Appendix A Early Adulthood Migration Drives Skill-Based Sorting

In this section, I show that differences between low and high skill workers in mobility and spatial distribution, particularly across the urban-rural divide, are driven by migration choices in early adulthood. This finding, which is absent of any quantitative connection to student debt, suggests that factors weighing heavily on individuals in early adulthood could impact key initial migration choices that are persistent into later adulthood. While a standard life-cycle model would suggest student debt, which is a small share of lifetime income, should have little impact on post-college choices like migration and career, debt repayments as a share of income are largest in early adulthood, and Rothstein and Rouse (2011) find that student debt causes individuals to choose initially higher-salary jobs.[45] This departure from the expected behavior of life-cycle agents is consistent with evidence from the Education Longitudinal Study of 2002 showing that additional student debt is positively correlated with recent graduates taking a job outside of their field of study, having to work more than one job at the same time, and having to work more hours than desired (see Table A1). If student debt impacts early adulthood migration choices as well, which I will identify in Section 3, the findings in this section suggest it could have a persistent effect on the spatial distribution of workers.

I proceed in this section by tracking the location choices of all adults in the Panel Study of Income Dynamics that were born between 1940 and 1979, observing their location choices biennially from 1968 to 2019.[46,47] I classify workers by high (college-educated) and low (less than college degree) skill using their educational attainment through age 40. First, I show that high skill individuals (college graduates) are generally more geographically mobile than non-college educated individuals over their lifetime, a result consistent with previous literature[48], but that this overall mobility differential masks significant heterogeneity over the life cycle. High skill workers are more mobile in early adulthood, but less mobile in later periods – suggesting that early migration choices are particularly persistent for high skill workers. I then examine educational mobility differentials over the life cycle in relation to skill-based geographic sorting across various definitions of the urban-rural divide. When considering migration from one PSID wave to the next, high skill individuals are 2-3 percent more likely than low skill individuals to move from a 'rural' county to an 'urban' one for waves in which the individual is younger than 30 (mean sample probability of 2.2 percent).

Appendix A.0.1 Data

The PSID is a longitudinal survey with a nationally representative sample that collects location information at a high frequency and allows me to understand migration across the urban-rural divide over the life cycle. The survey has been conducted since 1968, making it the longest-running longitudinal survey in the world. The sample follows original households and their descendants with occasional 'refresher' samples to ensure the PSID remains representative of the U.S. population.

[45] They attribute this departure from standard life-cycle model predictions to credit constraints after schooling, but cannot rule out debt aversion.

[46] Sample was limited by date of birth for three reasons: (1) so that each cohort is observed for a considerable period of their adult life (20 years); (2) at least once before the age of 30; and (3) so that they are individuals in their prime working years during the time period of interest (1980-2019) for skill-based geographic sorting. Results are robust to the addition of all cohorts in the PSID.

[47] The PSID conducted surveys annually from 1968 through 1997, then biennially. To provide consistency across the sample, the 1968 through 1997 subsample was trimmed so that location is observed at most biennially starting from an individual's first year in the sample.

[48] See Ladinsky (1967); Greenwood (1975); Wozniak (2010); Malamud and Wozniak (2012).

Households were surveyed annually through 1997 and biennially thereafter. To maintain consistency across the sample, the 1968 through 1997 data was trimmed so that individuals are observed at most biennially starting from an individual's first year in the sample. I limit the sample population to the reference person and spouses of households, i.e. the adult population.[49] Furthermore, I restrict the sample to adults age 18 to 65 that were born between 1940 and 1979 for three reasons: (1) each cohort is observed for a considerable period (20 years) of their adult life; (2) at least once before the age of 30 ('early adulthood'); and (3) they are individuals in their prime working years during the time period of interest (1980-2019) for skill-based geographic sorting. The resulting sample includes approximately 7,000 individuals, observed on average for 25 years (12.5 waves).

In addition to location, the PSID collects information on a variety of other topics including education, employment, income, and wealth. Location is observed at the county level for each wave. Individuals are also asked to report the "county where they grew up", hereinafter referred to as the origin location – a factor that has been shown to be important in migration decisions (Diamond, 2016). The PSID also includes a wealth of other information on an individual's personal characteristics, including sex, race, education attainment, and marital status. Although I only highlight a few here that are likely to impact mobility and migration behavior, a variety of other demographic and economic variables are included in different robustness specifications below.

Appendix A.0.2 Overall Mobility

I use the following framework to estimate educational differentials in general mobility:

$$\mathbb{1}[\text{Moved}_{it}] = \alpha + \beta_1 \mathbb{1}[\text{Skill}_i = \text{High}] + \sum_c \delta_c \, \mathbb{1}[\text{Age}_{it} \in c] + \mathbf{X_i'}\boldsymbol{\beta_x} + \gamma_t + \gamma_c + \mu_{it} + \epsilon_{it} \ , \tag{25}$$

where $\mathbb{1}[\text{Moved}_{it}]$ is an indicator for whether an individual changed county of residence from wave $t-1$ to wave t. The coefficient of interest, β_1, captures whether an individual is a high skill (college educated) worker. Fixed effects control for age of individual in wave t by age categories (δ_c), wave (γ_t), and cohort (γ_c). As high and low skill workers likely establish independent households in the PSID at different ages[50], I also include a fixed effect for the number of waves that an individual has been an independent household (μ_{it}). This would capture common trends such as if there is frequent movement while initially setting up an independent household. A set of demographic characteristics are captured by the vector $\mathbf{X_i}$. These include sex, race, and marital status. Coefficients are estimated using a linear probability model with standard errors clustered by individual.

To understand how educational mobility differentials differ across the life cycle, I augment this framework by interacting the indicator for being a high skill worker with age categories:

$$\begin{aligned}\mathbb{1}[\text{Moved}_{it}] = \ & \alpha + \sum_c \beta_1^c (\mathbb{1}[\text{Skill}_i = \text{High}] \ \times \ \mathbb{1}[\text{Age}_{it} \in c]) + \sum_c \delta_c \, \mathbb{1}[\text{Age}_{it} \in c] + \\ & \mathbf{X_i'}\boldsymbol{\beta_x} + \gamma_t + \gamma_c + \mu_{it} + \epsilon_{it} \ ,\end{aligned} \tag{26}$$

where all other aspects of the estimating equation remain the same.

Results from the estimation of Equations 25 and 26 using a linear probability model are reported

[49] Individuals that are descendants of PSID families establish independent family units when they live in a separate household and are above the age of 18 (underage individuals can be independent households after 1993, but I continue prior definition for consistency). Individuals in an institution, including a college dormitory, are not labelled independent households.

[50] Individuals enrolled in post-secondary institutions are not considered independent households, and are thus not included in this sample until after graduation.

in Table A2. As shown in Column 1, high skill individuals are more geographically mobile than low skill workers overall. In any given wave, the probability that a high skill individual changed counties is 1.6 percentage points higher than for a low skill individual – a meaningful difference relative to a mean location change probability of 28.2 percent. This overall effect hides significant heterogeneity over the life cycle, as shown in Column 2. High skill workers are equally mobile for ages less than 22, then are significantly more mobile than low skill workers from 22 to 35. From age 36 to 56, geographic mobility of high skill workers is less than low skill workers, though this effect is insignificant outside of the 46-50 age range. These results can also be seen in Figure A1, Panel (a). As shown in Panel (b), these results are not driven by any single cohort.

Appendix A.0.3 Skill-Based Geographic Sorting

I next turn to analyzing mobility across the urban-rural divide. I use the same framework as above, but classify locations by four measures of 'urban': (1) indicator for top metropolitan county as defined by the Rural-Urban Continuum Codes[51]; (2) percentile of county population density (decade-specific); (3) indicator for county in 80$^{\text{th}}$ percentile of population density (decade-specific); and (4) indicator for county in 95$^{\text{th}}$ percentile of population density (decade-specific). As origin location has been shown to play a significant role in location preferences[52], I also augment the previous specifications by controlling for origin location urbanicity using the same measure as the dependent variable. The educational migration differential along the urban-rural divide is estimated using the following framework:

$$\begin{aligned}\mathbb{1}[\textbf{Live in } \text{`Urban'}_{it}] = \; & \alpha + \beta_1 \mathbb{1}[\text{Skill}_i = \text{High}] + \\ & \sum_c \delta_c \, \mathbb{1}[\text{Age}_{it} \in c] + \mathbb{1}[\text{Origin `Urban'}_i] + \mathbf{X_i'}\boldsymbol{\beta_x} + \\ & \gamma_t + \gamma_c + \mu_{it} + \epsilon_{it} \;\; ,\end{aligned} \tag{27}$$

where all other variables are as defined following Equation 25. Coefficients are estimated using a linear probability model with standard errors clustered by individual.

To understand how migration differentials across the urban-rural divide vary by age, I interact the indicator for individual skill level with age:

$$\begin{aligned}\mathbb{1}[\textbf{Live in } \text{`Urban'}_{it}] = \; & \alpha + \sum_c \beta_1^c (\mathbb{1}[\text{Skill}_i = \text{High}] \; \times \; \mathbb{1}[\text{Age}_{it} \in c]) + \\ & \sum_c \delta_c \, \mathbb{1}[\text{Age}_{it} \in c] + \mathbb{1}[\text{Origin `Urban'}_i] + \mathbf{X_i'}\boldsymbol{\beta_x} + \\ & \gamma_t + \gamma_c + \mu_{it} + \epsilon_{it} \;\; ,\end{aligned} \tag{28}$$

where all other aspects of the estimating equation remain the same. Taking advantage of the biennial frequency in the PSID data, I also conduct this analysis for moving to urban areas. For a movement indicator to be 1, an individual has to living in a 'rural' county in wave $t-1$ and then in an 'urban' county in wave t based on a given definition of urban and rural.

Results of this analysis on living in and moving to urban areas are reported in Tables A3 and A4, respectively. As shown in Columns 1 through 4 of Table A3, high skill individuals live in urban counties at higher rates than low skill individuals when using all 4 definitions of urban. This is

[51] A roughly decade-specific measure. RUCC were initially developed in 1974 and have been updated in 1983, 1993, 2003, and 2013. Codes used in this analysis are RUCC codes closest to PSID wave. Top metropolitan counties in the RUCC have a population of over 1 million.

[52] See, for example, Molloy et al. (2011).

consistent with the skill-based geographic sorting detailed in Figure 12. In Columns 5-8 of Table A3 and Figures A2-A5, I decompose this sorting by age category. For all definitions of urban, high skill workers are not living in urban counties at higher rates than low skill workers when they are less than 22. Between ages 22 and 35, this changes significantly as high skill workers increasingly choose to live in urban counties. Following age 35, the educational differential with respect to living in urban areas levels out and remains constant for the rest of the prime working-age years. Panel (b) of each figure shows that these estimates are not driven by any specific cohort.

The results in Table A4 and Figures A6-A9 more directly examine this sorting by observing movement to urban counties. As shown in Columns 1-4, high skill individuals are significantly more likely move to urban counties than low skill individuals. In Columns 5-8 and the accompanying figures, it is clear that migration in early adulthood (ages 22-35) drives this effect while there is minimal difference in migration across the urban-rural divide between skill groups after 35. Interestingly, the coefficients for living in and migrating to counties in the 80^{th} percentile by population density are smaller when compared to living in and migrating to counties in the 95^{th} percentile. This suggests that skill-based geographic sorting, and particularly that which is driven by early adulthood migration, is most prominent along the division between counties that are a top urban county and those that are not.

Appendix B Data: U.S. Census

To characterize skill-based sorting and run counterfactual simulations, I utilize data from the following U.S. Census samples: 1980 5 percent, 1990 5 percent, 2000 5 percent, 2010 ACS, and 2019 5-year ACS.[53] In each year, I characterize a location's labor force skill mix by examining a subpopulation restricted to: prime working-age (25-54), employed, earning a positive wage, and living in the contiguous U.S. Using this sample, I create measures of high and low skill populations (and high skill share as a function of the two) and population density in each location and year. For regressions weighted by total population, this includes all individuals in a location and was downloaded from IPUMS NHGIS (Manson et al., 2022). Additionally, I use data on wages and rents to estimate various measures of a location's 'urbanicity' as described in the next section.

[53] All download from IPUMS Ruggles et al. (2022).

Table A1: Importance of Student Debt in for Career Outcomes

	(1) $\mathbb{1}$[Took Job Outside of Field of Study]	(2) $\mathbb{1}$[Have to Work >1 Job At a Time]	(3) $\mathbb{1}$[Have to Work More Hours Than Desired]
Student Debt ($thds, log)	0.086***	0.085***	0.122***
	[0.014]	[0.013]	[0.014]
Observations	2,080	2,080	2,080
R-squared	0.141	0.125	0.144
Dep. Mean	0.363	0.255	0.346
Student Loan Mean	3.171	3.171	3.171
Student Loan S.D.	0.953	0.953	0.953
Demographic Controls	✓	✓	✓
ELS Major FEs	✓	✓	✓

*** $p< 0.01$, ** $p< 0.05$, * $p< 0.1$
NOTE: Estimates from a linear probability model. Dependent variable is a survey question asking directly about the influence of student debt. Demographic controls include sex, race, father's education, father's occupation, and high school GPA. ELS major fixed effects are reported academic major in postsecondary education. Number of observations rounded to the nearest 10 per IES restricted-use guidelines. Robust standard errors clustered by sample design (survey base-year high school) in parentheses.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Base Year", 2002, "First Follow-up", 2004, "High School Transcripts", 2005, "Second Follow-up", 2006, and "Third Follow-up", 2012.

Table A2: Educational Migration Differentials: Overall Mobility

	(1) I[Move]	(2) I[Move]
I[High-Skill]	0.016**	
	(0.005)	
× I[Age: <22]		0.005
		(0.012)
× I[Age: 22-25]		0.051***
		(0.011)
× I[Age: 26-30]		0.096***
		(0.011)
× I[Age: 31-35]		0.033***
		(0.011)
× I[Age: 36-40]		−0.007
		(0.010)
× I[Age: 41-45]		−0.013
		(0.010)
× I[Age: 46-50]		−0.035***
		(0.011)
× I[Age: 51-55]		−0.013
		(0.011)
× I[Age: 56-60]		0.002
		(0.012)
× I[Age: 61-65]		0.022*
		(0.013)
Observations	86,447	86,447
Unique Individuals	6,983	6,983
R-squared	0.113	0.135
Dep. Variable Mean	0.282	0.282
Demographic Controls	✓	✓
Age Category FEs	✓	✓
Wave FEs	✓	✓
Cohort FEs	✓	✓
PSID Age FEs	✓	✓

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

NOTE: Dependent variable is an indicator function for whether an individual changed counties from wave $t-1$ to t. Based on the estimation of Equations 25 and 26. PSID Age is the number of waves an individual has been in the PSID sample. Robust standard errors clustered by individual in parentheses. Demographic controls include sex, race, and marital status.

SOURCE: Panel Study of Income Dynamics, restricted use data. Produced and distributed by the Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI (2022).

Table A3: Educational Migration Differentials: Live in 'Urban'

	(1) Top RUCC	(2) Pop. Density Percentile	(3) 80^{th} Percentile Pop. Density	(4) 95^{th} Percentile Pop. Density	(5) Top RUCC	(6) Pop. Density Percentile	(7) 80^{th} Percentile Pop. Density	(8) 95^{th} Percentile Pop. Density
I[High-Skill]	0.088***	4.376***	0.104***	0.127***				
	(0.012)	(0.508)	(0.011)	(0.014)				
× I[Age: <22]					−0.001	0.451	−0.013	0.001
					(0.025)	(1.121)	(0.025)	(0.028)
× I[Age: 22-25]					0.041***	3.139***	0.071***	0.100***
					(0.014)	(0.607)	(0.013)	(0.017)
× I[Age: 26-30]					0.095***	4.244***	0.094***	0.129***
					(0.013)	(0.539)	(0.012)	(0.015)
× I[Age: 31-35]					0.095***	4.564***	0.110***	0.128***
					(0.013)	(0.555)	(0.012)	(0.015)
× I[Age: 36-40]					0.096***	4.627***	0.118***	0.136***
					(0.014)	(0.570)	(0.013)	(0.016)
× I[Age: 41-45]					0.098***	4.284***	0.113***	0.137***
					(0.015)	(0.601)	(0.014)	(0.017)
× I[Age: 46-50]					0.095***	5.046***	0.122***	0.133***
					(0.017)	(0.659)	(0.015)	(0.019)
× I[Age: 51-55]					0.087***	4.811***	0.112***	0.143***
					(0.019)	(0.748)	(0.017)	(0.020)
× I[Age: 56-60]					0.091***	4.747***	0.105***	0.136***
					(0.021)	(0.832)	(0.020)	(0.023)
× I[Age: 61-65]					0.092***	4.519***	0.101***	0.115***
					(0.025)	(0.987)	(0.023)	(0.026)
Observations	86,447	86,447	86,447	86,447	86,447	86,447	86,447	86,447
Unique Individuals	6,983	6,983	6,983	6,983	6,983	6,983	6,983	6,983
R-squared	0.294	0.354	0.307	0.062	0.294	0.354	0.307	0.062
Dep. Variable Mean	0.449	83.297	0.700	0.380	0.449	83.297	0.700	0.380
Demographic Controls	✓	✓	✓	✓	✓	✓	✓	✓
Age Category FEs	✓	✓	✓	✓	✓	✓	✓	✓
Wave FEs	✓	✓	✓	✓	✓	✓	✓	✓
Cohort FEs	✓	✓	✓	✓	✓	✓	✓	✓
PSID Age FEs	✓	✓	✓	✓	✓	✓	✓	✓
Origin Location Control	✓	✓	✓	✓	✓	✓	✓	✓

*** $p< 0.01$, ** $p< 0.05$, * $p< 0.1$.

NOTE: Dependent variable in Columns 1, 3, 4, 5, 7, and 8 is an indicator function for whether an individual lived in an 'urban' county in wave t based on the following definitions of urban: metropolitan area of 1 million+ (Top RUCC), in the 80^{th} percentile of counties by population density (wave-decade specific), and in the 95^{th} percentile of counties by population density (wave-decade specific). Columns 2 and 6 dependent variable is county percentile in population density (wave-specific). Robust standard errors clustered by individual in parenthesis.

SOURCE: Panel Study of Income Dynamics, restricted use data. Produced and distributed by the Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI (2022).

Table A4: Educational Migration Differentials: Move to 'Urban'

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Top	Pop. Density	80^{th} Percentile	95^{th} Percentile	Top	Pop. Density	80^{th} Percentile	95^{th} Percentile
	RUCC	Percentile	Pop. Density	Pop. Density	RUCC	Percentile	Pop. Density	Pop. Density
I[High-Skill]	0.010***	0.014***	−0.002	0.019***				
	(0.001)	(0.003)	(0.002)	(0.005)				
× I[Age: <22]					0.001	−0.001	0.002	−0.005
					(0.005)	(0.010)	(0.006)	(0.008)
× I[Age: 22-25]					0.025***	0.038***	0.015***	0.019***
					(0.005)	(0.008)	(0.006)	(0.007)
× I[Age: 26-30]					0.028***	0.045***	0.007	0.023***
					(0.004)	(0.007)	(0.005)	(0.007)
× I[Age: 31-35]					0.009***	0.015**	−0.004	0.027***
					(0.003)	(0.006)	(0.004)	(0.007)
× I[Age: 36-40]					0.004	0.009	−0.009**	0.020***
					(0.003)	(0.006)	(0.004)	(0.007)
× I[Age: 41-45]					0.004	−0.005	−0.009*	0.011
					(0.003)	(0.006)	(0.005)	(0.007)
× I[Age: 46-50]					0.004	−0.000	−0.004	0.018**
					(0.003)	(0.007)	(0.005)	(0.008)
× I[Age: 51-55]					−0.003	0.001	−0.009*	0.013
					(0.004)	(0.007)	(0.005)	(0.008)
× I[Age: 56-60]					0.011***	0.008	−0.003	0.023**
					(0.004)	(0.008)	(0.005)	(0.009)
× I[Age: 61-65]					0.000	0.022***	0.003	0.024**
					(0.004)	(0.008)	(0.006)	(0.009)
Observations	86,447	86,447	86,447	86,447	86,447	86,447	86,447	86,447
Unique Individuals	6,983	6,983	6,983	6,983	6,983	6,983	6,983	6,983
R-squared	0.017	0.101	0.014	0.013	0.018	0.102	0.015	0.014
Dep. Variable Mean	0.022	0.090	0.031	0.053	0.022	0.090	0.031	0.053
Demographic Controls	✓	✓	✓	✓	✓	✓	✓	✓
Age Category FEs	✓	✓	✓	✓	✓	✓	✓	✓
Wave FEs	✓	✓	✓	✓	✓	✓	✓	✓
Cohort FEs	✓	✓	✓	✓	✓	✓	✓	✓
PSID Age FEs	✓	✓	✓	✓	✓	✓	✓	✓
Origin Location Control	✓	✓	✓	✓	✓	✓	✓	✓

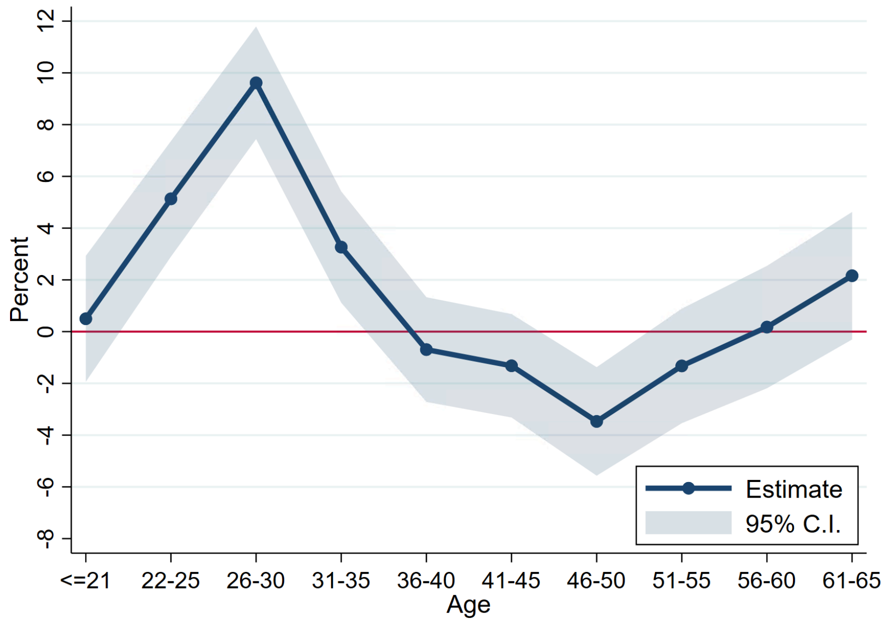
*** p< 0.01, ** p< 0.05, * p< 0.1.

NOTE: Dependent variable in Columns 1, 3, 4, 5, 7, and 8 is an indicator function for whether an individual moved to an 'urban' county in wave t based on the following definitions of urban: metropolitan area of 1 million+ (Top RUCC), in the 80^{th} percentile of counties by population density (wave-decade specific), and in the 95^{th} percentile of counties by population density (wave-decade specific). Columns 2 and 6 dependent variable is county percentile in population density (wave-specific). Robust standard errors clustered by individual in parenthesis.

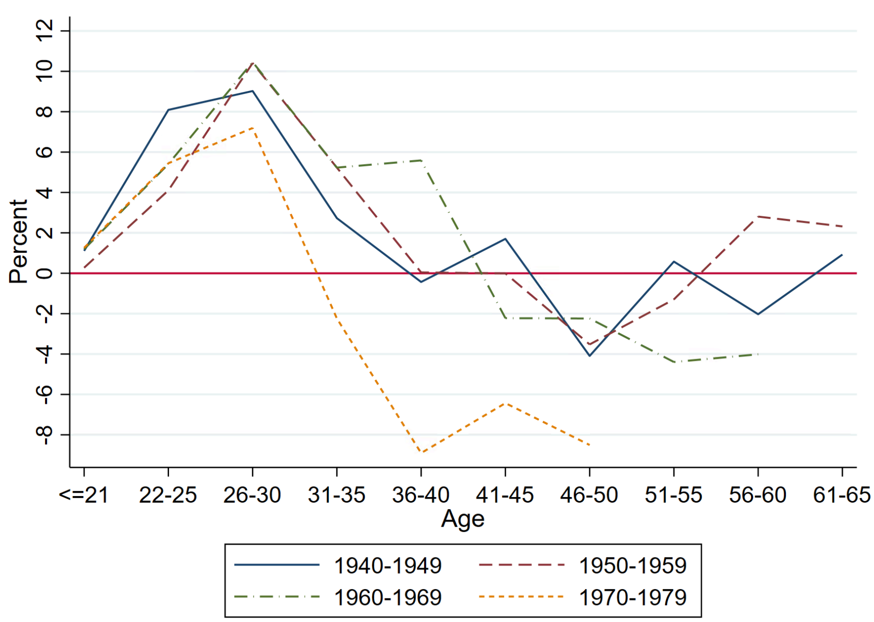
SOURCE: Panel Study of Income Dynamics, restricted use data. Produced and distributed by the Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI (2022).

Figure A1: Association Between Being High-Skill and
Probability of Changing Locations

(a) Aggregate Estimate



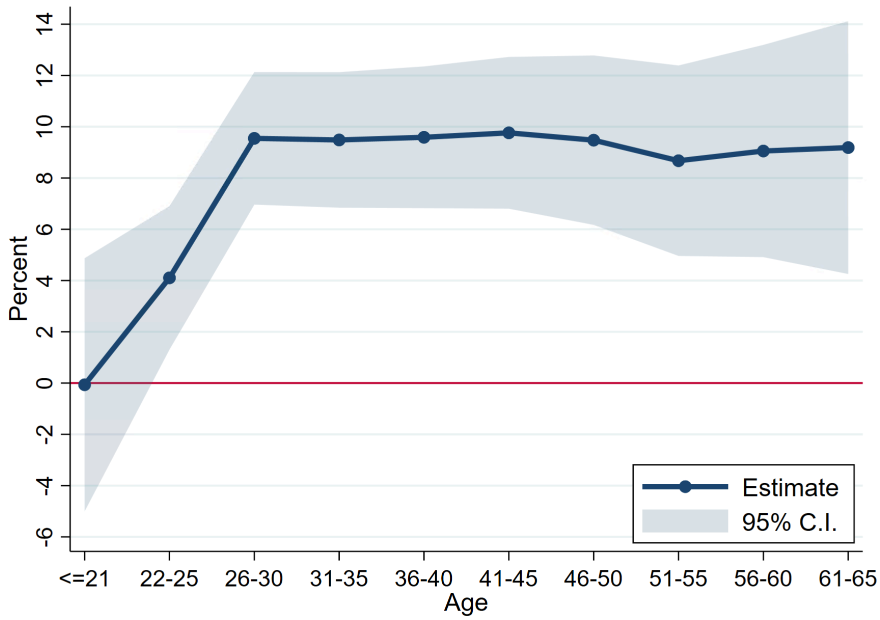
(b) By Cohort



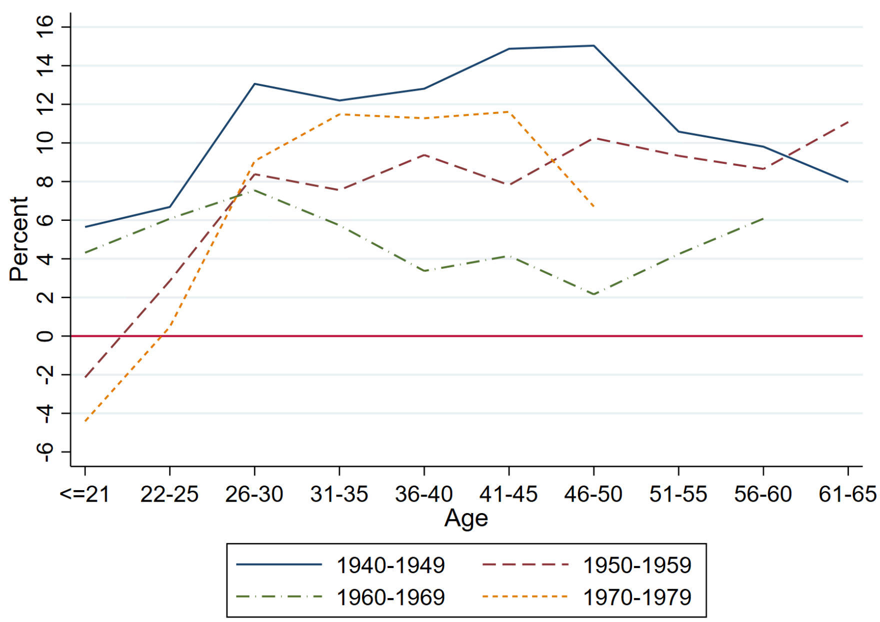
NOTE: Figure (a) shows the association between being high-skill and the probability of changing county of residence in a given PSID wave relative to the previous wave. Figure (b) shows this correlation for each cohort separately. Graphical representation of the results in Table A2 Column 2.
SOURCE: Panel Study of Income Dynamics, restricted use data. Produced and distributed by the Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI (2022).

Figure A2: Association Between Being High-Skill and
Probability of Living in Top Metropolitan County (RUCC-Defined)

(a) Aggregate Estimate



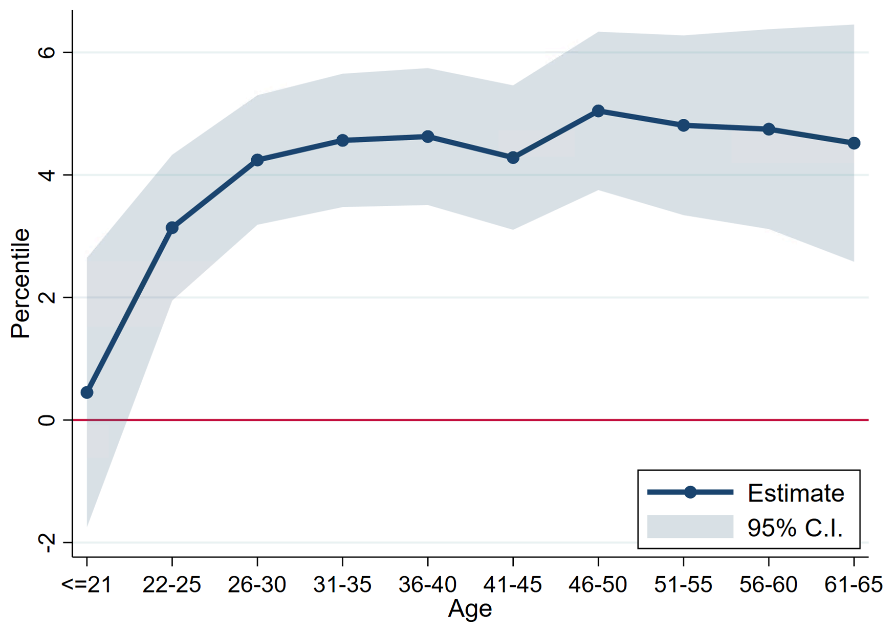
(b) By Cohort



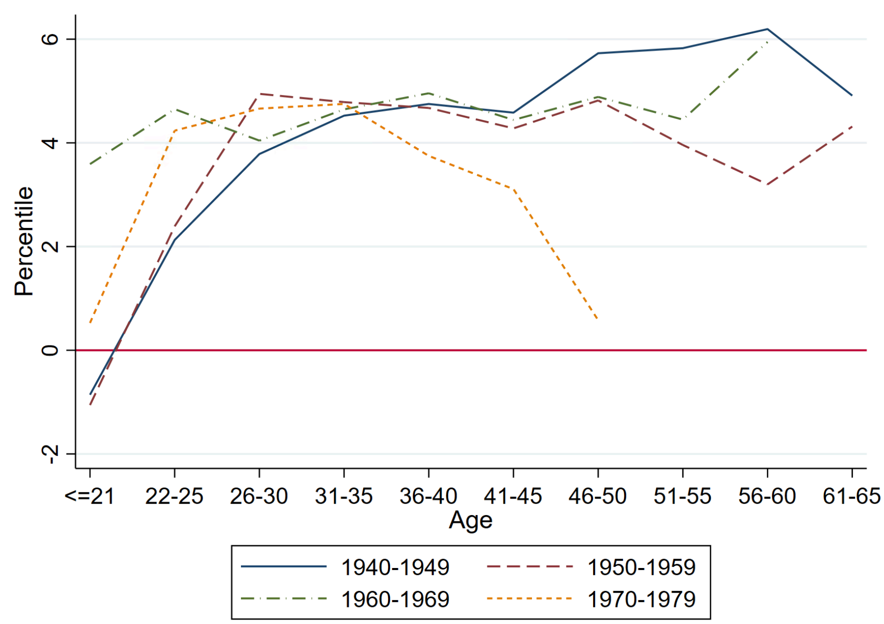
NOTE: Figure (a) shows the association between being high-skill and the probability of living in a top metropolitan county in a given PSID wave, as defined by the Rural-Urban Continuum Codes (decade-specific). Figure (b) shows this correlation for each cohort separately. Graphical representation of the results in Table A3 Column 5.
SOURCE: Panel Study of Income Dynamics, restricted use data. Produced and distributed by the Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI (2022).

Figure A3: Association Between Being High-Skill and Percentile of County Population Density

(a) Aggregate Estimate



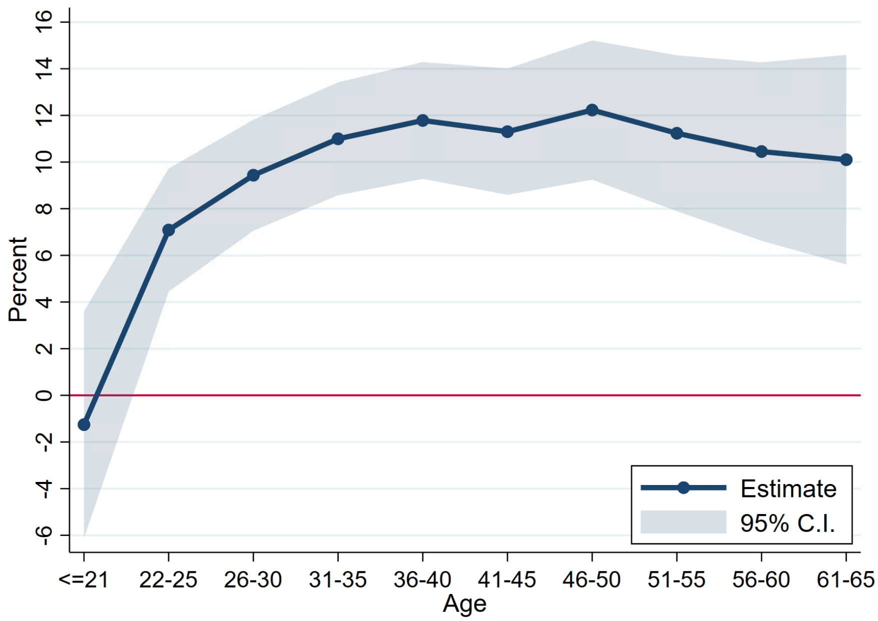
(b) By Cohort



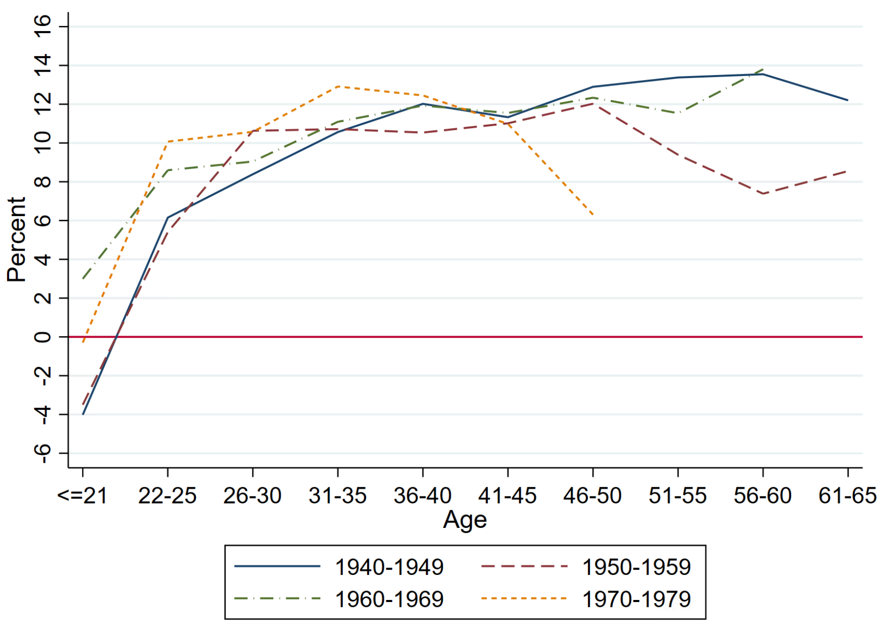
NOTE: Figure (a) shows the association between being high-skill and the percentile of county population density (decade-specific) in a given PSID wave. Figure (b) shows this correlation for each cohort separately. Graphical representation of the results in Table A3 Column 6.
SOURCE: Panel Study of Income Dynamics, restricted use data. Produced and distributed by the Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI (2022).

Figure A4: Association Between Being High-Skill and
Probability of Living in 80th Percentile Population Density County

(a) Aggregate Estimate



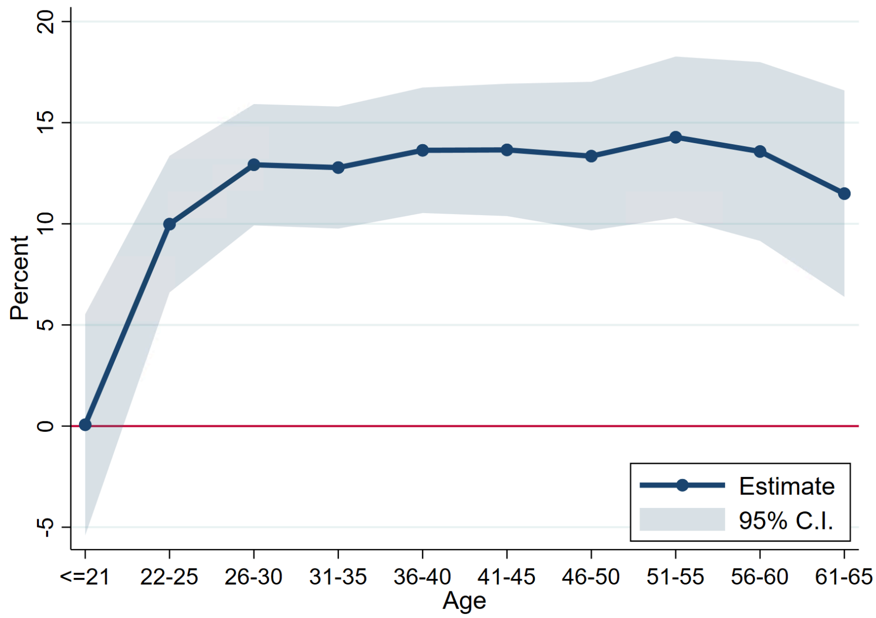
(b) By Cohort



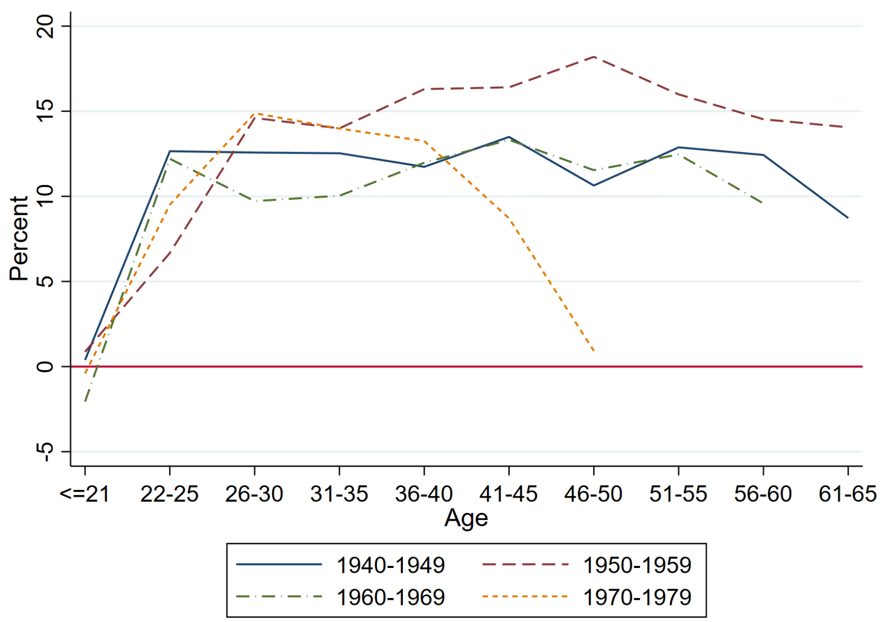
NOTE: Figure (a) shows the association between being high-skill and the probability of living in a county with a population density in the 80th percentile (decade-specific) in a given PSID wave. Figure (b) shows this correlation for each cohort separately. Graphical representation of results in Table A3 Column 7.
SOURCE: Panel Study of Income Dynamics, restricted use data. Produced and distributed by the Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI (2022).

Figure A5: Association Between Being High-Skill and Probability of Living in 95th Percentile Population Density County

(a) Aggregate Estimate



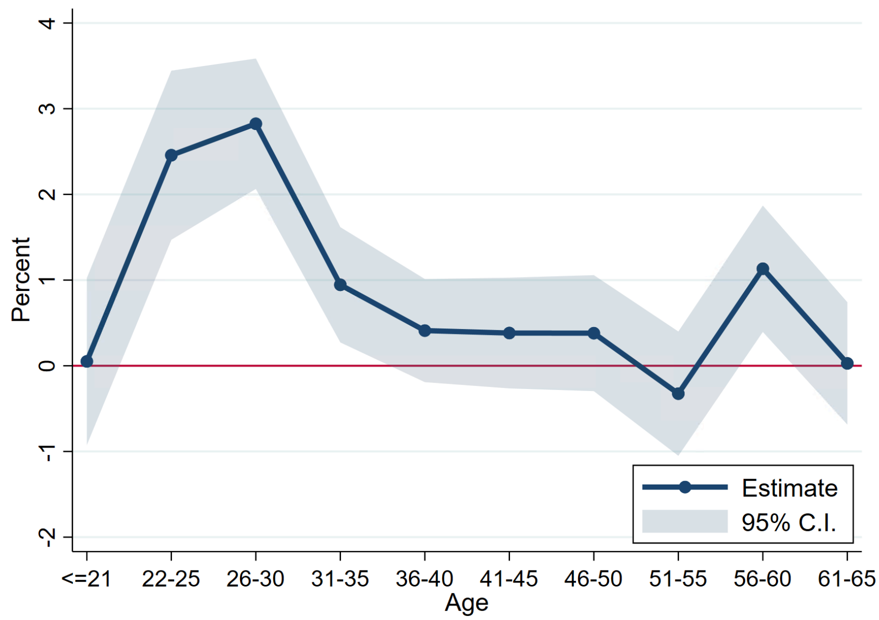
(b) By Cohort



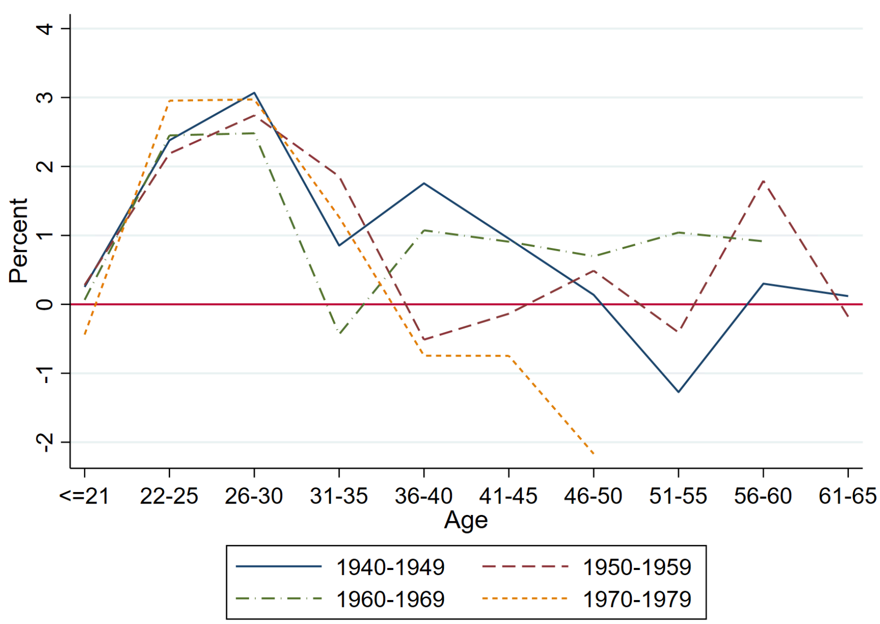
NOTE: Figure (a) shows the association between being high-skill and the probability of living in a county with a population density in the 95th percentile (decade-specific) in a given PSID wave. Figure (b) shows this correlation for each cohort separately. Graphical representation of results in Table A3 Column 8.
SOURCE: Panel Study of Income Dynamics, restricted use data. Produced and distributed by the Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI (2022).

Figure A6: Association Between Being High-Skill and
Probability of Moving to Top Metropolitan County (RUCC-Defined)

(a) Aggregate Estimate



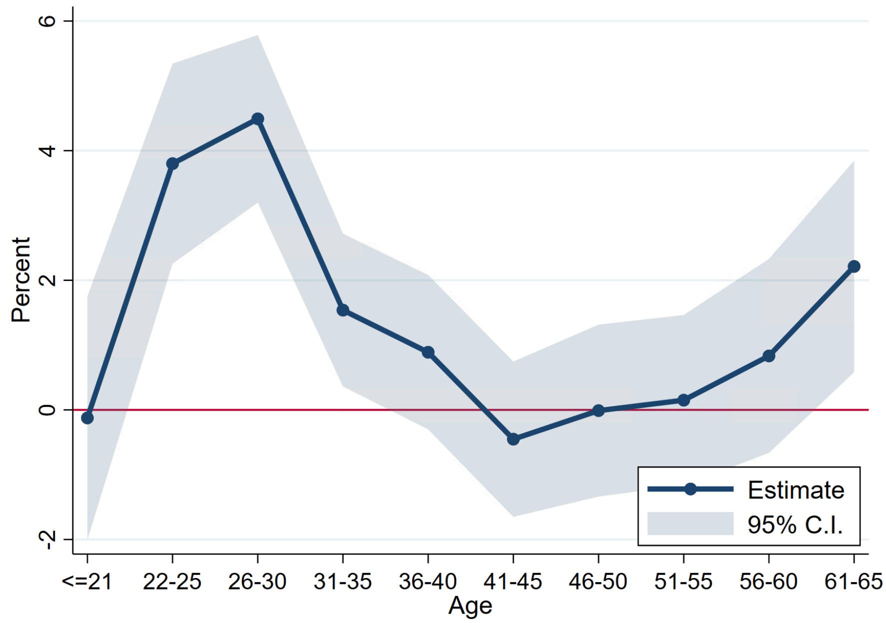
(b) By Cohort



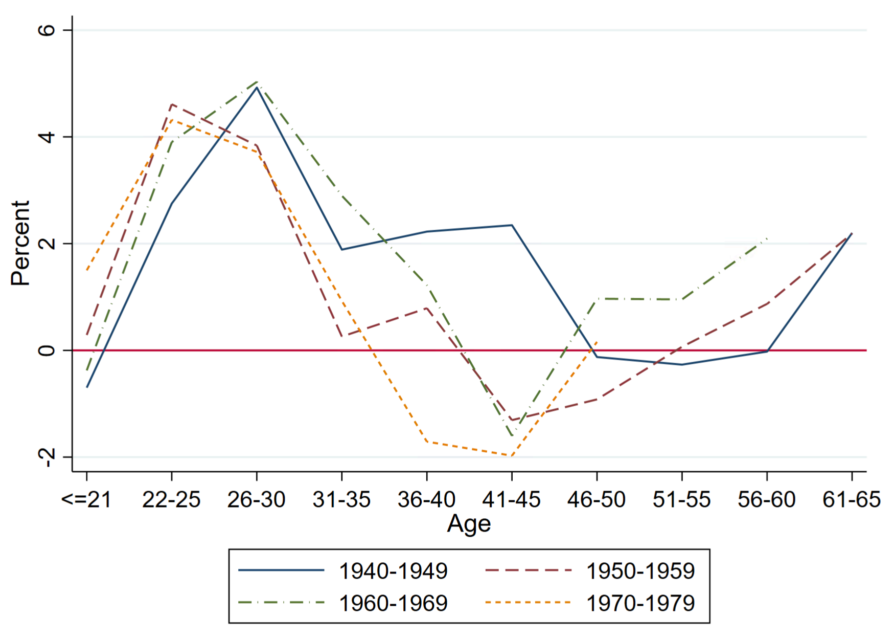
NOTE: Figure (a) shows the association between being high-skill and the probability of moving to a top metropolitan county in a given PSID wave, as defined by the Rural-Urban Continuum Codes (decade-specific). Figure (b) shows this correlation for each cohort separately. Graphical representation of results in Table A4 Column 5.
SOURCE: Panel Study of Income Dynamics, restricted use data. Produced and distributed by the Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI (2022).

Figure A7: Association Between Being High-Skill and
Probability of Moving to Higher Population Density County

(a) Aggregate Estimate



(b) By Cohort

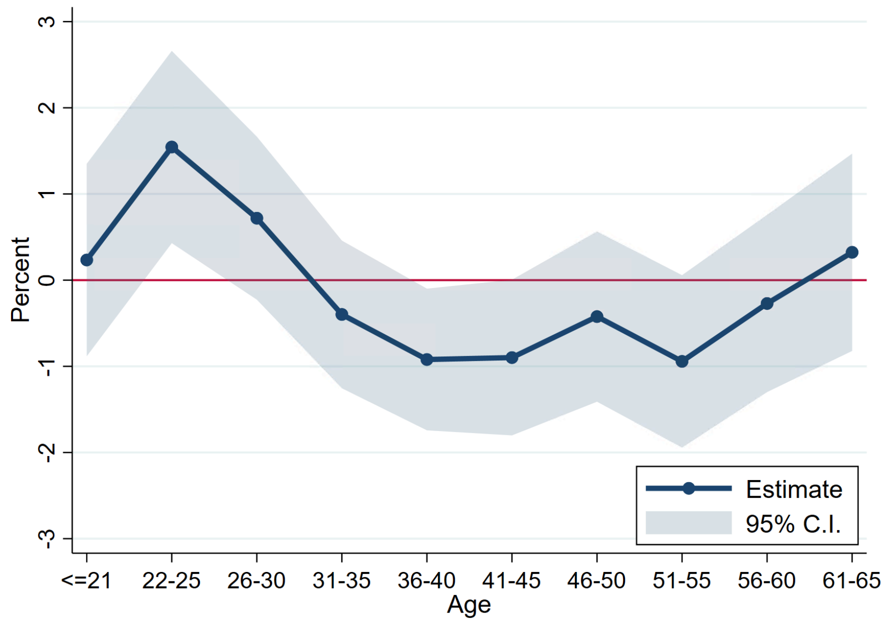


NOTE: Figure (a) shows the association between being high-skill and the probability of moving to a county with a higher percentile of county population density (decade-specific) in a given PSID wave, relative to the previous county of residence. Figure (b) shows this correlation for each cohort separately. Graphical representation of results in Table A4 Column 6.

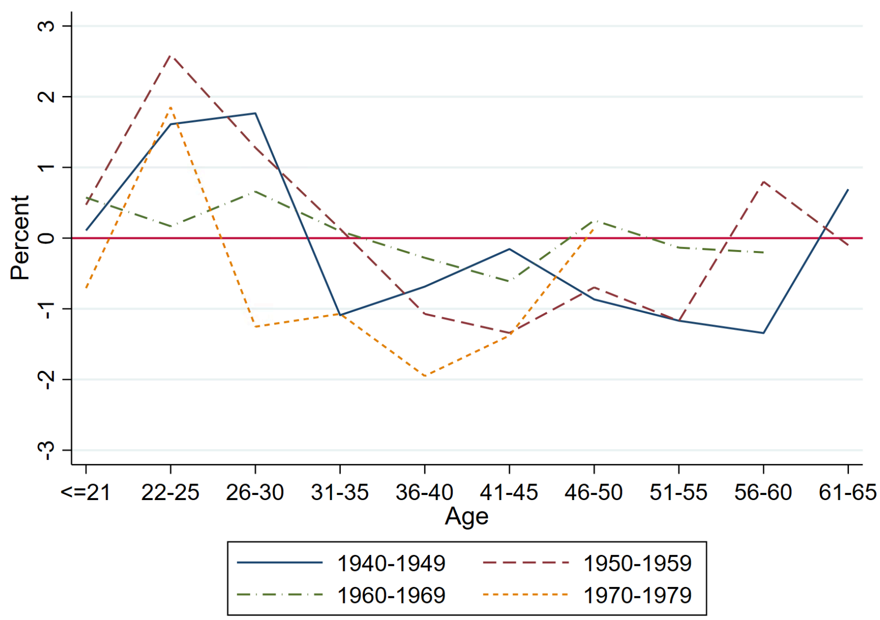
SOURCE: Panel Study of Income Dynamics, restricted use data. Produced and distributed by the Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI (2022).

Figure A8: Association Between Being High-Skill and Probability of Moving to 80^{th} Percentile Population Density County

(a) Aggregate Estimate



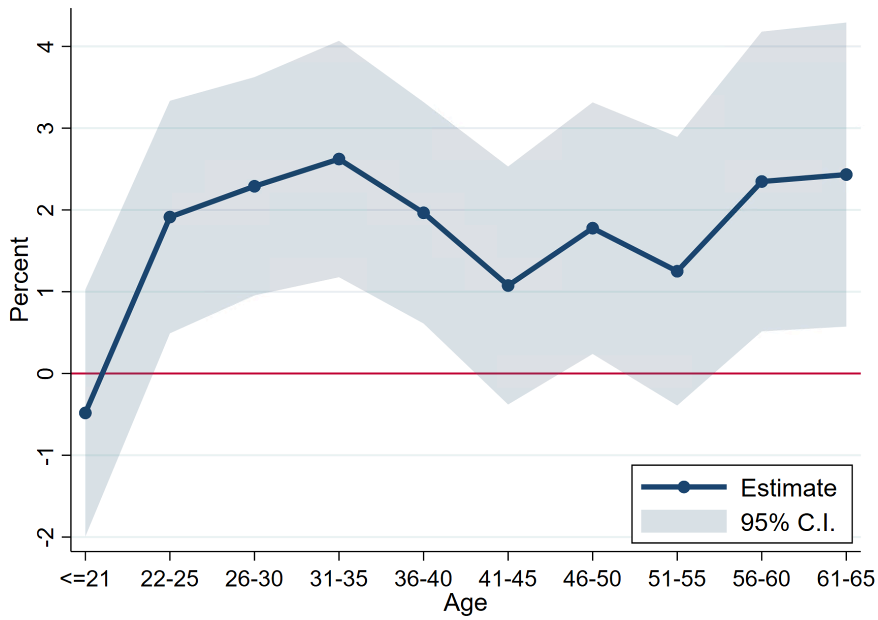
(b) By Cohort



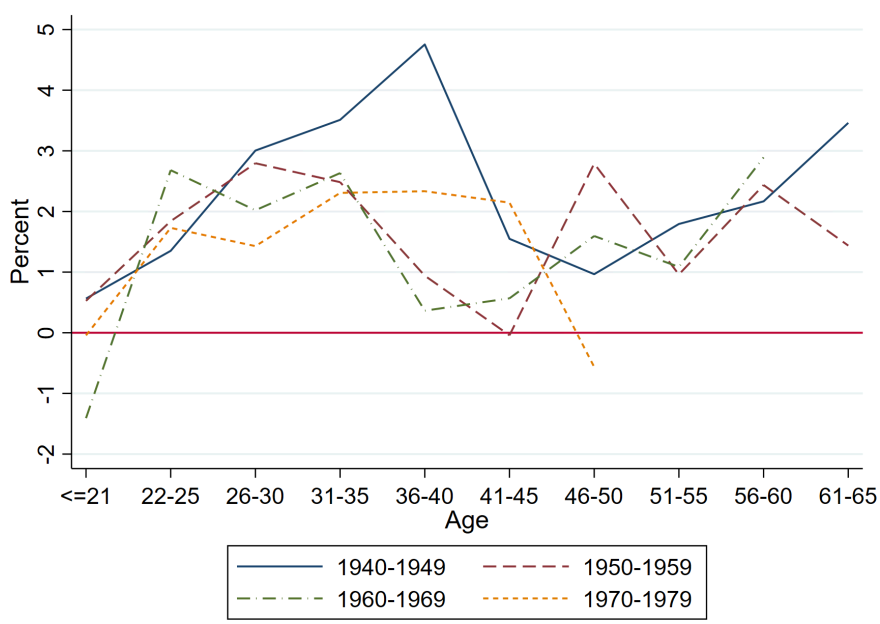
NOTE: Figure (a) shows the association between being high-skill and the probability of moving to a county with a population density in the 80^{th} percentile (decade-specific) in a given PSID wave. Figure (b) shows this correlation for each cohort separately. Graphical representation of results in Table A4 Column 7.
SOURCE: Panel Study of Income Dynamics, restricted use data. Produced and distributed by the Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI (2022).

Figure A9: Association Between Being High-Skill and
Probability of Moving to 95$^{\text{th}}$ Percentile Population Density County

(a) Aggregate Estimate

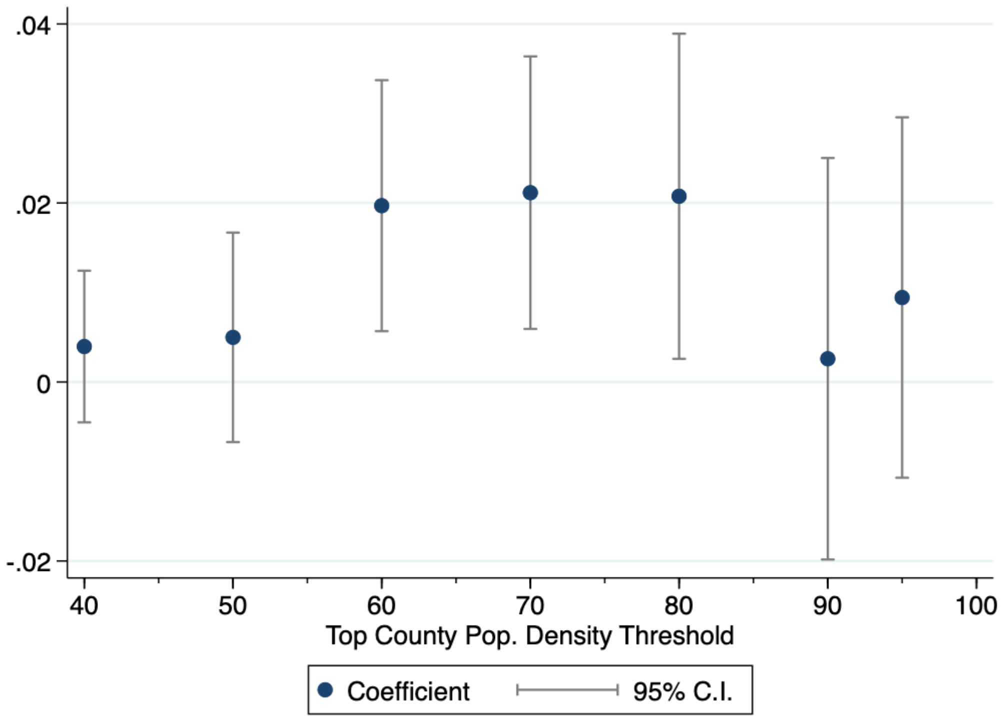


(b) By Cohort



NOTE: Figure (a) shows the association between being high-skill and the probability of moving to a county with a population density in the 95$^{\text{th}}$ percentile (decade-specific) in a given PSID wave. Figure (b) shows this correlation for each cohort separately. Graphical representation of results in Table A4 Column 8.
SOURCE: Panel Study of Income Dynamics, restricted use data. Produced and distributed by the Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI (2022).

Figure A10: Association Between Student Debt ($1,000s, log) and Probability of Post-College County Location Being in Top X Percentile Pop. Density



NOTE: From estimation of Equation 4 including origin and college county location controls (defined using same population density threshold as dependent variable), demographic controls (sex, race, father's education, father's occupation, and high school GPA), and ELS major fixed effects. Confidence intervals reflect robust standard errors clustered by sample design (survey base-year high school). As shown, the coefficient is near zero and insignificant when considering 40^{th}, 50^{th}, 90^{th}, and 95^{th} percentile thresholds. It is positive, significant, and nearly identical when considering location choice across the 60^{th}, 70^{th}, and 80^{th} percentiles. This is further suggestive evidence that student debt is pushing individuals who would locate in counties in the bottom 60^{th} percentile of population density (perhaps one definition of "rural") to locate in counties in the top 20^{th} percentile.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Base Year", 2002, "First Follow-up", 2004, "High School Transcripts", 2005, "Second Follow-up", 2006, and "Third Follow-up", 2012.