ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 22-580

Grads on the Go: Measuring College-Specific Labor Markets for Graduates

Johnathan G. Conzelmann
University of North Carolina at Chapel Hill

Steven W. Hemelt
University of North Carolina at Chapel Hill

Brad J. Hershbein
W.E. Upjohn Institute for Employment Research

Shawn Martin
University of Michigan

Andrew Simon
University of Chicago

Kevin M. Stange
University of Michigan

This paper introduces a new measure of the labor markets served by colleges and universities across the United States. About 50 percent of recent college graduates are living and working in the metro area nearest the institution they attended, with this figure climbing to 67 percent in-state. The geographic dispersion of alumni is more than twice as great for highly selective 4-year institutions as for 2-year institutions. However, more than one-quarter of 2-year institutions disperse alumni more diversely than the average public 4-year institution. In one application of these data, we find that the average strength of the labor market to which a college sends its graduates predicts college-specific intergenerational economic mobility. In a second application, we quantify the extent of "brain drain" across areas and illustrate the importance of considering migration patterns of college graduates when estimating the social return on public investment in higher education.

VERSION: May 2022

Suggested citation: Conzelmann, Johnathan G., Steven W. Hemelt, Brad J. Hershbein, Shawn Martin, Andrew Simon, and Kevin M. Stange. (2022). Grads on the Go: Measuring College-Specific Labor Markets for Graduates. (EdWorkingPaper: 22-580). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/y5p3-zs70

Grads on the Go:
Measuring College-Specific Labor Markets for Graduates[*]

Johnathan G. Conzelmann
University of North Carolina at Chapel Hill

Shawn Martin
University of Michigan

Steven W. Hemelt
University of North Carolina at Chapel Hill

Andrew Simon
University of Chicago

Brad J. Hershbein
W.E. Upjohn Institute for Employment Research

Kevin M. Stange
University of Michigan

May 2022

Abstract

This paper introduces a new measure of the labor markets served by colleges and universities across the United States. About 50 percent of recent college graduates are living and working in the metro area nearest the institution they attended, with this figure climbing to 67 percent in-state. The geographic dispersion of alumni is more than twice as great for highly selective 4-year institutions as for 2-year institutions. However, more than one-quarter of 2-year institutions disperse alumni more diversely than the average public 4-year institution. In one application of these data, we find that the average strength of the labor market to which a college sends its graduates predicts college-specific intergenerational economic mobility. In a second application, we quantify the extent of "brain drain" across areas and illustrate the importance of considering migration patterns of college graduates when estimating the social return on public investment in higher education.

Keywords: colleges, labor markets, postsecondary education, economic mobility

JEL Codes: I23, I25, J21, J40, J61

[*] This research was supported by the National Science Foundation (Collaborative Grant #1919387), Russell Sage Foundation (Grant #1811-09737), and Institute of Education Sciences (Grant R305B150012). We are grateful to Celeste Carruthers and Charlie Clotfelter for helpful comments. The data assembled on college-specific labor markets are available for research use via the Open ICPSR Archive at http://doi.org/10.3886/E170381. All errors and any opinions are our own.

I. Introduction

A principal aim of colleges is to equip students with knowledge, skills, and connections that will lead to labor market success and future wellbeing. A clear understanding of the labor markets in which a college operates stands to inform institution-level decision-making as well as broader questions about links between college-going and economic development, mobility, and inequality (e.g., Chetty et al., 2020).[1] However, most work in this area focuses on characterizing markets that colleges face for incoming students (e.g., Hoxby & Turner, 2019; Monarrez & Washington, 2020; Sá et al., 2004) rather than the markets where students ultimately live and work. Even when the latter is the focus, due to data limitations, the surrounding state or nearest metropolitan statistical area (MSA) is often a crude proxy. Given the wide range of institutions that speckle the United States, such simplifications may mismeasure the relevant labor markets for many colleges.

The ability to characterize relevant labor markets for an institution's graduates has implications for our understanding of the supply of college-educated workers, gaps in skill demand, institutional planning, and public finance concerns about the loss of homegrown graduates to other labor markets (e.g., Kelchen & Webber, 2018; Winters, 2020). To facilitate the study of these topics, we develop a new measure of such markets using data on alumni from publicly accessible institutional webpages on LinkedIn (LI). These data provide aggregate information on the geographic locations of former students for nearly all public and private non-profit colleges and universities in the United States. We have made these data available for research use via the Open ICPSR Archive at http://doi.org/10.3886/E170381 (Conzelmann et al.,

[1] We are focused on the labor markets for the graduates of colleges, specifically undergraduates. Institutions also face labor markets for staff and faculty, each of which vary in scope according to the research intensity of the institution.

2022). Various validation exercises demonstrate a strong correspondence between LI and government data sources in the number and location of graduates, lending support to our measure. However, the LI data cover a more extensive range of institutions and are available at finer levels of geography than any current source of such information.

We operationalize college labor markets for 2,600 public and private non-profit institutions that offer at least an associate's degree and participate in the Title IV federal financial aid programs. For each institution, we construct a set of weights from the LI data that represent the share of the institution's alumni living in geographic units that map to one or more Core-Based Statistical Areas (CBSAs).

Borrowing from the literature on market concentration, we calculate a Herfindahl-Hirschman Index (HHI) to quantify each institution's geographic clustering of graduates across the United States. We also compute the average distance traveled by alumni of each institution. Taken together, these statistics permit us to characterize colleges' labor markets in a more detailed fashion and for a wider group of institutions, relative to existing data sources. In general, we find that graduates of more selective institutions appear in more distant and geographically diverse locations. However, there is appreciable variation in the HHI and distance metrics across groups of institutions defined by level (2-year/4-year) and selectivity. Indeed, we find that 28 percent of 2-year institutions boast a diversity of graduates' locations that surpasses the average among public 4-year institutions, even though the typical 2-year graduate tends not to venture far from the institution she attended.

The labor markets served by colleges differ substantially from the markets of their self-identified peer institutions. We find limited overlap (about 7 percent) in the peers identified by institutions—primarily competitors in terms of incoming students and faculty talent (Fuller &

O'Leary, 2012)—and the peers measured by the geographic locations of graduates. An improved understanding of the labor markets to which their graduates flow may allow institutions to better equip students with the knowledge and skills necessary to compete in those markets or find ways to adapt to reach new markets.

We demonstrate the utility of our labor market measure through two empirical applications. Our first application examines the relationship between college markets and rates of intergenerational economic mobility. Chetty et al. (2020) find that the colleges with the highest mobility rates "do not differ substantially from other colleges on institutional characteristics like public-versus-private status, instructional expenditures, or endowments" (p. 1570), which points to the need for further exploration of such differences based on measures of student outcomes. We posit that one way a college generates upward mobility for its students is through the labor market connections it provides. Institutions with well-established networks to robust labor markets across the country may more easily facilitate such mobility for their graduates from low-income backgrounds.

Using wage data from the American Community Survey (ACS), we calculate the wage for bachelor's degree recipients averaged across a college's labor markets, weighted by our college-specific labor market shares. We find that this measure of the strength of the labor markets to which a college sends its graduates meaningfully predicts variation in the bottom-to-top-quintile economic mobility rates across colleges, even after conditioning on a range of institutional and student-body characteristics as well as geography fixed effects that capture access to a common local labor market. A 10 percent increase in the average bachelor's degree wage of the relevant labor market for an institution is associated with a 14.2 percent increase in the likelihood that a student from a household in the bottom income quintile reaches the top income quintile herself.

In our second application, we explore the role of migration in understanding the social return to public investment in higher education. Understanding and quantifying the migration of college graduates has implications for the financing of higher education. Bound et al. (2019) suggest that increased mobility of college graduates over time has contributed to the decline in state appropriations over the past 40 years, and Hurst et al. (2022) provide additional evidence that taxpayers' support for public funding on higher education depends on the return on investment. We show that the social benefits of public investment in postsecondary education disproportionately accrue to high-wage and urban/suburban areas due to graduate mobility, which we now can quantify at sub-state and institution-specific levels. We also find that regional public universities tend to produce the greatest number of 4-year college graduates who remain and work in-state per dollar of state funding.

More broadly, in research that aims to study the migration of college-educated workers and its determinants (e.g., Molloy et al., 2011) or estimate the responsiveness of human capital investments to demand shocks (e.g., Acton, 2020; Blom, Cadena, & Keys, 2020; Weinstein, 2020), our measure of college labor markets serves as a key ingredient for properly characterizing labor and skill demand. Our data should also be useful for studying spatial policies—for example, computing the marginal value of public funds of education policies (Hendren & Sprung-Keyser, 2020) in a federal system (Agrawal et al., 2022), which involves measuring changes in the tax base from new college graduates across space. These examples highlight the wide and policy-relevant potential uses of our new measure of college-specific labor markets.

The paper unfolds as follows. The next section describes the construction of our new measure of college-specific labor markets and discusses results from a series of validation checks. Section III uses these new data in a range of descriptive analyses that characterize the labor markets for college graduates. Section IV presents findings from our application on colleges and intergenerational economic mobility. Section V discusses results from our application on "brain drain" and the social return to public investment in higher education. Section VI concludes.

II. Measuring College Markets

A. Current Data Sources for Characterizing Destinations of College Graduates

Few existing data sources contain the requisite information to tie an individual's current area of residence to the college she attended; moreover, such sources are not typically representative of entire institutions or cover only a small subset of them. For instance, longitudinal datasets from the National Center for Education Statistics (NCES) include detailed geographic information on sample members up to 10 years after the completion of a bachelor's degree—however, institution-level estimates are not feasible and even state-level estimates are unreliable because of limited sample sizes.[2]

A new resource from the Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) program, the Post-Secondary Employment Outcomes (PSEO) project, contains data on employment outcomes for graduates of about 200 public 4-year institutions in 17 states. The data include counts of graduates employed in each Census Division, as well as in the institution's own state. At present, PSEO covers less than a third of public 4-year institutions and has sparse

[2] The same issues arise with datasets from the National Science Foundation, such as the National Survey of College Graduates. Moreover, information on the specific institution a worker attended for their bachelor's degree is often either unavailable or is a restricted-use variable.

coverage of private institutions in just three states (Indiana, Ohio, and Virginia). Further, the relatively coarse geographic information available is not sufficiently specific for measuring college labor markets, especially to the extent that former students congregate in specific metropolitan areas or nearby states due to the nature of local job opportunities.

B. A New Measure of College Labor Markets

We introduce a comprehensive measure of the geographic dispersion of college graduates from nearly all public and private non-profit institutions in the United States. Using the Integrated Postsecondary Education Data System (IPEDS), we define a population of 2,832 public and private non-profit colleges that are located in the 50 U.S. states or DC, offer at least an associate's degree, and participated in the Title IV federal financial aid program every year from 2010 to 2018.

For each institution we obtain publicly available information on college alumni webpages from LI, the popular professional social networking platform.[3] Nearly every college and university in the U.S. has claimed an official page—which houses aggregate, college-level counts of users who self-report having attended the school, counts from the 15 most common geographies where alumni reside,[4] and counts of other top-15 pieces of information, including employers, industry, skills, and college majors. These pages can be further filtered by years of college attendance and by individual geographies outside the top-15 list.

Of the 2,832 institutions in our population, we located and obtained LI geographic data for 2,600 (approximately 92 percent). These institutions account for 99 percent of the associate's and bachelor's degrees awarded from 2010 through 2018, per IPEDS counts. For each of the

[3] The aggregate college-level information is viewable to any user with an LI account, which is free to obtain.
[4] The entire United States is the top geography reported for each institution; thus, we observe 14 sub-state locations for each institution. Sub-state areas (e.g., Greater New York City Area) roughly correspond to one or more Core-Based Statistical Area (CBSA) from the U.S. Census Bureau.

2,600 schools with a valid page, we obtain alumni counts using year (of attendance) filters for 2010 through 2015. Our target population is bachelor's and associate's degree recipients from each college in our sample between 2010 and 2018. We collect data based on attendance through 2015 to minimize the number of current students in the counts. Because we are using aggregate data from institutional pages, we cannot explicitly limit our LI search to (bachelor's or associate's) graduates, and thus the year filters capture all individuals who report attendance that overlapped with the specified date range (in our case 2010–2015) for any degree program.

For each institution in our sample, we first collect the number of alumni in each of the non-country-wide LI geographies within its top-15 locations. Across institutions, this accounts for about 82 percent of alumni in the U.S., a known figure since the top geography on each institution's page in our sample is the whole United States. We then supplement these data by searching over specific institution-geography pairs. For each institution, we next find the number of alumni in the remaining labor markets in their own state. Finally, we augment each institution's set of locations with missing geographies from a pooled group of three matched peer institutions, determined using a Mahalanobis distance algorithm.[5] That is, for a given institution, we obtained counts from locations found in the top-15 LI geographies of three peer institutions that did not appear in the focal institution's initial top-15 list. After these additions, our dataset covers about 84 percent of U.S. alumni. Since we always capture all graduates residing in an institution's own state, the remaining 16 percent of graduates reside outside the state.

We work with this analytic sample of 2,600 institutions and 278 U.S.-specific LI geographies throughout the rest of the paper. The final dataset contains observations uniquely identified by institution-geography pairs, and each pair is accompanied by both a count and a

[5] Variables include each institution's state, control (public/private), level (2-year/4-year), percent of freshman attending from in-state, basic Carnegie classification, and Barron's selectivity rating.

share of graduates residing in that location.[6] Within institution, the shares sum to 1 and the raw graduate counts sum to the institution's total U.S. alumni for whom a location was identified.[7] Because LI geographies map closely to one or more CBSAs,[8] one can supplement the dataset of shares with characteristics of the geographies where graduates of an institution reside. For example, in our first empirical application, we use this crosswalk to develop a proxy for the "average strength" of the labor markets to which an institution's graduates flow based on data at the CBSA level from various government agencies (e.g., Census and Bureau of Labor Statistics). Appendix B provides more details the about the data collection process and analytic steps necessary to arrive at our final dataset.

C. *Validation of LinkedIn Coverage*

We assess the validity of the LI data for measuring college markets through several validation analyses. Participation in LI is voluntary, and we can only speculate on how students decide to create a profile and the information they publicize.[9] Thus, if participation is reflective of certain types of students or institutions, these data may generate a skewed characterization of college labor markets. We validate the LI data against several official government data sources to alleviate such concerns.

To assess the overall coverage of the LI data, we first compare counts of bachelor's and associate's degrees awarded by each institution (between 2010 and 2018; from IPEDS) to numbers of LI students we find residing in the United States. Panel A of Figure 1 plots these two measures overlaid with a simple linear regression line, where each observation is an institution

---

[6] For all 2,600 institutions, we take the alumni counts from each of the 278 available U.S.-specific geographies in LI and divide them by the institution's total number of alumni residing in the United States. These shares sum to 1 within institution after we add an observation with the count and share of "unlocated" U.S. graduates.
[7] This requires that we renormalize the shares after ignoring unlocated graduates. See Appendix B for details.
[8] Appendix B explains the construction of the CBSA-LI crosswalk.
[9] Users are unlikely to misrepresent easily verifiable information such as educational and employment histories (Guillory & Hancock, 2012).

weighted by its completion count. We see a strong positive relationship between IPEDS and LI, with an R-squared of 0.94 and a slope of 0.66. This suggests the LI data cover about 66 percent of graduates found in the true counts, on average, while the R-squared suggests this coverage is rather homogenous across institutions. Private non-profit institutions exhibit slightly lower coverage compared to public institutions. The slope of a regression line including only private institutions is smaller, at 0.58.[10] However, the R-squared from a regression on this subsample is comparable to the full sample, at 0.93.

Beyond aggregate coverage, geographical selection into LI would also undermine the validity of our measure. For example, the use of LI may be more common in certain parts of the country than others—and thus students with an LI profile might systematically differ from those without in terms of geography. We address this concern using data from the PSEO on 209 public 4-year institutions spread across 17 states[11] that are also in our analytic sample (U.S. Census Bureau, 2022).[12]

In Panel B of Figure 1, we compare PSEO data on the percentage of bachelor's degree graduates employed, five years after graduating, in the same state as the institution they attended to the percentage of alumni residing in their institution's state as reported in our LI data. We observe a strong positive relationship between the government data source and LI, with an R-squared of 0.75 and a slope of 0.93. Shares are about two percentage points lower, on average, in

[10] Among publics, coverage is higher at 4-year institutions (slope = 0.68) than 2-year institutions (slope = 0.58).
[11] These states are Alabama, Arizona, Colorado, Connecticut, Indiana, Iowa, Louisiana, Maine, Michigan, Missouri, New York, Ohio, Pennsylvania, Texas, Utah, Virginia, and Wisconsin.
[12] We note that PSEO covers only employed graduates, based on state unemployment insurance wage records. Since we are interested in labor markets for graduates, this restriction is not a problem for our purposes.

LI than in PSEO.[13] Encouragingly, we see little variation between flagship institutions and non-flagship public institutions.[14]

While overall coverage of LI is quite high and we find little evidence of bias in coverage of in- versus out-of-state students, non-random selection into LI based on major or field of study could still be a concern. As a check on the representativeness of the LI data along this dimension, we compare the aggregate distribution of majors in the LI data among the 4-year institutions in our sample to IPEDS degree completions between 2010 and 2018 by two-digit Classification of Instructional Programs (CIP) codes. Across institutions, some fields like Business are clearly overrepresented in LI, with 32 percent of graduates versus 18 percent of bachelor's degrees awarded in IPEDS. Others, like Health and Education, are underrepresented (6 vs. 10 percent and 2 vs. 5 percent, respectively).[15] The full distributions of reported majors from both sources, along with their differences, appear in Appendix Table A1.

In Figure 2, we explore the degree to which the under- and over-representation of different fields in the LI data alters basic conclusions about the labor markets of institutions' graduates. In Panel A, we compare institution-level PSEO estimates of the share of graduates employed in-state, built from a weighted-average across programs of study (y-axis), to the estimates we would get if we altered the weights to reflect the distribution of majors observed in

[13] Because we cannot restrict the LI data to undergraduate students, we also used the PSEO data to examine the association between the shares of undergraduate completers and graduate completers working in-state 5 years after graduation. The correlation coefficient weighted by the total number of employed graduates (across both degree levels) is 0.83, suggesting that any inclusion of graduate students in the LI data does not create much bias for capturing the mobility of undergraduates. It is also important to note that several professional schools (e.g., law) have their own LinkedIn pages, where graduate students may affiliate on the platform. Doing so would effectively remove them from the undergraduate counts on the overall institution's page.

[14] Several of the exceptions occur in one state—Pennsylvania—which has many regional colleges near state borders. As the PSEO data are based on employer location, not residence, it is possible that interstate commuters help explain the discrepancy.

[15] Aggregate information on majors is listed separately from information on alumni geographic locations in the LI data. The average share of total graduates from an institution for which we can observe major is 58 percent. This is due to the observation only of the top-15 most listed majors and likely also to students who do not list major on their profiles. Together, this suggests caution when interpreting magnitudes of these field-specific estimates.

our LI data for each institution (x-axis). We see that the share of graduates employed in-state changes very little when we allow for over- and under-representation of majors based on the LI data. The linear regression coefficient is 0.95 with an R-squared of 0.99.

Panel B of Figure 2 suggests that, when calculating institution-level labor market shares, the overrepresentation of certain fields in the LI data is offset by underrepresentation of other fields. Indeed, after reweighting the PSEO data to reflect the composition of majors in the LI data, the overall percentage of graduates employed in-state falls by about 1 percentage point, from 68.9 to 67.9 percent. Although we cannot rule out bias that may arise at finer levels of geography, such as sorting of students across different metro areas, our validation results imply that this bias, if it exists, is likely to be small when estimating college-specific labor market shares. All told, when benchmarked against reliable governmental data sources, the LI data stand up well to assessments of coverage and validity.

III. Describing College Labor Markets

A. Binary Definitions of College Graduates' Locations

Our college labor market data can be used to address simple questions related to where recent college graduates live and work. For instance, economic developers and state and local officials often want to know how well an institution's own state or closest metropolitan area retains its graduates. State policymakers in particular often have concerns about loss of homegrown, college-educated talent to other states (i.e., brain drain), which motivates policy interest in these measures (Bound et al., 2004). Indeed, many state merit aid programs are explicitly aimed at retaining talented college graduates (Fitzpatrick & Jones, 2016; Nguyen, 2019).

We find that the nearest metro area (i.e., LI geography) and own state capture 50 percent and 67 percent of all graduates, respectively. Figure 3 disaggregates these results by various institutional characteristics, showing the percentage of graduates living in their institution's nearest metro area with blue bars and the percentage living within the institution's home state in pink bars.[16] Even among 2-year institutions, where we expect graduates to cluster locally, nearly 30 percent of graduates live and work outside their home metro area. There is also substantial variation in these two measures across both Census region and Barron's selectivity level. Institutions in the Northeast appear to retain about half of their graduates in the nearest metro area, and 63 percent remain within the home state. In contrast, the percentage of graduates remaining in-state in the West is significantly higher, at 74 percent. In terms of selectivity, we see a near-linear negative relationship with the percentage of students residing in-metro-area or in-state. The more selective an institution, the less likely its graduates live and work nearby. The most selective 4-year institutions retain only about 36 and 43 percent of graduates within metro area and state, respectively, while institutions in the least selective category have 64 percent of their graduates living within the nearest metro area and 78 percent living within the same state.

B. Concentration and Distance of Graduates' Destinations

While informative, the percentage of graduates living and working in-state (or in-metro-area) is a blunt measure that does not capture the full range of a college's labor market. We thus use our data to compute a Herfindahl-Hirschman Index (HHI) to quantify the degree of geographic concentration of an institution's graduates across the United States.

[16] Because some metro areas cross state lines, our calculations assign graduates who reside in an institution's own or nearest metro area to the institution's "in-state" count. We explored several alternative approaches, including apportioning graduates based on population sizes of the states that share a given metro area. However, none of our alternative approaches produced statistics that performed better than the current approach in our validation analyses.

In our context, the HHI equals the sum of the squared percentages of an institution's graduates residing in each of the 278 possible LI geographies. A maximum value of 10,000 implies that 100 percent of an institution's graduates live in one single LI geography. Lower numbers imply that graduates are more dispersed across place. The weighted average of the HHI for the full sample of 2,600 institutions is 4,796.[17] The largest contribution to the HHI comes from graduates living and working in an institution's own or nearest LI geography. However, these graduates account for just half of all graduates in our sample. This suggests that recent college graduates migrate substantially.

Since HHI captures dispersion of alumni locations regardless of distance from the institution, equal shares of graduates across metro areas within a short drive can yield the same value as equal shares of graduates spread throughout the country. We therefore also compute the average distance traveled by employed graduates of each institution, weighted by labor market shares. The weighted average is 198 miles, with the typical graduate of the most selective 4-year institutions traveling about 5.5 times as far as the average community college graduate.

The granularity of the HHI and distance measures permit rich characterization of an institution's geographic labor market. Panel A of Figure 4 shows a clear negative relationship between HHI and selectivity, while Panel B shows a clear positive relationship between average distance traveled and selectivity. The gap in the average HHI between the most selective 4-year institutions and 2-year institutions is 3,746 (or 1.6 standard deviations) and the analogous gap in average miles traveled by alumni is 405 (or 2.5 standard deviations). Less-selective schools tend to have higher concentrations of graduates in fewer areas that are physically nearby, whereas the most-selective schools send their graduates farther and to a greater diversity of locations.

---

[17] We weight by each institution's number of graduates. The unweighted average (median) HHI equals 4,470 (4,486).

Panels C and D of Figure 4 show density plots of the HHI and distance measures, respectively, for three groups of institutions: public 4-year, private 4-year, and 2-year.[18] In Panel B, the HHI distributions for public and private 4-year institutions look quite similar, with the mean for 4-year privates falling a bit below that for 4-year publics. However, in testament to the appreciable variation in colleges' labor markets, we find that 28 percent of community colleges exhibit a geographic dispersion of graduates that surpasses the average public 4-year institution.[19] Yet, only a bit more than 8 percent of community colleges send their typical graduate farther than the average 4-year public institution.

Contrasting institutions in terms of these measures can shed light on the nature of their labor markets. For example, the University of Southern California (USC) and Shasta College (a 2-year institution in California) both have near-median HHI values of 4,400, suggesting that their graduates live and work in a similarly diverse set of locations. However, the average distance traveled by USC alumni is nearly 3.5 times as far as alumni of Shasta College (i.e., 345 miles and 99 miles, respectively), implying that these institutions realize comparable levels of alumni geographic dispersion within markedly different geographic reaches. Another pair of institutions, SUNY Buffalo State and SUNY Cortland, both send their graduates an average of 140 miles from campus. However, the HHI for SUNY Buffalo (4,520) is more than 50 percent larger than the HHI for SUNY Cortland (2,891). Thus, although their graduates travel similar distances, SUNY Buffalo alumni tend to be more concentrated in fewer geographic areas, compared to SUNY Cortland graduates.

[18] Nearly all of the 2-year institutions are public.
[19] The estimates in Figure 2 are weighted by the number of alumni and thus reflect the experiences of the average student (within a given group of institutions). Unweighted results show the same punchlines.

While we do not explore mechanisms for these patterns in this paper, we note that the underlying data provide the groundwork for future study of candidate explanations. For instance, it is an open question how closely linked the destinations of recent graduates are to the original areas from which they entered college. The data presented here could be used with other data sources that capture geographic information on incoming students to assess the strength of this relationship and heterogeneity across institutions.[20]

C. Macro Movements of College Graduates

Beyond quantifying how graduates of individual institutions migrate after college, we can use our data to explore the distribution of graduates across geographic boundaries at a more macro level. Such an exercise is relevant to studies that explore how state-level financial aid policy, like broad-based merit aid, influences the retainment of skilled workers within a state (e.g., Fitzpatrick & Jones, 2016; Sjoquist & Winters, 2014). We approximate the extent to which areas experience net in- or out-flows of graduates by aggregating initial counts of graduates by institution to some higher level of geography, and then compare this figure to the total number of students found residing in that geography after college from *any* institution.

In Figure 5, we depict the net flow of graduates at the state level, calculated for each state s as follows:

$$\frac{Living_s - Graduated_s}{Graduated_s} \cdot 100$$

[20] To illustrate this potential, we plot for each institution a measure of the percentage of first-time, degree-seeking freshmen from in-state (taken from IPEDS) against our institution-level measure of graduates living in-state after college. We present results at both the institution level (Appendix Figure A1) and state level (Appendix Figure A2). Appendix Figure A1 suggests that public institutions, as a group, tend to attract relatively high shares of in-state freshmen but retain less than proportional shares of graduates within their states. We observe the opposite phenomenon for private institutions in our sample, suggesting they may have greater “stickiness,” but also prompting a need for further investigation.

In the spirit of Bound et al. (2004), we define $Graduated_s$ as the sum of LI users who graduated from institutions in state s, and $Living_s$ as the sum of graduates from any institution whose LI location places them in state s in 2021 (i.e., when the data were obtained). Positive numbers indicate that more students now reside in that state than initially graduated there, suggesting more graduates were drawn to (i.e., "imported") or retained in that state than were "exported" to other states. A negative number suggests the opposite—more students left the state compared to the number of students who were drawn there or retained.

In terms of broad patterns, many Western states like California and Washington seem to import many students on net (green colors), most likely due to larger cities with ample job opportunities for college graduates, such as Seattle, the Bay Area, and Los Angeles. On the other side of the country, in the Northeast, large metropolitan areas like New York City and Washington, D.C. are also prominent importers of graduates, likely in part from neighboring states, as suggested by the large net outflow (dark purple) for the adjacent states of Delaware, Virginia, and Connecticut.

D. Who is my peer? Institutional Peers and the Destinations of College Graduates

Many polices that define or seek to influence the market for higher education require considerations about institutions' peers. Colleges themselves choose peers to use for benchmarking purposes and strategic planning (June, 2022; Poser, 2022). For example, when schools report yearly to IPEDS, they have the option to specify the institutions to which they would like to be compared in terms of metrics related to admissions, tuition and fees, graduation and retention rates, and financial measures.[21] Over half the institutions in our sample (1,645 of

[21] Acton et al. (R. K. Acton et al., 2022) use IPEDS data to show that during the early phases of the COVID-19 pandemic an institution's reopening decisions were shaped by the behavior of their self-identified peers.

2,600) provided "custom peer lists" varying in number from as few as 1 (a military institute) to as many as 100, with an average of 19.

The definition of a peer is likely to vary based on the analytic aim of the entity constructing the list. For example, a campus leader may bring quite different considerations to the selection of peers (e.g., competitors for incoming students and faculty) than a state policy leader (e.g., state economic development and "brain drain"). As suggested by several states' announced goals for educational attainment and state workforce development,[22] the concerns of the latter often anchor on geography and the mobility of college graduates. Thus, we explore how closely colleges' self-identified IPEDS peers overlap with peers defined by where their graduates live and work.

We calculate a cosine similarity measure between each institution, *i*, and all possible peers, *p*, in our sample, comparing respective shares of graduates living in each LI geographic area:[23]

$$Similarity_{i,p} = \frac{\sum_{g=1}^{279}(\mathrm{Share}_{i,g} \cdot \mathrm{Share}_{p,g})}{\sqrt{\sum_{g=1}^{279} \mathrm{Share}_{i,g}^2}\sqrt{\sum_{g=1}^{279} \mathrm{Share}_{p,g}^2}}$$

where *g* represents a specific LI geography. Because shares are non-negative, the similarity score is bounded by 0 and 1, where 0 is complete dissimilarity in the vector of shares and 1 is a perfect match, meaning that the pair of institutions had the same proportions of graduates in all LI geographies.

To generate a comparable geography peer list, we sort each institution's list of similarity scores in descending order and retain the same number of peers listed on the custom peer list

[22] Lumina Foundation tracks these state goals: https://www.luminafoundation.org/stronger-nation/report/static/States_with_Higher_Education_Attainment_Goals.pdf.
[23] We include the "unlocated" share in this exercise; hence the total number of LI geographies is 279 (not 278).

from IPEDS.[24] We constrain geography peers by level, where 4-year institutions match only with other 4-year institutions, and the same for 2-year institutions. However, we permit institutions that list at least one peer of each type (i.e., 4-year and 2-year institutions) to have geography peers of either level.

We compare the similarity of the two peer lists (IPEDS and LI geography) for each institution using a basic overlap statistic, the Jaccard Index:

$$J_i = \frac{Peer_{11}}{Peer_{01} + Peer_{10} + Peer_{11}}$$

Here, *Peer* is the count of institutions in each of following three groups, denoted with subscripts: the institution appeared on both lists ($Peer_{11}$), the institution appeared only on the LI list ($Peer_{01}$), or the institution appeared only on the IPEDS list ($Peer_{10}$). This is equivalent to taking the intersection of the two sets of peers divided by their union. The result, which ranges from 0 to 1, can be thought of in percentage terms. A *J* of 0.5, for example, indicates a 50 percent overlap between the two lists.

Figure 6 depicts the results of this exercise, wherein we aggregate *J* by basic Carnegie Classification, weighting each institution equally. The overlap between the lists of IPEDS peers and geography peers is minimal, at about 7 percent on average. The degree of overlap varies by institution type, with research-intensive universities exhibiting the lowest level of overlap, at 3 percent, and 2-year colleges boasting the highest, at nearly 12 percent.

To make these differences concrete, consider two institutions in the state of North Carolina, the flagship University of North Carolina at Chapel Hill (UNC), with a peer list overlap statistic of 3.4 percent, and Alamance Community College (ACC), with an overlap

[24] The complete list of similarity scores for each institution pairing is available upon request.

statistic of 33.3 percent. UNC's list of IPEDS peers spans the entire United States and is composed mainly of other elite public institutions (e.g., University of Michigan) and highly selective private institutions outside of North Carolina (e.g., Johns Hopkins University). The one exception, and consequently the one institution that is on both the IPEDS and LI peer lists, is Duke University. The institutions sending graduates to similar geographic locations as UNC are exclusively in North Carolina: North Carolina State University as well as some smaller private institutions like Elon University, neither of which appears on UNC's IPEDS peer list.[25]

In contrast, ACC exclusively listed North Carolina community colleges on its IPEDS peer list, seven of which were also identified as labor market peers in our analysis. Two community colleges over the Virginia border were identified as having similar labor markets as ACC. Thus, limiting IPEDS peers to North Carolina institutions misses these neighboring competitors. While the Virginia institutions face a different state policy environment and likely different pools of incoming students than ACC, their graduates end up in similar locations.

Lack of attention to the destinations of graduates in the selection of peer institutions could stem from a variety of reasons. Institutional leaders may focus more on potential incoming students and competition in the admissions market or on faculty recruitment, resources related to areas of institutional strength, or a variety of other factors likely to differ by type of institution. It could also be that IPEDS data do not lend themselves well to comparisons of student outputs such as labor market outcomes; hence, the capacity to designate peers based on such constructs is low. Thus, our geography-based measure provides a conceptually useful complement to peer lists based primarily on inputs.

[25] Please see Appendix Table A2 for the detailed lists of peers.

IV. Application 1: Understanding Economic Mobility of College Graduates

In our first illustrative research application, we use our measure of colleges' labor markets to examine the variation in institution-specific economic mobility rates, a relatively new marker of college success introduced by Chetty et al. (2020). The authors use comprehensive federal income tax data on children and their parents to construct intergenerational bottom-to-top income quintile mobility rates for colleges and universities in the United States. These mobility rates constitute a novel, accessible measure of the degree to which colleges promote economic opportunity for low-income students.

Aside from the measure's substantive value, the authors find that basic institutional characteristics—such as control (public/private) and measures of instructional and endowment spending—have limited capacity to predict bottom-to-top-quintile mobility rates (Chetty et al., 2020, pp. 1570). This finding highlights the promise of using data on student outcomes to explore other avenues through which colleges might promote intergenerational economic mobility. We use our data to explore a reasonable hypothesis that builds on this finding: one way a college may generate upward mobility for its students is through the labor market connections it provides. Institutions with connections to robust labor markets may facilitate greater mobility for their graduates.

We test whether the relative strength of the labor markets where an institution sends its college graduates can explain the variation in economic mobility rates across institutions. We measure the "strength" of the labor market of each college by calculating average hourly wages earned by bachelor's degree recipients within LI geographic areas, and then aggregating across areas using the institution-specific labor market shares as weights. More specifically, we calculate this wage measure by aggregating data from yearly waves (2010–2018) of the

American Community Survey (ACS) public-use datasets (from Ruggles et al. 2021) on individuals with a bachelor's degree, not currently enrolled in school, ages 24–35, to capture a population likely to reflect the alumni in our LI data. We aggregate wages to the LI-geography level using the CBSA designations of ACS respondents and our CBSA-LI crosswalk.[26] We then multiply each LI geography's wage rate by the share of graduates living in that area for each institution and sum these values within institution. Since the shares sum to 1, the new wage value is a weighted average unique to each institution and a function of the areas to which its graduates flow.

Our primary measure of intergenerational economic mobility is the proportion of students in low-income families who reach the top quintile of income as adults (P(Child in Q5|Parent in Q1)), or what Chetty et al. (2020) define as the "success rate," from the Opportunity Insights website.[27] Among the 1,913 institutions for which we have both labor market and mobility rate data, the weighted average bottom-to-top quintile "success rate" is 0.18, meaning that 18 percent of students who came from families in the bottom quintile of the income distribution and made it to the top quintile by age 30.[28]

Table 1 presents results from a series of regressions of the log of the success rate on the log BA wages, with progressively richer sets of controls. The baseline specification, with no controls, shows a strong, statistically significant positive relationship between the average wages

---

[26] We map ACS observations to CBSAs (2013 vintage) using the county field in the ACS microdata and a county-to-CBSA crosswalk available from the U.S. Census Bureau (https://www.census.gov/geographies/reference-files/time-series/demo/metro-micro/delineation-files.html). For the approximately 30 percent of observations where county is missing, we assign observations to county using a PUMA-to-county crosswalk from the Missouri Census Data Center's Geocorr 2018 (https://mcdc.missouri.edu/applications/geocorr2018.html).

[27] Source: https://opportunityinsights.org/data/. This task required creation of a crosswalk from IPEDS *unitid* to the 6-digit *opeid*. The Chetty et al. (2020) data further group some institutions into "super opeids" based on tax data and institutional reporting. They provide a crosswalk for this purpose.

[28] Weights equal the number of students in a cohort with parents in the bottom income quintile, from Chetty et al. (2020). The unweighted average is 0.21. We drop three institutions with a success rate of zero from the analysis.

of an institution's labor market and economic mobility. The addition of basic institutional characteristics in column 2 reduces the focal coefficient, although it remains highly statistically significant. Chetty et al. (2020) found that sociodemographic characteristics of colleges' student bodies explained some of the variation in mobility rates across colleges. Thus, in column 3, we add a vector of student characteristics, including shares of undergraduates of different races/ethnicities, the share female, the share over age 25, and median parental income. The estimated elasticity drops a bit further to 1.05 but remains statistically significant ($p < 0.01$). This suggests that the strength of the average labor market to which a college sends its graduates meaningfully predicts rates of intergenerational economic mobility across institutions, over and above key observable characteristics of colleges and the students they educate.

In the fourth and final column of Table 1, we add fixed effects for the institution's (nearest) LI Geography, constraining comparisons to institutions located within or near the same metropolitan areas. The elasticity remains statistically significant and positively related to mobility rates ($p < 0.01$). In this most stringent specification, a 10 percent increase in the average bachelor's degree wage of the relevant markets for an institution is associated with a 14.2 percent increase in the "success rate," or roughly a 2.6 percentage point increase relative to the sample mean of 0.18. The magnitude of this relationship is about three times as a large as a 10 percent increase in median parental income of the student body, which has an elasticity of 0.46.[29] Hence, our results suggest that institutions are likely to create more economic mobility when they have stronger links to robust labor market networks than other nearby institutions with similar access to the local market.

[29] Appendix Table A3 presents results for the full models. Unweighted regressions produce qualitatively similar conclusions.

These results are an important first step toward better understanding links between geographic and economic mobility of low-income students in American higher education. While some institutions may do well by their low-income students by fostering close ties to their local labor markets, it appears students also gain from institutional networks outside the most proximate labor market. Of course, disentangling student preferences for different locations, demand for individual majors and skills in different areas, and the institution's contribution to ultimate student outcomes requires micro-level data. Our analyses highlight the potential merit of such explorations by establishing a relationship between economic mobility and a novel measure of the strength of the labor markets in which graduates of institutions live and work.

V. Application 2: Brain Drain and the Value of Public Investment in Higher Education

In our second application, we use our data to show how migration affects the geographic incidence of public investment in higher education. In 2019, state governments collected more than $87 billion in tax revenue to fund public colleges and produce a college-educated workforce (Laderman & Heckert, 2021). Private colleges receive government funding as well, through programs like the federal Pell Grant, although on a smaller scale. Such revenue generally supports instructional expenditures by offsetting tuition costs (Webber, 2017), providing a transfer to college students. Increases in these transfers boosts degree attainment (Bound et al., 2010; Bound & Turner, 2007; Deming & Walters, 2017), benefiting both graduates and society by creating a social return in excess of the private benefits through salutary effects on crime, health, voting, and the labor market (Moretti, 2004).

However, the social rate of return on investment for a state depends on the spatial distribution of graduates across the country. States and other smaller localities may not receive the full value of their investment due to "brain drain." For instance, Bound et al. (2004) find that the number of baccalaureate degrees produced within state boundaries has little relationship to the stock of college-educated labor living and working in a state. High mobility rates decrease the value of public investment by states, potentially leading to an under-provision of the services as other states "free-ride." While federal investment is more agnostic to these issues, the mobility of graduates has equity implications for the geographic distribution of funds. In both state and federal investment, mobility determines which taxpayers subsidize and benefit from government-financed human capital production.

We use the number of 4-year college graduates working in an area divided by the total amount of public (state, local, and federal) funding transferred to colleges in that area as a proxy for the local rate of return of public investment. This return varies by the level of tax revenue spent per student, the rate at which students graduate, and net migration. For instance, the return may be low because of tax revenue spent on students who do not graduate or because an area has high levels of outmigration. Similarly, the measure may be high because an area retains a high number of graduates or because it imports a large number from other areas. We also compare this measure to the number of 4-year college graduates produced per dollar of government funding in each area to understand how migration after college affects the spatial distribution of benefits.[30]

We calculate total government expenditure using information on both the amount each college spends per student from government sources each year as well as the number of years students enroll. Using first-time full-time bachelor's degree seeking cohorts who began school in 2009 or 2010 from IPEDS, we estimate the total government dollars spent on the cohort in each

[30] Previous work on the incidence of public investment in higher education has implicitly ignored the social returns, and instead focused on private benefits, or subsidy, to those who enroll (e.g., Johnson, 2006). Bound and Simon (2021) also consider the effects on aggregate human capital, but not across different labor markets, as we do here.

year through 2018 as a function of the number of students who enroll initially along with retention, graduation, and drop-out rates.[31] Finally, we sum the dollar values across institutions located in each LI geography to produce our measure of graduates per total government expenditure for each area.

The maps in Figure 7 depict two measures of graduates per total government expenditure at the local level. Panel A shows the number of graduates produced by institutions in each area per \$100,000 of total government spending at those institutions, which is akin to what the return on taxpayer expenditures would look like without post-graduation migration. Panel B shows the number of 4-year graduates retained or received from other geographies per increment of government spending at colleges in each area. The measures in Panels A and B have the same denominator but differ in how they handle graduates in the numerator. In Panel A, graduates are assigned to the location of their college; in Panel B, they move to the labor market where they work after graduating based on the LI data. Comparing the two panels, mobility more than doubles the variance—and inequality—across areas in college-educated labor per dollar spent from 0.08 to 0.17.

Though both approaches result in considerable variability across areas, the specific areas identified as high versus low return differ across panels. To characterize these patterns, we estimate the extent to which labor market and institutional characteristics predict our two measures of the local rate of return. Panel A of Table 2 presents coefficients from a regression of graduates produced per \$100,000 in government funding on local area college graduates'

[31] We pool the 2009 and 2010 full-time, first-time (FTFT) freshman cohorts from IPEDS for approximately 1,490 4-year institutions, then estimate the number of cohort members enrolled in a given year between 2010 and 2018 using 1-year retention rates and 4-, 6-, and 8-year graduation rates. We then multiply the total federal, state, and local appropriations and grants per FTE (of all students) by the number of FTFT cohort members still enrolled each year and sum across years to get the total state spending for the cohort.

earnings, urbanicity, institutional control and selectivity. Including Census Division fixed effects nets out potential regional differences that may be due to geographic factors, although this has a minimal effect on our estimates.[32]

There is no difference between high- versus low-wage areas or urban/suburban versus rural areas in the number of graduates produced per dollar. Since this outcome captures transfers to students while they are still in college, it is perhaps unsurprising that labor market features—such as wages and urbanicity—do not predict that outcome. Also unsurprising, areas with larger shares of private, and especially selective, institutions have a greater number of graduates produced per dollar of government funding (Panel A). Such institutions tend to have higher graduation rates on average, driving up the numerator, and rely more on tuition, rather than public funding, lowering the denominator.[33]

Panel B of Table 2 examines the social return inclusive of migration (as captured in Panel B of Figure 7). Now the labor market features matter substantially. That is, areas with high wages for college-educated workers and urban areas have high local returns because they attract more graduates regardless of where the students completed college. Finally, the dependent variable in Table 2, Panel C is the difference between graduates produced and graduates retained or received (Figure 7 Panel A minus Panel B, respectively). The implication is that the social benefits of public investment in postsecondary education disproportionately accrue to high-wage and urban/suburban areas due to graduate mobility, which we now can quantify precisely.

[32] For example, the distance between major cities in New England is much smaller than in the Pacific Census Division, which likely influences migration patterns.

[33] When we estimate the social returns separately for state and federal funding, we find that the large positive coefficient on share of private institutions is driven by state funding, consistent with this finding (Column 1, Appendix Table A4).

A related question is how many graduates a state retains in-state for each public dollar spent at different types of institutions. This might be the relevant question of concern to a state higher education agency allocating funding across different institutions. State flagships, despite having much higher graduation rates, have greater out-of-state migration and higher spending, lowering the number of graduates retained in-state per state dollar spent.

Figure 8 presents the distribution of graduates retained per $100,000 in state funding for public 4-year institutions across the United States, including separate statistics for two subgroups of institutions.[34] There is substantial variation. Institutions with modest spending, high graduation rates, and low migration—such as selective regional public universities—rank highest. Flagships rank among the lowest. An important caveat is that this metric does not reflect benefits accruing from institutions' ability to attract graduates from out-of-state nor other local economic spillovers (e.g., Andrews, n.d.; Valero & Van Reenen, 2019).

VI. Conclusion

Research on the role of colleges in economic growth and mobility, workforce and skill evolution, and the social return to public investment in education—to name a few—has been hampered by an absence of data on where students from individual institutions actually go after graduating. Prior work has typically assumed—explicitly or implicitly—that graduates remain nearby or has relied on small samples of students or institutions. This paper introduces new data on the metro-level labor markets served by individual institutions drawn from alumni data contained on LinkedIn institutional pages. These data, validated with several official government sources, are available for almost every college and university in the United States.

[34] These statistics are unweighted.

We use these data to reveal several new descriptive facts about the mobility of college graduates. Recent college graduates tend to be quite mobile, traveling an average of nearly 200 miles, though there is wide variation across and within institutions' level, selectivity, and control. For instance, 28 percent of 2-year institutions have a geographic spread of alumni that surpasses that of the typical public 4-year institution. Clearly, assuming that graduates remain in-state or that in-state retention rates are similar across colleges is erroneous. Relatedly, flows of graduates across state or metro-area boundaries are far from uniform, as some areas tend to be strong net-importers of recent college graduates, where others are net-exporters of college-educated workers. These patterns are likely related to state and federal policies and could be explored in more depth in future work. Finally, colleges rarely consider the labor markets entered by their graduates when defining "peer institutions," and instead focus on their main competitors for incoming students. While appropriate for some analyses, the markets for incoming students and for outgoing graduates are quite distinct.

Two empirical applications showcase the research potential of our novel dataset, though many other uses are possible. First, we use these data to examine the relationship between the strength of the labor markets to which an institution's graduates flow and the economic mobility of low-income students at that institution. We find that college-specific, bottom-to-top quintile income mobility rates, a benchmark of institutional success, are statistically linked to the geographic destinations of colleges' graduates. This suggests that one of the key ways institutions propel upward economic mobility for their low-income students is through connections to robust labor markets across the country.

Our second application characterizes the spatial distribution of both graduates and public spending to illustrate the importance of accounting for migration when measuring the return on public investment in higher education. Federal and state transfers to college students create both direct benefits in a college's own area as well as spillovers from migration. Focusing only on where the money starts, rather than how it follows students after graduation, misrepresents who benefits from public investment. Specifically, higher-wage and urban or suburban areas tend to have more college graduates per public dollar spent due to inequitable migration of graduates. From the lens of a state policymaker, we show that regional public institutions tend to produce the greatest number of graduates who stay and work in-state per dollar of state funding.

These are a few of many uses of new data on the destinations of graduates from individual colleges in the United States. Postsecondary institutions play an important role in labor markets, constitute a target of substantial public investment, and are crucial cultural and political institutions. Simply knowing where an institution's graduates end up living after graduation is a hurdle many analysts struggle to surmount. Our novel measure helps to fill this analytic gap.

References

Acton, R. (2020). Community College Program Choices in the Wake of Local Job Losses. *Journal of Labor Economics*, 712555. https://doi.org/10.1086/712555

Acton, R. K., Cook, E. E., & Luedtke, A. (2022). The influence of peer institutions on colleges' decisions: Evidence from fall 2020 reopening plans. *Journal of Economic Behavior & Organization*, *195*, 288–302. https://doi.org/10.1016/j.jebo.2022.01.016

Agrawal, D. R., Hoyt, W. H., & Ly, T. (2022). *The Marginal Value of Public Funds in a Federation* (SSRN Scholarly Paper No. 4009872). Social Science Research Network. https://papers.ssrn.com/abstract=4009872

Andrews, M. J. (n.d.). How Do Institutions of Higher Education Affect Local Invention? Evidence from the Establishment of U.S. Colleges. *American Economic Journal: Economic Policy*. https://doi.org/10.1257/pol.20200320

Blom, E., Cadena, B., & Keys, B. J. (2020). Investment over the Business Cycle: Insights from College Major Choice. *Journal of Labor Economics*, 712611. https://doi.org/10.1086/712611

Bound, J., Braga, B., Khanna, G., & Turner, S. (2019). Public Universities: The Supply Side of Building a Skilled Workforce. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, *5*(5), 43–71.

Bound, J., Groen, J., Kézdi, G., & Turner, S. (2004). Trade in university training: Cross-state variation in the production and stock of college-educated labor. *Journal of Econometrics*, *121*(1–2), 143–173. https://doi.org/10.1016/j.jeconom.2003.10.012

Bound, J., Lovenheim, M. F., & Turner, S. (2010). Why Have College Completion Rates Declined? An Analysis of Changing Student Preparation and Collegiate Resources. *American Economic Journal: Applied Economics*, *2*(3), 129–157. https://doi.org/10.1257/app.2.3.129

Bound, J., & Simon, A. (2021). *College Choice, Private Options, and The Incidence of Public Investment in Higher Education* (Working Paper No. 29353; Working Paper Series). National Bureau of Economic Research. https://doi.org/10.3386/w29353

Bound, J., & Turner, S. (2007). Cohort crowding: How resources affect collegiate attainment. *Journal of Public Economics*, *91*(5), 877–899. https://doi.org/10.1016/j.jpubeco.2006.07.006

Chetty, R., Friedman, J. N., Saez, E., Turner, N., & Yagan, D. (2020). Income Segregation and Intergenerational Mobility Across Colleges in the United States. *The Quarterly Journal of Economics*, *135*(3), 1567–1633. https://doi.org/10.1093/qje/qjaa005

Conzelmann, Johnathan, Hemelt, Steven, Hershbein, Brad, Martin, Shawn, Simon, Andrew, and Stange, Kevin. (2022). Grads on the Go: Measuring College-Specific Labor Markets for Graduates [US, 2010-2018]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2022-05-18. https://doi.org/10.3886/E170381V3

Deming, D., & Walters, C. (2017). *The Impact of Price Caps and Spending Cuts on U.S. Postsecondary Attainment* (No. w23736; p. w23736). National Bureau of Economic Research. https://doi.org/10.3386/w23736

Fitzpatrick, M. D., & Jones, D. (2016). Post-baccalaureate migration and merit-based scholarships. *Economics of Education Review*, *54*, 155–172. https://doi.org/10.1016/j.econedurev.2016.07.002

Fuller, A., & O'Leary, B. (2012, September 10). Who Does Your College Think Its Peers Are? *Chronicle of Higher Education*. https://www.chronicle.com/article/who-does-your-college-think-its-peers-are/

Guillory, J., & Hancock, J. T. (2012). The Effect of Linkedin on Deception in Resumes. *Cyberpsychology, Behavior, and Social Networking*, *15*(3), 135–140.

Hendren, N., & Sprung-Keyser, B. (2020). A Unified Welfare Analysis of Government Policies*. *The Quarterly Journal of Economics*, *135*(3), 1209–1318. https://doi.org/10.1093/qje/qjaa006

Hoxby, C. M., & Turner, S. (2019). *Measuring Opportunity in U.S. Higher Education* (No. w25479). National Bureau of Economic Research. https://doi.org/10.3386/w25479

Hurst, W. R., Ricks, M. D., & Simon, A. (2022). *Public Good Misperceptions and Support: Evidence from Higher Education Appropriations* [Working Paper]. https://static1.squarespace.com/static/5efccf8ce24dd229b522f785/t/625d759878777267891dc0c4/1650292120945/Hurst_Ricks_Simon_4_22.pdf

Johnson, W. R. (2006). Are Public Subsidies to Higher Education Regressive? *Education Finance and Policy*, *1*(3), 288–315. https://doi.org/10.1162/edfp.2006.1.3.288

June, A. W. (2022, March 24). *'An Art and a Science': Colleges' Tricky Task of Selecting Peers*. The Chronicle of Higher Education. https://www.chronicle.com/article/an-art-and-a-science-colleges-tricky-task-of-selecting-peers

Kelchen, R., & Webber, D. A. (2018). Examining the Interstate Mobility of Recent College Graduates. *Educational Researcher*, *47*(3), 213–215. https://doi.org/10.3102/0013189X17753124

Laderman, S., & Heckert, K. (2021). SHEF: State Higher Education Finance, FY 2020. In *State Higher Education Executive Officers*. State Higher Education Executive Officers. https://eric.ed.gov/?id=ED614994

Molloy, R., Smith, C. L., & Wozniak, A. (2011). Internal Migration in the United States. *Journal of Economic Perspectives*, *25*(3), 173–196. https://doi.org/10.1257/jep.25.3.173

Monarrez, T., & Washington, K. (2020). Racial and Ethnic Representation in Postsecondary Education. Research Report. In *Urban Institute*. Urban Institute. https://eric.ed.gov/?id=ED606027

Moretti, E. (2004). Human Capital Externalities in Cities. In J. V. Henderson & J.-F. Thisse (Eds.), *Handbook of Regional and Urban Economics* (Vol. 4, pp. 2243–2291). Elsevier. https://doi.org/10.1016/S1574-0080(04)80008-7

Poser, S. (2022, April 18). *How a College Decides Who Its Peers Are*. The Chronicle of Higher Education. https://www.chronicle.com/article/how-a-college-decides-who-its-peers-are

Sá, C., Florax, R. J. G. M., & Rietveld, P. (2004). Determinants of the Regional Demand for Higher Education in The Netherlands: A Gravity Model Approach. *Regional Studies*, *38*(4), 375–392. https://doi.org/10.1080/03434002000213905

Sjoquist, D. L., & Winters, J. V. (2014). Merit aid and post-college retention in the state. *Journal of Urban Economics*, *80*, 39–50. https://doi.org/10.1016/j.jue.2013.10.003

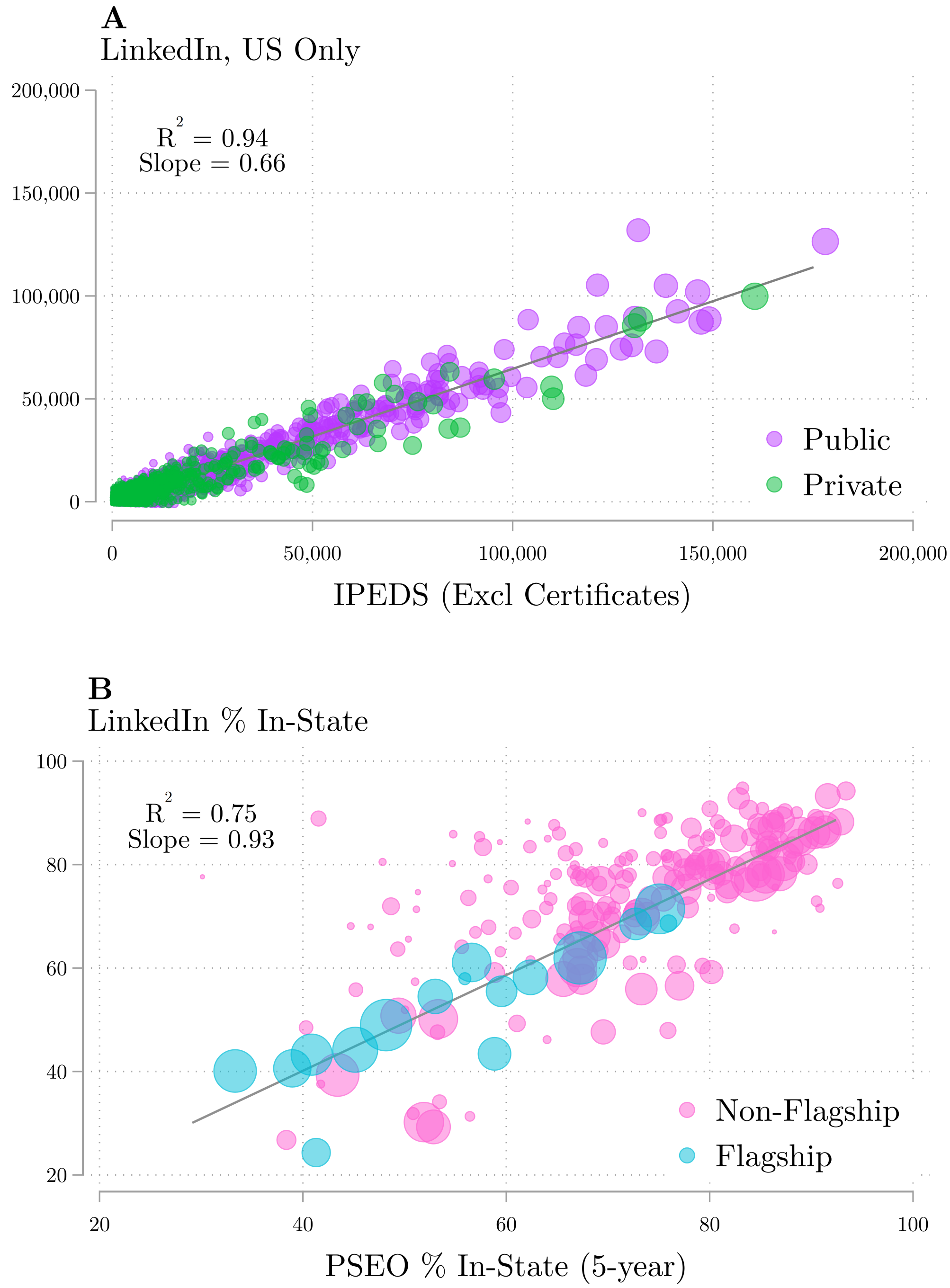
Valero, A., & Van Reenen, J. (2019). The economic impact of universities: Evidence from across the globe. *Economics of Education Review*, *68*, 53–67. https://doi.org/10.1016/j.econedurev.2018.09.001

Webber, D. A. (2017). State divestment and tuition at public institutions. *Economics of Education Review*, *60*, 1–4. https://doi.org/10.1016/j.econedurev.2017.07.007

Weinstein, R. (2020). Local Labor Markets and Human Capital Investments. *Journal of Human Resources*, 1119-10566R2. https://doi.org/10.3368/jhr.58.1.1119-10566R2

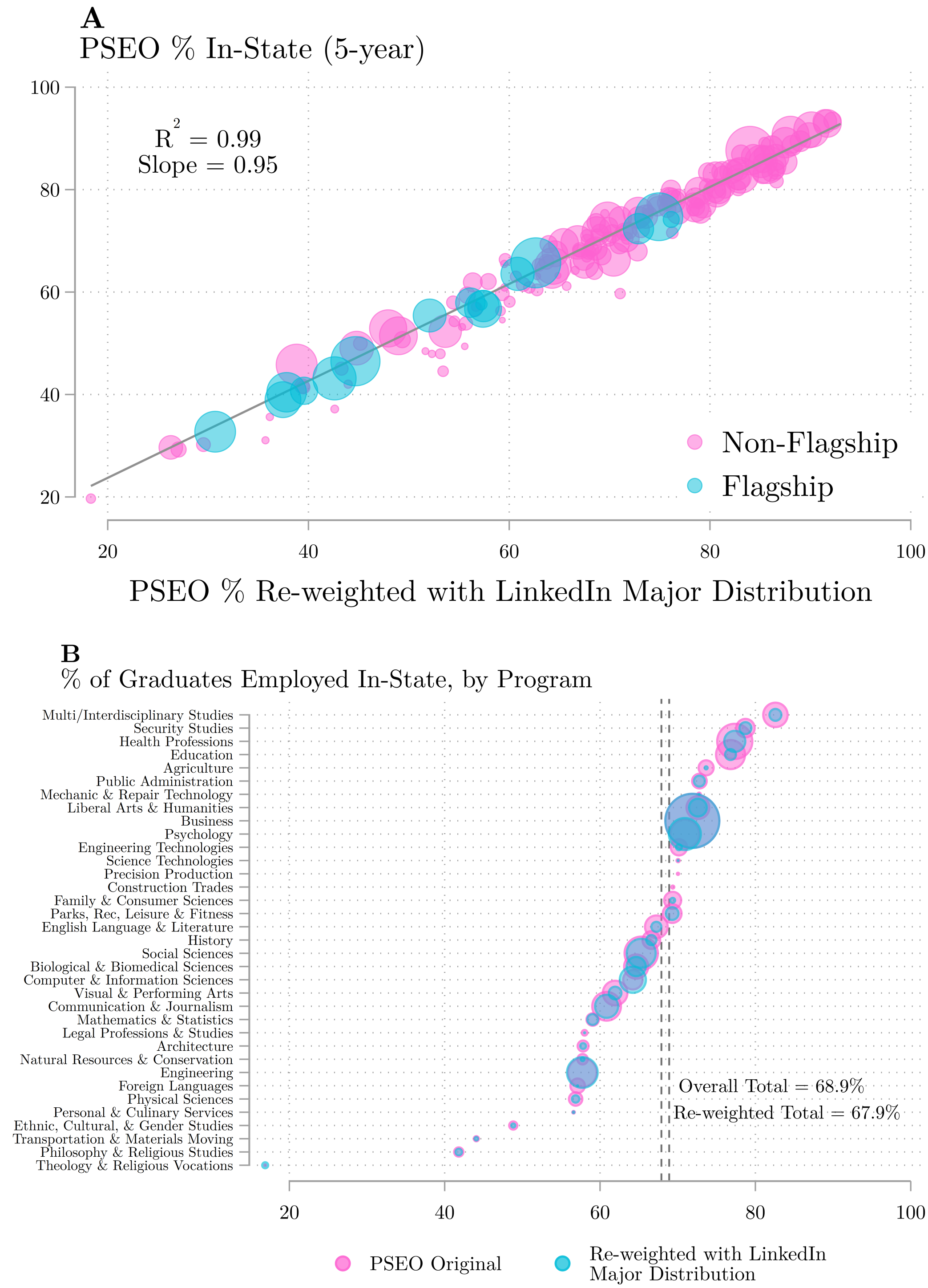
Winters, J. V. (2020). In-State College Enrollment and Later Life Location Decisions. *Journal of Human Resources*, *55*(4), 1400–1426. https://doi.org/10.3368/jhr.55.4.0916-8255R2

Figure 1. LinkedIn Validation Exercises: Comparisons to IPEDS Completion Counts and PSEO Percent of Students Residing Within Institution's State



Notes: Panel A compares IPEDS degree counts (excluding certificates) awarded between 2010 and 2018 to counts of alumni found at the same institution in LinkedIn targeting the same time frame. The R-squared and slope are from a linear regression of the LinkedIn count on the IPEDS count, weighted by institutional enrollment. Panel B compares data from PSEO institutions on the percent of their graduates residing in state 5 years after graduation to the percent of LinkedIn alumni residing in state. The R-squared and slope are from a regression analogous to that in Panel A.

Figure 2. Limited Influence of Compositional Differences in Distribution of Majors in LinkedIn and PSEO on Measures of College-Specific Labor Markets

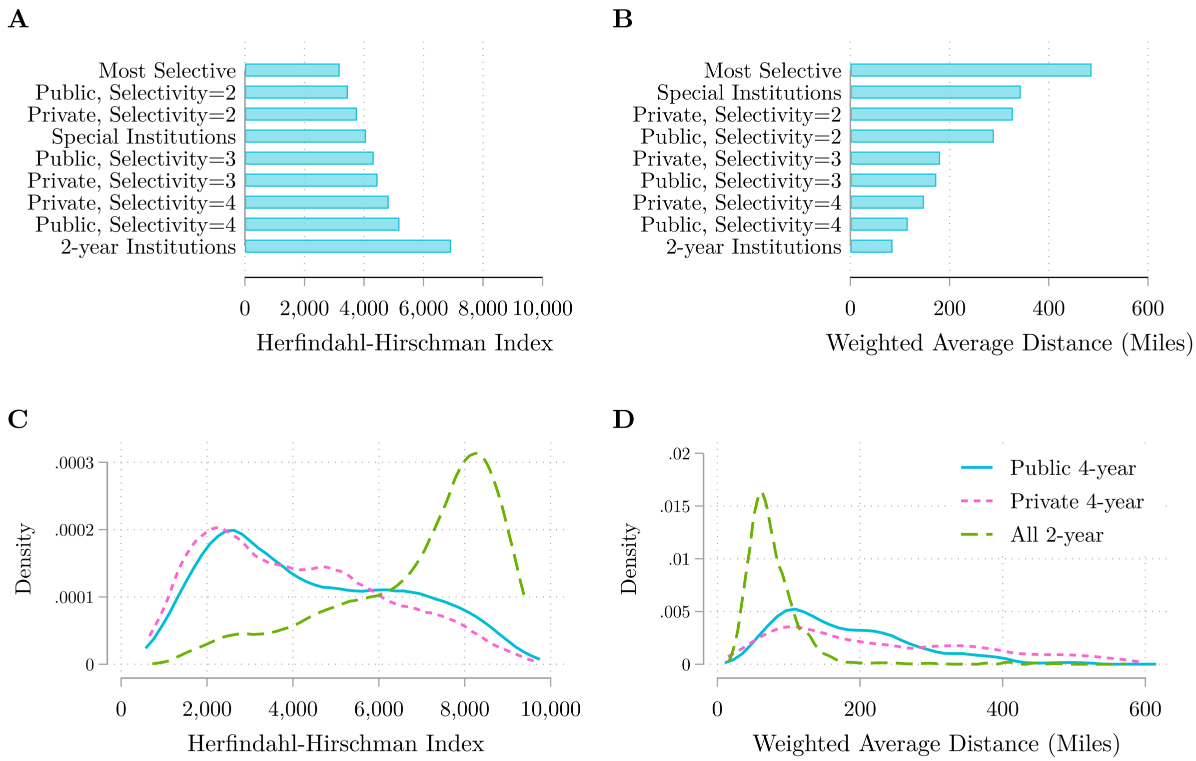


Notes: Institutions represented in this figure are 4-year public colleges in states with PSEO data (see main text for list) that also appear in our main analytic sample. Dots in Panel A are weighted by the total number of graduates across programs at a given institution found in PSEO data. Dots in Panel B are weighted by the total number of degrees awarded in the corresponding 2-digit CIP code in PSEO, overlaid with the re-weighted number that reflects the major distribution reported in LinkedIn at the same set of institutions.

Figure 3. Percent of Institution's Graduates Residing in Same or Nearest Metro Area and Same State, by Institution Type, Census Region, and Barron's Selectivity Index

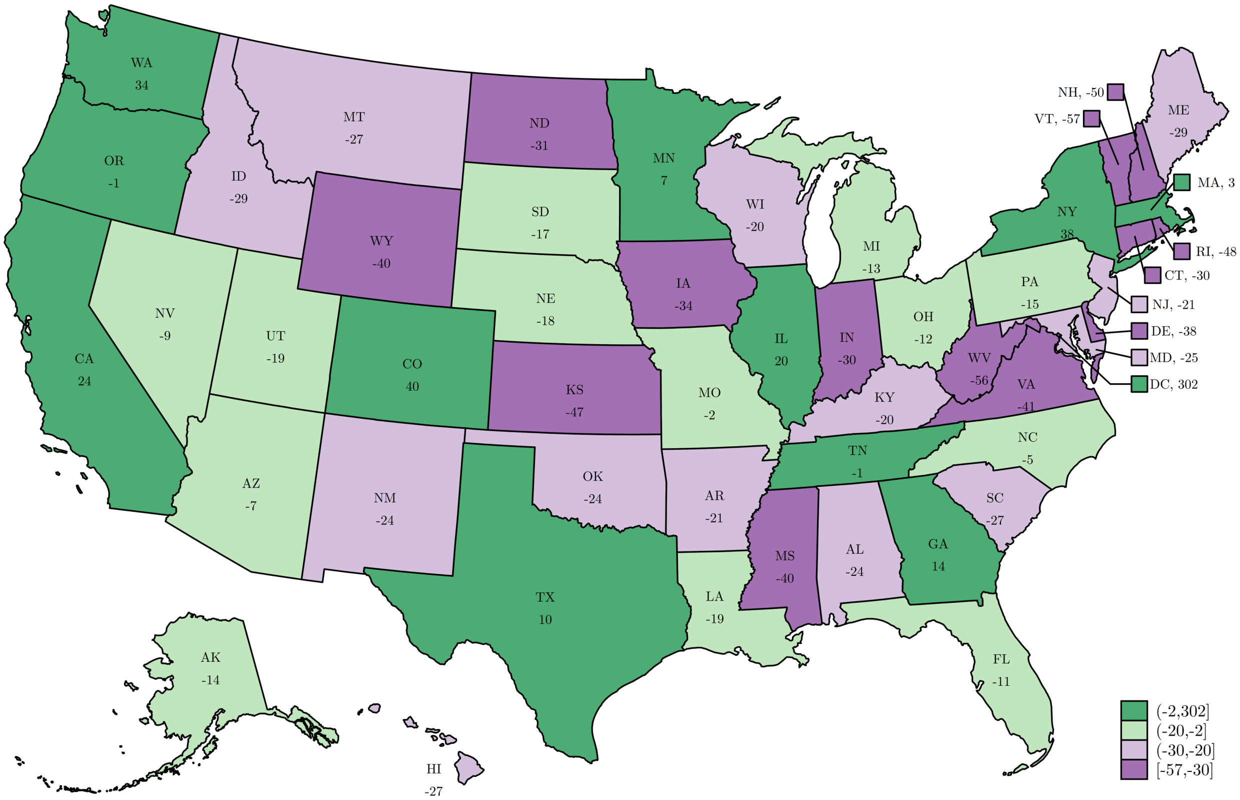
Notes: The Barron's Selectivity panel reflects 4-year institution counts only (i.e., excluding 2-year counts). The top panels include the full sample of institutions. "Less" includes all categories below "Very" except for "Special." Institutions not located in a metro-area are assigned the nearest one based on driving distance to the metro area's geographic center. Average proportions are implicitly weighted by the number of graduates.

Figure 4. Herfindahl-Hirschman Index and Weighted Average Distance Traveled of Alumni, by Institution Type



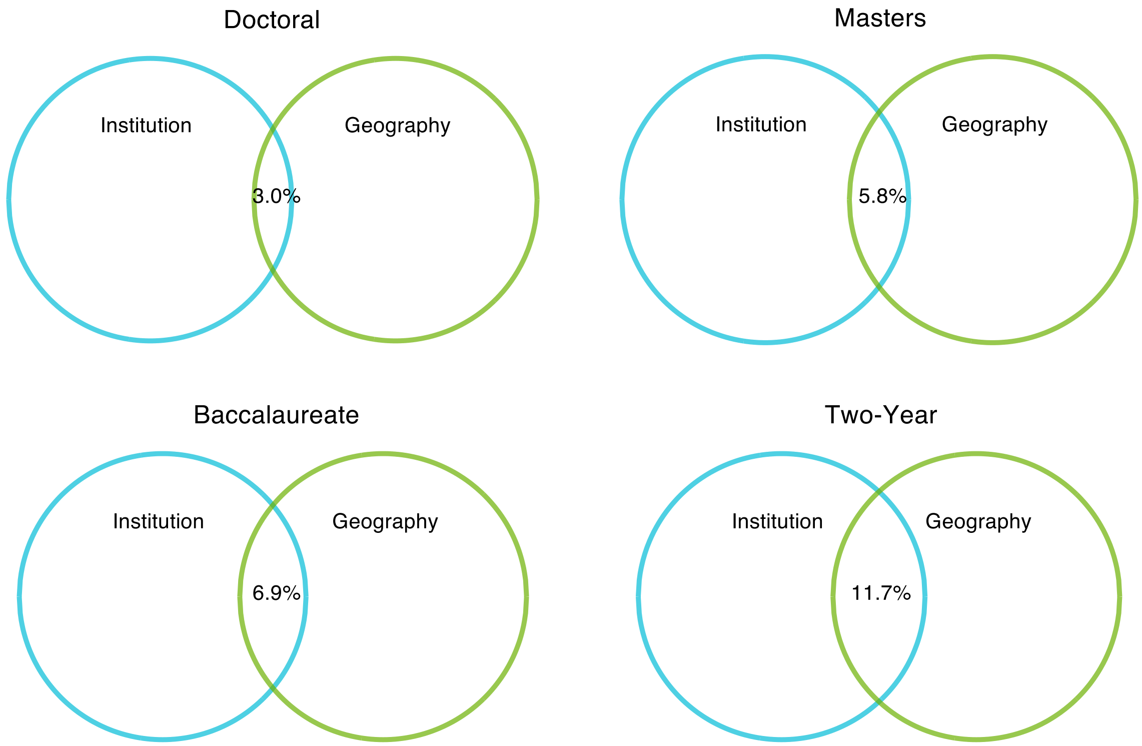
Notes: Selectivity categories are based on Barron's competitiveness index; 1=Most Competitive, 2=Highly Competitive, 3=Very Competitive and Competitive, 4=Less and Non-competitive. The Herfindahl-Hirschman Index (HHI) is calculated for each institution by summing the squared shares of graduates (multiplied by 100) living in each LinkedIn (LI) geography. A maximum of 10,000 implies 100 percent of the institution's graduates reside in one area, whereas lower numbers imply greater dispersion across the country. The average distance for each institution is calculated by taking the crow-flies distance from each institution to the geographic center of each main LI geography multiplied by the share of graduates residing in that geography, then summed within institution. The HHI and average distance measure for each group is a weighted average, where each institution's value is weighted by the number of alumni.

Figure 5. Net Import and Export of 4-Year College Graduates Across States



Notes: Underlying data are restricted to 4-year institution counts. Numbers are rounded to the nearest percentage point, calculated with the following formula, $\frac{Living_s - Graduated_s}{Graduated_s} \times 100$, where *Graduated* refers to the total number of graduates observed in LinkedIn from institutions in state *s,* and *Living* refers to the total number of graduates from any institution who were living within state *s* in 2021 per LinkedIn. Green shades imply higher levels of importing (or retainment of own-state graduates) and purple shades imply higher levels of exporting.

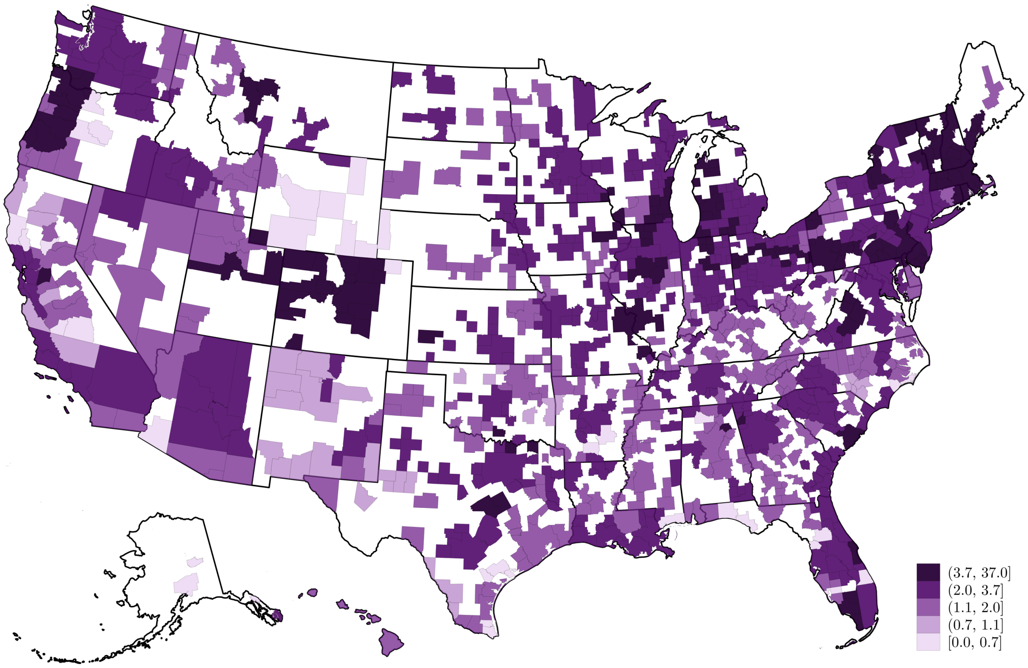
Figure 6. Overlap Statistics for Institution-Identified Peers and College Market Geography Peers, by Basic Carnegie Classification Type



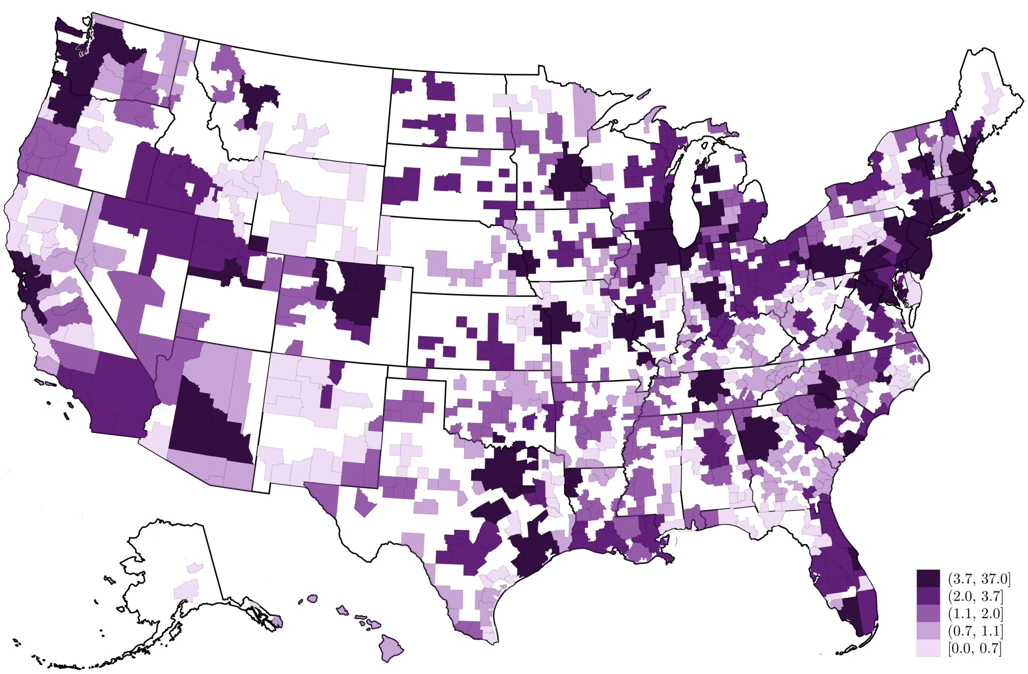
Note: Sample includes only institutions that supplied "Custom" peer lists for IPEDS Data Feedback Report purposes (N=1,645). "Special focus" 4-year institutions (e.g., Arts or Music Academies) are not included in any of the panels above. College Market Geography peers were identified using cosine similarity measures calculated for each pair of institutions in the sample based on the share of their LinkedIn graduates residing in each of the 279 LinkedIn geographies. After ranking based on the highest similarity, we took the same number of peers based on geography as listed by the institution on their custom list and compared the two lists by calculating the Jaccard Index (overlap).

Figure 7. The Mobility of 4-Year College Graduates and Public Investment in Higher Education

A. Graduates Produced per \$100,000 Invested

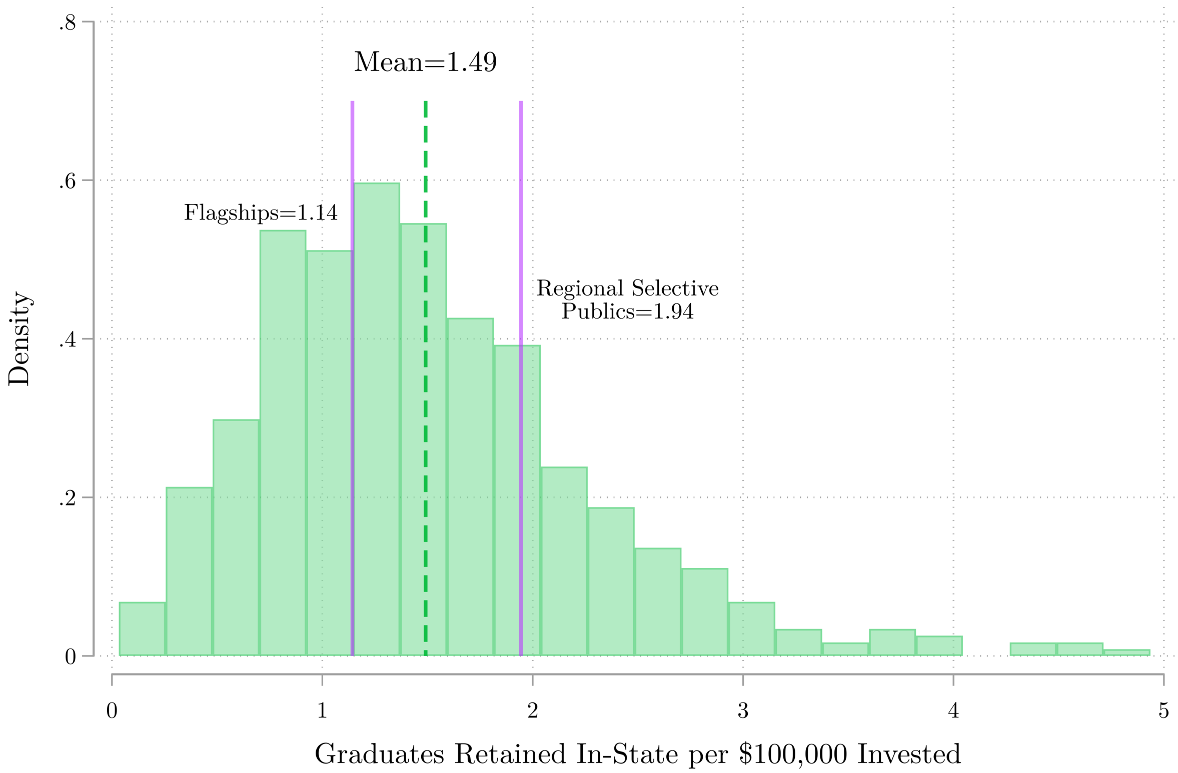


B. Graduates Retained or Received per \$100,000 Invested



Notes: Panel A presents the number of 4-year college graduates produced by all colleges in a given LI geography divided by the total amount of state, local, and federal government spending (i.e., appropriations and grants) by those colleges. Panel B contains a similar measure but the numerator is the total number of 4-year college graduates that move to or stay in a given LI geography after graduation, estimated with our college-specific labor market shares. Spending and college graduation data are from IPEDS and migration rates are based on the LI data. Student counts from IPEDS are based on 0pooled 2009 and 2010 first-time full-time bachelor's degree seeking cohorts.

Figure 8. Public 4-Year Graduates Retained In-State Per $100,000 State Expenditures



Notes: The analytic sample includes public 4-year institutions in the U.S. that appear in our broader sample of institutions. Flagship public institutions are the most selective, research-intensive institutions in each state. Regional selective publics are doctoral and master's institutions within the top three selectivity bins of Barron's data that fall outside the "very high research activity" (R1) Carnegie classification. State expenditures include state appropriations and state grants from IPEDS.

Table 1. Economic Mobility and College Labor Markets

	Outcome = Log(P(Child in Q5|Parent in Q1))			
Independent variable	(1)	(2)	(3)	(4)
Log BA degree wages	2.5566***	1.5876***	1.0535***	1.4213***
	(0.2271)	(0.1510)	(0.1519)	(0.3445)
School Characteristics	N	Y	Y	Y
Student Characteristics	N	N	Y	Y
LI Geography Fixed Effects	N	N	N	Y
Adjusted R-squared	0.289	0.696	0.793	0.830
Observations	1,913	1,913	1,913	1,913

Note: Observations are weighted by the number of students in a cohort with parents in the bottom income quintile, from Chetty et al. (2020). Wages for each institution's individualized labor market were calculated using the average hourly wages for bachelor's degree recipients in each CBSA from pooled 2010-2018 ACS estimates and aggregated to the LI-geography level. These averages were then multiplied by the share of an institution's graduates residing in each area and then summed within institution. School characteristics include control, level, Barron's selectivity, urbanicity, HBCU designation, log of net price, and log instructional expenditures per FTE. Student body characteristics include share of students who are White, Black, Hispanic, Asian, female, above age 25, and log median parent's income. Full regression results are available in Appendix Table A2. Standard errors in parentheses are clustered at the institution's LI-geography level. LI = LinkedIn. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

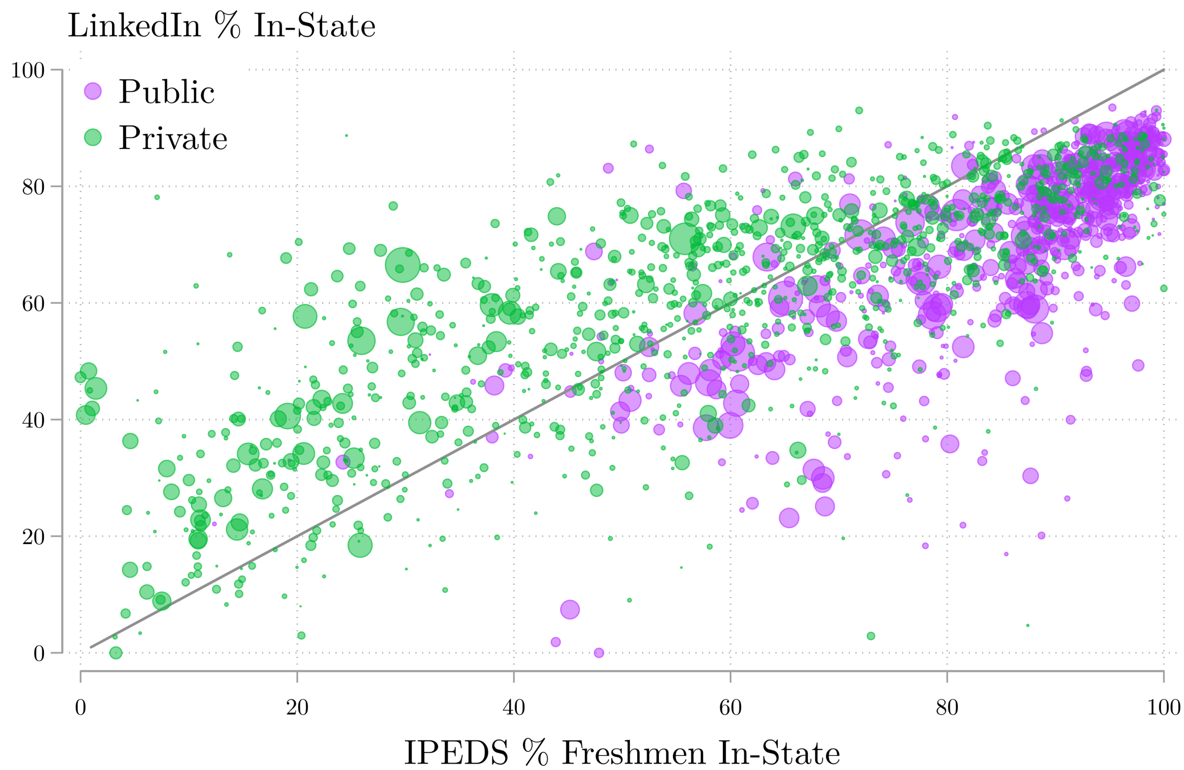
Table 2. Variation in the Social Return to Public Investment in Higher Education

	Panel A: Graduates Produced per $100K		Panel B: Graduates Retained or Received per $100K		Panel C: Net Graduates Retained or Received per $100K	
Independent variable	(1)	(2)	(3)	(4)	(5)	(6)
Log hourly BA degree wage	-0.0059	0.0208	2.3883***	2.6520***	2.3942***	2.6312***
	(0.2696)	(0.2792)	(0.4870)	(0.5122)	(0.4023)	(0.4530)
Share enrollment in towns or rural areas	0.0242	-0.0353	-0.5300***	-0.5603***	-0.5541***	-0.5251***
	(0.1020)	(0.0953)	(0.1646)	(0.1549)	(0.1330)	(0.1333)
Share enrollment in private institutions	1.1789***	1.0172***	1.9773***	1.8894***	0.7984***	0.8722***
	(0.1475)	(0.1330)	(0.2127)	(0.2261)	(0.1875)	(0.2124)
Share enrollment in selective institutions	0.1952*	0.1355	-0.1781	-0.2213	-0.3733***	-0.3568***
	(0.1059)	(0.1016)	(0.1551)	(0.1554)	(0.1397)	(0.1363)
(Share enrollment in private institutions)*(Share enrollment in selective institutions)	0.6550**	0.6409**	0.0519	0.1185	-0.6032*	-0.5224
	(0.2675)	(0.2568)	(0.4234)	(0.4132)	(0.3531)	(0.3594)
Constant	0.5244	0.5866	-6.6956***	-7.4007***	-7.2200***	-7.9874***
	(0.8021)	(0.8266)	(1.4144)	(1.4815)	(1.1691)	(1.2993)
Adjusted R-squared	0.473	0.537	0.479	0.505	0.328	0.351
Census Division Fixed Effects	No	Yes	No	Yes	No	Yes
Observations	263	263	263	263	263	263

Notes: Robust standard errors appear in parentheses. The outcome for Panel A is the number of 4-year college graduates produced at institutions in a given LI geography divided by the total amount of state, local, and federal government spending by those colleges. The outcome for Panel B contains a similar measure but the numerator instead estimates the total number of 4-year college graduates in a given LI geography's labor force after graduation based on the number of graduates produced and our college-specific labor market shares. LI geographies without any 4-year college enrollment are excluded from these regressions, since our measures are not defined for those areas. Average wages are from pooled ACS 2009-2018 data and estimated using employed bachelor's degree recipients ages 23-32. Selectivity is defined as the share of the LI geography's enrollment from institutions in categories 1-3 of Barron's Competitiveness Index. Student counts from IPEDS are based on pooled 2009 and 2010 first-time full-time bachelor's degree seeking cohorts. All regressions additionally control for the log of FTE enrollment in each area.

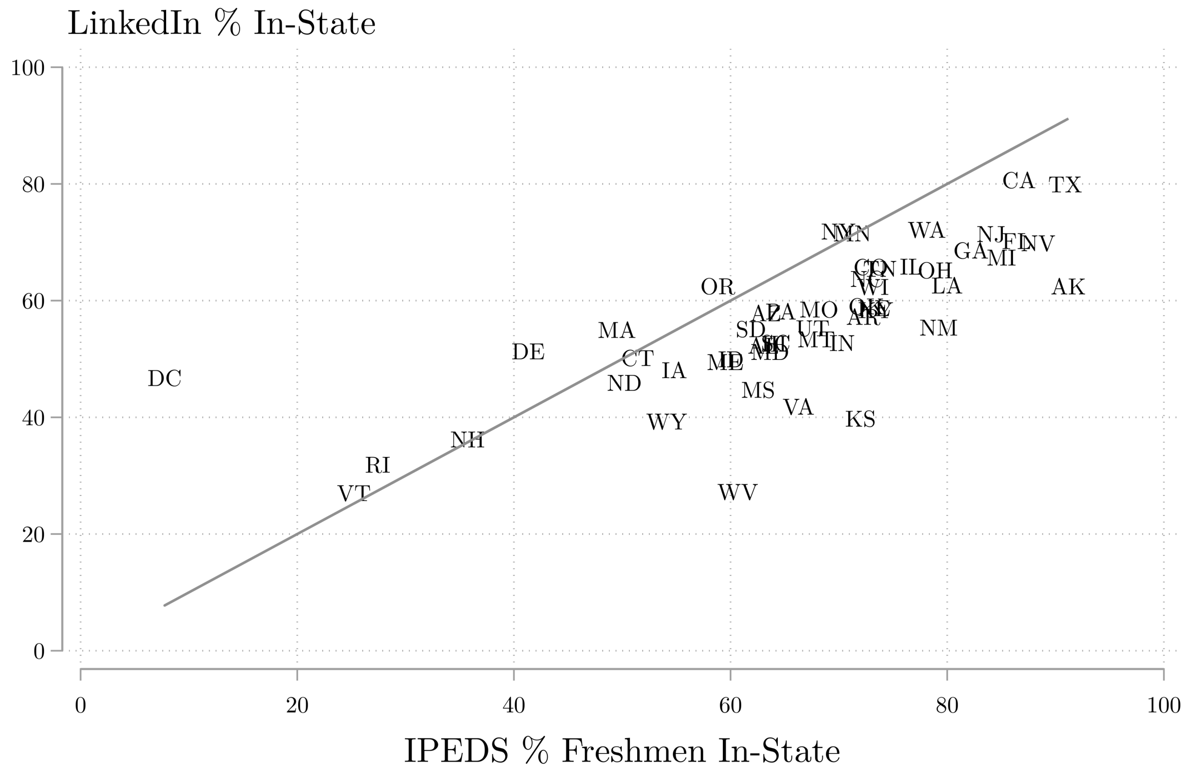
Appendix A. Supplemental Figures and Tables

Figure A1. In-State Freshmen vs. College Graduates Living in State, by Institution



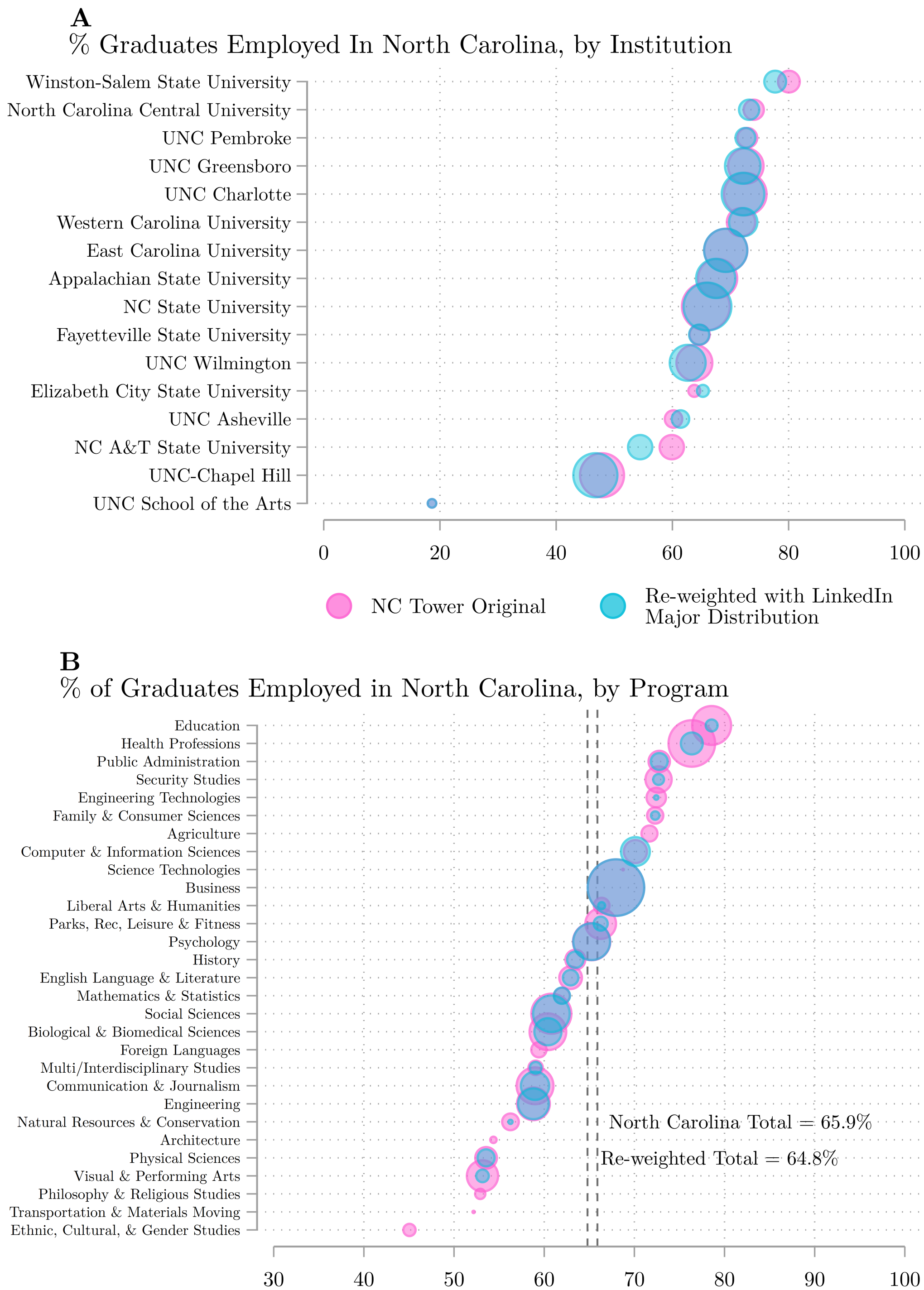
Note: Includes 45-degree line. Each dot represents an institution weighted by enrollment size. See text for details on construction of the analytic sample.

Figure A2. In-State Freshman vs. College Graduates Living in State, by State



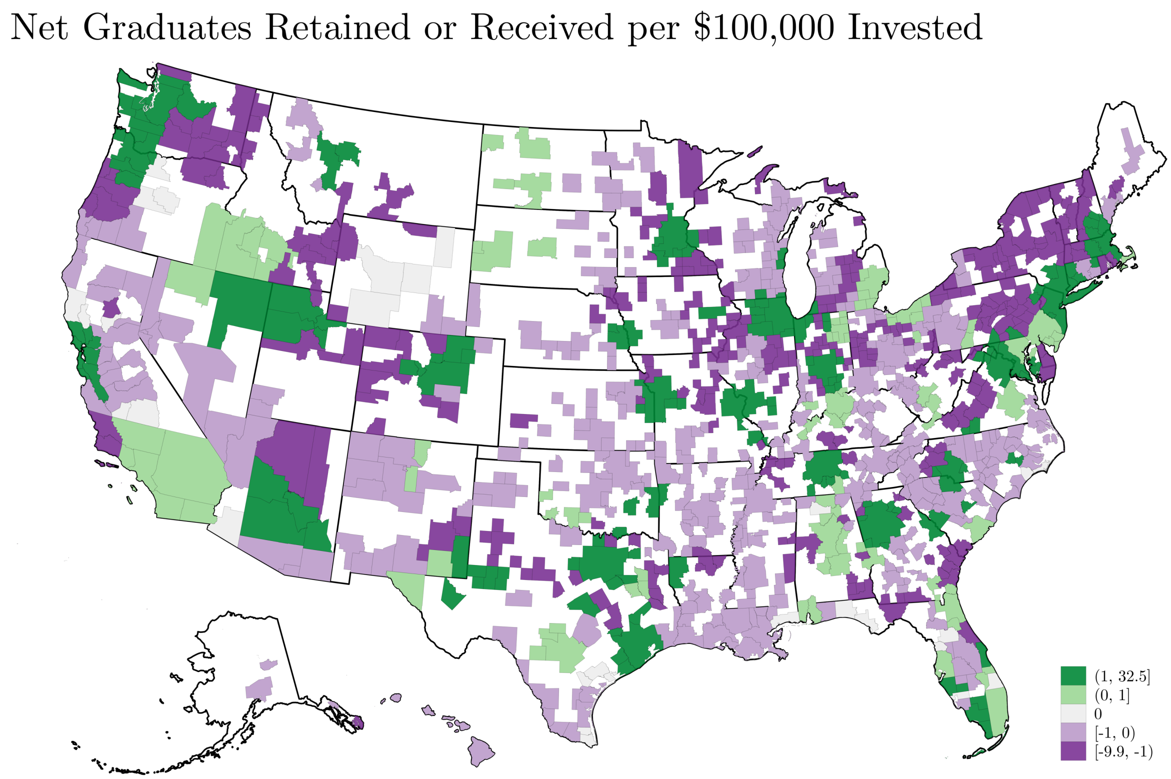
Note: Includes 45-degree line. Each state captures freshman (IPEDS) or graduates (LI) associated with institutions in our analytic sample of public and private institutions. See text for details on sample construction.

Figure A3. Influence of Compositional Differences in Distribution of Majors in LinkedIn and NC Commerce Data on Measures of College-Specific Labor Markets, North Carolina



Note: Dots in panel A are weighted by the total number of graduates across programs at that school found in NC Tower data from the North Carolina Department of Commerce. The Dots in panel B are weighted by the total number of degrees awarded in the corresponding 2-digit CIP code in NC Tower, overlaid with the re-weighted number corresponding to the major distribution reported in LinkedIn at the same set of institutions.

Figure A4. Net Inflow of 4-Year College Graduates and Public Postsecondary Investment



Notes: Figure plots the net number of 4-year college graduates retained or received per dollar of total government spending by institutions in a given geography (i.e., retained/received less produced). Spending and college graduation data are from IPEDS and migration rates are based on the college-specific labor market data produced in this paper. Student counts from IPEDS are based on pooled 2009 and 2010 first-time full-time bachelor's degree seeking cohorts.

Table A1. Program Distributions in LinkedIn and IPEDS (Bachelor's degree completions)

Program (2-digit CIP Code)	LinkedIn Share	IPEDS Share	Difference
Business	0.3166	0.1788	0.1378
Health Professions	0.0553	0.1002	-0.0449
Social Sciences	0.0931	0.0781	0.0150
Liberal Arts & Humanities	0.0299	0.0690	-0.0390
Psychology	0.1008	0.0584	0.0424
Biological & Biomedical Sciences	0.0397	0.0544	-0.0147
Education	0.0184	0.0498	-0.0313
Engineering	0.0683	0.0478	0.0204
Communication & Journalism	0.0527	0.0473	0.0054
Visual & Performing Arts	0.0252	0.0467	-0.0215
English Language & Literature	0.0168	0.0260	-0.0092
Computer & Information Sciences	0.0636	0.0256	0.0380
Security Studies	0.0122	0.0252	-0.0130
Multi/Interdisciplinary Studies	0.0136	0.0232	-0.0096
Parks, Rec, Leisure & Fitness	0.0187	0.0228	-0.0042
History	0.0129	0.0167	-0.0038
Physical Sciences	0.0078	0.0154	-0.0076
Public Administration	0.0124	0.0151	-0.0027
Foreign Languages	0.0003	0.0143	-0.0140
Mathematics & Statistics	0.0111	0.0123	-0.0012
Family & Consumer Sciences	0.0030	0.0121	-0.0091
Engineering Technologies	0.0025	0.0103	-0.0077
Agriculture	0.0010	0.0100	-0.0089
Natural Resources & Conservation	0.0029	0.0085	-0.0056
Philosophy & Religious Studies	0.0026	0.0073	-0.0047
Ethnic, Cultural, & Gender Studies	0.0023	0.0057	-0.0034
Architecture	0.0040	0.0047	-0.0007
Theology & Religious Vocations	0.0039	0.0042	-0.0003
Transportation & Materials Moving	0.0031	0.0026	0.0006
Legal Professions & Studies	0.0010	0.0024	-0.0014
Personal & Culinary Services	0.0019	0.0019	-0.0000
Mechanic & Repair Technology	0.0005	0.0018	-0.0012
Science Technologies	0.0000	0.0005	-0.0005
Construction Trades	0.0000	0.0005	-0.0004
Precision Production	0.0001	0.0003	-0.0003
Military Technologies & Applied Sciences	0.0005	0.0001	0.0004
Library Science	0.0003	0.0000	0.0003
Basic Skills & Development	0.0003	0.0000	0.0003
High School/Secondary Diplomas	0.0001	0.0000	0.0001
Military Science/Leadership	0.0001	0.0000	0.0001
Citizenship Activities	0.0000	0.0000	0.0000
Health-Related Knowledge/Skills	0.0000	0.0000	0.0000

Notes: IPEDS Shares are calculated using the cumulative total of bachelor's degree completions between 2010 to 2018 from IPEDS for the sample of 4-year institutions that are also present in our LinkedIn sample. LinkedIn shares come from the top-15 majors and counts listed for each institution, for which the text was crosswalked to CIP code titles.

Table A2. IPEDS Peer and Geography-Based Labor Market Peer Overlap for Two Institutions in North Carolina

Institution	Peer State	Geography Peer	IPEDS Peer	Cosine Similarity
A. University of North Carolina at Chapel Hill (3.4% overlap)				
Duke University	NC	1	1	0.879
North Carolina State University at Raleigh	NC	1	0	0.972
Campbell University	NC	1	0	0.946
North Carolina Central University	NC	1	0	0.937
Shaw University	NC	1	0	0.935
Meredith College	NC	1	0	0.926
William Peace University	NC	1	0	0.912
Saint Augustine's University	NC	1	0	0.885
Apex School of Theology	NC	1	0	0.867
East Carolina University	NC	1	0	0.828
University of Mount Olive	NC	1	0	0.775
Barton College	NC	1	0	0.759
Elon University	NC	1	0	0.751
North Carolina Wesleyan College	NC	1	0	0.736
St. Andrews University	NC	1	0	0.723
University of Virginia-Main Campus	VA	0	1	0.430
Johns Hopkins University	MD	0	1	0.415
University of Pennsylvania	PA	0	1	0.312
Northwestern University	IL	0	1	0.308
University of Michigan-Ann Arbor	MI	0	1	0.293
University of Wisconsin-Madison	WI	0	1	0.263
University of Maryland-College Park	MD	0	1	0.261
University of Pittsburgh-Pittsburgh Campus	PA	0	1	0.224
University of Southern California	CA	0	1	0.212
University of California-Berkeley	CA	0	1	0.203
University of California-Los Angeles	CA	0	1	0.192
The University of Texas at Austin	TX	0	1	0.172
University of Washington-Seattle Campus	WA	0	1	0.122
University of Minnesota-Twin Cities	MN	0	1	0.118
B. Alamance Community College (33.3% overlap)				
Randolph Community College	NC	1	1	0.983
Surry Community College	NC	1	1	0.973
Forsyth Technical Community College	NC	1	1	0.972
Davidson County Community College	NC	1	1	0.970
Rockingham Community College	NC	1	1	0.927
Sandhills Community College	NC	1	1	0.361
Durham Technical Community College	NC	1	1	0.319
Guilford Technical Community College	NC	1	0	0.976
Wilkes Community College	NC	1	0	0.721
Montgomery Community College	NC	1	0	0.608
Wytheville Community College	VA	1	0	0.519
Louisburg College	NC	1	0	0.459
Patrick Henry Community College	VA	1	0	0.452
Piedmont Community College	NC	1	0	0.380
Central Carolina Community College	NC	0	1	0.314
Vance-Granville Community College	NC	0	1	0.298
Johnston Community College	NC	0	1	0.280
Wayne Community College	NC	0	1	0.108
Caldwell Community College and Technical Institute	NC	0	1	0.094
Rowan-Cabarrus Community College	NC	0	1	0.089
South Piedmont Community College	NC	0	1	0.061

Note: IPEDS = Integrated Postsecondary Education Data System

Table A3. Economic Mobility and College Labor Markets, Full Linear Regression Results

Independent variable	Baseline (1)	Add School Characteristics (2)	Add Student Characteristics (3)	Add Geography FEs (4)
Log average 4-year degree wages	2.5566***	1.5876***	1.0535***	1.4213***
	(0.2271)	(0.1510)	(0.1519)	(0.3445)
Control: Private		-0.1086**	0.0317	0.0743
		(0.0421)	(0.0464)	(0.0576)
Level: 2-year		-0.4575***	-0.3225***	-0.3192***
		(0.0377)	(0.0331)	(0.0492)
Level: Mix (Super OPEID)		-0.2178**	-0.1893***	-0.2241*
		(0.0855)	(0.0643)	(0.1198)
Selectivity: Highly		0.1118**	0.1174**	0.2261***
		(0.0548)	(0.0518)	(0.0526)
Selectivity: Very		-0.1992***	0.0665	0.2085***
		(0.0671)	(0.0633)	(0.0623)
Urbanicity: Suburban		0.0241	-0.0179	-0.0376
		(0.0252)	(0.0263)	(0.0295)
Urbanicity: Town/Rural		-0.0773***	0.0084	0.0041
		(0.0286)	(0.0284)	(0.0359)
Urbanicity: Mix (Super OPEID)		0.0364	-0.0623**	-0.0333
		(0.0325)	(0.0307)	(0.0342)
Historically Black College/University		-0.3420***	0.1101	0.1656*
		(0.0759)	(0.0770)	(0.0940)
Log net price of attendance (in $1,000s)		0.1315***	0.0491**	0.0129
		(0.0306)	(0.0245)	(0.0387)
Log instr. expend per FTE (in $1,000s)		0.2944***	0.1366***	0.1858***
		(0.0349)	(0.0310)	(0.0383)
Log Avg 2010-2018 FTE (in 1,000s)			0.0456***	0.0610***
			(0.0163)	(0.0202)
Share of UG White			-0.1175	-0.2230
			(0.1793)	(0.2520)
Share of UG Black			-0.2741	-0.5480*
			(0.2074)	(0.3202)
Share of UG Hispanic			0.3922**	0.1180
			(0.1846)	(0.3421)
Share of UG Asian			1.0630***	1.0117**
			(0.2541)	(0.4506)
Share of UG Female			-0.9553***	-0.6850***
			(0.1837)	(0.1313)
Share of UG age 25 and above			-0.6620***	-0.5313***
			(0.1404)	(0.1233)
Log median parents' income (in $1,000s)			0.4469***	0.4649***
			(0.0859)	(0.1272)
Constant	-9.9818***	-7.2795***	-6.7525***	-8.2241***
	(0.7121)	(0.5123)	(0.5135)	(0.8493)
Adjusted R-squared	0.289	0.696	0.793	0.830
Region Fixed Effects	No	Yes	Yes	No
LI Geography Fixed Effects	No	No	No	Yes
Observations	1913	1913	1913	1913

Note: Observations are weighted by the number of students in a cohort with parents in the bottom income quintile. Wages for each institution's individualized labor market were calculated using the average hourly wages for bachelor's degree recipients in each CBSA from pooled 2010-2018 ACS estimates and aggregated to the LI-geography level. These averages were then multiplied by the share of an institution's graduates residing in each area and then summed within institution. Standard errors in parentheses are clustered at the institution's LI-geography level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A4. Variation in the Social Return to Public Investment in Higher Education By Funding Source

	Panel A: Graduates Produced per $100K		Panel B: Graduates Retained and Received per $100K		Panel C: Net Graduates Retained and Received per $100K	
	State and Local	Federal	State and Local	Federal	State and Local	Federal
	(1)	(2)	(3)	(4)	(5)	(6)
Log hourly BA degree wage	0.1436	-0.0214	2.8216***	2.6097***	2.6780***	2.6312***
	(0.4300)	(0.2417)	(0.6347)	(0.4523)	(0.4652)	(0.4530)
Share enrollment in private institutions	1.7030***	0.0374	2.5694***	0.9096***	0.8664***	0.8722***
	(0.1790)	(0.1242)	(0.2375)	(0.2332)	(0.2225)	(0.2124)
Share enrollment in selective institutions	-0.0189	0.6796***	-0.3741**	0.3227**	-0.3551**	-0.3568***
	(0.1142)	(0.0926)	(0.1594)	(0.1413)	(0.1371)	(0.1363)
(Share enrollment in private institutions)*(Share enrollment in selective institutions)	1.0253**	-0.0139	0.4474	-0.5363	-0.5779	-0.5224
	(0.4224)	(0.2797)	(0.5154)	(0.4112)	(0.3865)	(0.3594)
Share enrollment in towns or rural areas	-0.0577	-0.0562	-0.5715***	-0.5812***	-0.5138***	-0.5251***
	(0.1450)	(0.0805)	(0.1899)	(0.1463)	(0.1372)	(0.1333)
Adjusted R-squared	0.598	0.489	0.580	0.376	0.348	0.351
Census Division Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	261	263	261	263	261	263

Notes: Robust standard errors appear in parentheses. The outcome for Panel A is the number of 4-year college graduates produced at institutions in a given LI geography divided by the total amount of state, local, and federal government spending by those colleges. The outcome for Panel B contains a similar measure but the numerator instead estimates the total number of 4-year college graduates in a given LI geography's labor force after graduation based on the number of graduates produced and our college-specific labor market shares. LI geographies without any 4-year college enrollment are excluded from these regressions, since our measures are not defined for those areas. Average wages are from pooled ACS 2009-2018 data and estimated using employed bachelor's degree recipients ages 23-32. Selectivity is defined as the share of the LI geography's enrollment from institutions in categories 1-3 of Barron's Competitiveness Index. Student counts from IPEDS are based on pooled 2009 and 2010 first-time full-time bachelor’s degree seeking cohorts. All regressions additionally control for the log of FTE enrollment in each area.

Appendix B. Data Collection and Crosswalk Creation

We begin with 2,832 public and private non-profit 2- and 4-year institutions in the Integrated Postsecondary Education Data System (IPEDS) and identify, to the extent possible, every school's official LinkedIn (LI) landing page. Appendix Table B1 assesses our success, using both unweighted and weighted IPEDS counts of associate's and bachelor's degrees granted between 2010 and 2018. We obtained LI geographic data for 2,600 institutions (approximately 92 percent). These institutions account for 99 percent of the associate's and bachelor's degrees awarded from 2010 through 2018, implying little information loss, mostly due to missing or unclaimed LI pages for a small number of schools. We observe small, but practically insignificant, differences in our ability to locate 4-year versus 2-year institutions, with 2-year institutions being somewhat less likely to have a claimed LI page.

For each of the 2,600 schools with a valid page, we obtain alumni counts using year (of attendance) filters 2010 through 2015. Our target population is bachelor's and associate's degree recipients from each college in our sample between 2010 and 2018. Because we are using aggregate data from institutional pages, we cannot explicitly limit our LI search to *graduates*, and thus the year filters capture all individuals who report attendance that overlapped with the specified date range (in our case 2010-2015). As a concrete example, a student who reports attending College A from 2014-2018 would be captured in our search, whereas another who reports attending from 2016-2020 would be excluded. We exclude years 2016-2018 in our filter to minimize the number of former or current students who completed or were scheduled to complete their degrees after 2018.[1]

[1] We also collected data based on a date rate of 2010-2018. Relative to the 2010-2015 sample, the wider window produced larger counts of alumni working for the institution itself and residing in the institution's CBSA and state, suggesting this filter captures more current students and very recent graduates (i.e., post-2018). Students with LI profile years that predate 2016 appear less likely to be currently attending the affiliated institution.

Geographic units in LI tied to alumni counts are identified with a unique text string. Each sub-state LI geography roughly corresponds to one or more Core-Based Statistical Areas (CBSA) from the Census Bureau. After visual inspection of the LI geography text strings, we concluded that no single CBSA designation (e.g., metropolitan areas only) would properly capture the underlying population found in each LI geography. While most LI geographies approximate a metropolitan statistical area (e.g., San Antonio, Texas Area), a small number map to a micropolitan statistical area (e.g., Danville, Virginia Area), and others likely represent a Combined Statistical Area (CSA), capturing a group of two or more CBSAs with economic ties (e.g., Greater Chicago Area). We map all CBSAs from the 2013 Census definitions to an LI geography using an algorithm described in subsection B below.

A. Coverage Assessment and Supplemental Geography Search

The data from the top 15 geographies on each institution's page account for 81.8 percent of the total number of alumni in our sample who report residing in the United States.[2] Coverage varies by institution type—with somewhat higher coverage of 2-year institutions (90 percent) than 4-year institutions (80 percent).

To supplement the top 15 locations, we also obtain counts for the full list of each institution's own-state geographies by individually searching the in-state locations that were excluded from the initial top-15 lists. Finally, we augment each institution's set of locations with missing geographies from a pooled group of three matched peer institutions, determined using a Mahalanobis distance algorithm.[3] That is, for a given institution, we obtained counts from

[2] The non-foreign U.S. alumni count for each institution is known because the "United States" is included as the top destination for each school. Aside from foreign geographies, all the other geographies are a subset of the total U.S. count.

[3] Variables include each institution's state, control (public/private), level (2-year/4-year), percent of freshman attending from in-state, basic Carnegie classification, and Barron's selectivity rating.

locations found in the top-15 LI geographies of three peer institutions that did not appear in the focal institution's initial top-15 list. For example, the University of Michigan (UM) was identified as a peer of Michigan State University (MSU). UM listed the "Greater Philadelphia Area" in its initial top 15 list, but MSU did not, leading us to search for the count of MSU graduates residing in Philadelphia, which we then added to MSU's list. After these additions, our dataset covers about 84.0 percent of U.S. alumni. Since we always capture counts or shares of graduates residing in an institution's own LI geography or own state, the remaining 16 percent of graduates come from LI geographies outside the state.[4]

We ignore this unlocated fraction of alumni in our analyses. However, we explore differences in this unlocated share across types of institutions and hold this limitation in mind as we interpret our findings. Most notably and perhaps unsurprisingly, unlocated graduates are more common among 4-year institutions than 2-year institutions (17 and 10 percent, respectively), and the most selective 4-year institutions have lower average coverage in our data (21 percent unlocated) than less selective institutions (16 percent unlocated). This suggests institutions with high admissions standards tend to send graduates to more diffuse geographic locations, a pattern revealed by our other analyses as well.

B. Building a Crosswalk from CBSA to LinkedIn Geography

The institution-level LI data and a manual search of the LI website revealed 286 total US sub-state geographies. The 2013 vintage of CBSA codes from the Census Bureau contains 917 areas, excluding Puerto Rico, consisting of 381 Metropolitan Statistical Areas (MSAs) and 536 micropolitan statistical areas. There are also 166 CSAs, which are groups of two or more CBSAs that have close economic ties, according to Census. Some micropolitan statistical areas are

[4] These do not include foreign geographies. The 16 percent is among those known to be residing in the United States. See footnote 2.

clearly present in the LI data, and many geographies are likely larger than a single MSA. To accommodate these differences between the two data sources and to facilitate a high degree of flexibility for future users of these data, we built a crosswalk that maps each CBSA to one LI-geography code using the following steps:

Step 1. We begin by fuzzy string matching each LI text string to CBSAs based on primary city names within CBSA and state. Each LI code can be parsed to obtain both pieces of information. After one iteration of this process, we were able to confidently match 278 out of 286 LI codes to at least one CBSA. The remaining 8 were coded or recoded into a main CBSA by hand.

This matching procedure revealed that a small number of LI geographies were more specific (or smaller) than a single CBSA. For instance, in the Los Angeles area, LI has codes both for the "Greater Los Angeles Area" and "Orange County, California Area" yet Census has just a single CBSA, the "Los Angeles-Long Beach-Anaheim, CA" MSA capturing these areas. To eliminate many-to-one matches from LI to CBSA, we combined LI codes exhibiting this phenomenon. In total, we collapsed 14 codes into 6 new codes. To use the previous example, we created a new code for "Los Angeles/Orange County Combined, California Area." We flag institutions in our dataset without alumni counts from all component areas of the new code, as the counts will not represent the full number of graduates residing in the combined area, only a part. The resultant measurement error is likely small, as the missing location is most always the smaller one(s), like Orange County rather than the Greater Los Angeles Area.

At this stage, the interim data product contains 281 observations where each LI code is matched to (at least) one main CBSA; 275 codes match to one CBSA, and 3 matched to two (e.g., the "Akron/Cleveland, Ohio Area" has two distinct CBSA codes). These 281 "main match" CBSAs are depicted graphically in the top panel of Appendix Figure B1.

Step 2. To assign CBSAs without a main match to an LI geography, we first use CSA definitions from Census. Of the 166 CSAs, 107 had component CBSAs matched to a single LI geography code in Step 1. We assign the remaining, unmatched component CBSAs to the same LI geography code, effectively mapping these LI codes to a CSA. To provide a concrete example, the "Greensboro/Winston-Salem, North Carolina Area" LI code was matched to the "Winston-Salem, NC" and "Greensboro-High Point, NC" CBSA codes in Step 1. In Step 2, we assign the "Mount Airy, NC" and "Burlington, NC" micropolitan CBSAs to the same LI geography because they are part of the same CSA.

Step 3. Some CSAs contained CBSAs that matched to two distinct LI codes in Step 1. For example, the "San Francisco Bay Area" and "Stockton, California Area" are two distinct LI codes with unique CBSA codes, but both are also in the larger CSA, "San Jose-San Francisco-Oakland, CA." In these cases, we assign the remaining CBSAs in the CSA to the already matched CBSA with the largest population. Hence, the "San Jose-Sunnyvale-Santa Clara, CA" along with several other MSAs in the CSA get assigned to the "San Francisco Bay Area" LI code, rather than to "Stockton," based on size.

Step 4. At this stage, 356 CBSAs remain, most of them micropolitan statistical areas that fall outside a CSA represented in LI and do not have their own distinct LI geography code. We assign these CBSAs to an LI code using the shortest driving distance from their geographic centers to the geographic centers of the main-match CBSAs from Step 1. We calculate driving distance using the Stata package, *georoute* (Weber & Péclat, 2017).

The final dataset contains 917 observations, one for each CBSA code mapped to an LI geography. We include a flag for the match type which takes values 1 through 4, corresponding to each of the four steps described above. The lower panel of Appendix Figure B1 graphically depicts the crosswalk process, combining Steps 2 and 3 into one class for CSA-based matches.

C. Final Dataset of Institution-Geography Shares

For all 2,600 institutions, we take the alumni counts from each of the 278 available U.S.-specific geographies in LI and divide them by the institution's total number of alumni residing in the United States. These shares sum to 1 within institution after we add an observation with the count and share of "unlocated" U.S. graduates. We also include a renormalized column of shares that excludes the unlocated graduates. A wide version of the dataset is also available upon request, where each row represents an institution, rather than an institution-geography pair, and the columns contain shares for each of the 278 geographies.

Users can pair these datasets with the CBSA-LI crosswalk to obtain additional information about the places where the graduates of institutions live and work. Our use of average wages from the American Community Survey aggregated from CBSA to LI geography is one example.

Appendix B References

Weber, S., & Péclat, M. (2017). A Simple Command to Calculate Travel Distance and Travel Time. *The Stata Journal: Promoting Communications on Statistics and Stata*, *17*(4), 962–971. https://doi.org/10.1177/1536867X1801700411

Appendix B. Data Creation Details

Figure B1. LinkedIn Geographies Mapped to Core-Based Statistical Areas (CBSAs)

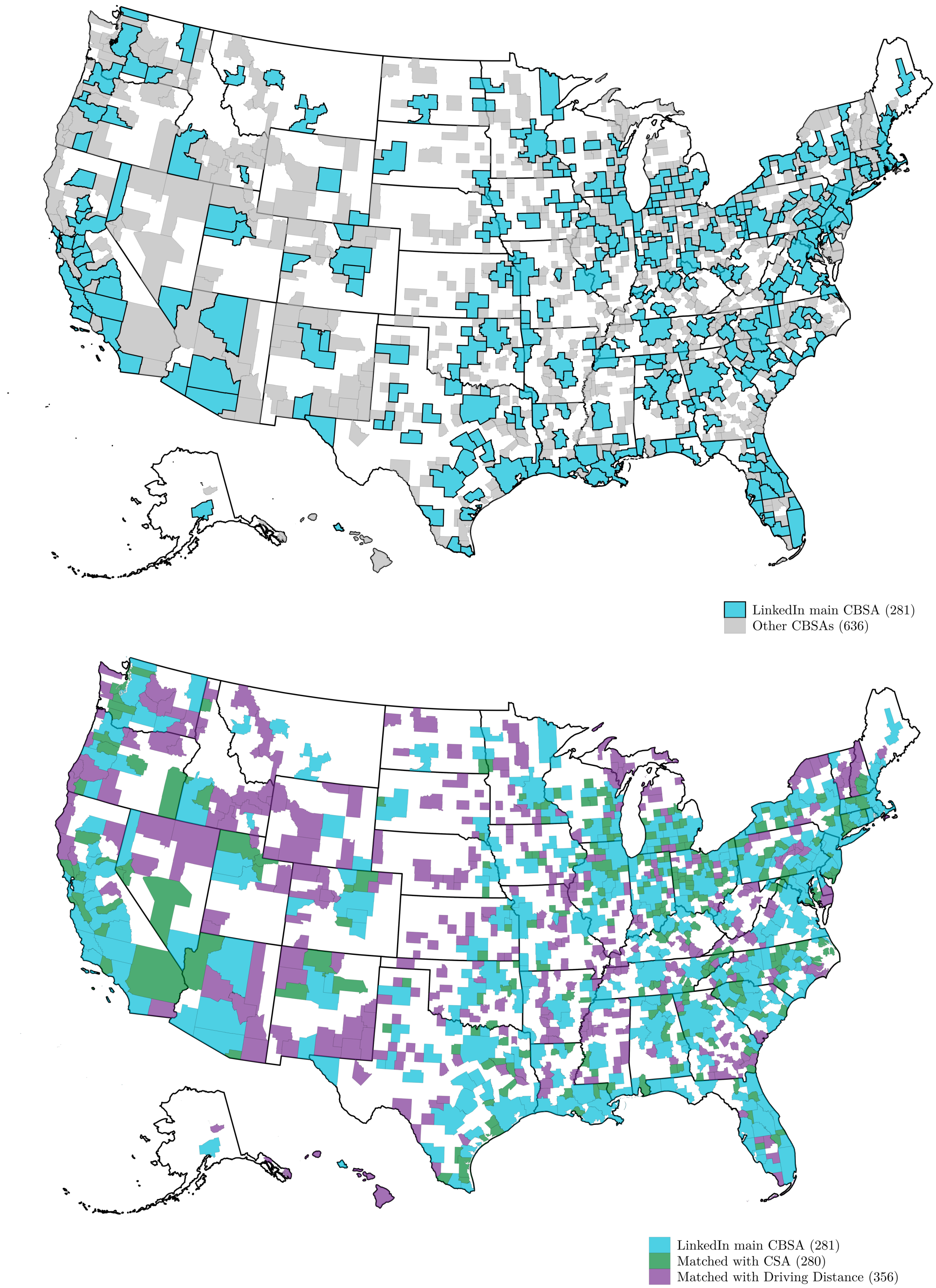


Table B1. LinkedIn Institution Page Discovery and Coverage Rates (Percentages)

LinkedIn Page Status	All schools (Unweighted)	All schools (Weighted)	4-year schools (Unweighted)	4-year schools (Weighted)	2-year schools (Unweighted)	2-year schools (Weighted)
Page Found (no issues)	91.81	98.98	92.45	99.33	90.54	97.87
Unclaimed Page	5.05	0.60	4.04	0.28	7.05	1.59
Less than 100 alums	1.69	0.07	1.91	0.07	1.26	0.06
Page Not Found	1.45	0.36	1.59	0.32	1.16	0.48
N	2,832	25,803,448	1,881	19,525,432	951	6,278,017

Notes: Weighted columns use IPEDS total associate's and bachelor's degree counts for each institution between 2010 and 2018.