# The Spillover Effects of Labor Regulations on the Structure of Earnings and Employment: Evidence from Occupational Licensing

Samuel Dodini[1]

**Abstract**

This paper measures how labor regulations affect the structure of earnings and employment in other occupations in the context of occupational licensing. Using a state boundary discontinuity design, I estimate the market spillovers of licensing on other occupations with similar skills, which I classify using hierarchical clustering techniques on skills data from O*NET. I find evidence of negative earnings and employment spillovers, with the largest earnings effects concentrated among women, black, and foreign-born Hispanic workers. These effects lead to greater earnings inequality. The results are most consistent with licensing changing skill- and industry-specific labor demand and with a monopsony model where licensing increases search costs and reduces workers' outside options.

**Keywords:** Occupational Licensing, Labor Demand, Spillovers, Monopsony, Labor Regulations, Inequality

**JEL Codes:** J21, J24, J31, J42, J62, D63

[1]Department of Economics, Norwegian School of Economics, Helleveien 30, 5045 Bergen, Norway. samuel.dodini@nhh.no

# 1. Introduction

Occupational licensing is state-sanctioned permission to work in a particular occupation. These regulations are typically passed in pursuit of protecting the health, safety, and well-being of consumers. Across the United States and Europe, licensing has grown during the last fifty years from applying to approximately 5% of workers to over 20% (Cunningham, 2019; Koumenta et al., 2014; Koumenta and Pagliero, 2019). As licensing grows, it becomes increasingly important to understand how these regulations affect workers and the structure of employment and earnings, particularly if they contribute to new adverse labor market frictions.

The prior literature suggests that occupational licensing regulations in the US, many of which differ across states, have significant effects on the labor market dynamics of licensed occupations[1] and on certain occupations that perform substitutable functions.[2] However, except for studies that examine occupations that perform overlapping duties, the prior literature has not considered how licensing regulations in one occupation spill over to directly affect the labor market experience of workers in other occupations. In particular, it is important to consider these questions: for workers that would have entered a licensed occupation *but for* the requirements of the license, where do they go, what are their earnings and employment rates, and how does that affect the labor market generally? The answer can inform economists, workers, and policymakers about the important ways in which occupational regulations may exacerbate income inequality and reduce economic efficiency.

This study addresses these questions by testing for the presence of earnings and employment spillovers of occupational licensing on a set of plausible counterfactual occupations. The lack of clear or systematic definitions of "counterfactual occupations" has been a key limitation in the literature. I address this by defining these as occupations that use similar skills, which I measure using data from the Occupational Information Network (O*NET) database. I propose the application of an unsupervised machine learning procedure for creating skill groups. I then use variation in the licensing environment across US states to empirically identify the scope of earnings spillovers as well as the mechanisms behind them.

I empirically test for spillovers in three steps. First, in order to define a set of counterfactual occupations, I use data from the O*NET database and a non-parametric hierarchical agglomerative clustering (HAC) technique to group together occupations into clusters that require similar levels of key skills. The skills upon which I base these clusters come from Acemoglu

[1] Licenses reduce overall labor supply into licensed occupations (Blair and Chung, 2019; Kleiner and Soltas, 2019), change the composition of workers (Bailey and Belfield, 2018; Blair and Chung, 2018; Redbird, 2017), increase prices for goods and services produced by licensed workers (Adams III et al., 2002; Wing and Marier, 2014), generate a wage premium in licensed occupations (Kleiner and Krueger, 2013; Gittleman et al., 2018; Kleiner and Vorotnikov, 2017; Kleiner and Soltas, 2019; Pizzola and Tabarrok, 2017; Thornton and Timmons, 2013; Carollo, 2020b), and reduce interstate labor migration and occupational mobility (see Johnson and Kleiner (2017); Kugler and Sauer (2005); Bae and Timmons (2023) and Kleiner and Xu (2020)). There is also some evidence that occupational licensing increases earnings inequality because the wage returns to having a license appear at the upper end of the education or income distribution (Kleiner and Krueger, 2013; Gittleman et al., 2018; Zhang and Gunderson, 2020).

[2] See Cai and Kleiner (2016); Kleiner and Park (2010); Kleiner et al. (2016).

and Autor (2011) and represent combinations of non-routine, routine, manual, cognitive, and interpersonal skills.[3] This is the first study to use this novel, skill-based approach to study the effects of labor market regulations. This approach provides a roadmap for future work to expand the set of applications for data on occupational skills.

Second, using data from the Current Population Survey, I use the share of individual workers that indicates they are required to have a license as a proxy for the regulatory environment for each occupation in each state (Kleiner and Soltas, 2019). As the key treatment variable, I calculate the share of workers licensed within a state-skill cluster cell *outside* one's own occupation (which I call the "focal occupation") to measure licensure exposure from other relevant occupations. My empirical strategy then uses microdata from the American Community Survey in a state boundary discontinuity design to systematically compare the earnings of observationally equivalent workers in the same occupation in local labor markets that arbitrarily face different licensing exposures depending on which side of the state border they reside in. Because the licensing environment is defined at the state level, licensing shares are exogenous to local labor market factors. I also test for heterogeneous effects across different subgroups, particularly across gender, race/ethnicity, and nativity. Using my estimates, I then calculate the counterfactual distribution of earnings within occupations if licensing were eliminated altogether. This exercise demonstrates the total effect that licensing has on earnings inequality when accounting for these spillovers in addition to any positive earnings effects of licensing one's own occupation.

Third, I estimate the effects of licensing exposure on local labor market employment and demographic composition in each focal occupation. The direction of this employment effect informs the underlying mechanism behind earnings spillovers. A positive employment spillover on other occupations is consistent with a direct labor supply mechanism, while a negative employment spillover is consistent with other explanations such as shifting labor demand or a monopsony mechanism.

Consistent with the prior literature, I find an average earnings premium of approximately 8% in occupations required to have a license in their state relative to the same occupation in non-licensed jurisdictions on the other side of a state border. Conversely, I find that a 10 percentage point increase in the share of licensed workers in the same skill cluster outside a worker's own occupation (approximately one standard deviation) is associated with earnings that are 1.6-2.3% *lower*. These negative effects are stronger for women, non-Hispanic black, and foreign-born Hispanic workers and are most concentrated in the non-tradable sector of the economy. Because these demographic groups are in the lower portion of the income distribution, these effects imply that licensing and related regulation externalities contribute to local

[3]The application of this clustering approach to occupational skills evolved concurrently with Dodini et al. (2023). An alternative to this approach would be to use empirically observed job transitions to cluster occupations. However, job-to-job transitions are endogenously determined by the structure of the labor market, including licensing laws. Skill clusters, therefore, measure a set of counterfactual options independent of the structure of local labor markets.

income inequality. My counterfactual exercise suggests that eliminating occupational licensing would reduce earnings inequality within occupations by 2-4% across various measures such as the 90/10 and 10/50 percentile earnings ratios, while the overall Gini coefficient within occupations would fall by as much as 7%. Eliminating licensing would increase mean earnings, particularly for black workers. In overall counterfactual exercises, for every extra $1 earned via the own-occupation licensing premium, approximately $2.23 is lost via spillover effects.

I find no evidence of a direct labor supply increase to the focal occupation. On the contrary, I find a statistically significant decline in employment in the focal occupation as a result of licensing in other occupations. I also find that as a cluster outside the focal occupation becomes more licensed, the share of workers in the focal occupation who are women falls, as does the average years of completed education in the focal occupation. The share of workers in the focal occupation that is Hispanic or foreign-born rises. Following Oster (2019), I test the importance of sorting along these characteristics in explaining the earnings effects I observe. Occupational composition in observed characteristics collectively accounts for approximately one-third of the observed earnings effects; under the assumption of the equal importance of sorting on unobserved and observed characteristics, total sorting on both accounts may explain approximately half of the total spillover effects.

I augment my analysis with a variety of robustness tests, including specifications with localized labor market fixed effects that limit identifying variation to areas that border multiple states, using simple cross-state variation in licensing for identification, sequentially eliminating skill groups, and using a different data source for licensing rules. I also perform a placebo exercise in which I randomly assign occupations to skill clusters and recompute my estimates. This exercise rules out correlated shocks coming through mechanisms other than occupational skill similarity.

I then discuss some of the mechanisms that may explain spillovers across occupations. Spillovers may come through several candidate mechanisms. First, licenses can have negative wage spillovers by raising costly barriers to entry in one occupation and redirecting and increasing labor supply to unlicensed occupations. However, the results that I find for employment are not consistent with this explanation.

Second, other work on licensing suggests that the rate of new firms opening in particular locations is responsive to the licensing environment (Plemmons, 2022; Zapletal, 2019). If firms avoid costly licensing rules via location choices, this may drive down local demand for labor in general or for labor in particular skill groups or industries. Though I do not find overall reductions in labor demand, I do find some support for the industry-specific labor demand effect by separating within-industry licensing exposure from total exposure. Both types of exposure have important effects, meaning that, while industry-specific labor demand may be a mechanism behind some portion of the total effect, industry sorting cannot explain the entire effect.

Third, a contraction in employment in licensed occupations might also reduce employment in occupations that are complements to licensed occupations. Measuring complementarity exposure as a weighted share of licensing intensity within narrow industries, I find a limited role for this channel in explaining the earnings and employment effects in my analysis.

Fourth, in a state of imperfectly competitive labor markets, occupational licensing may increase firm monopsony power because licenses make outside options costlier to enter or harder to find. In such a setting, employment in other occupations may fall because firms with market power have the ability to hire fewer workers and pay lower wages (Ashenfelter et al., 2010). This can happen directly by increasing adjustment costs (e.g. time, money) or by indirectly reducing the number of outside firms hiring workers with similar skills through firm location choices (Plemmons, 2022). Both of these may lead to a decrease in labor supply elasticity to the firm and increase firm market power. Though not conclusive without firm-level data, this analysis provides evidence of this channel being a likely explanation.

This paper contributes to the growing empirical literature on the effects of labor market regulations on workers and the labor market as a whole. In addition to the growing field of research on occupational licensing, recent studies have examined the effects of enforcing non-compete and non-disclosure agreements and have found that enforcement of these agreements reduces worker wages (Starr, 2019; Starr et al., 2021; Lipsitz and Starr, 2022), including possible negative effects on those unconstrained by these agreements (Starr et al., 2019). The main mechanisms are a combination of a decrease in the number of outside options for a worker, an increase in search and switching costs, and a decline in worker mobility, each of which increases firm monopsony power. This recent literature connects to a well-established theoretical literature on the role of labor market institutions and occupational crowding. For example, discriminatory hiring practices against black workers, particularly in stable skilled "craft" professions, may lead to occupational segregation and lower earnings for black workers by pushing them into lower-paid service jobs (Bergmann, 1971, 1974). Similarly, unionization in industries may have negative effects on the earnings of non-union labor in the presence of crowding or complementarities (Neumark and Wachter, 1995). This study contributes to this literature by considering licensing as a particularly costly barrier that positively affects the "in-group" while negatively affecting the "out-group."

On the topic of occupational licensing, recent studies suggest there are sizable wage premiums associated with occupational licensing on the order of 7–30% (Kleiner and Krueger, 2013; Gittleman et al., 2018; Kleiner and Vorotnikov, 2017; Kleiner and Soltas, 2019; Thornton and Timmons, 2013; Carollo, 2020b; Zhang and Gunderson, 2020). Synthetic control and other panel estimates of licensing wage premiums within occupations are approximately 7–10% and are similar to my estimates (Pizzola and Tabarrok, 2017; Thornton and Timmons, 2013; Carollo, 2020b).

The main mechanism for these wage premiums is a reduction in labor supply and employment in licensed occupations (approximately 20% as in Blair and Chung (2019); Kleiner and Soltas (2019)), with some exceptions in occupations like nursing (DePasquale and Stange, 2016). Licensed workers also work more hours (Bailey and Belfield, 2018; Kleiner and Soltas, 2019), and prices in the product market for licensed workers rise (Adams III et al., 2002; Wing and Marier, 2014). In addition, the composition of workers shifts with licensing, with more women and black workers entering licensed occupations (Bailey and Belfield, 2018; Redbird, 2017), possibly to take advantage of the signal value of a license (Blair and Chung, 2018). Work on migration suggests that licenses decrease interstate migration by as much as 36% (Johnson and Kleiner, 2017). Importantly, occupational licensing reduces labor market fluidity as measured by job changes and can explain nearly 8% of the total reduction in occupational mobility over the last twenty years (Kleiner and Xu, 2020). This paper contributes to this literature by showing that the wage premium to licensing also comes with a wage penalty for other workers, which may arise through these various frictions.

There are two strands of the literature on licensing related to spillovers. First, some studies estimate a wage premium for having a license using binary indicators for licensure as the treatment variable (e.g. Kleiner and Krueger (2013); Gittleman et al. (2018); Zhang and Gunderson (2020)). While the estimates without occupation fixed effects may contain some information about possible "cross-status" spillover effects of licensing—i.e. licensed on the unlicensed–evaluating that information is not the focus of those studies. These coefficients cannot separate occupation selection effects from cross-occupation spillovers. Finally, these studies primarily rely on selection on observables approaches with individual licensing status as the treatment, which complicates a causal interpretation. My boundary discontinuity design addresses these selection issues, and my clustering approach provides a novel method for directly identifying spillovers while accounting for occupation characteristics.

Second, a few important papers find notable direct effects of licensing requirements on occupations that perform substitutable functions. Licensing and credentialing requirements for physical therapists can negatively affect occupational therapists' wages because many services are substitutable between the two (Cai and Kleiner, 2016). When nurse practitioners are given broader scope for their practice, physicians' wages fall, while nurse practitioners' wages rise (Kleiner et al., 2016; Dillender et al., 2022). Kleiner and Park (2010) examine the effects of broadening the scope of practice for dental hygienists. They find that as regulations that allow hygienists to be self-employed are implemented, wages and employment for hygienists both increase, while earnings and employment for dentists both fall. The authors contextualize this result in a monopsony model in which tighter scope of practice regulations grant monopsony power to dentists, who tend to own their own practices and often house the services of hygienists. This paper expands upon this work by broadening the possible role of spillovers and labor market power to a wider set of occupations.

This paper contributes to our understanding of the operation of regulated labor markets by identifying the broad effects of occupation-specific regulations on other occupations. This paper demonstrates that strict entry regulation comes at a cost to other workers: lower labor market earnings and employment for those in occupations that use similar skills. This study also shows that these wage externalities are not consistent with a pure labor supply shift to unlicensed occupations. Though industry- and skill-specific labor demand that responds to licensing rules may play a factor, my analysis suggests this cannot explain the entirety of the effects I find. Occupational entry restrictions increase labor market rigidity and influence firm location choices, both of which intensify firm labor market power. My analysis also sheds light on who bears the largest costs of these occupational regulations and shows that the costs disproportionately load on workers that are already more likely to be lower in the income distribution, resulting in an increase in earnings inequality both within and across occupations. This analysis deepens our understanding of the trade-offs between the consumer protection benefits of entry regulations and the dispersed costs of licensure as they are imposed upon consumers and workers in general, most of whom play no part in the legislative negotiations that ultimately determine the scope of these regulations.

# 2. Data

To empirically test for spillover effects of occupational licensing, I bring together three main data sources: the 2015–2018 Current Population Survey (CPS) for state-specific licensing requirements for individual occupations; the Occupational Information Network (O*NET) dataset for details on the skill requirements of occupations; and microdata samples from the American Community Survey (ACS) from 2014–2017 for data on individual earnings, occupations, demographics, and sub-state geographic identifiers.

## 2.1. Current Population Survey (CPS)

One major challenge to estimating the effects of occupational licensing is a lack of clear data on licensing requirements at the national or state level. In 2015, the CPS began asking individual workers questions regarding licensing and certification, which helps address these measurement challenges. I consider a worker licensed if the worker in the survey indicates 1) that they have an active professional certification or state or industry license; and 2) that any of those certifications were issued by a federal, state, or local government. This classification yields estimates of national licensing shares of approximately 22 percent, consistent with other surveys (Blair and Chung, 2019) as well as other papers using the same measure (Kleiner and Soltas, 2019; Cunningham, 2019).

Using CPS data from 2015–2018, I construct two key measures for my analysis as proxies for the policy environment within each state. First, following Kleiner and Soltas (2019), as a measure of policies affecting a single occupation, I calculate the state-occupation cell share of workers that are licensed. This abstracts away from individual determinants of receiving a license, which may be subject to selection issues. This exercise also allows me to incorporate differences in sub-occupational licensing status into broader occupational categories in the CPS. For example, the "physician assistant" occupation in the CPS includes "physician assistant," "anesthesiologist assistant," and "family practice physician assistant," which often have different licensing rules in different states (Vargo et al., 2020). If these sub-categories are differently licensed across states, my aggregated measure will capture this variation across states within a single occupation code.[4]

Second, using individual licensing status, for every occupation, I calculate the share of workers in the same skill cluster *outside* the excluded occupation (the focal occupation) that is licensed. This measure characterizes "exposure" to licensing from other similarly skilled occupations and is defined at the state level. Every state-occupation cell experiences a different measure of licensing exposure within its own cluster across states. This is the key treatment variable for my analysis. Notably, the approach using individual license shares as a proxy for the regulatory environment is validated in Kleiner and Soltas (2019) and is highly correlated with licensing laws from other sources when the laws are well-defined and understood. Recent work (Carollo, 2020b) suggests that in some cases there may be undercounting of licensing rates in the CPS when compared to well-defined licensing laws for occupations licensed in all US states. Because these measures are aggregations of binary values at the individual level, measurement error in the aggregations will be mean-reverting, which induces attenuation bias. This means my estimates plausibly represent lower bounds.

## 2.2. American Community Survey

To construct my boundary discontinuity sample, I use data from the American Community Survey with geographic identifiers for Public Use Microdata Areas (PUMAs) (Ruggles et al., 2019) on or near each state border for 2014-2017.[5] Figure A1 shows maps of my border PUMAs in four Census Divisions. PUMAs map within states and across counties and are intentionally coincident with Metropolitan Statistical Areas in densely populated areas, and each contains at least 100,000 people and no more than 200,000. I categorize workers into 2010 Census occupation codes to match the licensing shares in the CPS. The dataset also contains data on sex, race/ethnicity, and nativity. For my main outcome of interest, I consider log weekly earnings.[6]

[4]To give an example of the measurement challenge, using OES employment weights and the statutes in the Northwestern Licensing Database (NLD) (Redbird, 2016) to calculate the share of a state-occupation cell licensed, I compare the NLD to the CPS data. The correlation in licensing shares is 0.6. More than *half* of workers in the CPS who say they are required to have a license for their occupation in the CPS would not be required to have a license under a binary (50% cutoff) licensing rule in the NLD, meaning it significantly underestimates regulation intensity within and across occupations. I have replicated my baseline cross-sectional analysis using the list of regulations for 2015 to 2018 in the Northwestern Licensing Database (Redbird, 2016) as a measure of licensing intensity. The result is a set of estimates in a similar direction to my base estimates, though both are attenuated toward zero, particularly in measuring an occupation's own wage premium. See Figure A10.

[5]The ACS and CPS samples are constructed in order to provide sufficient sample sizes within occupations and states to generate occupation-state cell license shares and estimate my boundary discontinuity model. The lack of perfect overlap is not a significant concern so long as there are no substantive shifts in licensing across the non-overlapping years, which is unlikely. The effect of such shifts may be to add a marginal amount of noise to the estimates.

[6]One limitation of the ACS is that part-time workers systematically under-report hours in the survey, leading to implausibly large estimates of their hourly wages (Baum-Snow and Neal, 2009). I follow others in the literature by dropping those with allocated/imputed earnings and using log weekly earnings as the outcome variable rather than hourly wages (Busso et al., 2013). While earnings are a function of wages and hours, when analyzing regular weekly work hours as the dependent variable in my models, the effects are

Table 1: Summary Statistics by Sample

| | Border Sample | | Full Sample | |
|---|---|---|---|---|
| | (1) Mean | (2) SD | (3) Mean | (4) SD |
| Log Weekly Earnings | 6.51 | 0.83 | 6.55 | 0.85 |
| Female | 0.47 | 0.50 | 0.47 | 0.50 |
| NH-White | 0.74 | 0.44 | 0.62 | 0.49 |
| NH-Black | 0.11 | 0.31 | 0.12 | 0.32 |
| Hispanic | 0.10 | 0.30 | 0.18 | 0.38 |
| Asian/Pacific Islander | 0.03 | 0.17 | 0.06 | 0.23 |
| Foreign Born | 0.11 | 0.32 | 0.19 | 0.39 |
| Age | 40.48 | 12.95 | 40.04 | 12.80 |
| High School/Less | 0.37 | 0.48 | 0.34 | 0.47 |
| Associate's Degree | 0.09 | 0.29 | 0.09 | 0.28 |
| Bachelor's Degree | 0.19 | 0.40 | 0.22 | 0.41 |
| Master's Degree | 0.08 | 0.28 | 0.09 | 0.29 |
| PhD/Professional Degree | 0.02 | 0.13 | 0.02 | 0.14 |
| Share Own Occupation Licensed | 0.18 | 0.19 | 0.17 | 0.18 |
| Share Cluster Licensed Outside Focal Occ. | 0.21 | 0.11 | 0.21 | 0.11 |
| Occupation Employment Shares (1-digit) | | | | |
| 0 Management, Business, and Financial | 0.15 | | 0.16 | |
| 1 Computers, Engineers, Scientists | 0.05 | | 0.06 | |
| 2 Social Services, Legal, Education, Arts | 0.11 | | 0.11 | |
| 3 Healthcare, Protective Services | 0.07 | | 0.07 | |
| 4 Food, Building Maintenance, Personal Care, Sales | 0.23 | | 0.23 | |
| 5 Office and Administrative Support | 0.14 | | 0.15 | |
| 6 Natural Resources, Construction, Extraction | 0.07 | | 0.06 | |
| 7 Installation, Maintenance, Production | 0.07 | | 0.06 | |
| 8 Production | 0.05 | | 0.04 | |
| 9 Transportation and Material Moving | 0.05 | | 0.05 | |
| Number of respondents | 1,337,103 | | 4,578,382 | |
| PUMAs | 650 | | 2351 | |
| Occupations | 410 | | 410 | |
| State Boundary/Border Pairs | 110 | | N/A | |

Source: Author's calculations of ACS, CPS, and O*NET data.
Notes: Clusters are based on the description in Section 3.1. ACS samples are from 2014-2017 corresponding with CPS individual licensing data from 2015-2018.

I limit my sample to those ages 18–64 who are in the labor force and report positive weekly earnings. Following Gittleman et al. (2018) and Kleiner and Soltas (2019), I eliminate all "universally" licensed occupations like physicians, lawyers, etc. because they contribute nothing to identification across states, though my estimates are robust to their inclusion. This

small, consistent with Kleiner and Soltas (2019), and not statistically significant in my model.

also helps avoid the possible undercoverage problem mentioned previously (Carollo, 2020b).[7] As seen in Table 1, my border estimation sample contains 1.3 million individuals across the 48 contiguous US states and the District of Columbia in 650 PUMAs, 110 boundary pairs, and 410 Census-defined occupations. The border sample is similar to the overall ACS sample along most dimensions except in the share of the population that is Hispanic or Asian or Pacific Islander and the share that is foreign-born. This is primarily driven by the exclusion of coastal California, which has highly concentrated Asian and Hispanic populations, as well as cities in central and southern Texas. There is also a small difference in the share with a Bachelor's degree. Importantly, these sample areas are very similar in terms of their licensed shares, both within the focal occupation and within clusters.[8] There is, therefore, no significant composition difference in the distributions of occupational employment in these areas relative to the country as a whole with regard to licensing or broad occupation distributions. My measure of licensing exposure in my border sample has a mean of 21%, a standard deviation of 11%, and an interquartile range of 11% to 33%.

## 2.3. O*NET

The Occupational Information Network (O*NET) database is the result of a survey funded and directed by the US Department of Labor. Incumbent workers and occupation experts are surveyed about over 400 attributes of each occupation. These include the abilities required to perform the job, the type of tasks performed, and the skill level of the job. Survey respondents rate the importance of each component as it relates to their single occupation on a 1-5 scale. O*NET then generates a single score for each occupation and each component. I standardize these to be mean zero with a standard deviation of one following Acemoglu and Autor (2011).

Using the 2017 O*NET data, I classify the levels of six important latent skill areas as defined in Acemoglu and Autor (2011) for each occupation. Conceptually, these measures are used elsewhere in the literature to explain skill and work task polarization, but they are also useful in this context to classify occupations by overall skill and task types (Autor and Dorn, 2013; Autor, 2014). These are: non-routine cognitive/analytical; non-routine cognitive/interpersonal; routine cognitive; routine manual; non-routine manual/physical; and non-routine interpersonal adaptability.

The components in the O*NET questionnaire used to define these skills are listed in Table A1. Following Acemoglu and Autor (2011), I generate skill composite measures by summing the value of all the input components. These skills capture important characteristics about each occupation beyond educational requirements and characterize the abilities, either acquired or endowed, that are essential for someone to be able to perform that occupation. Someone working in an occupation that requires routine, manual work is unlikely to easily transition to a job requiring intense non-routine cognitive skills. The imposition of a license to perform

[7] The inclusion of universally licensed occupations only makes the estimates less precise.

[8] My analysis using the full CPS sample (Figure A5) suggests the treatment effects are similar, meaning that the particular composition of the border sample is not a significant concern.

a job heavy in routine, manual work may influence the labor market for workers whose jobs heavily rely on the same underlying skill.

The O*NET data are collected according to SOC code definitions. Following (Acemoglu and Autor, 2011), I use Occupation Employment Statistics (OES) at the national level to create a weighted average of these O*NET skill characteristics at the 2010 Census occupation level to match the occupation categories in the CPS. The final figure is a national employment-weighted average skill content for each Census occupation code across these six skill measures. In addition to these, I also calculate the median log wage for the national distribution of wages in each occupation from the 2015–2018 CPS as an additional clustering criterion. I include this to set bounds on the relevance and feasibility of cluster members that may be in a worker's choice set. A worker's reservation wage may exclude a set of lower-wage (and thus, irrelevant) occupations even if their skill requirements were similar. Many higher-wage occupations may be infeasible to enter (and thus, are also irrelevant). I discuss this further in Section 3.1.

# 3. Empirical Approach

## 3.1. Occupation Clustering

To classify occupations into similar groups based on their skill content and define a set of counterfactual occupations, I use a hierarchical agglomerative clustering technique (HAC) (Sokal and Michener, 1958) because of its non-parametric properties and intuitive interpretation.[9] This approach begins with all occupations in their own cluster then merges the closest occupations together based on the remaining "distance" between occupations and places them in the same cluster. As the allowed distance between cluster members increases, fewer clusters will form. Eventually, all occupations will be grouped in a single cluster. This non-parametric procedure forms a dendrogram (or tree) of these various cluster merges. The researcher using the approach has the option of choosing "cut" points to trim the tree at a set number of clusters or a maximum distance between cluster members. HAC flexibly does not require an occupation to be a member of a larger cluster and has the advantage of being able to handle varying densities across clusters, which is a noticeable feature of the O*NET skills data.

Figure 1 presents a toy example of HAC. The left pane represents data points along two dimensions, and the right pane represents the dendrogram of the hierarchy. First, groups 5 and 6 merge to form the purple cluster. Next, this purple cluster merges with group 4 to form the blue cluster. Next, groups 1 and 2 merge to form the yellow cluster. Then group 0 merges with the yellow cluster to form the red cluster. Finally, group 3 is merged with the blue cluster to form a green cluster. Along the progression of these merges, the analyst may choose either a maximum allowable distance between cluster members (a "cut" point, which is the y-axis measure of distance between points when they are first connected by a horizontal bar) or by selecting a set number of clusters (the number of vertical lines intersecting with a horizontal

[9] The popular K-means clustering algorithm is another alternative, though HAC clustering is more replicable because HAC does not require the selection (ad-hoc or random) of start points for the clusters to begin forming. HAC marginally outperforms K-means in nearly all my diagnostic tests on the O*NET data.

line at some cut point). Depending on the technique chosen to validate a number of clusters as "optimal" or the institutional details known to the researcher, there could be anywhere from 2 to 6 clusters in this example.

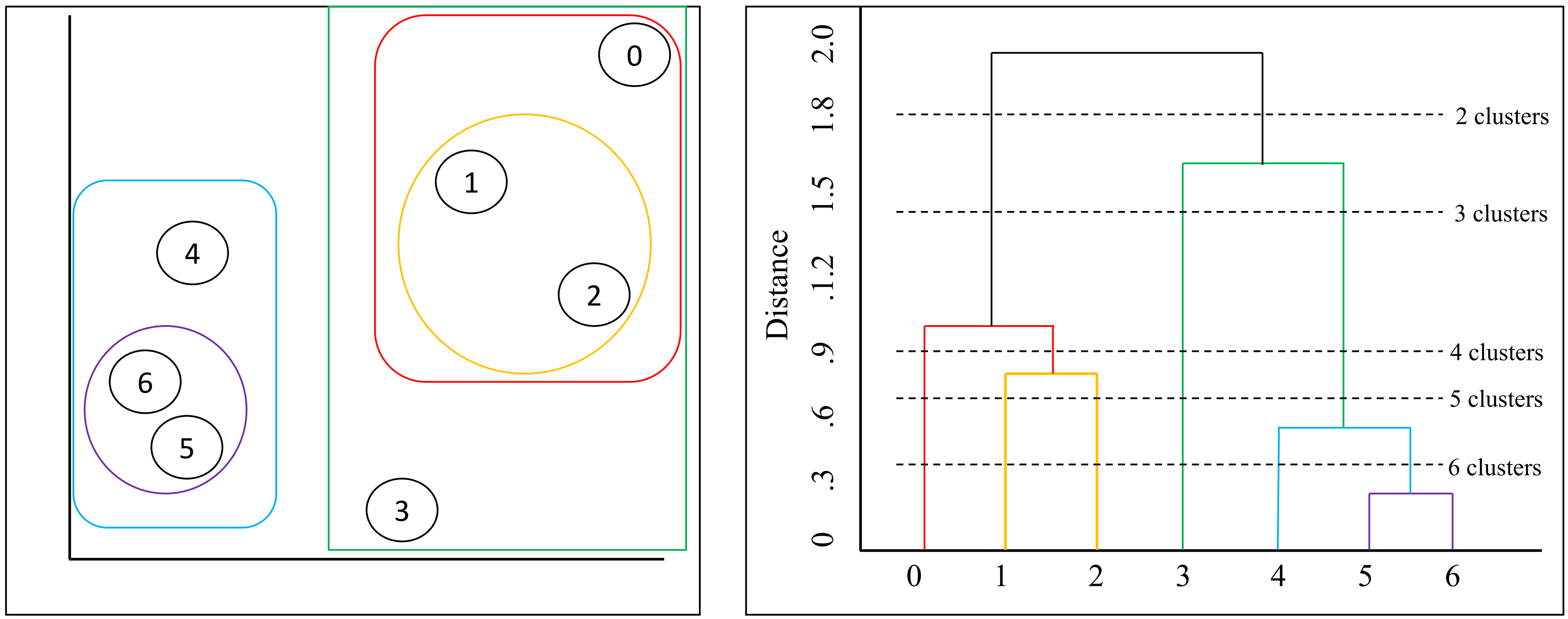
With this technique in mind, I pursue the following steps: first, I calculate the correlative distance between each occupation across these six occupational skill characteristics as well as the national median log wage for each occupation. This distance is simply one minus the Pearson correlation coefficient between occupations on all seven measures. The advantage of this measure is that it is not sensitive to the scales of the inputs as a Euclidean or other distance measure would be. This may be important given that not all my skill measures have the same scale after construction. The result is a single matrix with a range [0,2] for every occupation-occupation dyad. Second, with this matrix of dissimilarity, I use the HAC algorithm to group together occupations based on their distances step by step and form a dendrogram of the relationships. Third, I calculate a data-driven "optimal" number of clusters and select the corresponding cut point. I use the subsequent cluster definitions in my empirical models.

There are three main researcher choices that must be made when performing any HAC exercise. The first choice relates to which input characteristics to use. The literature on skills and trends in wages has focused much attention on the six skills I use in my clustering analysis (Acemoglu and Autor, 2011). These skills prove useful not only in examining wage trends but also in classifying the skills used across occupations. I provide detailed justification—including empirical tests—for using these particular skills for clustering rather than other aspects of an occupation available in the O*NET data (such as its principal components) in Appendix D, particularly in Figure D1 and Table D2. Put briefly, the computer science literature states that in many cases, the principal components of the data, while capturing the greatest variation across the attributes, do not capture the *cluster* structure of the data as well as using a well-reasoned subset of the variables (Yeung and Ruzzo, 2001).[10]

Because the goal of the algorithm is to characterize a set of alternatives that are viable outside options, I include the national median wage in each occupation in the algorithm to minimize the possibility of matching occupations whose labor market returns vastly vary. For example, being a professional athlete is not a viable outside option for a freight laborer (and vice versa). Some comparison of the labor market returns to a package of skills would, of course, be a relevant consideration for any worker. Including the national median wage bounds the distance between cluster members in wages conditional on being close on their skill components—serving the function of trimming wage outliers from the group and classifying them in another.

[10]An alternative to the non-parametric structure is to allow a parametric relationship between occupations based on skill similarity. This approach is explained in Appendix D, and Figure D3 shows substantively similar results to the HAC approach.

Figure 1: Hierarchical Agglomerative Clustering Example




Notes: Moving up the dendrogram, clusters are sequentially merged. The researcher can then choose cut points at a certain number of clusters or at a maximum distance value within clusters. First, groups 5 and 6 merge to form the purple cluster. Next, this purple cluster merges with group 4 to form the blue cluster. Next, groups 1 and 2 merge to form the yellow cluster. Then group 0 merges with the yellow cluster to form the red cluster. Finally, group 3 is merged with the blue cluster to form a green cluster.

The second choice is what parameter of distance to choose when merging two clusters that have already formed. I use what is called "average linkage distance" (or "unweighted pair-group method with arithmetic means" (UPGMA)), which uses the mean data value of all points in formed clusters when determining the distance between clusters, i.e. from cluster mean to cluster mean or cluster mean to singleton (yet unclustered) occupation. Unlike measures such as "single" or "complete" linkages, which, respectively, use the nearest or the furthest unit of the cluster to calculate distances between clusters, the average linkage approach is more robust to outliers within clusters, though the distinction makes little difference in my estimates.

The third choice is how many clusters to use in the final analysis. To support the choice of twenty clusters for my main analysis, I use four validation measures common to clustering applications to validate the cluster structure: Silhouette (Rousseeuw, 1987); Dunn's index (Dunn, 1974); SD index (Halkidi et al., 2000); and the C index (Hubert and Levin, 1976), though there are dozens from which to select. The first two measures are based on maximizing their index values, while the latter two are based on minimizing their values. It is also useful to look for structural breaks in the index values. The visual results of these tests are in Figure D2. Based on the totality of these tests, there is considerable overlap in the optimal number from the mid-teens to twenty. For transparency, I calculate and plot a range of estimates across the number of clusters from 4 to 20 to report the coefficients of interest under larger, less compact clusters (4) relative to smaller, more compact clusters (20) in Figure 4 and 8. As the number of clusters gets larger, cluster size falls, making the occupations more narrowly related along skill dimensions, but identifying variation within the cluster will also fall. In my estimates, treatment effects above approximately ten clusters are robust to increasing the number of clusters and consistent across my outcomes of interest. In my main estimates, I present the results at twenty clusters.

I present the five most frequent occupations in each cluster at their most compact (20 clusters) in Table D1. The definitions are sensible from a human perspective, and many occupations, though a part of separate industries or Census occupation groups, make logical companions to each other. For example, personal care aides may use similar interpersonal, management, cognitive, and physical skills as waiters, though they are separated by industry definitions. Police detectives and private investigators use similar investigative, cognitive, and management skills as construction and building inspectors despite being in very different industries. A childcare worker can personally attest to taking on multiple roles as a fitness/recreation worker, coach, and umpire—often simultaneously.

On a conceptual level, it is useful to compare this approach to more mainstream methods of grouping occupations together, namely Census occupation groups. I provide a useful comparison in Appendix E. Fundamentally, Census occupation groups represent groups that are homogeneous in sub-industry activities (though not skills) and licensing. Generally, the use of these groups may mechanically curtail the effects of licensing by forcing comparisons to those that have already sorted based on niche sub-industries and on having overcome (or being on a path to overcome) barriers generated by licensing. The results suggest that licensing exposure in Census occupation groups has either a zero or a positive effect on earnings, and the relationship is sensitive to model specification and sample. The spillover effects are not robust to taking this wholly different approach using Census occupation groups. However, skill clusters may capture more substantive underlying relationships between occupations, at least with respect to licensing, though the clustering approach is admittedly more of a black box. When instead considering licensing spillovers within major industry groups, the measured effects again are negative and significant. Census occupation groups are the outlier among the grouping options I examine.

## 3.2. Boundary Discontinuity Design

To estimate the direct and spillover effects of occupational licensing, I construct a matched border sample of Public Use Microdata Areas (PUMAs) from the American Community Survey with proximity to a common state border. Importantly, PUMAs are sub-state geographic areas, and observations are at the individual level—some workers living in one part of the state and other workers in another. This structure allows me to leverage the fact that local workers or local economic conditions do not endogenously determine licensing laws because licensing is determined at the state level, and all workers in the state face the same licensing rules. Workers in different parts of each state near each state border have neighbors across the boundary that share the same local economic conditions. Despite shared economic environments, local workers face different state laws for reasons determined by the state, not by their local economic conditions or local government.

My estimating equation includes state, occupation, and boundary fixed effects that generate these narrow comparisons:

$$y_{iocms} = \beta_0 + \beta_1 LicensedShare_{os} + \beta_2 LicensedShare_{cs}^{-o} + X_i'\beta_3 + \delta_o + \gamma_s + \tau_m + \varepsilon_{iocms} \quad (1)$$

This equation characterizes outcome $y$ for individual $i$ in occupation $o$ in skill cluster $c$ in state $s$ whose PUMA is on the state-state boundary $m$ (e.g. the "California-Oregon" border, including all workers along both sides of the border).[11] Outcome $y$ is log weekly earnings. $LicensedShare_{os}$ is the share of occupation $o$ in state $s$ reporting in the CPS that they are required to have a license. $LicensedShare_{cs}^{-o}$ is the share of workers in skill cluster $c$ excluding occupation $o$ (the focal occupation) in state $s$ that reports a licensing requirement in the CPS, scaled to 10 percentage point units, which is approximately one standard deviation. Coefficient $\beta_1$, therefore, captures the earnings effect of licensing individual $i$'s own occupation (the "own-occupation" effect), whereas $\beta_2$ captures the effect on occupation $o$ of a ten percentage point increase in licensure for all other workers in cluster $c$ *outside* of occupation $o$ (the spillover

[11] For PUMAs that share borders with multiple states, I stack the sample and divide the sample weights by the number of boundaries. Because I am using multiple years of data so as to increase statistical power, one might be concerned about the need for year fixed effects. This leads to essentially identical estimates.

effect). I cluster my standard errors at the occupation level.[12] Importantly, $\beta_2$ captures spillover effects across occupations net of any earnings effects within occupations, therefore, characterizing total spillover effects within and across the "licensed" vs "unlicensed" status groups, which has not been captured in the prior literature. $X$ is a set of individual controls for sex, race/ethnicity, age, and age squared. I omit other controls which may be directly affected by licensing such as education to avoid collider bias. If licensing rules require an additional year of schooling, for example, controlling for that additional year of schooling will bias the estimate of the effects of the rule and the effects will load on education, but the reason for the education *is* the rule. This is particularly important because I am examining effects within occupations. For spillovers, if education requirements in a person's outside options increase with licensing, those with less education that are deterred by the costs of education may disproportionately filter into other occupations in the skill cluster. Given that this is a direct effect of licensing, controlling for education would strip away a mechanism through which licensing might affect earnings. I show evidence in Section 4 that spillovers do partially operate through education composition, and the average age in an occupation is not affected in a statistically significant way.[13]

### 3.2.1. Identifying Variation and Assumptions

The econometric challenge of identifying the causal effects of occupational licensing on earnings and employment using observational data is two-fold: first, state selection into licensing occupations may be related to other underlying economic factors in a state that also influence earnings such as statewide industry agglomeration and industry lobbying; second, licensing statutes may also be correlated with other state policies that influence earnings such as minimum wages, collective bargaining and unionization regulations, or tax policy.

The boundary discontinuity design overcomes this obstacle by comparing workers in the same occupation on two sides of the same state border where the state boundary creates differences in their occupational licensing status and the status of other occupations in their skill cluster. Because of the state fixed effects, identifying variation for each occupation comes from having *multiple* borders in each state that differ in their occupational licensing rules

[12]This clustering level makes my standard errors slightly more conservative than the prior literature (e.g. Kleiner and Soltas (2019)). An alternative approach to the boundary discontinuity design would be to instead add interacted fixed effects at the occupation-by-boundary level. This specification does not take into account other unobserved characteristics that are shared between all workers in the same state because is no flexible control for unobserved state-level variables. Nevertheless, the results are nearly identical (see Figure A2).

[13]Kleiner and Krueger (2013) include education in their regressions to control for selection for *individuals* stating that they have a license, much as the literature on the union wage premium does, because, as the authors state, instruments for individual licensure are rare. The CPS questions provide one such measure but were only implemented beginning in 2015. Controlling for education when including occupation fixed effects may exacerbate the possible collider bias and underestimate the total wage premium. Kleiner and Soltas (2019) notably do not control for education because they show that education is likely an intermediate outcome (or a "bad control") to wages affected by licensing rules. My results suggest the same for spillovers (see Panel B of Figure 7).

across each specific boundary. Because licensure is determined at the state level, not the local level, differences in licensure across each specific border are due to processes conditionally unrelated to other determinants of worker earnings—or in other words, for arbitrary reasons with regard to the economic conditions and individual characteristics at the boundary.

One way of conceptualizing this method is as an analog to the difference-in-discontinuities approach (Grembi et al., 2016). In the basic setup, there is a threshold that is exogenous to individual characteristics in both a treatment and control group. Across this threshold, there is a discontinuity for treated units that differs from the discontinuity for the control units. The difference in the discontinuity across the two groups is then attributable to the difference in treatment status. My boundary discontinuity design performs the same function with two key differences: treatment status is continuous, and there are hundreds of treatment/control comparisons: one for every state boundary and occupation. This design is the difference-in-discontinuity design broadened to include information from more spatial dimensions. Across multiple states and multiple borders within each state with multiple occupations, I compare the discontinuity in average earnings for workers as a function of the differences in licensing exposure across the border. Like the basic difference-in-discontinuities approach, the boundary discontinuity approach relies simply on the conditional orthogonality of the boundary with regard to local market and individual characteristics.

The fixed effects that create this conditional orthogonality in my models are of particular importance. The occupation fixed effects are what force the statistical comparison to be within occupations, and they control for systematic differences across occupations across all sample states. Importantly, the state fixed effects hold constant all shared or systematic attributes (even unobserved) of a worker's state that affect the distribution and dynamics of earnings and employment in all occupations and PUMAs in the state. These include regulatory conditions (e.g. minimum wage laws, education regulations, tax policy, overall state propensity to license, lobbying strength in particular industries, etc.) as well as statewide shared economic conditions (e.g. industrial composition, historical comparative advantage, etc.). The remaining earnings variation is residualized earnings variation. It is also key to remember that licensing intensity in my data is constructed at the state level, and occupation clusters cross industry lines.

The boundary fixed effects hold constant systematic differences across small geographic regions such as the spatial distribution of employment and labor demand. They also force (conditional on state fixed effects) the operating comparison in my regressions to be between workers within shared geographic regions after accounting for policies and economic conditions common across all PUMAs in a state. Importantly, the variation in earnings used in the border comparisons is residual variation after accounting for each of these three separate layers.[14] Table A2 presents correlations between several occupation and spatial variables before and after residualizing on these fixed effects. This is an analog to the correlation between the steepness

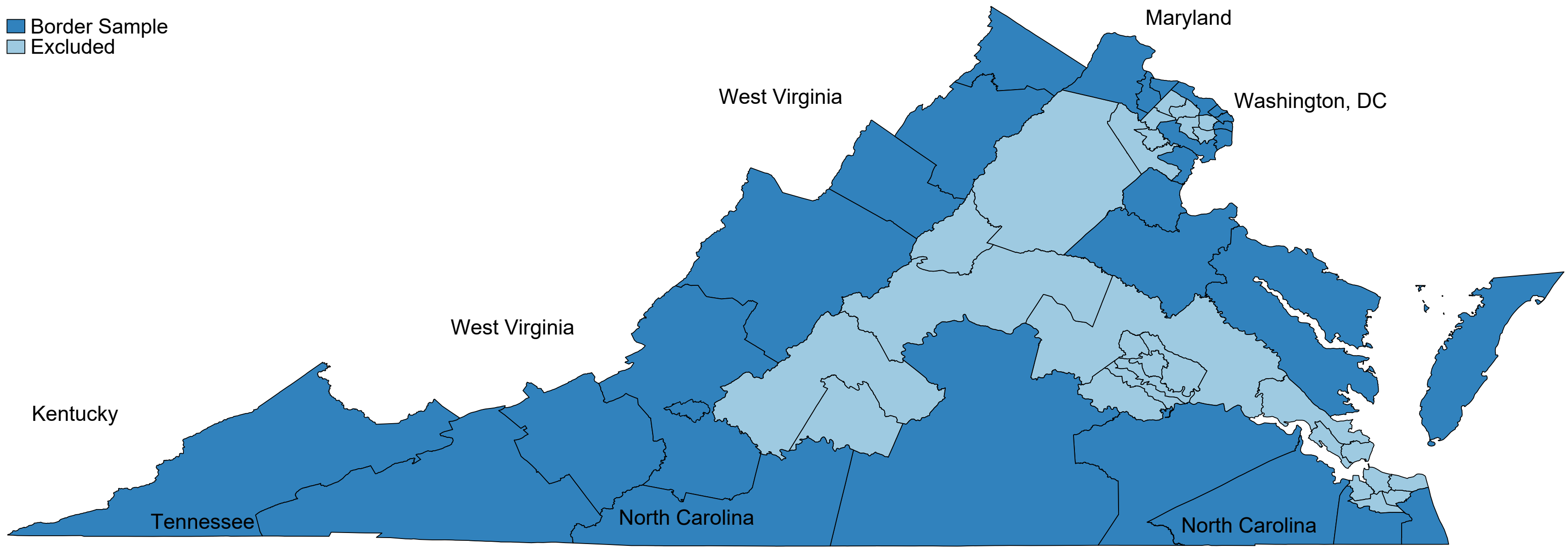
[14]Many studies use a boundary discontinuity approach, such as Black (1999), who first used the design to examine the effects of school quality on house prices, and some work in the occupational licensing literature (Blair and Chung, 2019).

of the cross-border gradient in exposure and the various characteristics of the PUMA, industry mix, and individual workers. After the fixed effects, there is no meaningful relationship between these various characteristics and treatment intensity.

As a concrete example to demonstrate the boundary comparisons, consider PUMAs in the state of Virginia in Figure 2. In my regression, the occupation fixed effects ensure that comparisons come from variation in licensing exposure and earnings within occupations. In other words, I am comparing, for example, carpenters to carpenters, where these may differ across state lines in licensing requirements to be a carpenter and the requirements for other occupations that use similar skills to a carpenter. In Figure 2, the states that border Virginia are written in the corresponding areas in the state map. The PUMAs in northern Virginia border the District of Columbia, Maryland, and West Virginia. PUMAs in the south border North Carolina, while the PUMA farthest to the southwest borders Kentucky and Tennessee. The state fixed effects in my regression take into account unobserved factors that affect carpenters in every PUMA in Virginia such as certification requirements, the strength of state lobbying efforts related to carpenters and construction, and other policies. The remaining variation in each outcome must then mechanically come at the sub-state level, either at the regional or individual level.

The boundary fixed effects take into account regional variation in economic conditions related to earnings and employment for all workers, including carpenters. For example, PUMAs that border Maryland may have different economic conditions than PUMAs that border North Carolina. These fixed effects ensure that carpenters on the Virginia side are compared to carpenters on the other side of the specific border, e.g. that carpenters in southern Virginia are compared to carpenters in northern North Carolina, while carpenters in northern Virginia are compared to carpenters in western Maryland. All pairwise combinations of carpenters across shared borders contribute to identification, and, conditional on the various fixed effects, cross-border differences in licensure across each specific border occur for arbitrary reasons orthogonal to unobserved determinants of worker earnings and employment. The average treatment effect on earnings for a carpenter in Virginia is a weighted average of the difference in earnings between carpenters as a function of their differences in licensing exposure across each border pair. In my full sample, this process continues across every contiguous US state.

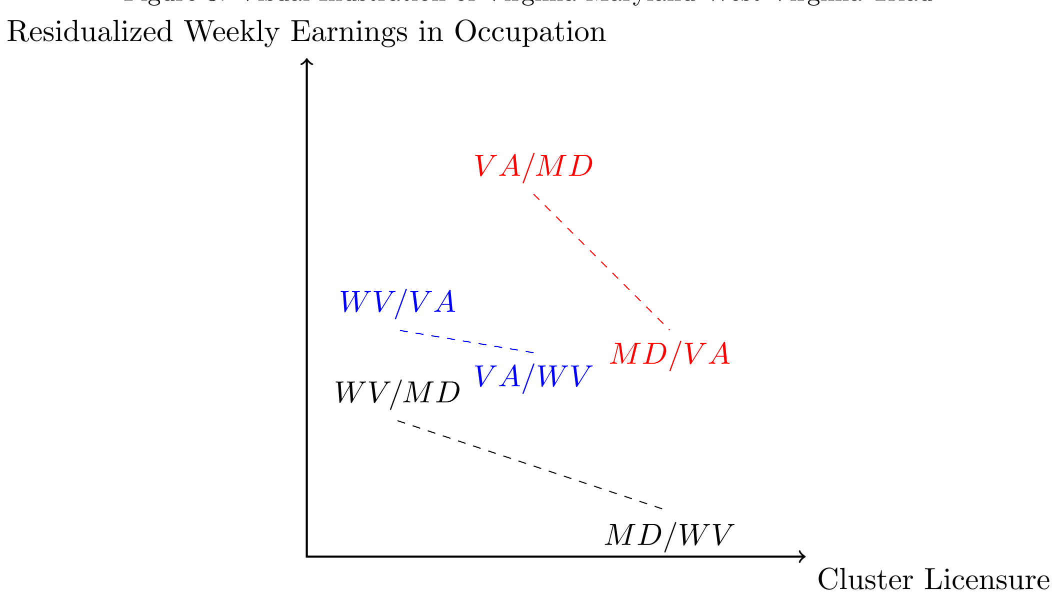
Figure 2: Boundary Discontinuity Design Example: Virginia




Source: Author's mapping of 2010 ACS Public Use Microdata Areas in the state of Virginia.
Notes: My border fixed effects in Equation 1 ensure that workers on the Virginia side are compared to workers in the same occupation on the other side of the specific border, i.e. that workers in southern Virginia are compared to workers in the same occupation in northern North Carolina, while workers in northern Virginia are compared to workers in the same occupation in Maryland, West Virginia, and DC. All pairwise combinations of share borders contribute to identification: Virginia-Tennessee, Virginia-Kentucky, Virginia-North Carolina, Virginia-Maryland, and Virginia-DC. The average treatment effect for Virginia for earnings is a weighted average of the difference in earnings between workers in the same occupation conditional on their differences in licensing exposure in southern Virginia vs northern North Carolina, eastern Kentucky and northeastern Tennessee vs southwestern Virginia, and northern Virginia vs eastern West Virginia, western Maryland, and the District of Columbia.

To make this final point clearer, consider Figure 3 and suppose that there are six hypothetical carpenters in the Virginia, Maryland, and West Virginia areas. Carpenter 1 lives in Northern Virginia on the Maryland border ("VA/MD"). Carpenter 2 lives in Northern Virginia near the border with West Virginia ("VA/WV"). Carpenter 3 lives in Maryland on the Virginia border ("MD/VA"). Carpenter 4 lives in Maryland near the West Virginia border ("MD/WV"). Carpenter 5 lives in West Virginia near the Virginia border ("WV/VA"), and Carpenter 6 lives in West Virginia near the Maryland border ("WV/MD"). Through the state fixed effect, any state-level wage regulations, lobbying, or industrial policies that affect Carpenter 1 will also affect Carpenter 2, including occupational licensing rules. Similarly, any unobserved regulations or rules that affect Carpenter 3 will also affect Carpenter 4. The same is true of Carpenters 5 and 6. After the state fixed effect, the remaining variation in earnings between Carpenter 1 and 2 (as well as 3 and 4, 5 and 6) is residual earnings unexplained by common state characteristics. These residualized earnings are the y-axis values in Figure 3. Note that each point with the same first state (e.g. "VA" in the "VA/MD" and "VA/WV" points) has the same x-axis value because each state-occupation cell has the same licensing intensity.

Figure 3: Visual Illustration of Virginia-Maryland-West Virginia Triad




Source: Visual illustration of the average treatment effect for licensure spillover in hypothetical data.
Notes: The y-axis represents residual variation in log weekly earnings after state and occupation fixed effects. The first state abbreviation denotes the person's state of residence, and the full combination denotes the border. For example, "VA/MD" corresponds to someone living in Virginia on the Maryland border, while "MD/VA" corresponds to someone living in Maryland on the Virginia border.

The boundary fixed effect then imposes a comparison of this remaining variation of Carpenter 1 to 3 (the red line), Carpenter 2 to 5 (the blue line), and 4 to 6 (the black line) as a function of the differences in licensing rules for other (non-carpenter) occupations with skill requirements similar to carpenters (the spillover effect). The effect of interest in this example is the weighted average of the slopes of the red, blue, and black lines. In the canonical difference-in-discontinuities approach, the difference in the y-axis values for points sharing the same color reflects the "difference-in-discontinuities" coefficient if the x-axis difference is a binary treatment variable. In this setting, the difference in the x-axis values is a continuous scale of treatment intensity.[15]

One may be concerned that states that have different rules regarding the licensing of carpenters (or other occupations close in skills) at the state level may also have other characteristics that affect carpenters through other channels (for example, lobbying). However, because of the construction of this boundary discontinuity design, my estimates will capture unbiased measures of the causal effect as long as there are no statewide factors related to licensure and earnings that systematically vary across *individual* state boundary pairs. In other words, for an unobserved state-level factor to explain my results, it would have to systematically and disproportionately affect earnings for carpenters in the regions in the state that *happen* to border states with lower licensing exposure for carpenters. In the logic of Figure 3, that factor would, for example, not move licensing exposure levels for "VA/WV" or "VA/MD" along the x-axis (because these are fixed at the state level), but would have to alter the y-axis values for "VA/MD" but not for "VA/WV." A similar factor would have to systematically affect the y-axis of "WV/VA" differently than the y-axis value of "WV/MD" and so forth across multiple state border pairs in the sample. When considering the full sample of states, that pattern would have to extend across all state borders. Such a specific pattern of spatially cascading factors across state border pairs across the United States is unlikely.

In Section 4.6, I explore a large battery of robustness tests and specifications that vary the assumptions of the model in order to rule out alternative explanations. These include using all states as implicit controls, a placebo exercise, the addition of PUMA fixed effects, the sequential elimination of clusters in estimation, using different data sources for licensing data and/or outcomes of interest, and leveraging variation in licensing rules across states over time for identification. All my various tests support the results of my main approach. These exercises also place further restrictions on what any unobserved factor would have to entail in order to be driving or biasing my results.

Taken together, these exercises suggest that, in addition to the specific pattern of cascading border relationships mentioned above, for an unobserved factor to drive my results, such a factor must differentially affect earnings and employment in specific boundary areas in a knife-edge case where such effects are equivalent when considering all states as controls. It must also hold across the types of occupations being considered (as I sequentially eliminate skill clusters from the sample). It would have to hold even when controlling for conditions in the local labor market (as I include PUMA fixed effects). It must be correlated with

[15]This is where my approach diverges from the simple difference-in-discontinuities approach because treatment is continuous rather than binary. The "zero" treatment condition is from a fitted line extended back to the y-axis intercept rather than being an average of the simple binary differences, so the treatment effect is, by definition, zero at this value.

cluster-specific propensities to license in a way that is *specifically* correlated with skill cluster structure from O*NET (when I randomly assign placebo clusters). It must be correlated with licensing measures from a separate database (when I use the Northwestern Licensing Database in my boundary discontinuity design). Tax policy, minimum wages, industrial relations and unionization policies, industry lobbying efforts, etc., do not fit that description. Indeed, it is difficult to imagine such a policy, actor, or economic condition that would generate this relationship if not the licensing environment itself.[16]

### 3.2.2. Heterogeneous Effects, Composition, and Employment

To examine heterogeneous treatment effects, I interact my measures of own-occupation licensure and cluster licensure outside the focal occupation with demographic indicators for sex, racial/ethnic groups, and nativity (native- vs foreign-born) in separate models. This allows the effects of licensure to vary across groups.

To measure composition effects, I calculate the share of each PUMA-occupation cell belonging to each demographic group as outcome variables. These outcomes include the share that is female, non-Hispanic white, non-Hispanic black, and Hispanic as well as the share that is foreign-born. I also calculate each cell's average age and average years of education. Finally, I estimate the share of each cell having a set of broad education credentials (i.e. Associate's, Bachelor's, Master's degrees, and a PhD/professional degree) as well as the share that is age 18-25 to see where the effects may be most concentrated. These effects are important to measure, as other types of labor regulations like the minimum wage do appear to shift up the education and age distribution of minimum wage workers (Clemens et al., 2021).

The estimation equation is:

$$y_{opcms} = \beta_0 + \beta_1 LicensedShare_{os} + \beta_2 LicensedShare_{cs}^{-o} + \delta_o + \gamma_s + \tau_m + \varepsilon_{opcms} \qquad (2)$$

, where $y_{opcms}$ represents the share of occupation $o$ in PUMA $p$ belonging to each of the groups listed above. These models communicate the change in the conditional probability that a worker in an occupation is a member of a particular demographic group as occupational licensing rules change and therefore capture the compositional sorting effects of these regulations. The coefficient on $\beta_2$ tells us the average change in the share of workers in occupation o that is, for example, a woman as a result of occupation o's cluster becoming more licensed.

To examine the overall employment effects of licenses and licensing spillovers, I estimate employment in the occupation-PUMA cell (occupation $o$ in PUMA $p$) as the outcome variable assuming a Poisson distribution, with mean:

$$\lambda_{opcms} = exp(\beta_0 + \beta_1 LicensedShare_{os} + \beta_2 LicensedShare_{cs}^{-o} + \delta_o + \gamma_s + \tau_m) \qquad (3)$$

[16]Furthermore, such a factor would also have to affect earnings and employment in a similar pattern when considering (albeit imperfect) measures of licensure over time (when I estimate panel models in the CPS using the Northwestern Licensing Database in Appendix C).

This is solved via maximum likelihood from the log-likelihood function:[17]

$$\ell(\beta, \delta, \gamma, \tau)] = \sum[EMP * Log(\lambda) - \lambda] \quad (4)$$

The own-occupation effect ($\beta_1$) measures the effects of licensure on employment in that occupation itself, while $\beta_2$ captures the employment spillovers. A pure labor supply explanation for the earnings effects I find would predict a negative $\beta_1$ and a positive $\beta_2$ coefficient as licensing pushes workers into other occupations using similar skills. A negative $\beta_2$ spillover coefficient on employment in the focal occupation is suggestive of another mechanism if earnings effects are also negative in my individual models in Equation 1, which I explore in Section 5.

# 4. Results

## 4.1. Earnings Premium and Spillovers

I first present the results for the overall earnings effects of widespread occupational licensure. Figure 4 plots the coefficients and confidence intervals for occupation spillovers for 4 to 20 clusters. Spillover coefficients are for a 10 percentage point increase in licensure in one's own cluster, which is approximately 1 standard deviation (11 percentage points). Two standard deviations around the mean of approximately 21% exposure would reach, at most, 40-43%. The figure indicates that having one's own occupation licensed leads to an earnings premium of approximately 8%, an effect that is remarkably consistent with estimates that leverage cross-state policy variation over time in certain occupations (Carollo, 2020b; Pizzola and Tabarrok, 2017) or that incorporate occupation fixed effects (e.g. Kleiner and Krueger (2013) and Gittleman et al. (2018), whose estimates are 12% and 5.7-8%, respectively, with these fixed effects). On the other hand, increasing licensing rates in all other occupations in one's own skill cluster by 10 percentage points reduces weekly earnings in the focal occupation by 1.6–2.3% on average. The confidence intervals rule out average effects smaller than -0.25% and effects larger than -3%. Given that the validated optimum number of clusters is somewhere in the 13–20 range, the effects are concentrated around 1.6-2.3%.[18] To ease interpretation, I present the rest of my estimates using 20 skill clusters, the most conservative set of estimates.

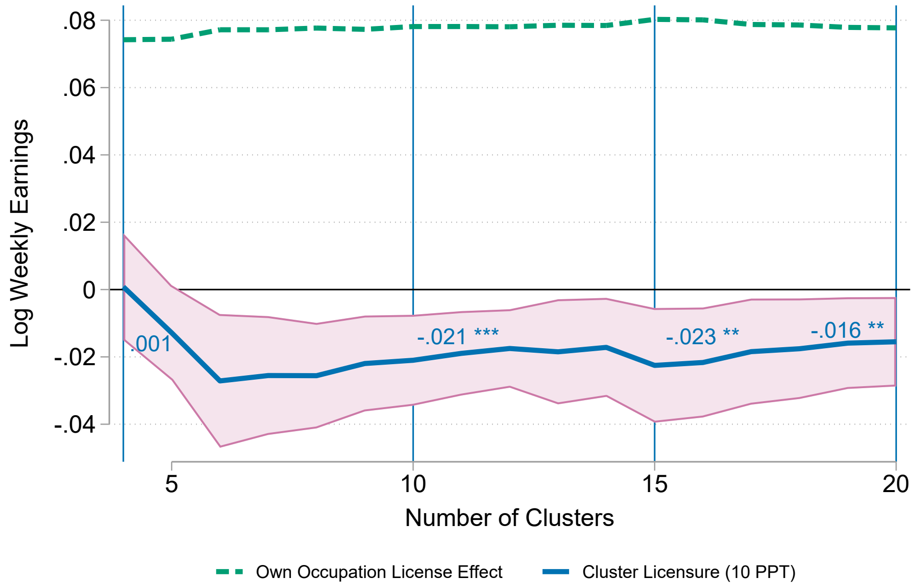
I find substantial heterogeneity in the own-occupation and spillover effects of licensure across gender as well as race/ethnicity and nativity as detailed in Figure 5.[19] Panel A shows the effects of licensure in one's own occupation, while Panel B shows the spillover effects of

[17] In practice, the estimation of multiple fixed effects requires the use of Iteratively Reweighted Least Squares and the Frisch-Waugh-Lovell theorem to remove the fixed effects. See Correia et al. (2020). The results of this Poisson model are similar to a more straightforward log-linear specification.

[18] For ease of reading, I plot the point estimates of "own-occupation" effects without standard errors because these are not necessarily the estimates of interest, but are instructive for the validity of comparing my point estimates to other studies. The 95% confidence intervals cover the interval [0.0392, 0.1163]. Though the point estimates appear constant over the number of clusters, there are trivially small differences not captured visually. Cluster licensure is a leave-one-out measure and does not vary with own-occupation licensure, so we would not expect differences in the point estimates for the own-occupation effect except for random noise.

[19] For a table of these estimates, see Table A3. Subgroup estimates varying the number of clusters at 4, 10, 15, and 20 are in Table A4, and each follows a similar pattern to the overall estimates in Figure 4.

Figure 4: Coefficients of Log Weekly Earnings by Number of Clusters

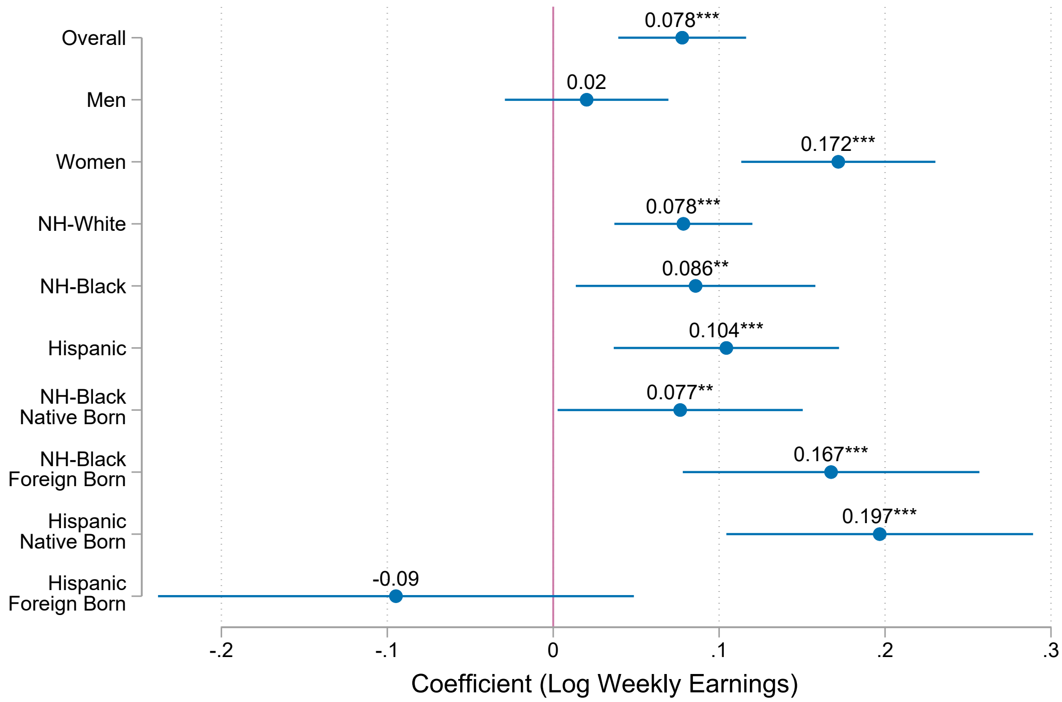



Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: *** p<0.01, ** p<0.05, * p<0.1. Coefficients are generated from the boundary discontinuity design detailed in Equation 1. Spillover coefficients are based on a 10 percentage point increase in licensure of an occupation's cluster outside their own occupation. Standard errors are clustered at the occupation level. 95% confidence intervals in red. Models include occupation, state, and boundary fixed effects and controls for race/ethnicity, sex, age, and age squared. Vertical bars and coefficients are for clusters at 4, 10, 15, and 20.

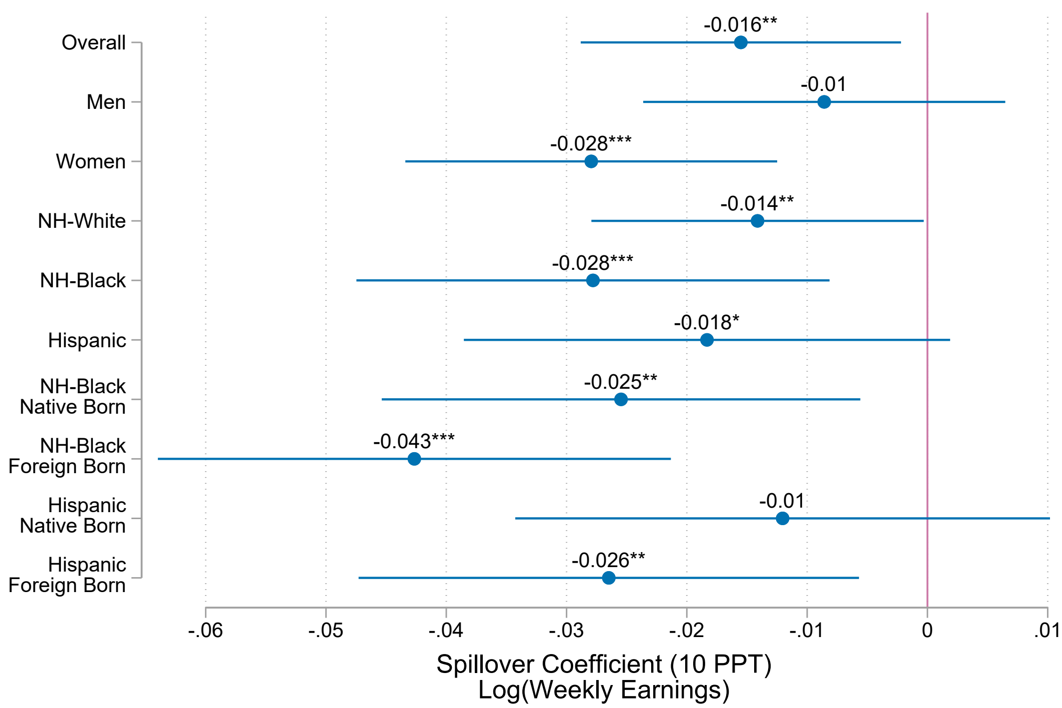
cluster licensure. While women in licensed occupations receive a larger earnings premium than men, they also experience a larger earnings spillover penalty. Women receive an earnings premium of 17% in licensed occupations relative to other women in the same occupation that are not licensed, but increasing skill cluster licensing requirements by 10 percentage points leads to a reduction in their earnings of approximately 2.8%. The same coefficient is less than 1% for men (the p-value of the difference is 0.02). Non-Hispanic black workers and Hispanic workers experience larger earnings spillovers than their non-Hispanic white counterparts. The point estimate for Hispanic workers is -1.8% for a ten percentage point increase in cluster licensure compared to -1.4% for non-Hispanic white workers, though the difference is not statistically different at conventional levels. Non-Hispanic black workers experience the largest penalty, with a point estimate of 2.8% (different from non-Hispanic whites at the 15% level). The large relative penalty for non-Hispanic black workers may be due to licensing requirements that prohibit those who have been convicted of a felony from obtaining a license (Blair and Chung, 2018, 2019). As a result, the set of occupations in which someone with a set of skills may work after conviction narrows. The returns to obtaining a license as an ability or past history signal may be higher in this case (Blair and Chung, 2018).

Figure 5: Coefficients of Log Weekly Earnings by Subgroup at 20 Clusters

Panel A: Own Earnings Effects




Panel B: Within-Cluster Spillover Effects




Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: *** p<0.01, ** p<0.05, * p<0.1. Coefficients are generated from the boundary discontinuity design detailed in Equation 1 using 20 skill clusters. Dots represent point estimates, and the bands represent the 95% confidence interval clustered at the occupation level. Spillover coefficients are based on a 10 percentage point increase in licensure of an occupation's cluster outside their own occupation for 20 total defined clusters. Models include occupation, state, and boundary fixed effects and controls for race/ethnicity, sex, age, and age squared.

Most of the negative earnings spillover effect on Hispanic workers is driven by foreign-born Hispanic workers. The estimates of the own-occupation effect are noisy for foreign-born Hispanic workers, perhaps due to statistical power issues: a small share of Hispanic immigrants can obtain a license when compared to other immigrant groups, be it for education, language, or legal status reasons. Spillover effects for a ten percentage point increase in cluster licensure are 2.6% compared to just over 1% for native-born Hispanic workers (the pairwise difference has a p-value of 0.023). Given the young age, relatively low educational attainment, and migrant status of foreign-born Hispanic workers, other outside options for foreign-born Hispanic workers may be lower than their native-born counterparts. In particular, citizenship or permanent residency requirements for many licenses may preclude many foreign-born Hispanic workers from entering a variety of occupations, which strongly limits their choice set.

Given the presence of possible statistical or taste-based discrimination against non-Hispanic black workers (as in the occupational crowding literature) as well as the additional imposition of citizenship or residency requirements for foreign-born workers, I expect spillover effects to be the largest for foreign-born black workers. The estimates show that this is, indeed, the case. Native-born black workers experience spillovers of 2.5% with a ten percentage point increase in cluster licensure, while foreign-born black workers experience spillover effects of 4.3%. This is significantly different from the effect on native-born whites, whose effect is nearly identical to the overall non-Hispanic white effect, at the 1% level.

These heterogeneous results suggest that there are substantial earnings spillovers of widespread occupational licensing within a worker's skill cluster and that the effects are highly concentrated among those that already are disproportionately lower-income and that are less likely to be able to absorb the costs of licensing requirements.

#### 4.1.1. Total Cross-Status Effects

When considering the effects of licensure, the prior literature sometimes estimates the effect of having a license on earnings without controlling for occupation fixed effects. These coefficients may include information about the possible wage premium for having a license and some information about the spillover effects of licensure to quantify the total "cross-status" gap, or in other words, the gap between workers with a status of "licensed" in comparison to a status of "unlicensed." However, these comparisons may be confounded by occupational sorting and positive selection on licensing status across occupations. To solve this selection issue, other studies include occupation fixed effects. But this limits our ability to separate the wage premium effects of licensing from cross-occupation spillover effects because only within-occupation variation in own-license status is considered, and other variation is accounted for in the model residual.

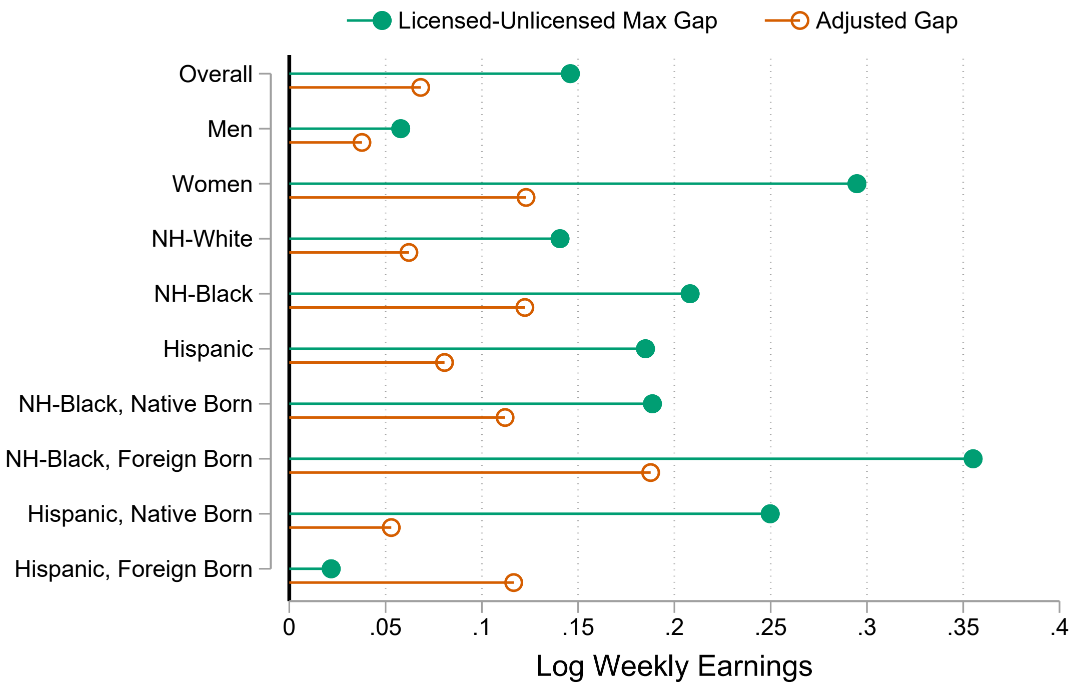
My clustering framework allows me to include occupation fixed effects while quantifying the gaps between the licensed and unlicensed *net* of spillovers. To do this, I estimate Equation 1 and generate predictions from the model. I compare mean predicted earnings in three conditions: (1) setting own-license status to 100% and spillover exposure to zero; (2) setting own-license status to 100% and spillover exposure to 44%; and (3) setting own-license status to 0% and spillover exposure to 44%, which is about two standard deviations above the mean near the 99th percentile. The gap between (1) and (3) tells us the maximum earnings gap would be between the licensed and unlicensed status labels incorporating spillovers on the unlicensed while shielding the licensed from spillovers. The gap between (2) and (3) tells us the earnings premium between the licensed and unlicensed statuses at maximum exposure to spillovers, or in other words, a net-of-spillover gap when allowing for spillovers to affect both the licensed and unlicensed. When considering gaps by demographic group, I perform this exercise when including interactions for each demographic group as in Figure 5. These gaps account for both the different marginal effects of licensure and spillovers in Figure 5 and also differences in the *level* of exposure across groups.

The results of this exercise are in Figure 6. Were those with a "licensed" status shielded from spillovers, the average earnings premium for holding a license would be approximately 15% (the maximum gap). When adjusting exposure to spillovers in both statuses to a common exposure rate of 44% (about two standard deviations above the mean and equal to the 99th percentile), the average overall gap shrinks to approximately 7%. This overall gap is primarily driven by women, whose maximum gap is nearly 30 log points (35%). After adjusting for spillovers, the earnings gap for women falls to 12%. In a similar vein, the maximum gap for black and Hispanic workers is larger than it is for non-Hispanic white workers. The difference between maximum and adjusted gaps is also larger for these groups, meaning that spillovers have disproportionate total effects on the earnings of *licensed* black and Hispanic workers compared to licensed non-Hispanic white workers and on licensed women compared to licensed men. The group with the largest difference between the maximum and adjusted gaps is native-born Hispanic workers, while the maximum returns to holding a license are the largest for foreign-born black workers. These various gaps are consistent with the differential returns to license-holding for black workers and women (Blair and Chung, 2018) and with the differential marginal spillover effects on these groups.

Overall, the maximum gaps in this exercise of 15-35% are largely consistent with the estimates of the prior literature for the total cross-status gap (without occupation fixed effects), which are approximately 22% (Gittleman et al., 2018), 20% for those above the median in Canada (Zhang and Gunderson, 2020), and 30% (Kleiner and Krueger, 2013). A significant share of the observed cross-status gap is attributable to spillovers.

Together, these results make it clear that licensing substantially increases the earnings gap between licensed and unlicensed workers, which mechanically increases inequality within groups and likely across groups. However, the majority of workers in the United States do not hold licenses, meaning they are experiencing spillovers without compensation for having a license themselves. I further explore a counterfactual world in which licensing was entirely eliminated for all workers to characterize the total mean earnings effect and conditional distributional

Figure 6: Maximum and Adjusted Gaps between the Licensed and Unlicensed




Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Values are generated from the predictions from the boundary discontinuity design detailed in Equation 1 in three conditions: (1) own-license status = 100% and spillover exposure = 0; (2) own-license status = 100% and spillover exposure = 44%; and (3) own-license status = 0% and spillover exposure = 44% (two standard deviations above the mean). The gap between (1) - (3) is the maximum earnings gap, while (2) - (3) is the adjusted gap when allowing for spillovers to affect both licensed and unlicensed statuses. Estimates for group-specific gaps account for interactions in the base model.

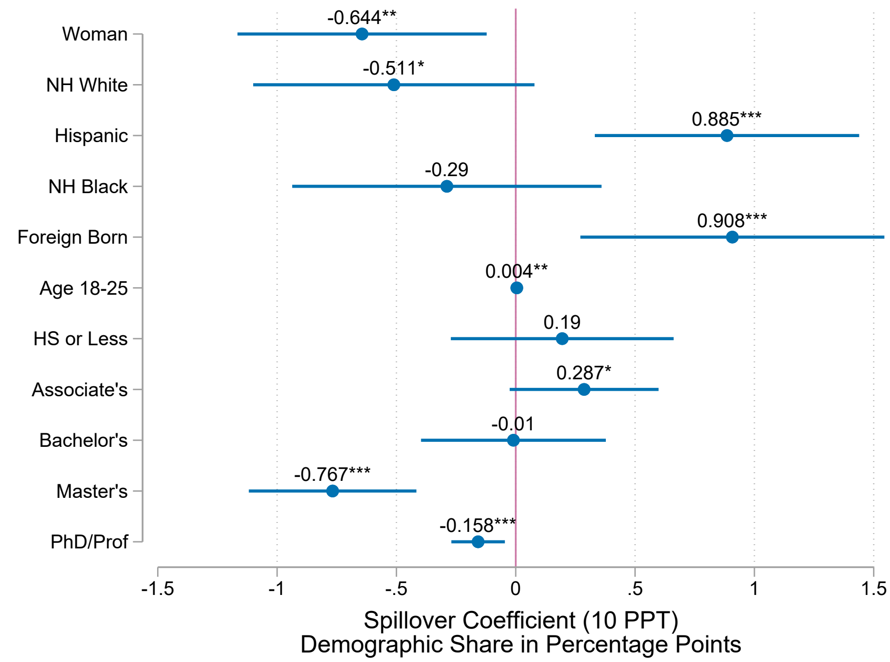
effects of licensing in Section 4.5.

## 4.2. Composition Effects

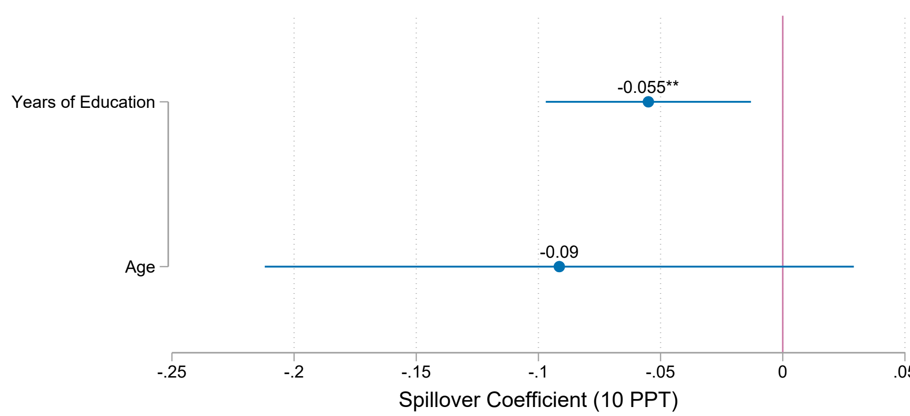
One mechanism underlying the earnings effects above is that licensing spillovers may shift the distribution of workers within occupations in terms of educational attainment, sex, nativity, or race/ethnicity. Figure 7 shows the results of estimating shares of each occupation-PUMA cell in different classifications of sex, educational categories, race/ethnicity groups, and nativity. Panel A shows that as other occupations in the cluster become more licensed by 10 percentage points, the likelihood that a worker in the focal occupation is a woman declines by nearly two-thirds of one percentage point. The probability of having a Master's degree declines by approximately three-quarters of a percentage point. Workers in the focal occupation are 0.89 and 0.91 percentage points more likely to be Hispanic or foreign-born, respectively. In Panel B, in terms of years of education, the average effect across all education levels is a reduction of approximately 0.055 years. There is also a decline in the age of workers in the focal occupation of 0.09 years, though the effect is not statistically significant at conventional levels. This exercise provides evidence that excluding education from the list of controls is appropriate (see Section 3.2). Given that sorting on education is a direct effect of licensing, controlling for it would cut out the portion of the total effect operating through this channel.

Figure 7: Composition Effects of Licensing Spillovers, 20 Clusters

Panel A: Demographic Groups




Panel B: Age and Education (Continuous)




Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: *** p<0.01, ** p<0.05, * p<0.1. Coefficients are generated from the boundary discontinuity design detailed in Equation 2. Dots represent point estimates, and the bands represent the 95% confidence interval clustered at the occupation level. Spillover coefficients are based on a 10 percentage point increase in licensure of an occupation's cluster outside their own occupation. Models include occupation, state, and boundary fixed effects.

These results indicate that as other occupations in the skill cluster become more licensed, there is not a large influx of those with lower levels of education (e.g. high school graduates without a college degree) shifting across occupations. However, there is a negative effect on the education distribution on average. In general, the larger shifts are in the gender and race/ethnicity composition of the focal occupation. Widespread licensing appears to shift some men (women) out of (into) licensed occupations as the share of women in the focal occupation shifts downward. Hispanic workers and foreign-born workers filter out of licensed occupations in the skill cluster.

In a Mincer-style earnings regression on my sample and conditioning on occupation, state, and boundary, an additional year of education yields an increase in earnings of approximately 5.22%, while an additional year of experience increases earnings by 1.58% at the sample mean. Given the coefficients in Figure 7 of -0.055 for years of education, sorting over education may account for 0.00287 log points (0.0522*0.055) of the 0.016 log point total spillover effect, or about 19% of the effect. Despite the age composition effect not being statistically significant, if we add this effect linearly to the education effect, sorting on these two dimensions collectively accounts for as much as 0.0043 log points ($\approx$ 28%) of the total spillover effect. In addition to sorting on observed characteristics, there may also be sorting on unobservables, which I explore below.

To explore how important these composition changes are collectively for my observed earnings effects, I follow the procedure in Oster (2019). Table A5 shows the effects when excluding all demographic information from my estimation models and the change in the coefficient and $R^2$ after including controls for gender, age, education (linear in years), race, and nativity.[20] The coefficients change by 0.0037 log points when these demographic controls (or mediators) are included. In the Oster (2019) approach, the identified set at $\delta = 1$ and a maximum R-squared of 1.3 times the controlled R-squared tells what the effects would be if sorting on unobservables was as important as sorting on observables; this set excludes zero. Furthermore, the unobservables would have to be 2.6 times more important than the observables to fully explain the effects I find. Assuming $\delta = 1$, the combined effect of sorting on observed and unobserved characteristics explains 52% of the total, uncontrolled negative earnings spillover effects.

This leads to three conclusions: first, sorting from licensing on education may explain a portion of the observed spillover effect I find, but not the majority of the effect; second, for unobserved factors such as unmeasured differences in productivity to explain these results, such factors would have to be more than 2.5 times more important than observed characteristics, which falls far beyond the conventional threshold of $\delta = 1$; third, assuming sorting dynamics on unobservables that match the observables (i.e. $\delta = 1$), the sorting effects of licensing may explain about half of the total average earnings spillover effect.

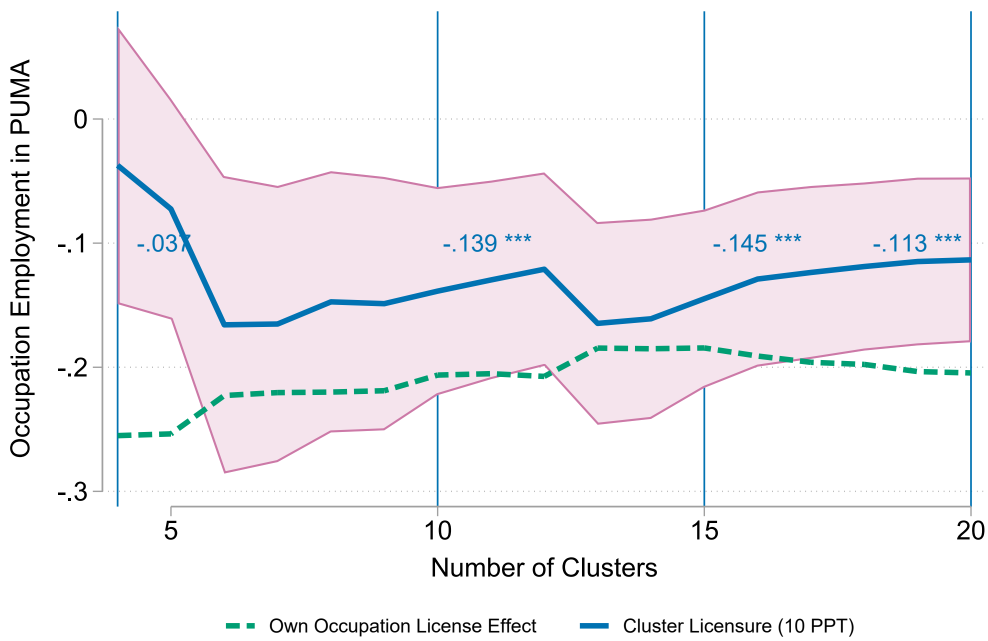
[20] I implicitly treat the fixed effects for state, occupation, and boundary as nuisance parameters, meaning I only consider changes in R-squared and coefficients after partialling out these factors.

### 4.3. Employment

To understand the other mechanisms underlying the earnings effects I observe, I estimate a boundary discontinuity model of employment (Equations 3 and 4) where observations are at the occupation-PUMA cell. These estimates yield employment pseudo-elasticities for the own-occupation effect of licensing and the cross-occupation spillover effect within skill clusters.

Figure 8 indicates that the employment elasticity for one's own licensing status is approximately -0.2, a finding consistent with other work (e.g. Blair and Chung (2019); Kleiner and Soltas (2019)). As licensure increases across the cluster by 10 percentage points outside the focal occupation, overall employment in the focal occupation *falls* by approximately 11%. For interpretation, for the licensing effect of cluster licensure to match the effect within one's own occupation, cluster-wide licensing exposure would have to increase by approximately two standard deviations. Shifting cluster-wide licensure to this degree represents an extreme change in the policy regime compared to changing the licensing status of a single occupation. In other words, it takes a larger relative change in state policy in order to reduce employment in the focal occupation through other channels to the same degree as changing the licensing status of the focal occupation itself.

Figure 8: Employment Effects of Licensing Spillovers by Number of Clusters




Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: *** p<0.01, ** p<0.05, * p<0.1. Coefficients are generated from the boundary discontinuity design detailed in Equations 3 and 4 with data in PUMA-occupation cells. Spillover coefficients are based on a 10 percentage point increase in licensure of an occupation's cluster outside their own occupation. Standard errors are clustered at the occupation level. 95% confidence intervals in red. Models include occupation, state, and boundary fixed effects. Vertical bars and coefficients are for clusters at 4, 10, 15, and 20.

Taken together, these results indicate that widespread licensing in a skill cluster leads to negative employment effects in the focal occupation. This observation is inconsistent with the idea that licensing simply diverts workers into the remaining occupations in the skill cluster. In Section 5, I discuss various mechanisms through which this negative employment effect may occur in conjunction with the negative earnings effects, including firm location choice, skill-based and industry labor demand, labor complementarities, and firm monopsony behavior.

### 4.4. Effects Across Sectors

A relevant consideration for understanding the mechanisms behind the heterogeneous earnings effects and the employment effects in this analysis is the tradability of the sector in which the worker is employed. Here, I examine the effects of licensing and licensing spillovers when considering the tradable goods sector (i.e. agriculture, manufacturing, and mining) and the non-tradable sector (all others).

This is interesting for several reasons. First, Burstein et al. (2020) suggest that labor substitution across workers (in that case, natives and immigrants) is more pronounced in the non-tradable sector than in the tradable sector; in other words, tradable sector labor is somewhat more protected from the dynamics of labor supply changes. In the context of licensing, this may imply that the effects of licensing spillovers could be less pronounced in the tradable goods sector. Second, the literature on "crowding" suggests that pushing certain workers (e.g. black workers) into service jobs and out of "craft" jobs in the tradable sector may be an important consequence of labor market discrimination and occupational segregation (Bergmann, 1971, 1974). This type of dynamic may partially explain the heterogeneous treatment effects over race and ethnicity. Third, private-sector unionization rates in the United States are higher in the tradable goods industries than in non-tradables, particularly compared to most services (US Bureau of Labor Statistics, 2021). These institutions may shield workers from the negative effects of monopsony power if that is a channel through which these effects operate (Dodini et al., 2021). Similarly, wage compression from unionization may lead to a smaller wage premium for having a license and mitigate spillovers. Alternatively, unionization may be a substitute for the job protections offered by occupational licensing as evidenced by the fact that the two have moved in the exact opposite direction over the last 40 years (Kleiner and Krueger, 2013). Fourth, men in my sample are far more likely to work in the tradable goods sector. If there are differences across sectors in the effects of licensing, this may explain the heterogeneous treatment effects I find across gender.

Table 2 shows the results of estimating Equation 1 for the tradable and non-tradable sectors. Though the own-occupation effect of licensing is smaller in the tradable sector, the difference between the two is not statistically significant. The main difference is that there is not a large or notable negative earnings spillover from licensing exposure within skill clusters in the tradable sector. Though the coefficient is positive, the standard errors mean that the model cannot rule out negative effects of up to approximately -1.8 percent. The difference between the two model estimates for the spillover effect has a p-value of 0.0577. These results provide some context for the heterogeneous treatment effects across gender, race/ethnicity, and nativity and provide suggestive evidence of the mechanisms through which the spillovers operate.

Table 2: Licensing Exposure Effects on Earnings by Sector

| VARIABLES | (1) Non-Tradable Industries | (2) Tradable Industries |
|---|---|---|
| Own Occupation Effect | 0.0870*** | 0.0507* |
| | (0.0222) | (0.0300) |
| Cluster Spillovers (10 PPT) | -0.0194*** | 0.00422 |
| | (0.00641) | (0.0110) |
| Observations | 1,459,416 | 305,864 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: Author's calculations of ACS, CPS, and O*NET data.
Notes: Estimates correspond to Equation 1 estimated in tradable sectors (agriculture, manufacturing, extraction) vs others. Standard errors are clustered at the occupation level. Estimates include fixed effects for occupation, state, and boundary.

## 4.5. Distributional Effects

My main results suggest that occupational licensing regulations have negative earnings spillovers for workers that use similar skills and that these effects are concentrated among those already likely to be lower-income workers. Because nearly 80% of workers in the sample are not required to have a license, exposure to spillovers will be more widespread than wage premiums for having a license, which has significant implications for the distribution of incomes.

To contextualize these effects, I present graphical evidence of the counterfactual kernel density distribution of predicted weekly earnings in my sample if licensing were set to zero for all workers, both for their own occupation and others in their skill clusters. Specifically, I estimate Equation 1 for raw earnings and predict individual earnings based on this model. Then, setting licensing for one's own occupation and cluster to zero, I use the same model coefficients to predict individual earnings. I then present the kernel density distributions of these two different predictions. The second prediction answers the question, "Given the earnings effects from licensure measured in the model, what would individual earnings be if workers had no licenses in their own occupation and no licenses in their cluster?" I also perform the same exercise for the distributions of race/ethnicity-specific earnings. I do this by estimating the version of my model with interactions for race/ethnicity and using the predictions from that model when presenting race-specific counterfactual earnings distributions.

The various fixed effects in the model remove variation over geographic space through the border fixed effect, occupation through the occupation fixed effect, and states through the state fixed effect, so the distributions I measure are conditional distributions. This partially explains the relatively uneven densities, which are not present in the unconditional distribution.[21] This counterfactual exercise assumes a constant distribution of employment, so it does not capture the effects of changes to employment across occupations or moving some workers out of employment altogether but holds constant the occupational and spatial distribution of employment in my sample. It also assumes constant effects (by subgroup and overall), so it will not account for dynamic self-selection on unobservables as in a Roy-type selection model that may shape general equilibrium effects. Nevertheless, the exercise is informative for the purposes of examining partial equilibrium effects on inequality.

The results are in Figure 9. Panel A shows that after eliminating licensing in the sample, predicted weekly earnings shift rightward across most of the conditional distribution. There is a general shift for earnings below $1,000 per week. More narrowly, there is a substantial change in the density moving earnings from approximately $500 per week to approximately $600 per week. There is also an increase from approximately $1,000 per week to over $1,100 per week and a decrease above $1,400 per week, suggesting a compression effect in the distribution of predicted earnings. Panels B and C show the distributions of predicted weekly earnings for non-Hispanic white workers in comparison to non-Hispanic black workers (Panel B) and Hispanic workers (Panel C). The results in Panel B suggest that if occupational licensing were eliminated, the conditional distribution of non-Hispanic black workers' earnings would nearly match the distribution of non-Hispanic white workers' earnings under the status quo due to the disproportionate spillover effects of licensing on black workers in comparison to white workers. There is a similar story among Hispanic workers, though there is still a gap in the predicted earnings for Hispanic workers in a "no licensing" world in comparison to white workers under the status quo.

To formalize and contextualize the comparison between the two distributions of predicted weekly earnings, I generate distributional statistics for the 90/10, 90/50, and 10/50 percentile ratios and the Gini coefficient for each distribution. Table 3, Panel A shows the comparison of these statistics between the distributions. There are significant changes in within-group inequality as a result of eliminating occupational licensing. The ratio of the 90th to 10th percentile of weekly earnings would fall by nearly 4%, and the ratio of the 90th to 50th percentiles would fall by 2.5% if licensing were eliminated in my sample. Much of the decline in the 90/10 ratio comes from increases in the 10/50 ratio, meaning that despite the median moving upward, the 10th percentile increases at a faster rate. Overall, the predicted Gini coefficient within the conditional distribution falls by nearly 7%.[22]

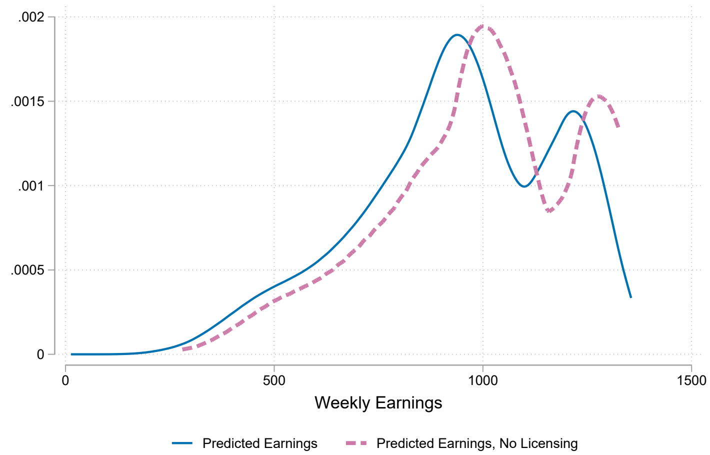
Given the larger own-occupation effect experienced by racial minorities and women in the sample, it is useful to compare the net effect of eliminating licensing on demographic group-specific mean predicted earnings. I do this in Panel B of Table 3. This exercise suggests that mean earnings for non-Hispanic whites and Hispanics in their occupations would rise

[21] This may also be an effect of the distribution of exposure having two noticeable peaks (see Figure A4)

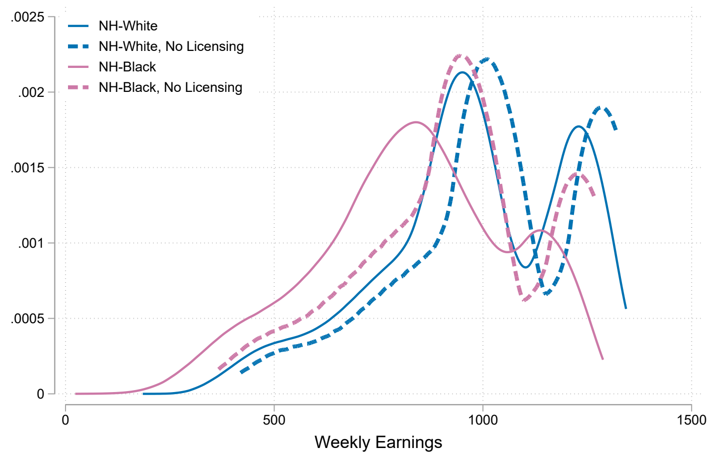
[22] As a point of comparison, unionization across all sectors in the United States is estimated to reduce inequality across the economy by approximately 10% (Card et al., 2020).

Figure 9: Kernel Density of the Counterfactual Distribution of Weekly Earnings

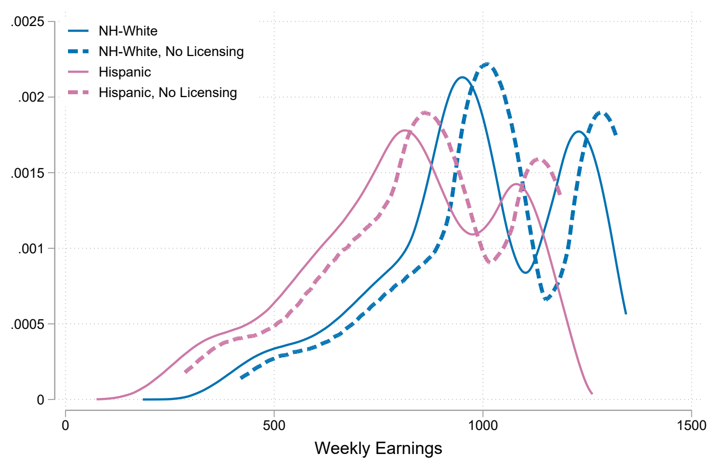
Panel A: Kernel Density of Predicted Earnings




Panel B: Non-Hispanic White vs Non-Hispanic Black




Panel C: Non-Hispanic White vs Hispanic




Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Values are generated by predictions from the boundary discontinuity design detailed in Equation 1 using 20 skill clusters. The model includes fixed effects for occupation, border pair, and state. Predicted earnings are for the status quo and for setting licensing rates to zero for one's own occupation and skill cluster.

Table 3: Distributional Statistics of Predicted Weekly Earnings with vs without Licensing

| Panel A: Overall Distributions | | | |
|---|---|---|---|
| | (1) Status Quo | (2) Without Licensing | (3) Percent Change |
| Ratio 90/10 | 2.062 | 1.982 | -3.88% |
| Ratio 90/50 | 1.312 | 1.279 | -2.52% |
| Ratio 10/50 | 0.636 | 0.645 | 1.42% |
| Ratio 75/25 | 1.438 | 1.394 | -3.06% |
| Gini Coefficient | 0.144 | 0.134 | -6.82% |

| Panel B: Mean Earnings by Race, Gender | | | |
|---|---|---|---|
| | (1) Status Quo (Mean) | (2) Without Licensing (Mean) | (3) Percent Change |
| Overall | 939 | 999 | 6.4% |
| NH White | 976 | 1031 | 5.6% |
| NH Black | 833 | 937 | 12.5% |
| Hispanic | 805 | 863 | 7.2% |
| Men | 1068 | 1074 | 0.6% |
| Women | 796 | 954 | 19.8% |

Source: Author's calculations of ACS, CPS, and O*NET data.
Notes: Distributional statistics are based on the predictions from Equation 1 as described in Section 4.5. Estimates include fixed effects for occupation, state, and boundary. In Panel A, the difference in means is significant at the 2% level or lower for all statistics except for Men, which is not statistically significant. All distributional statistics differences are significant at the 0.1% level using a bootstrap method.

by approximately 5.6% and 7.2%, respectively, in the absence of all licensing requirements. For non-Hispanic black workers, this increase is 12.5%. This implies that licensing has a net negative effect on mean earnings across racial groups but that this effect is most strongly negative for black workers. Similar patterns are evident for men and women. The net effect of eliminating licensing for women is an increase in weekly earnings of 19.8%, while the net effect is negligible and not statistically significant for men on average.

Using these models, we can compare how much money is gained through the own-license effect vs the loss from spillovers. If all occupations were licensed but spillovers were held at status quo levels, the overall gain in weekly earnings would be about \$27. However, eliminating licensure altogether leads to average gains of approximately \$60 per week, meaning that for every extra \$1 earned via the own-occupation licensing premium, approximately \$2.23 is lost via spillover effects.

Overall, this exercise implies that if a portion of existing licenses were eliminated, the distribution of earnings within occupations would be significantly higher, with many of those gains accruing to workers below the median, resulting in a decline in earnings inequality. At the mean, the average total effect of licensure on earnings is negative. This is even more pronounced for black workers. The results also imply that eliminating licenses for which there is not a national consensus on their usefulness (or where states differ in their licensing rules) would reduce earnings inequality in the unconditional distribution of earnings by pulling up the bottom of the distribution.

This finding that inequality increases with occupational licensing is consistent with other work. In particular, uneven returns to licenses across the education distribution or quantiles of the income distribution increase inequality across occupations (Zhang and Gunderson, 2020; Kleiner and Krueger, 2013; Gittleman et al., 2018). My analysis shows that a substantial share of the increase in inequality comes within occupations and is attributable to direct spillovers between occupations.

## 4.6. Robustness to Alternative Explanations

It is important to consider and rule out alternative explanations that may drive the relationships I have presented. Even though many policies change across state borders, the state fixed effects will account for any common factors across workers (regardless of location) in the same state that may relate to earnings such as state minimum wage laws, state propensity to unionize, state lobbying efforts, state-wide demand factors in the product market, state educational institutions and policies, state-level industrial composition, and a host of others.

Therefore, in order to be driving my estimates, any state-level condition that changes across borders must affect certain border PUMAs differentially, and that differential effect must be systematically related to the licensing regime on the *other* side of the state boundary. For example, state industry policies must affect PUMAs at the Virginia-Maryland border differently than they do at the Virginia-West Virginia border, and that differential effect must be correlated with the difference in licensing across the VA-MD and VA-WV borders, and so on for all boundaries in the United States. Such a cascade of hyper-specific local effects of statewide policies is unlikely.

We can place further bounds on the characteristics of such a policy or unobserved condition with further tests. For example, would the relationships I observe continue to hold if specific cluster assignments changed? If not, then we can rule out any policies or conditions that are general in scope that may affect earnings through mechanisms outside of within- and cross-industry skill similarity. This is because the treatment effects would have to specifically relate to the structure of the cluster assignments even though clusters are defined across industries and sectors—the level at which many policies are typically made and firms make major decisions.

To test this, I perform a placebo exercise in which I randomly assign with equal probability each occupation to be a part of one of twenty clusters. I use the CPS to calculate the licensed share of workers outside the focal occupation that is licensed within their placebo cluster. I then perform all of my main estimates using these shares as the treatment variable. If the licensing environment is correlated with state variables that are also correlated with the earnings effects I find, then the relationship between licensing exposure and earnings and employment should not significantly change.[23]

The results of this exercise are in Figure A3. Panel A shows that licensing exposure within placebo clusters results in zero overall earnings spillovers. This relationship holds across all subgroups except native-born Hispanic workers. If anything, the propensity to have high levels of licensure may be weakly positively correlated with earnings for racial/ethnic minorities (though not statistically significant) when considering placebo clusters, which runs counter to the large, negative effects noted in Figure 5. Similarly, Panel B shows that there is no relationship between occupational employment/composition and licensing exposure in placebo clusters.

This exercise strengthens the case for a causal interpretation of the spillover effects I have identified by showing that cross-state differences in the propensity to license their occupations that may be correlated with unobserved determinants of employment and earnings are not a significant driver of my results. In short, any unobserved joint determinant of PUMA-specific occupational earnings/employment and state-level licensing that is not licensing itself must be correlated with employment outcomes only in a way that is specific to the actual skill structure of occupations in the O*NET.

As another test, I estimate my boundary discontinuity sample including PUMA fixed effects in Section 4.6, making this the most restrictive of my models. In this specification, identification comes purely from PUMAs that share borders with multiple states such as those in northern Virginia that border Maryland *and* Washington, DC, or those that border both Kentucky *and* Tennessee. Results from this exercise show that my border design is robust to unobserved characteristics of the hyper-local labor market. Table A7 shows these results by number of clusters for the overall estimate, by sex, and by race/ethnicity, respectively. These estimates are nearly indistinguishable from my baseline estimates and indicate that unobserved determinants of wages in the local labor market, including differential effects of statewide policies across PUMAs, are not biasing or driving my baseline model.

To ensure that the population composition of my border sample is not driving my results and to minimize the threat of spatial spillovers, I use the Current Population Survey and simple cross-state variation in licensure to estimate the same models but without the border fixed effects. This exercise treats workers in the same occupation in all states as implicit controls rather than just workers on the other side of the state border. The results in Figure A5 show spillover estimates that are very similar to my boundary discontinuity design. It is, therefore, unlikely that sample selection in my border areas is driving my results.

As an additional check, I re-estimate my earnings regressions while sequentially eliminating one cluster at a time. This allows me to pinpoint if my results are driven by any particular

[23] Another implicit indication of this arises from the fact that the spillover effects within clusters are small and not statistically significant below five clusters when clusters are large and there is only minimal differentiation between the clusters.

cluster. These are in Figures A6–A9 and support my core results.

Abstracting away from the measurement issues discussed in Appendix D, I use the Northwestern Licensing Database (Redbird, 2016) to estimate my boundary discontinuity design and present those estimates in Figure A10. Despite slightly attenuated estimates, the pattern of the results strongly supports the results of my baseline method with a very different data source.

Taken together, these exercises show that for an unobserved factor—either related to underlying economic forces or related to endogenous policy adoption of occupational licensing statutes—to drive my results, such a factor must have five key characteristics. First, it must differentially affect earnings and employment in specific border areas, but that effect must be on a knife-edge case where such effects do not differ when considering all states in the US as controls. I conclude this because the simple cross-sectional model (using the CPS ORG dataset) and the border discontinuity model both yield similar results. Second, it must hold across the types of occupations being considered. This is based on the similarity of estimates when eliminating skill clusters from the sample. Third, it would have to hold even when controlling for conditions of the local labor market when I include PUMA fixed effects. Fourth, it must be positively correlated with cluster-specific propensities to license in a way that is specifically correlated with skill cluster structure from O*NET, even though skill clusters cross industry lines. This is because the random assignment of placebo skill clusters breaks the relationship between licensing exposure from other occupations and earnings. Fifth, it must be correlated with licensing from two different databases (both the CPS and the NLD). Tax policy, minimum wages, industrial relations and unionization policies, industry lobbying efforts, etc., do not fit that description. Indeed, it is difficult to imagine such a policy or economic condition that would generate this relationship if not the licensing environment itself.[24]

# 5. Discussion

The pattern of lower earnings and lower employment in the focal occupation as a result of cluster-wide licensure is not consistent with a simple redirection of labor supply into the focal occupation. These dynamics may be explained by a number of candidate mechanisms. Here, I consider four: firm location choices and overall labor demand; skill-based and industry-specific labor demand; labor complementarities; and monopsonistic firm behavior.

## 5.1. Firm Location Choice and Overall Labor Demand

Recent work proposes one important possible mechanism for these spillovers: firm location choice. Firms hiring workers with more onerous licensing requirements such as costly fees or long wait times before independent practice is allowed are less likely to locate in states with the costlier requirements (Plemmons, 2022; Zapletal, 2019). For example, Plemmons (2022) shows that for every 100 additional days (or more than \$100 in fees) required to obtain an occupational license, the probability of a firm affected by the licensing requirement opening in

[24] In Appendix C, I further explore another way of identifying the effects of licensing spillovers using variation in licensing rules over time.

that state declines by approximately 0.13 (1.5) percent across state borders. For certain types of firms such as barber shops, skin care and nail salons, commercial trucking, and daycare services, having licensing fees over $100 for workers in those firms reduces the probability of opening in the stringent state by as much as 11-23 percent. Firms affected by these rules also reduce the number of employees they hire. This firm location choice may, therefore, lead to an overall decline in labor demand and local economic activity. However, this total labor demand explanation is difficult to reconcile with the positive effect of licensing one's own occupation. If there is an overall labor demand effect, we might expect it to affect licensed occupations as well. That we observe positive own-occupation effects implies that there is still sufficient labor demand in licensed occupations to generate a wage premium. In addition, directly controlling for employment in the local PUMA does not change the estimates, making this general labor demand mechanism less likely (see Column 1 of Table A6).

Similarly, it may be that firm location choices disproportionately affect labor demand for specific skills rather than an overall reduction in labor demand. However, like in the case of controlling for overall PUMA employment, directly controlling for cluster-PUMA cell employment does not change the estimates for weekly earnings (see Column 2 of Table A6).

## 5.2. Skill-Based and Industry-Specific Labor Demand

A downstream effect of individual firm location choices may be that effects are coming through industry channels such as industry concentration and agglomeration as well as industry-related, skill-specific labor demand. That the effects on firm location choices in Plemmons (2022) may be more concentrated in services and the fact that the observed effects on earnings I find are stronger in the non-tradable sector both suggest that changes in labor demand may play a role. That my skill clusters can cross industry lines is an important strength of the clustering approach in this context.

To test the importance of cross- vs within-industry dynamics, I calculate leave-one-out exposure to licensure based on the combination of skill clusters and Census industry groups. Industry classifications in the CPS and ACS follow the North American Industry Classification System (NAICS), and there are 14 major industry groups. I calculate licensing exposure for each worker according to the share of workers in their skill cluster-major industry group cell licensed in their state. This measure captures licensing exposure for each worker as a function of the licensure of those that use similar skills and that also work in the same major economic sector. I then estimate a horse race regression between these industry-specific exposures and the overall exposure variable. If cross-industry effects are dominating, then within-industry group effects should be small and statistically and economically insignificant, while the overall skill cluster effects should dominate. This is because any decline in labor demand within an entire industry will be already accounted for through within-industry comparisons.

Table 4 shows the results of this exercise. Exposure of 10% to licensure inside one's own cluster-industry group cell decreases weekly earnings by approximately 0.9%. Including overall exposure in the model does not significantly change this coefficient. The coefficient on overall cluster exposure falls marginally from approximately -0.016 to -0.014, which is not significantly different from the base estimate. The main difference between this model and the base model is somewhat lower precision on the overall exposure variable. Adding the within-industry effect to the overall effect suggests a statistically significant combined effect of -2.2% for a 10 percentage point increase in total exposure.[25]

Table 4: Licensing Exposure Within Industries at 20 Clusters

| VARIABLES | (1)<br>Log(Weekly Earnings) | (2)<br>Log(Weekly Earnings) |
|---|---|---|
| Own Occupation Effect | 0.0968*** | 0.0998*** |
| | (0.0217) | (0.0212) |
| Cluster-Industry Exposure (10 PPT) | -0.0112*** | -0.0106*** |
| | (0.00354) | (0.00362) |
| Cluster Exposure (10 PPT) | | -0.0135 |
| | | (0.00854) |
| Observations | 1,658,443 | 1,658,443 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: Author's calculations of ACS, CPS, and O*NET data.
Notes: Estimates correspond to Equation 1 but include an additional measure of licensing exposure within major industry groups (NAICS) and skill clusters. Standard errors are clustered at the occupation level. Estimates include fixed effects for occupation, state, and boundary.

Overall, the within-industry effects are statistically and economically significant. There are separate and meaningful spillovers both within- and across industries, meaning that industry concentration of labor demand does not appear to be the primary mechanism for the observed effects on earnings in my main models. If these effects are contributing heavily to spillovers, it would be puzzling that the negative earnings effects would hold even conditioning on cluster-PUMA employment (Table A6). It is also not clear under this explanation why skill- or industry-specific changes in labor demand through spillovers would disproportionately negatively affect certain demographic groups more than others while the own-occupation effect simultaneously generates a larger wage premium for the same groups. However, the analysis cannot completely rule out the contributions of these effects.

## 5.3. Labor Complementarities

Another possible mechanism for these effects might be that workers in licensed occupations are complements of workers within their skill cluster (Neumark and Wachter, 1995). Falling employment in licensed occupations then leads to lower labor demand for the occupations that complement these labor inputs, i.e. for those that work alongside licensed workers. A way

[25] Estimating industry group exposure effects exclusively (see Table E2) leads to strikingly similar results, meaning that even outside of the skill clustering approach, spillovers across occupations that generally work in the same broad fields are negative and significant.

to measure the scope of these complementarities is to consider occupation $o$ as experiencing and contributing to licensing complementarities in industry $k$ via three avenues: first, their employment shares of industry $k$; second, the industry shares of employment in occupation $o$ working in industry $k$; and third, the licensing rules for occupation $o$ in the state.[26] In other words, there is more complementarity in an occupation-state cell if the occupation's industries have more licensing rules in the state, licensed occupations constitute a larger share of the industry's employment, and when employment in an occupation is concentrated in fewer industries.

I calculate complementarity exposure based on national occupation-industry composition combined with state-level variation in licensing shares for each occupation. Here, I define complementarity exposure in production for occupation $o$ in state $s$ as:

$$ComplementExposure_{os} = \sum_{o=1}^{p} \delta_{ok} \sum_{k=1}^{l} \omega_{ko} * LicensedShare_{os}$$

, where $LicensedShare_{os}$ is defined as in Equation 1 as the share of occupation $o$ in state $s$ that is licensed in the CPS. The value $\delta_{ok}$ represents the share of occupation $o$ working in industry $k$ at the national level, while $\omega_{ko}$ is the national share of industry $k$ employment in occupation $o$.

I then estimate the complementarity exposure analog of Equation 1 for person $i$ in occupation $o$ in industry $k$ in state $s$ at state border pair $m$:

$$y_{iokms} = \beta_0 + \beta_1 LicensedShare_{os} + \beta_2 ComplementExposure_{os} + X_i' \beta_3 + \delta_o + \gamma_s + \tau_m + \varepsilon_{iokms} \quad (5)$$

I also estimate the same model of employment as in Equations 3 and 4, except for replacing cluster licensure with complementarity exposure. Table 5 shows the results of these exercises when normalizing complementarity exposure in standard deviation units. For transparency, I estimate these models including and excluding own-occupation effects. A standard deviation increase in complementarity exposure to licensure in one's own industry raises log weekly earnings by approximately 1.7 percent. However, this effect does not persist when adding the effect for own-occupation licensure, and the coefficient drops to 0.7% and is not statistically significant. The direction of this effect is not consistent with the lower earnings I find within skill clusters. A standard deviation increase in complementarity exposure to licensing reduces employment by 2.8%, but this effect also does not persist when controlling for occupation-specific licensure. Overall, the results of this exercise suggest that direct complementarity spillovers reducing demand for non-licensed workers in the same industries as licensed workers is not a primary driver of my results.

The within-industry spillovers in Section 5.2 may be consistent with a complementarity view. However, because cross-industry effects appear to persist and within-industry effects

[26] I am grateful to an anonymous reviewer for this excellent suggestion.

Table 5: Complementarity Exposure Effects

| VARIABLES | (1) Log(Weekly Earnings) | (2) Log(Weekly Earnings) | (3) Log(Emp) | (4) Log(Emp) |
|---|---|---|---|---|
| Complementarity Exposure (1 SD) | 0.0170* | 0.00709 | -0.0280** | -0.00868 |
| | (0.0102) | (0.00947) | (0.0137) | (0.0162) |
| Own Occupation Licensing Effect | | 0.0613*** | | -0.217** |
| | | (0.0219) | | (0.0983) |
| Observations | 1,765,288 | 1,765,288 | 538,036 | 538,036 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: Author's calculations of ACS, CPS, and O*NET data.
Notes: Estimates correspond to Equation 5. Standard errors are clustered at the occupation level. Estimates include fixed effects for occupation, state, and boundary.

are smaller than the size of the total spillover, complementary labor inputs cannot explain the entire effect. Another likely explanation is that a worker's "outside options" are most relevant first within industries and then across industries in the same skill cluster, which lends itself to the monopsony explanation below. In addition, given the narrow definitions of Census occupation groups around industries and "people who work together," which I explore in Appendix E, we would expect large, negative effects when examining exposure in Census occupation groups. I do not find such effects.[27]

## 5.4. Monopsonistic Firm Behavior

The results above may be consistent with an increase in monopsony power in the local labor market. Monopsony power is not only a function of the costs of within-occupation switching across firms, but also of a worker's ability to leave the local labor market, switch occupations, or both. It is also a function of the number of viable outside options available for the worker. This view is supported by recent work that explores the use of more comprehensive definitions of a "local labor market" for workers that incorporates outside options (Schubert et al., 2021; Dodini et al., 2023).

It is clear from the past literature that licensing increases labor market rigidity across occupations (Kleiner and Xu, 2020). That transitions between occupations fall as a result of licensing logically fits into a monopsony framework in which incumbent workers are less able to credibly threaten to leave a low-paying firm.

Appendix F.2 presents a search model in which licensing costs decrease the share of firms

[27]Large negative effects on complementary labor inputs seems at odds with the fact that prices rise for products and services produced by licensed labor when occupations become licensed (Adams III et al., 2002; Wing and Marier, 2014). This means that the marginal revenue product of the complementary labor inputs would increase, thereby increasing labor demand.

that can legally hire an unlicensed worker, thereby increasing search costs and reducing a worker's reservation wage during the search, leading to lower wages. The model also rationalizes why the negative effects of an increase in firm market power and a decrease in outside options may disproportionately affect women, racial/ethnic minorities, and foreign-born workers. These differential effects may arise through direct discrimination, language requirements, citizenship or residency requirements, differential costs of pursuing a license, or blanket bans on those with a felony conviction from obtaining a license in some states and occupations.

Licensing increases this prospective cost of leaving a firm to pursue outside options within and across skill clusters and states, so the set of available options that are feasible for workers to enter shrinks, which may lower the elasticity of labor supply to the firm in highly licensed areas as well as drive down employment in those areas as firms scale back new hiring. In this case, the wage effect comes first, followed by the employment effect. Furthermore, if firms are deterred from locating in some local labor markets due to licensing effects (Plemmons, 2022), then it is not just the adjustment friction introduced by licensing in one's skill cluster that limits a worker's outside options. The overall number of firms available to compete over the worker in the skill-specific local labor market also falls. In this case, the employment effect comes first, followed by an increase in wage markdowns.

Altogether, an increase in firm labor market power from occupational licensing may arise from two separate but reinforcing mechanisms: an increase in the marginal cost of switching occupations and a simple reduction in the number of competing firms due to firm location choices. One is a function of switching costs conditional on the distribution of firms and occupations, while the other is a change in the distribution of firms. Both mechanisms would generate a steeper labor supply curve to the firm (see Appendix F.1). The results of my earnings and employment models above are consistent with both of these explanations, particularly in the patterns related to heterogeneous effects by gender, race/ethnicity, and nativity.

Do the results of other studies align with these explanations? There is evidence from the monopsony literature that the elasticity of labor supply to the firm is lower for women than it is for men, implying greater monopsony power in the labor markets employing women (Ransom and Oaxaca, 2010; Ransom and Lambson, 2011; Barth and Dale-Olsen, 2009; Hirsch et al., 2010). This is consistent with my findings of far greater earnings spillover effects for women, who are also more likely to be employed in the non-tradable sector. Likewise, the prior literature has shown that non-routine, cognitive tasks, which are more heavily used in the non-tradable and service sectors and which disproportionately employ women, are more exposed to monopsonistic behavior by firms (Bachmann et al., 2019; Dodini et al., 2023). The literature also suggests that immigrants supply labor to the firm much less elastically than their native-born counterparts, which leads to a predicted 7% wage penalty according to Hirsch and Jahn (2015). Taste-based discrimination may be far more consequential for wages in monopsonistic labor markets, affecting historically discriminated groups such as African Americans (Berson, 2016; Webber, 2015; Black, 1995) or women (Fanfani, 2018). Indeed, these explanations of limited outside options have been an important part of the "crowding" hypothesis since the 1960s regarding occupational segregation (Bergmann, 1971).[28] These two latter points together may partially explain why non-Hispanic black workers and women face the largest earnings penalties.

Overall, across these four proposed mechanisms, two stand out as the most plausible based on this analysis and the prior literature: changes in labor demand across industries and employer market power. Indeed, both may play a reinforcing role. Though the monopsony explanation appears most plausible in light of this analysis, more precisely pinpointing the monopsony mechanism with additional firm data is a promising area for future research.

Taken together, my analysis suggests that a reduction in licensing would increase incomes across the distribution, leading to a reduction in inequality both within and across occupations and within and across demographic groups.

# 6. Conclusion

This analysis has presented the first evidence of substantial direct labor market spillovers from occupational licensing in the United States using a boundary discontinuity design. I find that occupations that use similar skills to licensed occupations experience a fall in weekly earnings of approximately 1.6% as a result of a 10 percentage point (approximately 1 SD) increase in licensure. The earnings penalties are notably larger among women, foreign-born Hispanic workers, and Non-Hispanic black workers and are as large as 3.5-4% for a 10 percentage point increase in cluster licensure rates. I also find evidence of falling equilibrium employment and statistically significant increases in the share of workers that are women, foreign-born, and Hispanic as a result of licensure in other occupations. Eliminating or reducing the labor market frictions that come from licensing would increase earnings for many workers, particularly those at the bottom of the distribution.

These patterns are most consistent with two of the proposed explanations I consider here, though the balance of the evidence favors firm monopsony behavior as the leading explanation. While the analysis presented here is consistent with the industry labor demand and monopsony power explanations, I am limited in my ability to assess just how strongly these earnings penalties are correlated with markers of labor market power such as concentration or wage markdowns or measures of industry-specific labor demand. Future work in this area may attempt to directly measure the effects of occupational licensing on labor market power or industry concentration using firm-level or other data. Separating out these precise mechanisms with linked employer-employee data is a promising area for future research.

Occupational licenses are often justified by advocates as being in the best interest of consumer health and safety. One of the consequences of these regulations, intended or unintended,

[28]Bergmann, for example, explicitly argues this point: "A by-product of the limitation of most Negroes [sic] to menial occupations is the depressing of the wage rates of those relatively few Negroes [sic] who are hired for jobs in predominantly white occupations. Their low wage reflects at least in part their lower opportunity cost," (Bergmann, 1971), p. 298.

is a meaningfully large earnings premium for licensed workers. At the same time, raising barriers to entry across more and more occupations may have unintended consequences for other workers. As strict labor market regulation grows, workers who might otherwise choose to work in an occupation *but for* the existence of the license are made worse off, and these effects are most keenly felt by workers already more likely to be financially disadvantaged. As a result, occupational licensing significantly increases predicted earnings inequality.

The employment and earnings effects of licensing and other labor market regulations, if broadened to include more occupations, may lead to lower industry-specific labor demand or labor market conditions consistent with more pronounced monopsonistic behavior by firms. In that case, while some workers may be better off individually once they get a license, the imperfections induced by strict entry regulations lead to other workers having fewer opportunities for advancement, making most others unambiguously worse off due to the costs of the restrictions. These represent significant externalities. Policymakers should weigh the possible health and safety benefits of occupational licensing against the possible costs: the negative labor market effects of these regulations on workers that may not be a party to the negotiations between the professional or political entities involved.


# Acknowledgements

I am grateful to Michael Lovenheim, Maria Fitzpatrick, Evan Riehl, Seth Sanders, Alexander Willén, Nicholas Carollo, Bertil Tungodden, and Kjell Salvanes for helpful comments. I am also thankful to participants at the Knee Center Occupational Licensing Conference and seminar participants at Cornell University and the Norwegian School of Economics for helpful comments and discussion. This project was partially funded by the Research Council of Norway through its Centers of Excellence Scheme, FAIR project no. 262675.


# References


**Acemoglu, Daron, and David Autor.** 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." In *Handbook of Labor Economics*. 4: Elsevier, 1043–1171.

**Adams III, A. Frank, John D. Jackson, and Robert B. Ekelund Jr..** 2002. "Occupational Licensing in a "Competitive" Labor Market: The Case of Cosmetology." *Journal of Labor Research*, 23(2): .

**Ashenfelter, Orley C, Henry Farber, and Michael R Ransom.** 2010. "Labor Market Monopsony." *Journal of Labor Economics*, 28(2): 203–210.

**Autor, David.** 2014. "Skills, Education, and the Rise of Earnings Inequality Among the "Other 99 Percent"." *Science*, 344(6186): .

**Autor, David H., and David Dorn.** 2013. "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market." *American Economic Review*, 103(5): 1553–97.

**Bachmann, Ronald, Gökay Demir, and Hanna Frings.** 2019. "Labour Market Polarisation and Monopsonistic Competition."

**Bae, Kihwan, and Edward Timmons.** 2023. "Now You Can Take It with You: Effects of Occupational Credential Recognition on Labor Market Outcomes."

**Bailey, Thomas, and Clive R Belfield.** 2018. "The Impact of Occupational Licensing on Labor Market Outcomes of College-Educated Workers."

**Barth, Erling, and Harald Dale-Olsen.** 2009. "Monopsonistic Discrimination, Worker Turnover, and the Gender Wage Gap." *Labour Economics*, 16(5): 589–597.

**Baum-Snow, Nathaniel, and Derek Neal.** 2009. "Mismeasurement of Usual Hours Worked in the Census and ACS." *Economics Letters*, 102(1): 39–41.

**Bergmann, Barbara R.** 1971. "The Effect on White Incomes of Discrimination in Employment." *Journal of Political Economy*, 79(2): 294–313.

**Bergmann, Barbara R.** 1974. "Occupational Segregation, Wages and Profits When Employers Discriminate by Race or Sex." *Eastern Economic Journal*, 1(2): 103–110.

**Berson, Clémence.** 2016. "Local Labor Markets and Taste-Based Discrimination." *IZA Journal of Labor Economics*, 5(1): , p. 5.

**Black, Dan A.** 1995. "Discrimination in an Equilibrium Search Model." *Journal of labor Economics*, 13(2): 309–334.

**Black, Sandra E.** 1999. "Do Better Schools Matter? Parental Valuation of Elementary Education." *The Quarterly Journal of Economics*, 114(2): 577–599.

**Blair, Peter Q, and Bobby W Chung.** 2018. "Job Market Signaling Through Occupational Licensing." URL: http://www.nber.org/papers/w24791.

**Blair, Peter Q, and Bobby W Chung.** 2019. "How Much of Barrier to Entry is Occupational Licensing?" *British Journal of Industrial Relations*, 57(4): 919–943.

**Burstein, Ariel, Gordon Hanson, Lin Tian, and Jonathan Vogel.** 2020. "Tradability and the Labor-Market Impact of Immigration: Theory and Evidence From the United States." *Econometrica*, 88(3): 1071–1112.

**Busso, Matias, Jesse Gregory, and Patrick Kline.** 2013. "Assessing the Incidence and Efficiency of a Prominent Place Based Policy." *American Economic Review*, 103(2): 897–947.

**Cai, Jing, and Morris M. Kleiner.** 2016. "The Labor Market Consequences of Regulating Similar Occupations: The Licensing of Occupational and Physical Therapists." DOI: http://dx.doi.org/10.2139/ssrn.2802158.

**Card, David, Thomas Lemieux, and W Craig Riddell.** 2020. "Unions and Wage Inequality: The Roles of Gender, Skill and Public Sector Employment." *Canadian Journal of Economics/Revue canadienne d'économique*, 53(1): 140–173.

**Carollo, Nicholas A.** 2020a. "Historical Data on Professional and Occupational Credentialing Requirements in the United States."

**Carollo, Nicholas A.** 2020b. "The Impact of Occupational Licensing on Earnings and Employment: Evidence from State-Level Policy Changes."

**Charrad, Malika, Nadia Ghazzali, Veronique Boiteau, Azam Niknafs, and Maintainer Malika Charrad.** 2014. "Package 'nbclust'." *Journal of Statistical Software*, 61 1–36.

**Clemens, Jeffrey, Lisa B Kahn, and Jonathan Meer.** 2021. "Dropouts Need Not Apply? The Minimum Wage and Skill Upgrading." *Journal of Labor Economics*, 39(S1): S107–S149.

**Correia, Sergio.** 2018. "REGHDFE: Stata Module to Perform Linear or Instrumental-Variable Regression Absorbing Any Number of High-Dimensional Fixed Effects."

**Correia, Sergio, Paulo Guimarães, and Thomas Zylkin.** 2020. "Fast Poisson Estimation with High-Dimensional Fixed Effects." *The Stata Journal*, 20(1): 95–115, URL: https://doi.org/10.1177/1536867X20909691, DOI: http://dx.doi.org/10.1177/1536867X20909691.

**Cunningham, Evan.** 2019. "Professional Certifications and Occupational Licenses." *Monthly Labor Review* 1–38.

**DePasquale, Christina, and Kevin Stange.** 2016. "Labor Supply Effects of Occupational Regulation: Evidence from the Nurse Licensure Compact."

**Dillender, Marcus, Anthony T Lo Sasso, Brian J Phelan, and Michael R Richards.** 2022. "Occupational Licensing and the Healthcare Labor Market."

**Dodini, Samuel, Michael F Lovenheim, Kjell G Salvanes, and Alexander Willén.** 2023. "Monopsony, Job Tasks, and Labor Market Concentration." *NBER Working Paper no. 30823.*

**Dodini, Samuel, Kjell G Salvanes, and Alexander Willén.** 2021. "The Dynamics of Power in Labor Markets: Monopolistic Unions versus Monopsonistic Employers."

**Duncan, Otis Dudley, and Albert J Reiss.** 1961. "A Socioeconomic Index for All Occupations." *Class: Critical Concepts*, 1(1): 388–426.

**Dunn, Joseph C.** 1974. "Well-separated Clusters and Optimal Fuzzy Partitions." *Journal of Cybernetics*, 4(1): 95–104.

**Fanfani, Bernardo.** 2018. "Tastes for Discrimination in Monopsonistic Labour Markets." *Working Paper.*

**Gittleman, Maury, Mark A Klee, and Morris M Kleiner.** 2018. "Analyzing the Labor Market Outcomes of Occupational Licensing." *Industrial Relations: A Journal of Economy and Society*, 57(1): 57–100.

**Grembi, Veronica, Tommaso Nannicini, and Ugo Troiano.** 2016. "Do Fiscal Rules Matter?" *American Economic Journal: Applied Economics*, 8(3): 1–30.

**Halkidi, Maria, Michalis Vazirgiannis, and Yannis Batistakis.** 2000. "Quality Scheme Assessment in the Clustering Process." In *European Conference on Principles of Data Mining and Knowledge Discovery.* 265–276, Springer.

**Hauser, Robert M, and John Robert Warren.** 1997. "Socioeconomic Indexes for Occupations: A Review, Update, and Critique." *Sociological Methodology*, 27(1): 177–298.

**Hirsch, Boris, and Elke J Jahn.** 2015. "Is There Monopsonistic Discrimination Against Immigrants?" *ILR Review*, 68(3): 501–528.

**Hirsch, Boris, Thorsten Schank, and Claus Schnabel.** 2010. "Differences in Labor Supply to Monopsonistic Firms and the Gender Pay Gap: An Empirical Analysis Using Linked Employer-Employee Data from Germany." *Journal of Labor Economics*, 28(2): 291–330.

**Hubert, Lawrence J, and Joel R Levin.** 1976. "A General Statistical Framework for Assessing Categorical Clustering in Free Recall.." *Psychological Bulletin*, 83(6): , p. 1072.

**Johnson, Janna E, and Morris M Kleiner.** 2017. "Is Occupational Licensing a Barrier to Interstate Migration?." URL: http://www.nber.org/papers/w24107.

**Kleiner, Morris M.** 2000. "Occupational Licensing." *Journal of Economic Perspectives*, 14(4): 189–202.

**Kleiner, Morris M., and Alan B. Krueger.** 2013. "Analyzing the Extent and Influence of Occupational Licensing on the Labor Market." *Journal of Labor Economics*, 31(S1): S173–202.

**Kleiner, Morris M, Allison Marier, Kyoung Won Park, and Coady Wing.** 2016. "Relaxing Occupational Licensing Requirements: Analyzing Wages and Prices for a Medical Service." *The Journal of Law and Economics*, 59(2): 261–291.

**Kleiner, Morris M, and Kyoung Won Park.** 2010. "Battles Among Licensed Occupations: Analyzing Government Regulations on Labor Market Outcomes for Dentists and Hygienists."

**Kleiner, Morris M, and Evan J Soltas.** 2019. "A Welfare Analysis of Occupational Licensing in US States."(26283): .

**Kleiner, Morris M., and Evgeny Vorotnikov.** 2017. "Analyzing Occupational Licensing Among the States." *Journal of Regulatory Economics*, 55.

**Kleiner, Morris, and Ming Xu.** 2020. "Occupational Licensing and Labor Market Fluidity."

**Koumenta, Maria, Amy Humphris, Morris Kleiner, and Mario Pagliero.** 2014. "Occupational Regulation in the EU and UK: Prevalence and Labour Market Impacts." *Final Report, Department for Business, Innovation and Skills, School of Business and Management, Queen Mary University of London, London.*

**Koumenta, Maria, and Mario Pagliero.** 2019. "Occupational Regulation in the European Union: Coverage and Wage Effects." *British Journal of Industrial Relations*, 57(4): 818–849.

**Kugler, Adriana D, and Robert M Sauer.** 2005. "Doctors Without Borders? Relicensing Requirements and Negative Selection in the Market for Physicians." *Journal of Labor Economics*, 23(3): 437–465.

**Lipsitz, Michael, and Evan Starr.** 2022. "Low-Wage Workers and the Enforceability of Noncompete Agreements." *Management Science*, 68(1): 143–170, DOI: http://dx.doi.org/10.1287/mnsc.2020.3918.

**Mincer, Jacob.** 1974. "Schooling, Experience, and Earnings."

**Nakao, Keiko, and Judith Treas.** 1994. "Updating Occupational Prestige and Socioeconomic Scores: How the New Measures Measure Up." *Sociological Methodology* 1–72.

**Neumark, David, and Michael L Wachter.** 1995. "Union Effects on Nonunion Wages: Evidence from Panel Data on Industries and Cities." *ILR Review*, 49(1): 20–38.

**Oster, Emily.** 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business & Economic Statistics*, 37(2): 187–204.

**Pizzola, Brandon, and Alexander Tabarrok.** 2017. "Occupational Licensing Causes a Wage Premium: Evidence from a Natural Experiment in Colorado's Funeral Services Industry." *International Review of Law and Economics*, 50 50–59.

**Plemmons, Alicia.** 2022. "Occupational Licensing's Effects on Firm Location and Employment in the United States." *British Journal of Industrial Relations*, 60(4): 735–760.

**Ransom, Michael R, and Val E Lambson.** 2011. "Monopsony, Mobility, and Sex Differences in Pay: Missouri School Teachers." *American Economic Review*, 101(3): 454–59.

**Ransom, Michael R, and Ronald L Oaxaca.** 2010. "New Market Power Models and Sex Differences in Pay." *Journal of Labor Economics*, 28(2): 267–289.

**Redbird, Beth.** 2016. "Northwestern Licensing Database: Version 2016."

**Redbird, Beth.** 2017. "The New Closed Shop? The Economic and Structural Effects of Occupational Licensure." *American Sociological Review*, 82(3): 600–624, DOI: http://dx.doi.org/10.1177/0003122417706463.

**Rousseeuw, Peter J.** 1987. "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis." *Journal of Computational and Applied Mathematics*, 20 53–65.

**Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek.** 2019. "IPUMS USA: Version 9.0." DOI: http://dx.doi.org/https://doi.org/10.18128/D010.V9.0.

**Rumbaut, Rubén G, and Douglas S Massey.** 2013. "Immigration & Language Diversity in the United States." *Daedalus*, 142(3): 141–154.

**Schubert, Gregor, Anna Stansbury, and Bledi Taska.** 2021. "Employer Concentration and Outside Options." *Available at SSRN 3599454*.

**Shaw, Kathryn L.** 1987. "Occupational Change, Employer Change, and the Transferability of Skills." *Southern Economic Journal* 702–719.

**Siegel, Paul Matthew.** 1971. "Prestige in the American Occupational Structure." Ph.D. dissertation, The University of Chicago.

**Sokal, RR, and CD Michener.** 1958. "A Statistical Method for Evaluating Systematic Relationships."

**Starr, Evan.** 2019. "Consider This: Training, Wages, and the Enforceability of Covenants Not to Compete." *ILR Review*, 72(4): 783–817.

**Starr, Evan, Justin Frake, and Rajshree Agarwal.** 2019. "Mobility Constraint Externalities." *Organization Science*, 30(5): 961–980.

**Starr, Evan P., J.J. Prescott, and Norman D. Bishara.** 2021. "Noncompete Agreements in the US Labor Force." *The Journal of Law and Economics*, 64(1): 53–84, URL: https://doi.org/10.1086/712206, DOI: http://dx.doi.org/10.1086/712206.

**Thornton, Robert J, and Edward J Timmons.** 2013. "Licensing One of the World's Oldest Professions: Massage." *The Journal of Law and Economics*, 56(2): 371–388.

**US Bureau of Labor Statistics.** 2021. "A Look at Union Membership Rates Across Industries in 2020."Technical report, US Bureau of Labor Statistics, Accessed on March 10, 2023.

**US Department of Labor, and US Bureau of Labor Statistics.** 1999. "Revising the Standard Occupational Classification System."Technical Report Report 929, US Department of Labor and US Bureau of Labor Statistics.

**Vargo, Emily, Ethan Bayne, and Edward Timmons.** 2020. "Occupational Regulation Database. April 2020 version [dataset]." URL: https://csorsfu.com/find-occupations/.

**Webber, Douglas A.** 2015. "Firm Market Power and the Earnings Distribution." *Labour Economics*, 35 123–134.

**Wing, Coady, and Allison Marier.** 2014. "Effects of Occupational Regulations on the Cost of Dental Services: Evidence from Dental Insurance Claims." *Journal of Health Economics*, 34 131–143.

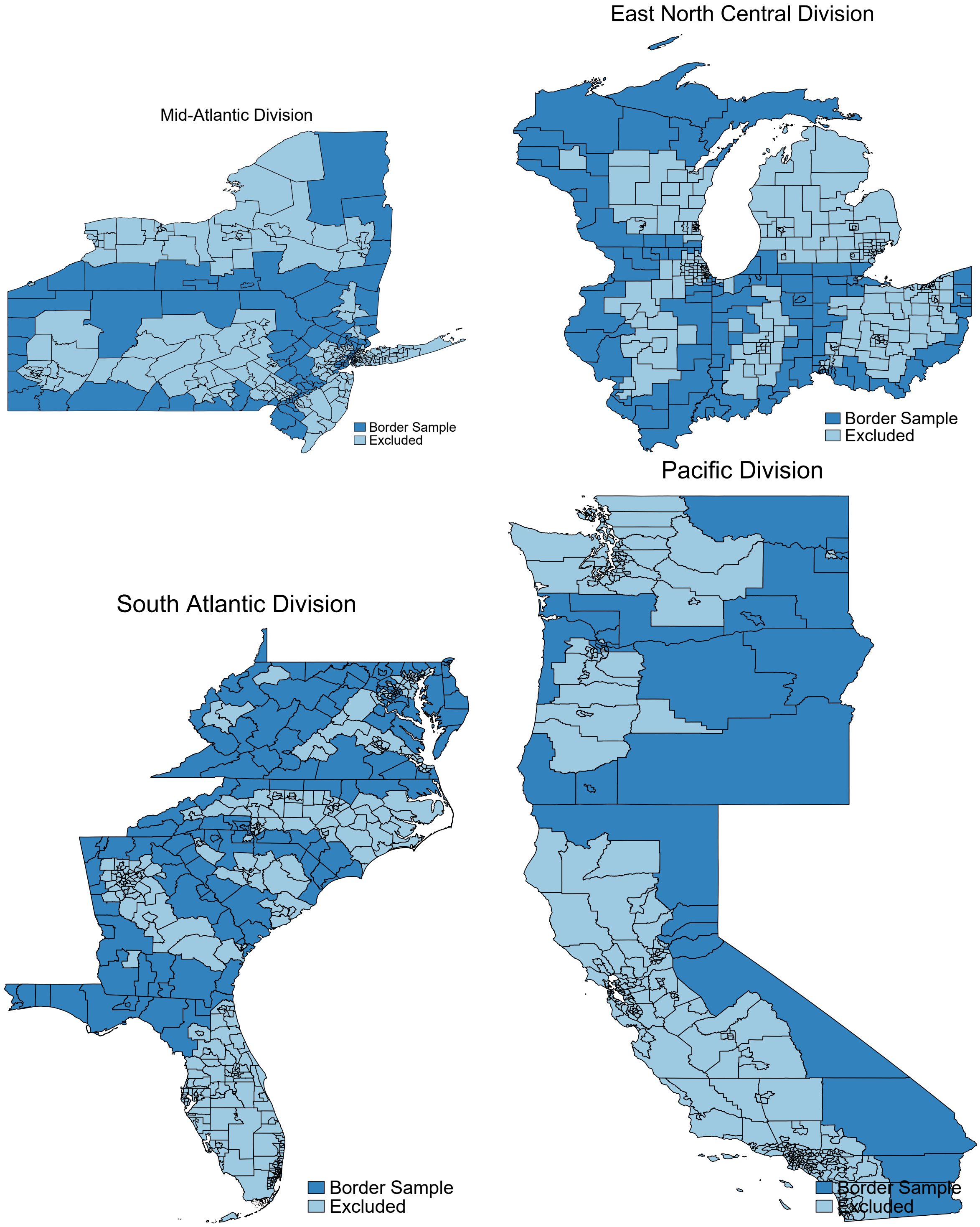
**Yeung, Ka Yee, and Walter L. Ruzzo.** 2001. "Principal Component Analysis for Clustering Gene Expression Data." *Bioinformatics*, 17(9): 763–774.

**Zapletal, Marek.** 2019. "The Effects of Occupational Licensing: Evidence from Business-Level Data." *British Journal of Industrial Relations*, 57(4): 894–918.

**Zhang, Tingting, and Morley Gunderson.** 2020. "Impact of Occupational Licensing on Wages and Wage Inequality: Canadian Evidence 1998–2018." *Journal of Labor Research*, 41(4): 338–351.

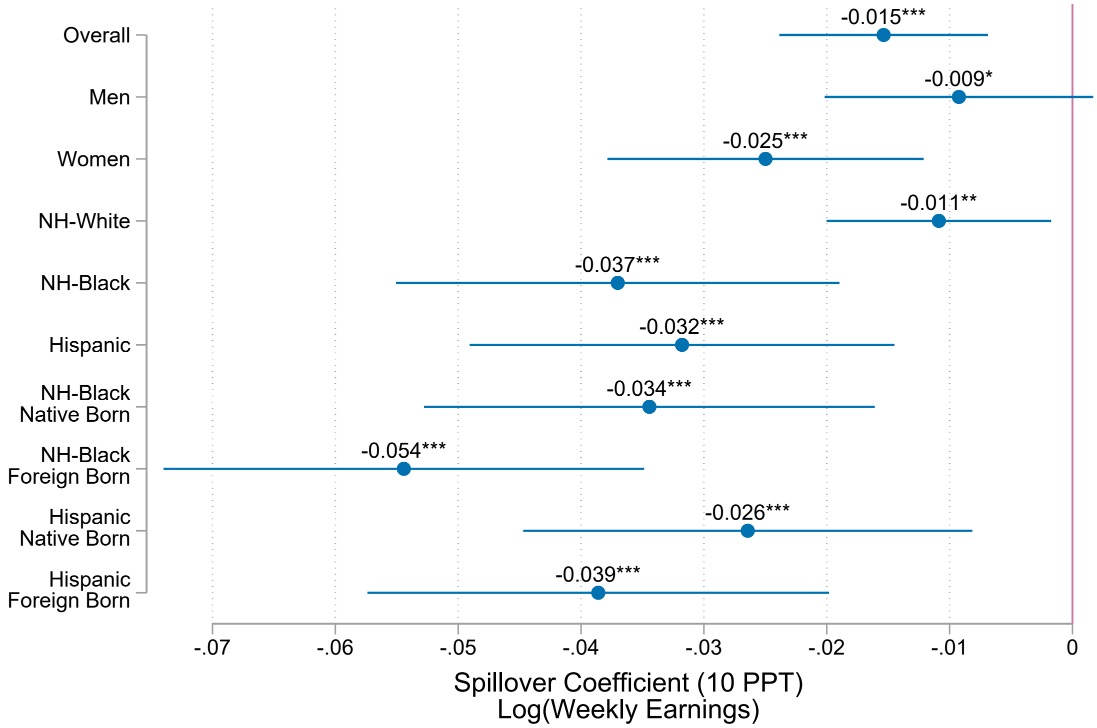
# Appendix A. Additional Figures and Tables

Figure A1: Border Sample PUMAs in Select Census Divisions



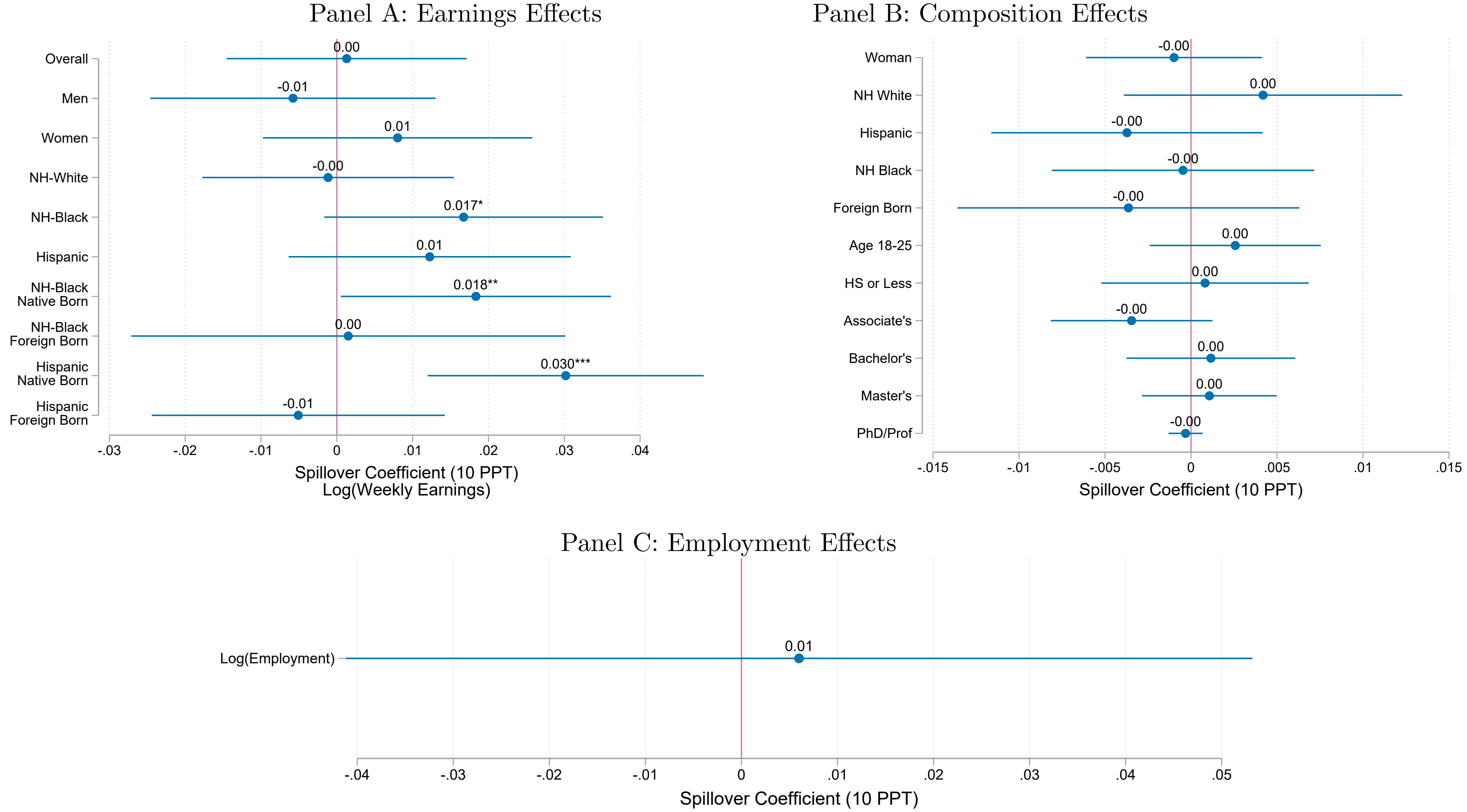

Source: Author's border sample of ACS Public Usa Microdata Areas (PUMAs)

Figure A2: Coefficients of Spillover Effects by Subgroup at 20 Clusters, Occupation by Border Fixed Effects



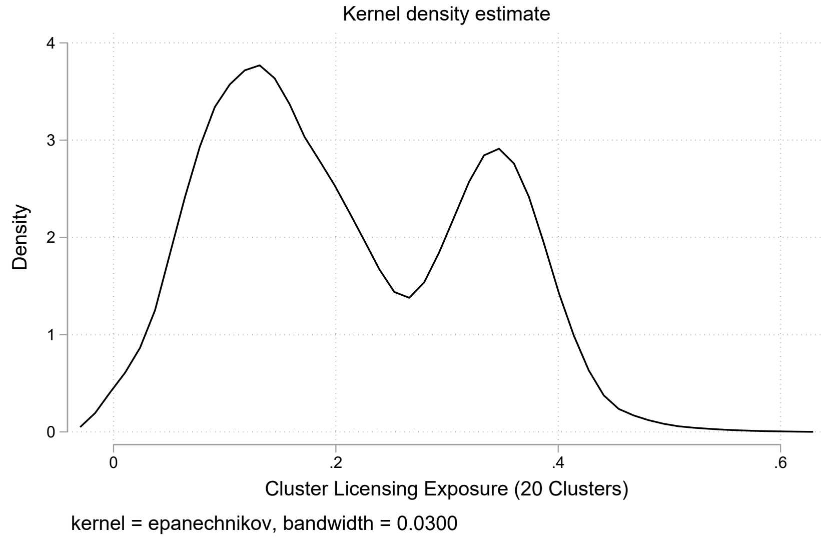

Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: *** p<0.01, ** p<0.05, * p<0.1. Coefficients are generated from the boundary discontinuity design with occupation-by-border-pair interacted fixed effects using 20 skill clusters. Dots represent point estimates, and the bands represent the 95% confidence interval clustered at the occupation level. Spillover coefficients are based on a 10 percentage point increase in licensure of an occupation's cluster outside their own occupation. Models include occupation-by-border-pair interacted fixed effects and controls for race/ethnicity, sex, age, and age squared.

Figure A3: Spillover Coefficients of Log Weekly Earnings, 20 Placebo Clusters



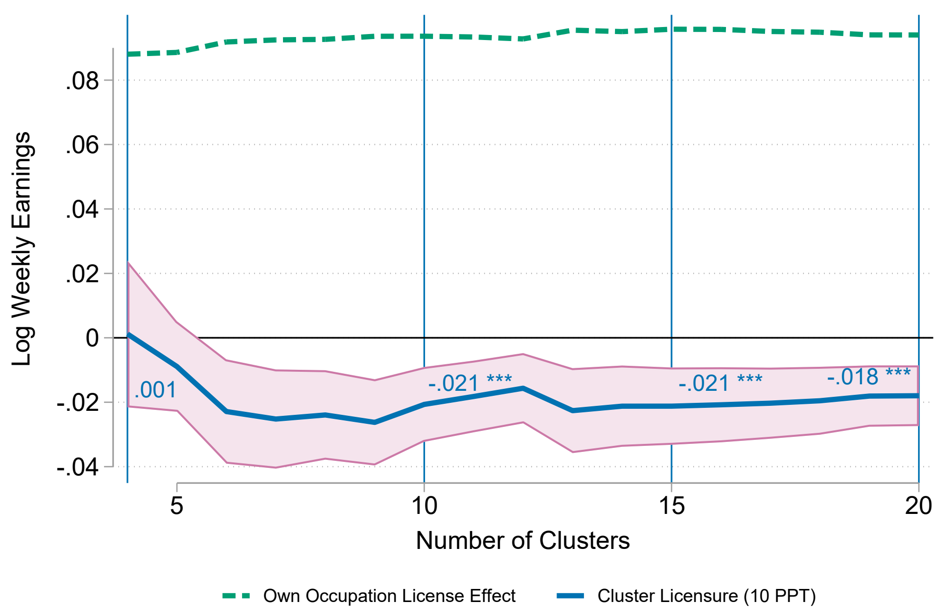

Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: *** p<0.01, ** p<0.05, * p<0.1. Coefficients are generated from $\beta_2$ in the boundary discontinuity design using 20 placebo skill clusters. Dots represent point estimates, and the bands represent the 95% confidence interval clustered at the occupation level. Spillover coefficients are based on a 10 percentage point increase in licensure of an occupation's cluster outside their own occupation. Models include occupation, state, and boundary fixed effects.

Figure A4: Distribution of Licensing Exposure within Clusters



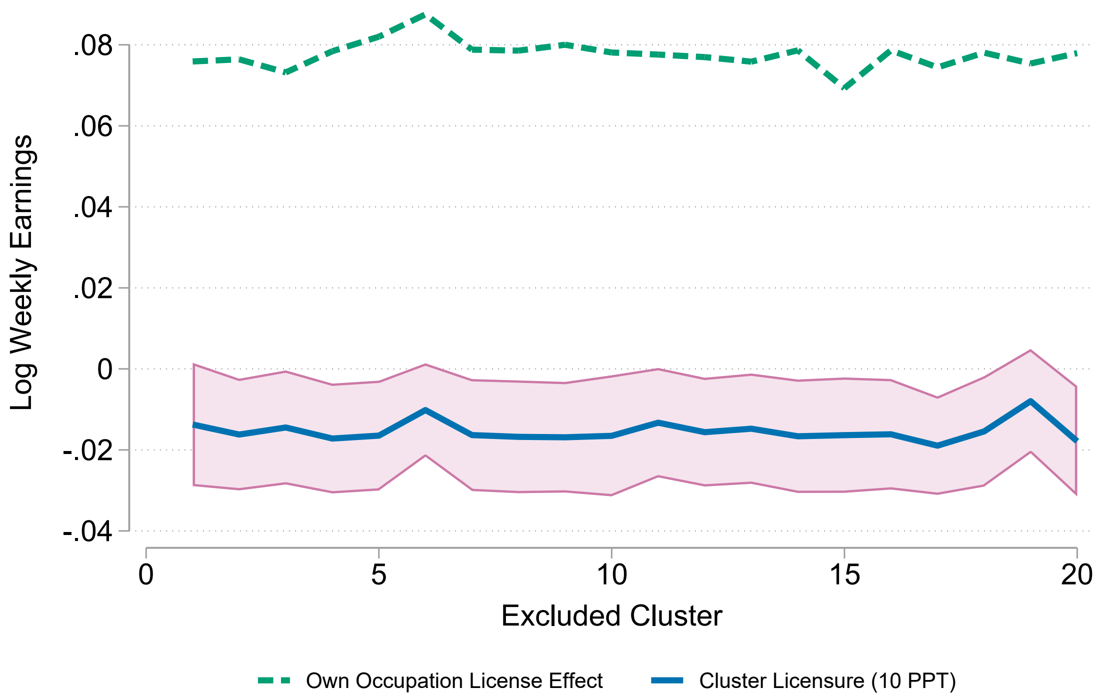

Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Kernel densities of exposure are calculated on the weighted unconditional distribution of cluster-wide licensure at 20 clusters in the border discontinuity sample.

Figure A5: Log Wage Effects with CPS 2015-2018
State and Occupation Fixed Effects




Source: Author's calculations of O*NET, and 2015-2018 CPS.
Notes: *** p<0.01, ** p<0.05, * p<0.1. Coefficients are generated from estimates of log hourly wage in the CPS on individual sex, race/ethnicity, age, age squared, and state and occupation fixed effects. Spillover coefficients are based on a 10 percentage point increase in licensure of an occupation's cluster outside their own occupation. Standard errors are clustered at the occupation level. 95% confidence intervals in red. Vertical bars and coefficients are for clusters at 4, 10, 15, and 20.

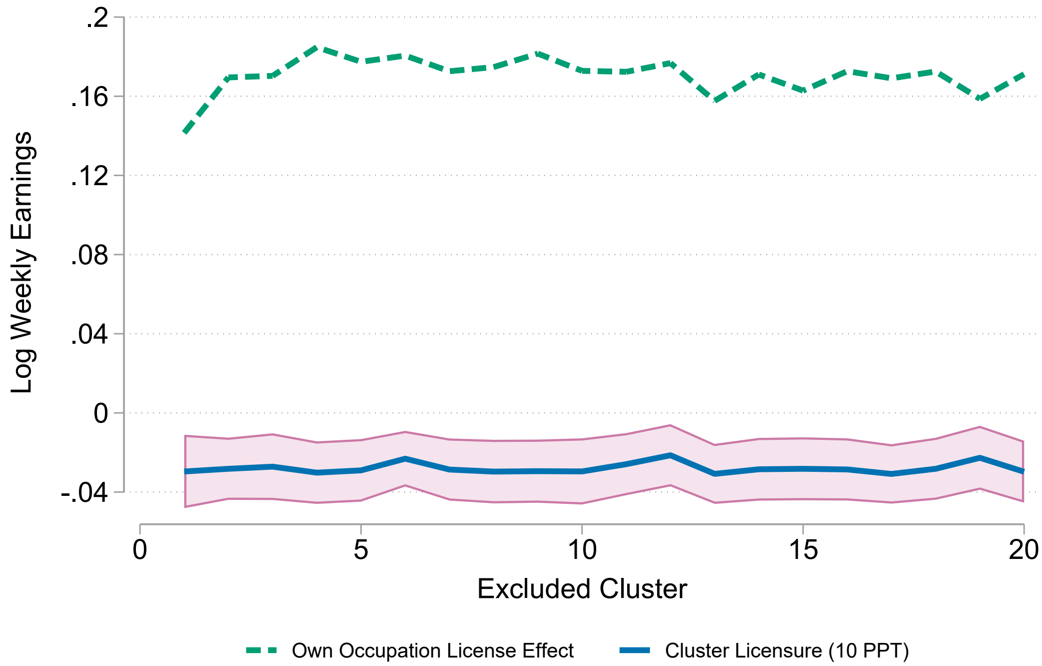
Figure A6: Earnings Effects, Sequentially Removing Clusters
20 Clusters



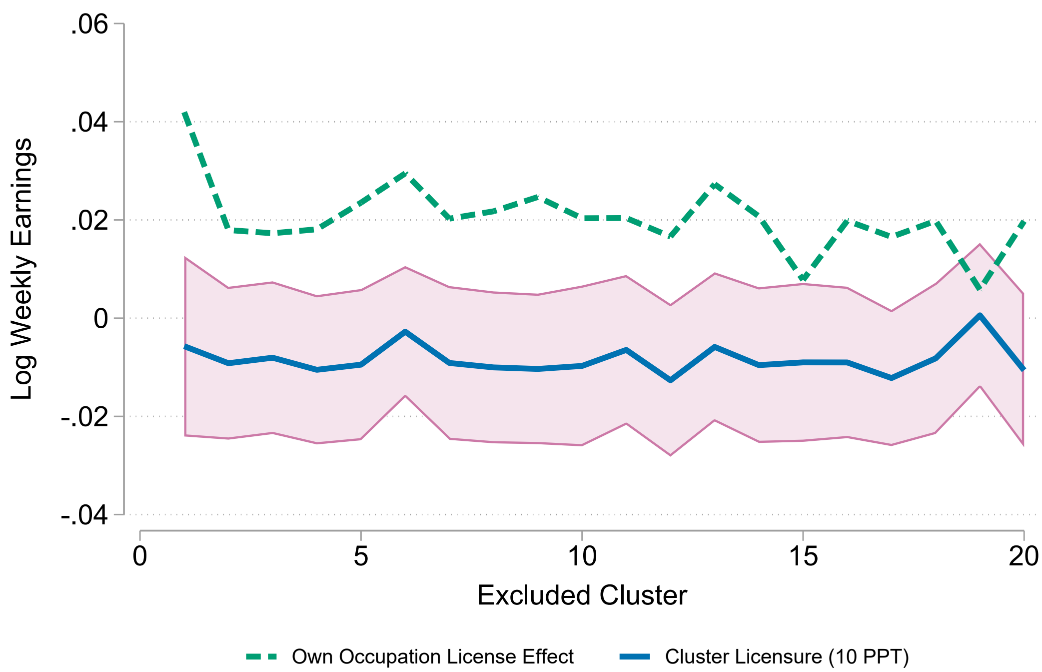

Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Coefficients are generated from the boundary discontinuity design detailed in Equation 1. Standard errors are clustered at the occupation level. Estimates include fixed effects for occupation, state, and boundary. 95% confidence intervals in red. Spillover coefficients are based on 10 percentage point increases in licensure of an occupation's cluster outside their own occupation for 20 total defined clusters, eliminating one cluster at a time.

Figure A7: Earnings Effects, Sequentially Removing Clusters
By Gender, 20 Clusters

Panel A: Women



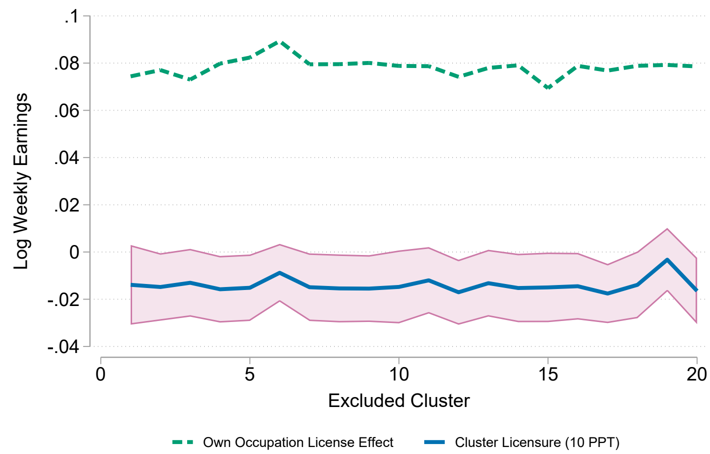

Panel B: Men



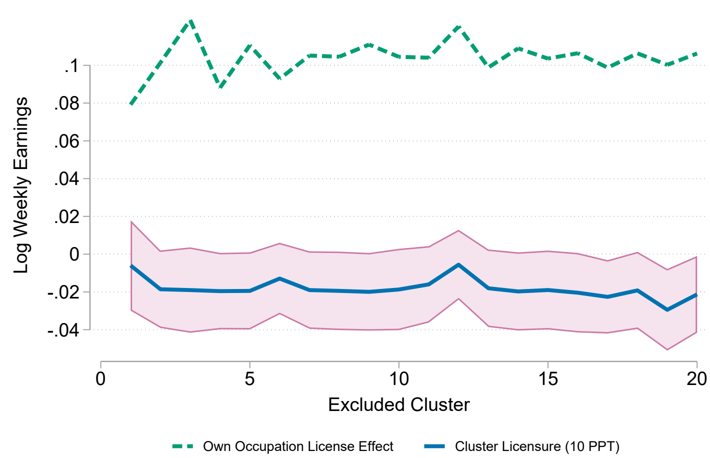

Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Coefficients are generated from the boundary discontinuity design detailed in Equation 1. Standard errors are clustered at the occupation level. Estimates include fixed effects for occupation, state, and boundary. 95% confidence intervals in red. Spillover coefficients are based on 10 percentage point increases in licensure of an occupation's cluster outside their own occupation for 20 total defined clusters, eliminating one cluster at a time.

Figure A8: Earnings Effects, Sequentially Removing Clusters
By Race/Ethnicity, 20 Clusters

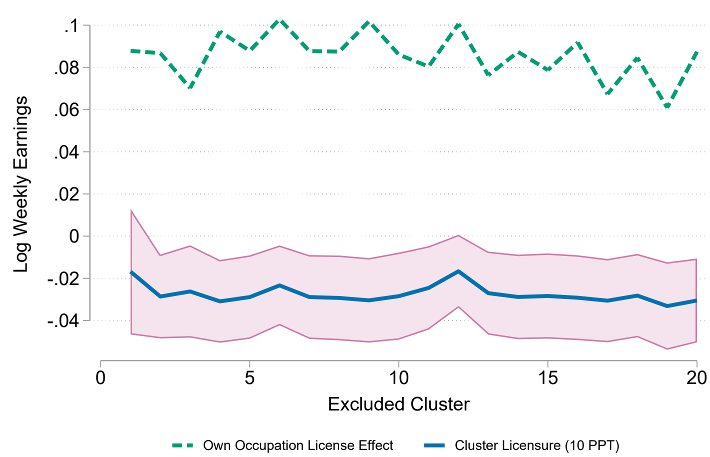
Panel A: Non-Hispanic White




Panel B: Hispanic

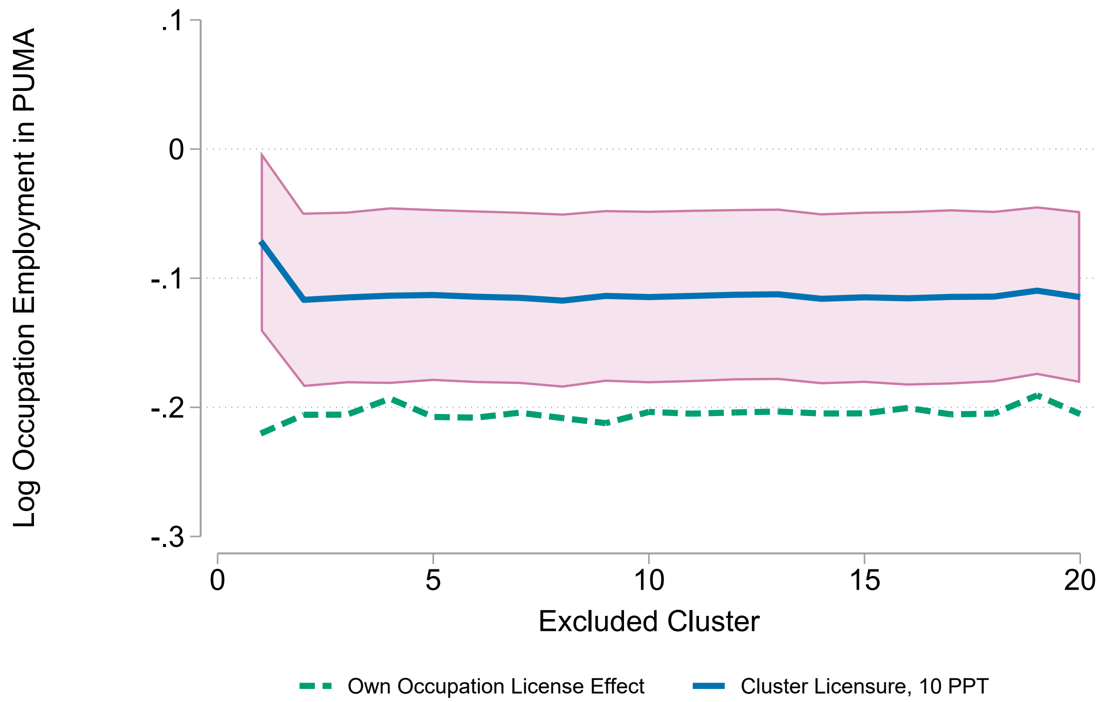



Panel C: Non-Hispanic Black



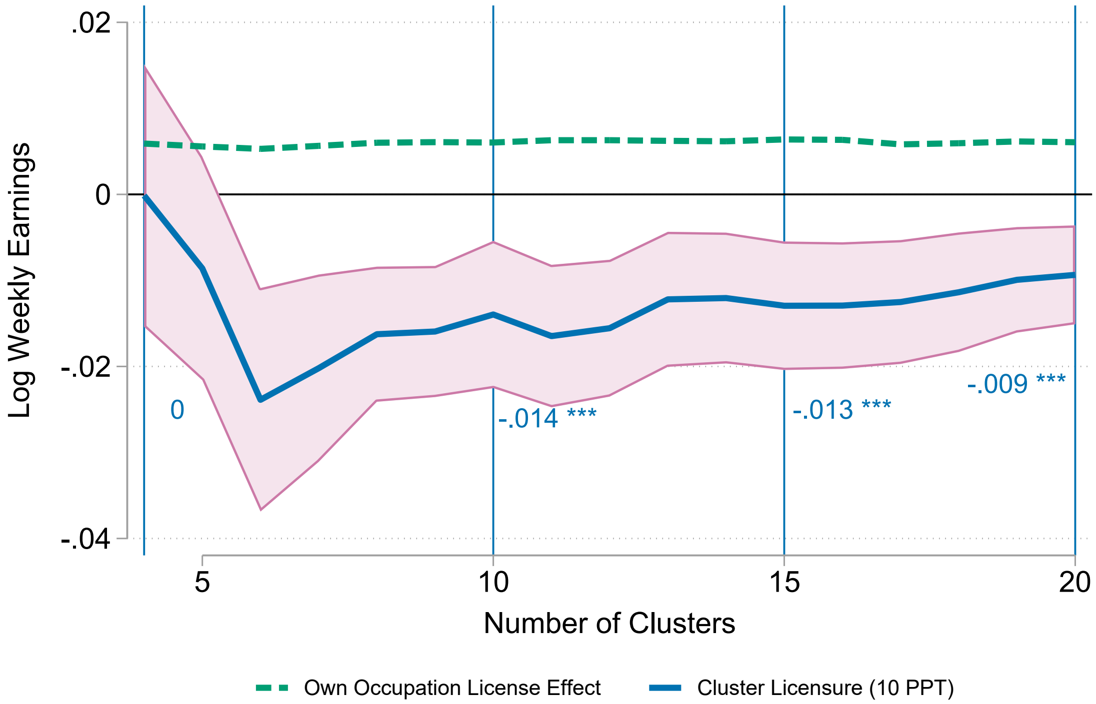

Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Coefficients are generated from the boundary discontinuity design detailed in Equation 1. Standard errors are clustered at the occupation level. Estimates include fixed effects for occupation, state, and boundary. 95% confidence intervals in red. Spillover coefficients are based on 10 percentage point increases in licensure of an occupation's cluster outside their own occupation for 20 total defined clusters, eliminating one cluster at a time.

Figure A9: Employment Effects, Sequentially Removing Clusters
20 Clusters




Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Coefficients are generated from the boundary discontinuity design detailed in Equations 3 and 4. Standard errors are clustered at the occupation level. Estimates include fixed effects for occupation, state, and boundary. 95% confidence intervals in red. Spillover coefficients are based on a 10 percentage point increase in licensure of an occupation's cluster outside their own occupation for 20 total defined clusters, eliminating one cluster at a time.

Figure A10: Log Earnings Effects using Northwestern Licensing Database




Source: Author's calculations of ACS, O*NET, and Northwestern Licensing Database (NLD) data.
Notes: *** p<0.01, ** p<0.05, * p<0.1. Coefficients are generated from the boundary discontinuity design detailed in Equation 1 using the Northwestern Licensing Database (NLD). Spillover coefficients are based on a 10 percentage point increase in licensure of an occupation's cluster outside their own occupation in the NLD. Standard errors are clustered at the occupation level. 95% confidence intervals in red. Models include occupation, state, and boundary fixed effects and controls for race/ethnicity, sex, age, and age squared. Vertical bars and coefficients are for clusters at 4, 10, 15, and 20.

Table A1: Components of Latent Skill Measurements

| Occupational Skill Area | O*NET Variables |
|---|---|
| Non-Routine, Cognitive, Analytical | "Analyzing data/information"<br>"Thinking creatively"<br>"Interpreting information for others" |
| Non-Routine, Cognitive, Interpersonal | "Establishing and maintaining personal relationships"<br>"Guiding, directing and motivating subordinates"<br>"Coaching/developing others" |
| Non-Routine, Manual,<br>Physical Adaptability | "Operating vehicles, mechanized devices, or equipment"<br>"Spend time using hands to handle, control or feel objects, tools or controls"<br>"Manual dexterity"<br>"Spatial orientation" |
| Routine, Cognitive | "Importance of repeating the same tasks"<br>"Importance of being exact or accurate"<br>"Structured v. Unstructured work (reverse)" |
| Routine, Manual | "Pace determined by speed of equipment"<br>"Controlling machines and processes"<br>"Spend time making repetitive motions" |
| Non-Routine, Interpersonal Adaptability | "Social Perceptiveness" |

Source: Version 22.0 of the O*NET database (2017) and Acemoglu and Autor (2011).

Table A2: Correlations with Cluster Share Licensed Outside Focal Occupation

| VARIABLES | (1) Unconditional | (2) Conditional |
|---|---|---|
| Own-Occupation Licensure | 0.303 | 0.0525 |
| Total PUMA Employment | 0.0504 | 0.00372 |
| PUMA Mean Earnings | 0.0995 | -0.0171 |
| State Top Marginally Tax Rate | -0.00975 | 3.65E-07 |
| Metropolitan Area | 0.0215 | 0.00356 |
| Pr(Agriculture, forestry, fishing, and hunting) | 0.0658 | -0.000761 |
| Pr(Mining) | -0.0218 | 0.0106 |
| Pr(Construction) | -0.042 | -0.000961 |
| Pr(Manufacturing) | -0.124 | -0.00754 |
| Pr(Wholesale and retail trade) | -0.177 | 0.0026 |
| Pr(Transportation and utilities) | -0.0267 | 0.00174 |
| Pr(Information) | 0.0635 | 0.00278 |
| Pr(Financial activities) | 0.104 | 0.000262 |
| Pr(Professional and business services) | 0.0647 | -0.00115 |
| Pr(Educational and health services) | 0.189 | 0.000429 |
| Pr(Leisure and hospitality) | -0.11 | 0.00349 |
| Pr(Other services) | -0.0139 | -0.00178 |
| Pr(Public administration) | 0.0992 | -0.00175 |
| Duncan Socioeconomic Index (SEI) | 0.6 | -0.00212 |
| Hauser and Warren Socioeconomic Index (HWSEI) | 0.648 | -0.000733 |
| Occupational income Score (Occscore) | 0.454 | -0.00161 |
| Siegel Occupational Prestige Score | 0.561 | -0.00224 |
| Nakao and Treas Occupational Prestige Score | 0.603 | -0.00184 |
| Married | 0.158 | 0.000848 |
| Spouse in the Labor Force | 0.137 | -0.00289 |
| Public Assistance Income | -0.0182 | 0.00101 |

Source: Author's calculations of ACS, CPS, and O*NET data.
Notes: Unconditional correlations are the pairwise correlation between licensing exposure in the focal occupation and individual or local area characteristics. Conditional correlations are for the residuals after regressing each variable separately on fixed effects for occupation, state, and state boundary. Clusters are based on description in Section 3.1. ACS samples are from 2014-2017 corresponding with CPS individual licensing data from 2015-2018. Socioeconomic and prestige indexes come from Duncan and Reiss (1961); Hauser and Warren (1997); Ruggles et al. (2019); Siegel (1971); Nakao and Treas (1994), respectively.

Table A3: Coefficients of Log Weekly Earnings by Subgroup at 20 Clusters

| VARIABLES | (1)<br>Own-Occupation Effect | (2)<br>Cluster Spillover Effect (10 PPT) |
|---|---|---|
| Overall | 0.0777*** | -0.0155** |
| | (0.0196) | (0.00677) |
| Men | 0.0201 | -0.00858 |
| | (0.0251) | (0.00766) |
| Women | 0.172*** | -0.0279*** |
| | (0.0298) | (0.00786) |
| NH-White | 0.0785*** | -0.0141** |
| | (0.0212) | (0.00703) |
| NH-Black | 0.0858** | -0.0278*** |
| | (0.0367) | (0.0100) |
| Hispanic | 0.104*** | -0.0183* |
| | (0.0345) | (0.0103) |
| NH-Black, Native Born | 0.0765** | -0.0255** |
| | (0.0376) | (0.0101) |
| NH-Black, Foreign Born | 0.167*** | -0.0427*** |
| | (0.0455) | (0.0108) |
| Hispanic, Native Born | 0.197*** | -0.0120 |
| | (0.0470) | (0.0113) |
| Hispanic, Foreign Born | -0.0948 | -0.0265** |
| | (0.0730) | (0.0106) |
| Observations | 1,765,288 | 1,765,288 |
| R-squared | 0.455 | 0.455 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: Author's calculations of ACS, CPS, and O*NET data.
Notes: Estimates correspond to those in Figure 5. Standard errors clustered at the occupation level. Estimates include fixed effects for occupation, state, and boundary.

Table A4: Coefficients of Log Weekly Earnings by Subgroup by Cluster Numbers

| VARIABLES | (1)<br>Own-Occupation Effect | (2)<br>Cluster Spillover - 4 | (3)<br>Cluster Spillover - 10 | (4)<br>Cluster Spillover - 15 | (5)<br>Cluster Spillover - 20 |
|---|---|---|---|---|---|
| Men | 0.0201 | 0.0208 | -0.0145* | -0.0146 | -0.0146 |
| | (0.0251) | (0.0126) | (0.00807) | (0.00986) | (0.00986) |
| Women | 0.172*** | -0.0208** | -0.0293*** | -0.0334*** | -0.0334*** |
| | (0.0298) | (0.0102) | (0.00878) | (0.00932) | (0.00932) |
| NH White | 0.0785*** | -0.00259 | -0.0207*** | -0.0211** | -0.0141** |
| | (0.0212) | (0.00854) | (0.00669) | (0.00882) | (0.00703) |
| NH Black | 0.0858** | -0.00337 | -0.0308*** | -0.0352*** | -0.0278*** |
| | (0.0367) | (0.0159) | (0.0113) | (0.0115) | (0.0100) |
| Hispanic | 0.104*** | 0.0214 | -0.0159 | -0.0271** | -0.0183* |
| | (0.0345) | (0.0156) | (0.0104) | (0.0117) | (0.0103) |
| NH Black, Native Born | 0.0754** | -0.00883 | -0.0268*** | -0.0286*** | -0.0242*** |
| | (0.0342) | (0.0123) | (0.00962) | (0.00912) | (0.00858) |
| NH Black, Foreign Born | 0.166*** | -0.0275* | -0.0446*** | -0.0453*** | -0.0415*** |
| | (0.0439) | (0.0145) | (0.0102) | (0.00926) | (0.00905) |
| Hispanic, Native Born | 0.196*** | 0.0216* | -0.00923 | -0.0163* | -0.0108 |
| | (0.0449) | (0.0111) | (0.00885) | (0.00977) | (0.00931) |
| Hispanic, Foreign Born | -0.0960 | -0.00501 | -0.0272*** | -0.0324*** | -0.0251*** |
| | (0.0736) | (0.0135) | (0.00952) | (0.00963) | (0.00882) |
| Observations | 1,765,288 | 1,765,288 | 1,765,288 | 1,765,288 | 1,765,288 |
| R-squared | 0.455 | 0.455 | 0.455 | 0.455 | 0.455 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: Author's calculations of ACS, CPS, and O*NET data.
Notes: Estimates correspond to Equation 2 by subgroups over select numbers of skill clusters for a 10 percentage point increase in cluster licensure. Standard errors clustered at the occupation level. Estimates include fixed effects for occupation, state, and boundary. The own-occupation effect estimates in column 1 are from the model estimated with 20 skill clusters.

Table A5: Contribution of Sorting to Observed Effects at 20 Clusters

| | |
|---|---|
| Baseline Effect, No Demographics ($\beta$) | -0.01655 |
| (SE) [$R^2$] | (0.00683) [0.364] |
| Controlled Effect ($\tilde{\beta}$) | -0.01284 |
| (SE) [$\tilde{R}^2$] | (0.00618) [0.470] |
| Identified Set [$\tilde{\beta}$, $\beta^*\|(1.3 * \tilde{R}^2, \delta = 1$]) | [-0.01284, -0.00791] |
| $\delta$ for $\beta = 0\|[1.3\tilde{R}^2]$ | 2.604 |
| Observed + Unobserved at $\delta = 1$, $1.3 * \tilde{R}^2$ | -0.00864 |
| Observed + Unobserved Share at $\delta = 1$, $1.3 * \tilde{R}^2$ | 52 % |

Source: Author's calculations of ACS, CPS, and O*NET data.
Notes: Procedures follow those suggested in Oster (2019). Controlled effects account for gender, age, education (linear in years), race/ethnicity, and nativity composition. ($1.3 * \tilde{R}^2 = 0.6117$).

Table A6: Estimates at 20 Clusters Controlling for Employment

| VARIABLES | (1) Log(Weekly Earnings) | (2) Log(Weekly Earnings) |
|---|---|---|
| Own Occupation Effect | 0.0782*** | 0.0775*** |
| | (0.0196) | (0.0196) |
| Cluster Exposure (10 PPT) | -0.0150** | -0.0159** |
| | (0.00675) | (0.00678) |
| Control for Log(PUMA Employment) | X | |
| Control for Log(PUMA-Cluster Employment) | | X |
| Observations | 1,765,288 | 1,765,288 |
| R-squared | 0.456 | 0.455 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: Author's calculations of ACS, CPS, and O*NET data.
Notes: Estimates follow Equation 1 but include controls for the log of total PUMA employment (column 1) and the log of total cluster employment (column 2). Standard errors are clustered at the occupation level. Estimates include fixed effects for occupation, state, and boundary.

Table A7: Coefficients of Log Weekly Earnings by Cluster Numbers
Adding PUMA Fixed Effects

| VARIABLES | (1) Own-Occupation Effect (20 Clusters) | (2) Cluster Spillover - 4 | (3) Cluster Spillover - 10 | (4) Cluster Spillover - 15 | (5) Cluster Spillover - 20 |
|---|---|---|---|---|---|
| Overall | 0.0771*** | -0.00194 | -0.0183*** | -0.0197** | -0.0135** |
| | (0.0194) | (0.00725) | (0.00647) | (0.00833) | (0.00639) |
| Men | 0.0206 | 0.0180 | -0.0117 | -0.0118 | -0.00665 |
| | (0.0249) | (0.0120) | (0.00787) | (0.00961) | (0.00736) |
| Women | 0.169*** | -0.0234** | -0.0267*** | -0.0306*** | -0.0259*** |
| | (0.0295) | (0.00971) | (0.00825) | (0.00890) | (0.00748) |
| NH White | 0.0778*** | -0.00548 | -0.0180*** | -0.0183** | -0.0121* |
| | (0.0210) | (0.00751) | (0.00636) | (0.00850) | (0.00668) |
| NH Black | 0.0775** | -0.00493 | -0.0276** | -0.0319*** | -0.0253*** |
| | (0.0366) | (0.0155) | (0.0109) | (0.0112) | (0.00970) |
| Hispanic | 0.106*** | 0.0205 | -0.0121 | -0.0234** | -0.0156 |
| | (0.0338) | (0.0153) | (0.0102) | (0.0114) | (0.0100) |
| NH Black Native Born | 0.0648* | -0.00857 | -0.0236*** | -0.0254*** | -0.0216*** |
| | (0.0339) | (0.0119) | (0.00907) | (0.00870) | (0.00816) |
| NH Black Foreign Born | 0.168*** | -0.0271* | -0.0418*** | -0.0422*** | -0.0388*** |
| | (0.0421) | (0.0141) | (0.00983) | (0.00899) | (0.00873) |
| Hispanic Native Born | 0.194*** | 0.0231** | -0.00487 | -0.0120 | -0.00710 |
| | (0.0417) | (0.0109) | (0.00855) | (0.00946) | (0.00899) |
| Hispanic Foregn Born | -0.0922 | -0.00463 | -0.0244*** | -0.0297*** | -0.0231*** |
| | (0.0710) | (0.0131) | (0.00927) | (0.00926) | (0.00850) |
| Observations | 1,765,288 | 1,765,288 | 1,765,288 | 1,765,288 | 1,765,288 |
| R-squared | 0.462 | 0.462 | 0.462 | 0.462 | 0.462 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: Author's calculations of ACS, CPS, and O*NET data.
Notes: Estimates correspond to Equation 2 by subgroups over select numbers of skill clusters for a 10 percentage point increase in cluster licensure. Standard errors clustered at the occupation level. Estimates include fixed effects for occupation, state, boundary, and PUMA. The own-occupation effect estimates in column 1 are from the model estimated with 20 skill clusters.

# Appendix B. Data Challenges

As mentioned in Section 2, one challenge to studying occupational licensing is the availability of high-quality data measuring licensing at the state or national level. Redbird (2016, 2017) organized a list of licensing requirements back to the 1970s based on the text of current state statutes in order to measure the effects of licensing on wages (the Northwestern Licensing Database). More recent improvements in historical data collection, however, have brought to light two challenges when handling such measures of licensure: the mapping of licensing laws into occupation codes in commonly used classifications, and the updating of licensing requirements over time. Carollo (2020b) develops what is arguably a more complete dataset on licensing rules based on multiple sources and finds an average long-run within-occupation licensing premium of approximately 7% after occupations become licensed, which is nearly identical to my boundary discontinuity estimate for the "own-occupation" effect.

Mapping the text of licensing laws onto occupational definitions as they are surveyed and coded by statistical agencies creates an important measurement challenge. Many licenses cover only a small subset of workers in what would be considered a larger occupation category. For example, in Alabama, "anesthesiologist assistant" is a licensed occupation, whereas next door in Mississippi, it is not. Even at the level of six-digit Standard Occupational Classification (SOC) code, the narrowest level at which many agencies collect occupational data, "anesthesiologist assistant" is grouped together under the "physician assistant" code with other occupations such as "family practice physician assistant." "Physician assistant" itself is also separately licensed in both Mississippi and Alabama as a different occupation involving particular responsibilities (Vargo et al., 2020). Who exactly is "treated" by a license within the SOC code is, therefore, a noisy measure. Some licenses cover some occupations only in certain industries, making the broad application of a statute across an entire occupation even more complex. Because licensure is binary, measurement error will lead to attenuation of the estimates relative to the true effect, especially when aggregating to even coarser coding systems (e.g. Census occupation codes). This tension between statistical occupation categories and legal definitions is not rare, and, in fact, becomes more complex as the number of occupations increases. This attenuation bias may be exacerbated by the fixed effects models used in this literature.

In addition to the challenge of code mapping is the issue of dynamic updating of rules. According to Carollo (2020a), measurement quality of occupational licensing status based on current state statutes deteriorates going back in time without careful updates of contemporaneous state codes and session laws. This is because state laws on the books today may not reflect historical changes from when licensing rules were first imposed. Replacements of rules with updates and changes from session laws are not captured in currently available data such as the NLD. I discuss this particular challenge further in Appendix C.

# Appendix C. Extension: Time-Varying Measures of Licensure

Notwithstanding the concern that attenuation due to measurement error may increase in panel fixed effects models—particularly if measurement quality deteriorates going back in time—I estimate a repeated cross-sectional model using the Current Population Survey Outgoing Rotation Group (ORG) dataset from 1983 to 2017 coupled with the Northwestern Licensing Database using cluster assignments at 20 clusters. Here, identification of the spillover effects and the within-occupation effect of licensing comes from variation in new licensing laws across states over time within occupations and within skill clusters.

To construct this analysis dataset, I first crosswalk occupations over time to 2010 Census occupation code equivalents. I use OES employment weights in each year from 1983 to 2017 to translate licensing rules from six-digit SOC codes in the NLD into these 2010 Census occupation cells to generate the core treatment variable: the share of workers in the skill cluster outside the focal occupation that must be licensed under the statutes in each state-year-occupation cell. I then estimate a model for outcome $y$ (hourly wages) for worker $i$ in occupation $o$ in cluster $c$ in state $s$ in year $t$. This is similar to Equation 1 but I include occupation, state, and year fixed effects:

$$y_{iocst} = \beta_0 + \beta_1 LicensedShare_{ost} + \beta_2 LicensedShare^{-o}_{cst} + X_i'\beta_3 + \delta_o + \gamma_s + \tau_t + \varepsilon_{iocst} \quad \text{(C.1)}$$

In this setting, the natural experiment is that there is an expansion of licensing in the occupations related to the focal occupation in a particular state, and the model then tracks the changes in wages over time in the affected state in comparison to changes in all other US states holding constant time-invariant characteristics of the state and occupation.

As in my earlier analysis, I also interact these time-varying measures of licensing with indicators for different demographic groups to examine heterogeneous treatment effects. To demonstrate the effect of measurement error being exacerbated over time, I vary the time periods over which I estimate my models with start dates in 1983, 1994, 2001, and 2010. To model the employment spillover effects, I use the CPS to calculate employment rates in each state-occupation-year cell for the outcome of interest. I then estimate a similar repeated cross-sectional model to the earnings equation above with observations at the state-occupation-year level:

$$Log(EMP)_{ocst} = \beta_0 + \beta_1 LicensedShare_{ost} + \beta_2 LicensedShare^{-o}_{cst} + \delta_o + \gamma_s + \tau_t + \varepsilon_{ocst} \quad \text{(C.2)}$$

These estimates deserve a word of caution. According to Carollo (2020a), the NLD's use of policy enactment dates from citations in currently listed state statutes can result in left-censoring of licensing enactment, regardless of the scope of the license. If laws have been updated, replaced, or repealed over time, or if requirements changed between certification, registration, and full licensure, measurement error in the NLD becomes more pronounced going back in time the earlier an occupation was first regulated (regardless of type of regulation). Therefore, the estimates that include the earliest years are likely to be subject to significant attenuation bias, while more recent years are likely subject to less attenuation.[29] However, given the challenge of mapping licenses to Census occupation codes discussed in Section 2,

[29] There is an additional subtlety to mention: some licenses in the NLD only apply to occupations in particular industries, and any expansion or contraction of the set of industries required to have a license will not be captured in the NLD.

attenuation bias from this source of measurement error is likely to be present in all years.

The results for the spillover coefficients are in Table C1.[30] Panel A shows the results for log hourly wages. Across all demographic groups, there is an average of a 4-7% wage penalty (depending on sample start dates) when a cluster is fully licensed outside the focal occupation or a 0.4–0.7% penalty for a 10 percentage point increase in cluster licensure. As the sample becomes more recent in columns 3 and 4, the size of the penalty increases as the level of measurement error likely decreases.

Given the fact that the NLD measure of licensing exposure has a (noisier) standard deviation of 0.162 compared to a standard deviation of 0.11 in the boundary discontinuity sample, I can rescale these estimates to match in terms of effects per standard deviation unit. Therefore, in these estimates, an increase of one standard deviation in cluster licensing exposure after 2010 decreases wages by 1.1% compared to a reduction in weekly earnings of approximately 1.8% in the boundary discontinuity estimates. Interestingly, the ratio of these estimates of 0.61 across data sources is virtually identical to the correlation between the NLD measure of licensure and the CPS measure of licensure in 2015–2018 (0.6) when the two samples overlap.

Looking at heterogeneous treatment effects, I find similar patterns of penalty differentials as I find in my boundary discontinuity estimates. The wage penalty after 2001 for non-Hispanic white workers was 0.64% for a 10 percentage point change in cluster licensure (or 1% for a standard deviation change), while the effects for non-Hispanic black and Hispanic workers were nearly 1.5% and 1%, respectively (or 2.4% and 1.6% for a standard deviation change). This doubling of the negative wage effect across racial groups closely matches what I find in Panel B of Figure 5 in my cross-sectional estimates. I also find similar patterns for wage penalties across gender lines: the spillover effect for women is larger than the spillover effect for men.

In Panel B of Table C1, I examine the employment effects of NLD licensure. A 10 percentage point increase in cluster licensure outside the focal occupation reduces employment in the focal occupation in the state by approximately 0.3–2.1% depending on the sample dates. The estimates in Panel B of Table C1 for my time-varying model are, therefore, smaller than those in my boundary discontinuity design.

In summary, although considerable measurement error may bias these estimates toward zero, I find consistent evidence of significant negative spillovers in hourly wages on similarly skilled occupations in the CPS using variation in licensing rules over time. This also extends to similar patterns in the relative size of the heterogeneous spillover effects across race/ethnicity and gender when compared to my boundary discontinuity estimates. I also find significant negative spillover effects of cluster licensure on employment in the focal occupation, though these estimates are subject to significantly more uncertainty. Overall, this exercise supports the totality of my boundary discontinuity results regarding spillovers.

[30] The own-occupation effects ($\beta_1$) are small and imprecisely estimated for all repeated cross-sectional estimates, so I omit these from the table. Nevertheless, the estimates range from -0.005 to 0.02 across specifications. Because occupation cells are smaller than cluster cells, measurement error in licensure will be more consequential in terms of attenuation bias for the own-occupation effects. I see the same phenomenon with the NLD in Figure A10, where the own-occupation effect in my boundary discontinuity design changes far more than the spillover effect.

Table C1: CPS Repeated Cross-Section Estimates from the NLD by Sample Start Year

| VARIABLES | (1)<br>1983 | (2)<br>1994 | (3)<br>2001 | (4)<br>2010 |
|---|---|---|---|---|
| Panel A: Log Wage Effects | | | | |
| Avg Cluster Spillover Effect (10 PPT) | -0.00398** | -0.00452*** | -0.00699*** | -0.00715*** |
| | (0.00167) | (0.00131) | (0.00135) | (0.00159) |
| Spillover Effect: NH White | -0.00338* | -0.00361** | -0.00640*** | -0.00661*** |
| | (0.00190) | (0.00149) | (0.00168) | (0.00193) |
| Spillover Effect: NH Black | -0.0106*** | -0.0128*** | -0.0146*** | -0.0139*** |
| | (0.00340) | (0.00360) | (0.00342) | (0.00349) |
| Spillover Effect: Hispanic | -0.00655** | -0.00761** | -0.00944*** | -0.0105*** |
| | (0.00259) | (0.00301) | (0.00301) | (0.00325) |
| Spillover Effect: Women | -0.00265 | -0.00498 | -0.00770** | -0.00825** |
| | (0.00387) | (0.00340) | (0.00310) | (0.00327) |
| Spillover Effect: Men | -0.00447** | -0.00427** | -0.00665*** | -0.00665*** |
| | (0.00198) | (0.00178) | (0.00161) | (0.00169) |
| Observations | 5,364,304 | 3,694,393 | 2,664,510 | 1,208,935 |
| Panel B: Log Employment Effects | | | | |
| Avg Cluster Spillover Effect (10 PPT) | -0.00345 | 0.00317 | -0.00635 | -0.0212*** |
| | (0.00726) | (0.00901) | (0.00757) | (0.00769) |
| Observations | 505,042 | 368,903 | 275,661 | 130,306 |
| Occupation FE | Yes | Yes | Yes | Yes |
| State FE | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: Author's calculations of CPS Outgoing Rotation Group, O*NET, and Northwestern Licensing Database (NLD) data.
Notes: Coefficients are generated from a repeated cross-sectional regression of log wages on individual characteristics and state-level licensing shares over time from the NLD. Estimates include occupation, state, and year fixed effects. Standard errors are clustered at the occupation level. Spillover coefficients are based on a 10 percentage point increase in licensure of an occupation's cluster outside their own occupation in the NLD. OES employment weights are used to create weighted averages of the share of 2010 Census occupation codes that must have a license in each state and year according to the statutes in the NLD as mapped to six-digit SOC codes. Employment is measured in the state-occupation cell.

# Appendix D. Clustering Appendix

Choice of Inputs: There are hundreds of skill, ability, and contextual variables that are a part of the O*NET database. In order to extract meaningful relationships between occupations, it is important to narrow down the set of candidate dimensions over which to cluster them. Failure to reduce the number of variables considered results in the "curse of dimensionality," particularly when attempting a clustering exercise.

One clear option for reducing dimensionality is a principal component analysis (PCA). Below in Figure D1, I present visual comparisons of the dissimilarity matrix using the first six principal components using all of the "skills" in the O*NET database along with the median wage of the occupation. I similarly present the cluster mapping of the first six principal components over all "context" variables in the O*NET database. This dissimilarity value is one minus the Pearson correlation coefficient over all 7 attributes.

The figure is a colored representation of the dissimilarity matrix. Each occupation is represented on both axes, and the diagonal of the matrix is the distance between each occupation with itself (zero). Darker blue regions represent small differences between occupations along the dimensions considered. In other words, these occupations are highly correlated. Lighter colors and white regions represent occupation pairs that are uncorrelated. The darkest orange areas represent occupations that are highly *negatively* correlated and therefore have the largest distance between them. Importantly, more consistent dark blue and dark orange regions represent more efficient separations or classifications for occupations because the characteristics better capture similarities and differences between occupations.

In turn, clustering over the skills in Acemoglu and Autor (2011) leads to more compact clusters. Table D2 below compares the "height" of the various dendrogram connections between occupations along the three measures considered in Figure D1. The heights represent the correlative distance between the two objects when they merge into a single cluster. Lower values of this height measure indicate tighter or more compact cluster definitions.

Overall, the measures in Acemoglu and Autor (2011) generate more compact clusters and greater separation between clusters than when clustering over the principal components of the O*NET data. The computer science literature bears this out, stating that in many cases, the principal components of the data, while capturing the greatest variation across the attributes, do not capture the *cluster* structure of the data as well as using a subset of the variables (Yeung and Ruzzo, 2001). As a result, what one would consider the "data-driven" approach to choosing attributes over which to cluster yields worse cluster matching. The alternative is either an ad hoc or a theory-driven choice of clustering attributes. The theoretical and empirical literature on worker skills supports the framework in my analysis, and the empirical exercise I present justifies using this approach over the principal-component approach.

In the clustering approach, individual component outliers may influence the measure of "distance" between prospective cluster members. However, my choice of clustering algorithm and "average linkage distance" (also called "unweighted pair-group method with arithmetic means" (UPGMA)) to form clusters is relatively robust to outliers within clusters. In addition, the inclusion of seven components helps mitigate outliers on any single component.

Optimal Number of Clusters:

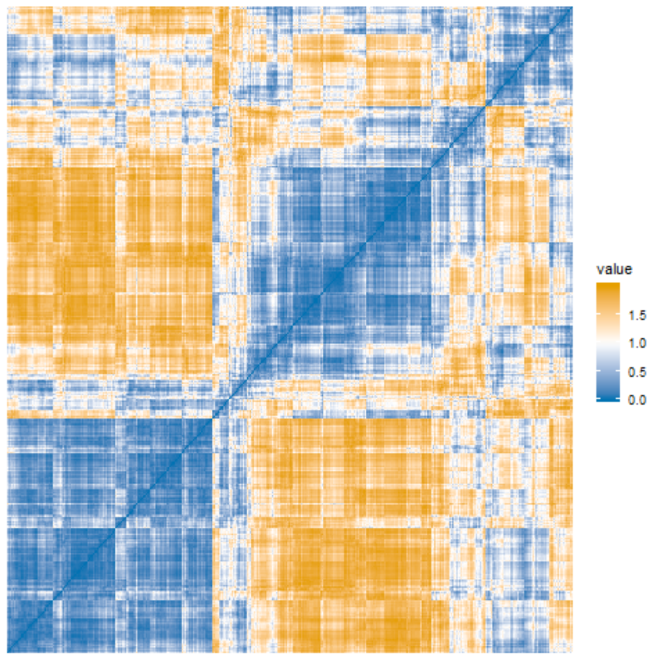
The number of clusters in the final analysis is an important choice variable. I use four validation measures common to clustering applications to validate the optimal number of clusters: Silhouette (Rousseeuw, 1987); Dunn's index (Dunn, 1974); SD index (Halkidi et al., 2000); and the C index (Hubert and Levin, 1976), though there are dozens from which to select. The first two measures are based on maximizing their index values, while the latter two are based on minimizing their values. It is also useful to look for structural breaks in the index values.

Figure D2 shows the results using these four measures. Panel A suggests that the optimal number of clusters is likely below 18, as the index bottoms out above this number, but is markedly higher at lower numbers of clusters and for clusters above 23. Panel B strongly suggests the optimal number of clusters is somewhere between 14 and 20. Panel C suggests the optimum ought to be below 13 or perhaps 19-22. Lastly, Panel D suggests the optimum is either 12–13 or 23–30, although the index values for 14–23 are stable and relatively low. These tests justify the use of 20 clusters in my analysis, though the results are robust to different possible choices of the optimal number.

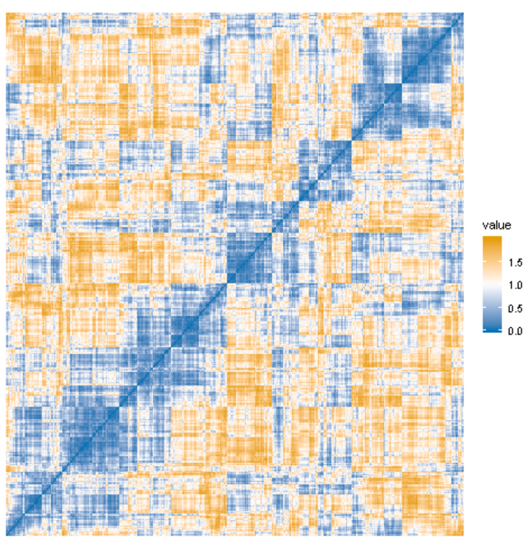
A Parametric Approach The HAC clustering algorithm imposes a non-parametric structure on the relationship between occupations based on skills. As an alternative to specifying the cluster structure without wages in the algorithm, which may introduce misclassification measurement error, I create an index of skill-distance weighted exposure to licensure (see Figure D3). I construct this by calculating licensure rates for every occupation-state cell and then define exposure to licensure from other occupations as the licensure rate of every other occupation in the state weighted by the skill similarity (Pearson correlation) of each occupation (excluding national wages). This imposes a linear parametric structure on skill distance rather than the non-parametric structure of the clustering approach, requiring stronger assumptions about the decay rate of relevant relative skill distance. In brief, this exercise treats all occupations as if they were in one "cluster" but directly weights licensing exposure from other occupations by how closely the licensed occupations match the focal occupation in skill content. The results generally confirm the rest of my analysis, but with relatively wider standard errors (that make heterogeneous effects noisy). The noisiness of the measure comes because possibly irrelevant alternatives are included in the treatment variable. This conceptually distinct approach to skill similarity and licensing exposure nevertheless leads to the same conclusion that there are significant negative spillovers from occupational licensing.

Figure D1: Correlative Distance Values Between Occupations

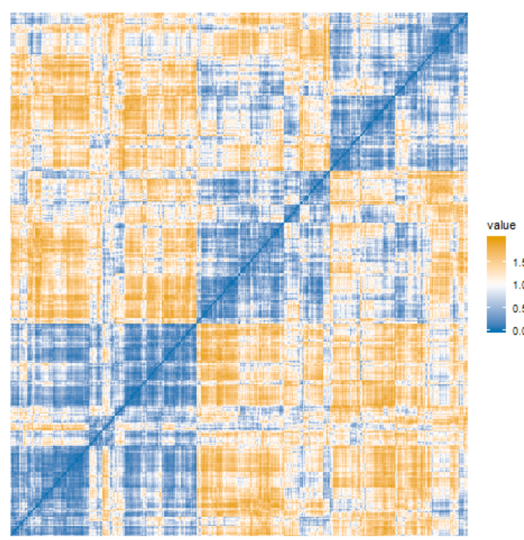
Panel A: Skills in Acemoglu and Autor (2011)




Panel B: PCA on O*NET "Skills"

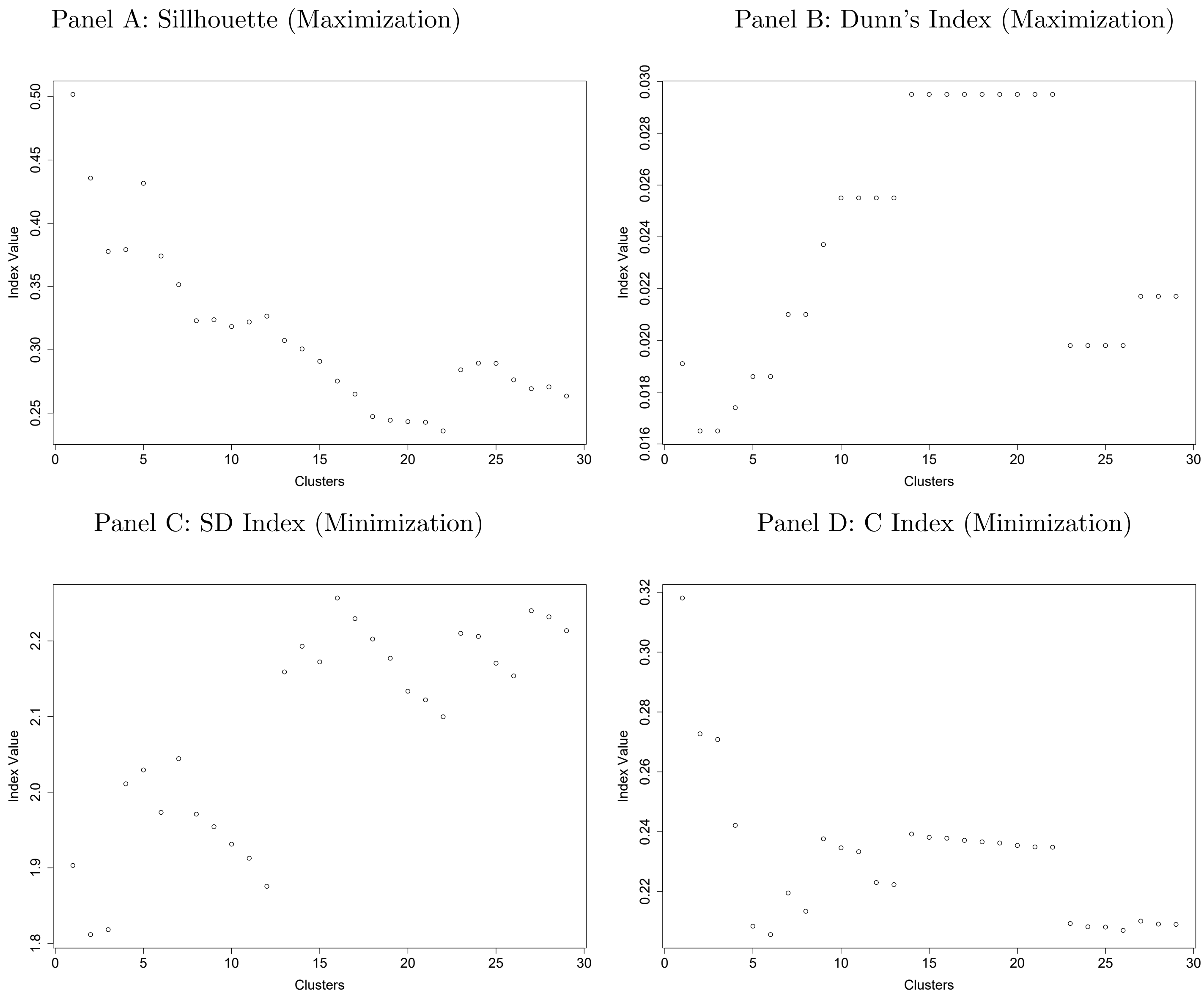



Panel C: PCA on O*NET "Context"




Source: Author's calculations of O*NET data.
Notes: Each panel is a matrix of the correlative distance in the seven attributes between each occupation pair, which is one minus the Pearson correlation coefficient. Darker blue represents the smallest differences between occupations along the dimensions considered, while the darkest orange colors represent the largest possible differences between the occupations. More consistent dark blue and dark orange regions represent better separations or classifications for occupations.

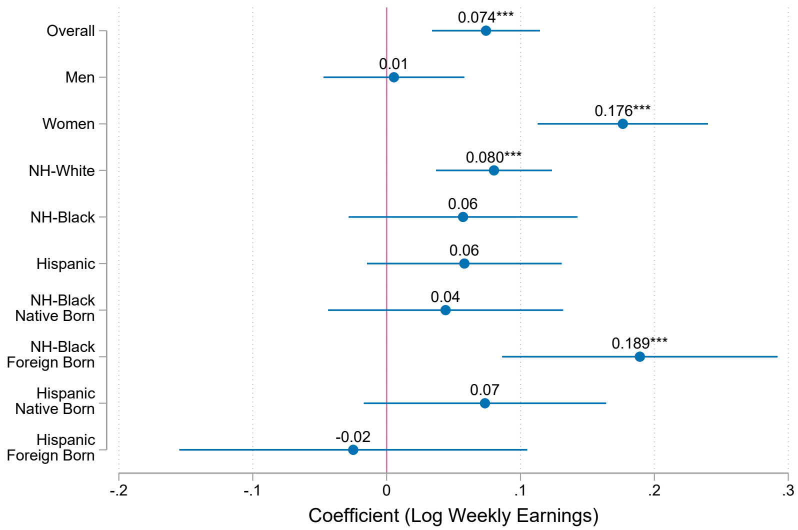
Figure D2: Cluster Validation Exercises



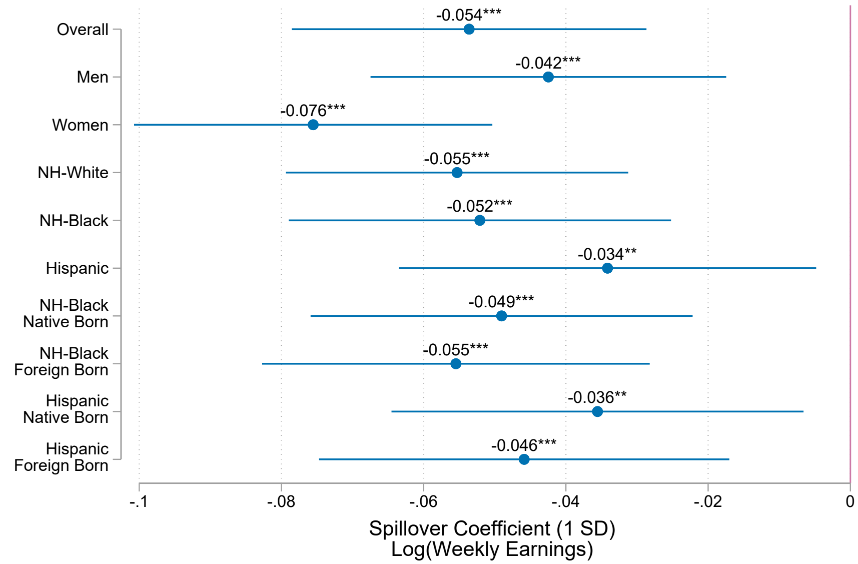

Source: Author's calculations of O*NET skills data following six skills in Acemoglu and Autor (2011) and median log wage.
Notes: Clusters are generated using the HAC approach detailed in Section 3.1.

Figure D3: Coefficients of Log Weekly Earnings by Skill-Distance Weighted Exposure

Panel A: Own Earnings Effects




Panel B: Skill-Distance Weighted Spillover Effects




Source: Author's calculations of ACS, O*NET, and CPS licensing data.

Notes: *** p<0.01, ** p<0.05, * p<0.1. Coefficients are generated from the boundary discontinuity design detailed in Equation 1, except for replacing cluster licensing exposure with skill-weighted exposure. Exposure to licensure is defined as licensing rates in every occupation in each state weighted by the skill similarity of each occupation to each other. Dots represent point estimates, and the bands represent the 95% confidence intervals clustered at the occupation level. Spillover coefficients are based on a one standard deviation increase in skill-weighted licensing exposure for all occupations in the state. Models include occupation, state, and boundary fixed effects and controls for race/ethnicity, sex, age, and age squared.

These estimates imply effects that are larger than the size of my estimates based on skill clusters for a standard deviation change in exposure (-5.4% vs -1.6%).

Table D1: Top 5 Focal Occupations by Cluster

| Occupation | Cluster | Freq | Rank |
|---|---|---|---|
| Managers, Nec (Including Postmasters) | 1 | 161944 | 1 |
| Elementary And Middle School Teachers | 1 | 161405 | 2 |
| Accountants And Auditors | 1 | 75768 | 3 |
| Postsecondary Teachers | 1 | 59665 | 4 |
| Computer Scientists And Systems Analysts/Network Systems Analysts/Web Developers | 1 | 59412 | 5 |
| Farmers, Ranchers, And Other Agricultural Managers | 2 | 15456 | 1 |
| Heating, Air Conditioning, And Refrigeration Mechanics And Installers | 2 | 13725 | 2 |
| Bus And Truck Mechanics And Diesel Engine Specialists | 2 | 12739 | 3 |
| Electronic Home Entertainment Equipment Installers And Repairers | 2 | 1356 | 4 |
| Home Appliance Repairers | 2 | 1092 | 5 |
| Chefs And Cooks | 3 | 90676 | 1 |
| Nursing, Psychiatric, And Home Health Aides | 3 | 71725 | 2 |
| Waiters And Waitresses | 3 | 70596 | 3 |
| Personal Care Aides | 3 | 44406 | 4 |
| Food Service And Lodging Managers | 3 | 36221 | 5 |
| Secretaries And Administrative Assistants | 4 | 136243 | 1 |
| Customer Service Representatives | 4 | 98746 | 2 |
| Receptionists And Information Clerks | 4 | 41797 | 3 |
| Medical Assistants And Other Healthcare Support Occupations, Nec | 4 | 33470 | 4 |
| Security Guards And Gaming Surveillance Officers | 4 | 33186 | 5 |
| Software Developers, Applications And Systems Software | 5 | 47609 | 1 |
| Computer Programmers | 5 | 16904 | 2 |
| Engineering Technicians, Except Drafters | 5 | 16298 | 3 |
| Paralegals And Legal Assistants | 5 | 15156 | 4 |
| Claims Adjusters, Appraisers, Examiners, And Investigators | 5 | 11561 | 5 |
| Police Officers And Detectives | 6 | 35700 | 1 |
| Editors, News Analysts, Reporters, And Correspondents | 6 | 9325 | 2 |
| Biological Scientists | 6 | 3498 | 3 |
| Construction And Building Inspectors | 6 | 3318 | 4 |
| Private Detectives And Investigators | 6 | 3071 | 5 |
| Radio And Telecommunications Equipment Installers And Repairers | 7 | 6138 | 1 |
| Surveying And Mapping Technicians | 7 | 2584 | 2 |
| Transportation Inspectors | 7 | 1690 | 3 |
| Electrical And Electronics Repairers, Transportation Equipment, And Industrial And Utility | 7 | 700 | 4 |
| Geological And Petroleum Technicians, And Nuclear Technicians | 7 | 674 | 5 |
| Data Entry Keyers | 8 | 13733 | 1 |
| Production, Planning, And Expediting Clerks | 8 | 13599 | 2 |
| Dental Assistants | 8 | 11273 | 3 |
| Agricultural And Food Science Technicians | 8 | 1763 | 4 |
| Prepress Technicians And Workers | 8 | 992 | 5 |
| Office Clerks, General | 9 | 51245 | 1 |

| | | | |
|---|---|---|---|
| Bookkeeping, Accounting, And Auditing Clerks | 9 | 47800 | 2 |
| Billing And Posting Clerks | 9 | 19213 | 3 |
| Diagnostic Related Technologists And Technicians | 9 | 15015 | 4 |
| Insurance Claims And Policy Processing Clerks | 9 | 14547 | 5 |
| Life, Physical, And Social Science Technicians, Nec | 10 | 8709 | 1 |
| Animal Control | 10 | 302 | 2 |
| Sales Representatives, Services, All Other | 11 | 23563 | 1 |
| Actors, Producers, And Directors | 11 | 6821 | 2 |
| Advertising Sales Agents | 11 | 6014 | 3 |
| Community And Social Service Specialists, Nec | 11 | 3616 | 4 |
| Eligibility Interviewers, Government Programs | 11 | 3191 | 5 |
| Cashiers | 12 | 106546 | 1 |
| Stock Clerks And Order Fillers | 12 | 59355 | 2 |
| Maids And Housekeeping Cleaners | 12 | 38977 | 3 |
| Food Preparation Workers | 12 | 31460 | 4 |
| Shipping, Receiving, And Traffic Clerks | 12 | 23018 | 5 |
| First-Line Supervisors Of Sales Workers | 13 | 156541 | 1 |
| Retail Salespersons | 13 | 111932 | 2 |
| Childcare Workers | 13 | 31616 | 3 |
| Recreation And Fitness Workers | 13 | 14611 | 4 |
| Athletes, Coaches, Umpires, And Related Workers | 13 | 9313 | 5 |
| First-Line Supervisors Of Construction Trades And Extraction Workers | 14 | 30855 | 1 |
| First-Line Supervisors Of Mechanics, Installers, And Repairers | 14 | 12044 | 2 |
| Photographers | 14 | 4030 | 3 |
| First-Line Supervisors Of Fire Fighting And Prevention Workers | 14 | 2197 | 4 |
| Electricians | 15 | 30869 | 1 |
| Aircraft Mechanics And Service Technicians | 15 | 7138 | 2 |
| Tool And Die Makers | 15 | 2481 | 3 |
| Precision Instrument And Equipment Repairers | 15 | 2166 | 4 |
| Security And Fire Alarm Systems Installers | 15 | 2037 | 5 |
| Painters, Construction And Maintenance | 16 | 14140 | 1 |
| Firefighters | 16 | 12405 | 2 |
| Dishwashers | 16 | 9838 | 3 |
| Roofers | 16 | 5876 | 4 |
| Electrical Power-Line Installers And Repairers | 16 | 5415 | 5 |
| Agricultural Workers, Nec | 17 | 34934 | 1 |
| Bus And Ambulance Drivers And Attendants | 17 | 20930 | 2 |
| Crossing Guards | 17 | 1671 | 3 |
| Motor Vehicle Operators, All Other | 17 | 1133 | 4 |
| First-Line Supervisors Of Production And Operating Workers | 18 | 39633 | 1 |
| First-Line Supervisors Of Housekeeping And Janitorial Workers | 18 | 7236 | 2 |
| Counter Attendant, Cafeteria, Food Concession, And Coffee Shop | 18 | 5017 | 3 |
| First-Line Supervisors Of Landscaping, Lawn Service, And Groundskeeping Workers | 18 | 4659 | 4 |
| First-Line Supervisors Of Farming, Fishing, And Forestry Workers | 18 | 2630 | 5 |

| | | | |
|---|---|---|---|
| Janitors And Building Cleaners | 19 | 87855 | 1 |
| Laborers And Freight, Stock, And Material Movers, Hand | 19 | 84622 | 2 |
| Construction Laborers | 19 | 53641 | 3 |
| Other Production Workers Including Semiconductor Processors And Cooling And Freezing Equipment Operators | 19 | 50467 | 4 |
| Assemblers And Fabricators, Nec | 19 | 39839 | 5 |
| Stationary Engineers And Boiler Operators | 20 | 3551 | 1 |
| Locksmiths And Safe Repairers | 20 | 761 | 2 |
| Electronic Equipment Installers And Repairers, Motor Vehicles | 20 | 302 | 3 |

Source: Author's calculations of ACS and O*NET data.

Notes: Clusters are based on description in Section 3.1. ACS samples are from 2014-2017.

Table D2: Comparison of Tree Height at Cutpoints

| Distance at Cluster Merge | Skills in AA (2011) | PCA Skills | PCA Context |
|---|---|---|---|
| Mean | 0.1137 | 0.1423 | 0.1556 |
| Min | 0.0008 | 0.0041 | 0.0040 |
| P25 | 0.0270 | 0.0459 | 0.0540 |
| P50 | 0.0594 | 0.0886 | 0.1078 |
| P75 | 0.1360 | 0.1662 | 0.1950 |
| Max | 1.3658 | 1.1667 | 1.2663 |

Source: Autor's calculations of version 22.0 of the O*NET database (2017) and Acemoglu and Autor (2011).

Notes: Summary statistics come from the shape of the dendrogram (tree) from the Hierarchical Agglomerative Clustering procedure. The "height" of the connection between occupations and clusters is the correlative distance between them when the two objects merge into a single cluster. Lower values of the height represent tighter or more compact cluster definitions and closer relationships between objects.

# Appendix E. Alternative Occupation Groupings

The skill clustering technique proposed in this paper is distinct from alternative approaches to grouping occupations in that it is a data-driven approach that imposes few assumptions about the work relationships between group members, career progression within or across groups, industry composition, and other considerations. By contrast, the statistical units in the United States maintain lists of occupations and major occupation groups for the purposes of statistical enumeration and analysis and other strategic national governance priorities. It is useful, therefore, to compare the two sets of occupational structures. Prior to this comparison, however, it is useful to provide context for the occupational classification systems used by the US Census Bureau that will serve as comparisons.

## Appendix E.1. Historical Context

The US Census Bureau has maintained its own occupational code list since 1850, though there has been harmonization with the Standard Occupational Classification (SOC) code system produced by the Office of Management and Budget (OMB) since 1980.[31] Importantly, the goals of OMB are not just statistical or descriptive in nature, but serve the interests of strategic governance of the Executive branch of the US federal government.[32] The original 1977/1980 SOC system was the result of an initial effort to harmonize data across household surveys, but the system underwent a substantial revision in the mid-1990s after more than a decade of weak funding and a lack of clear direction that prevented updates of the system from moving forward. During the revision process, several statistical agencies provided input, often with competing interests in mind to serve each agency's specific needs. In that process, the advisory panel of the Dictionary of Occupational Titles (DOT) suggested a skills-based approach to the new classification system. This may have been at one point the SOC committee's preferred approach as they commissioned projects by the Joint Program in Survey Methodology to explore this approach and also took great interest in survey work by the Census Bureau and BLS (see endnote 5 in US Department of Labor and US Bureau of Labor Statistics (1999)).

The SOC committee ultimately settled on a method of classifying occupations following the 1980 SOC system for practical reasons rather than what may have been more conceptually appealing approaches, focusing on the "type of work performed" with some minor provisions for "skills-based considerations" (US Department of Labor and US Bureau of Labor Statistics, 1999). This turn away from occupation skills and to "work performed" was because the SOC committee saw the development of measuring skills as not being sufficiently advanced or practically efficient to incorporate into the classification system in the 1990s after initial pilots. The SOC classification system and its hierarchies were finally based on a compromise position that incorporated practical feasibility (leaning on the "type of work performed" aspect), historical comparability with the old SOC system, and conformability across the disparate existing agency systems. The system incorporated occupational skills as considerations far lower in terms of priorities, and usually only to differentiate occupations deep into the hierarchical structure.

The hierarchical structure of the SOC system is based on "job families," i.e. occupational groups intended to "put all people who work together into the same group regardless of their skill level. So, for example, in the 2000 SOC, doctors, nurses, and health technicians are all in the same group instead of in different groups. Similarly, first-line supervisors are in the same

[31] See https://www.census.gov/topics/employment/industry-occupation/about/occupation.html and https://www2.census.gov/programs-surveys/demo/guidance/industry-occupation/overview2019.pdf (Accessed Feb 23, 2023).

[32] See, for example, https://www.whitehouse.gov/omb/ (Accessed Feb 23, 2023).

groups as the workers they supervise, and helpers are in the same groups as the workers they help."[33] This structure partially reflects a forking path of industries rather than what a worker has the capacity to do across industries. It also is suggestive of strong selection on ability for getting into particular groups in the first place.

The SOC system makes some curious choices in its structure. For example, the SOC system separates a broad set of managers into their own major occupation group despite the industry-specific definition of each management occupation (e.g. "Marketing and sales managers" vs "Construction managers" vs "Education administrators"). Thus, in the case of first-line supervisors and most other occupations, the SOC hierarchy imposes a strict industry relationship; but for those who are ostensibly near the top of the income distribution for their various industries, that relationship is cut off in the hierarchy. In addition, supervisors of production workers were divided into separate job classes in the hierarchy at a seemingly arbitrary cutoff of "20 percent of their time performing the same work as the workers they supervise," (US Department of Labor and US Bureau of Labor Statistics, 1999).

Overall, the hierarchy of the original SOC system and its 2000 update, which forms the basis for the Census occupation groups I analyze below, in the face of considerable effort, was not intended to reflect the theoretical considerations or broad skill sets of individual workers or their outside options but was formed with practical constraints in mind around work products and industry activities due to historical inertia, lack of skills data, and inter-agency compromise.

## Appendix E.2. Core Conceptual Differences

In general, Census occupation groups, which follow SOC hierarchies, represent a relatively rigid conception of labor markets that are typically constructed around industries as well as work type (i.e. "people who work together...regardless of their skill level"). This draws an essential distinction from my cluster-based framing of labor markets for several conceptual reasons.

Census occupation groups often reflect small subgroups within single industries subject to a significant amount of baseline selection, both in terms of spatial industry concentration and in terms of individual ability to enter each industry. For example, "Healthcare Practitioners" and "Legal Occupations" represent narrow groups within single industries. For comparability, it is, therefore, important to account for a worker's industry exposure to licensure when comparing my clustering approach to the group of SOC-based Census occupation groups, which I do below.

Because of this hierarchical structure, there is a high correlation between the focal occupation being licensed and the share of each Census occupation group that is licensed (0.63 for the 23 major occupation groups compared to only 0.31 for skill clusters). According to US Department of Labor and US Bureau of Labor Statistics (1999), required training and credentials were explicitly considered in the structure of the SOC, meaning that licensure itself may have played a role in shaping the definitions in the hierarchy, which adds an additional layer of complexity to any analysis of the effects of licensure. For example, one core difference between "Healthcare Support Occupations" and "Healthcare Practitioners and Technical Occupations" is precisely the need for more advanced certification. In addition, making comparisons within these occupation groups that are more homogeneous in licensing then selects based on having already cleared (or being on track to clear) the hurdle created by an occupational licensing requirement—a form of selection on the outcome.

[33] According to the FAQs page from the Census Bureau: https://www.census.gov/topics/employment/industry-occupation/about/faq.html (Accessed Feb. 23, 2023).

## Appendix E.3. Comparisons of Licensing Exposure

First, it is instructive to descriptively explore licensure in skill clusters compared to Census occupation groups. Figure E1 shows binned scatter plots of the relationship between cluster exposure at 20 clusters and the 23 major SOC/Census occupation groups. Panel A shows raw exposure values and reveals four key observations. First, there is generally a positive relationship between the two measures of licensing exposure below 18%. Second, there is a strong *negative* relationship between the two measures from approximately 18% of cluster licensing exposure until approximately 26%, which encompasses the sample mean exposure of 21%. Third, there is a noisy and unclear relationship between the two measures above approximately 30% cluster exposure. Fourth, there is far less variation within Census groups than within skill clusters: identifying variation is generally sparse in the Census occupation groups outside of the range of 10–30%, indicating that the Census occupation groups are more homogeneous in terms of the licensing regimes across states.

Panel B shows the relationship after accounting for fixed effects for occupation, state, and boundary. There is essentially no relationship between cluster exposure and Census occupation group exposure after accounting for these fixed effects. This suggests that the characteristics associated with occupations and states absorb a much larger share of the identifying variation in the treatment variable when considering Census occupation groups than clusters. It also is suggestive of substantial sorting in all three explanatory variables along Census occupation group lines in terms of space (border pairs and state), regulatory choices (state), and correlated within-occupation-group sorting. Industry homogeneity within groups may help explain part of this correlation. For this reason, I consider group by industry exposure to licensure in the treatment variables as in the skill cluster analysis in Section 5.2.

Table E1 shows the results of this exercise. When considering the effect of exposure to licensing within Census occupation groups in the same industry, there is no statistically significant effect on log weekly earnings. When considering licensing effects within occupation groups across industry groups, the effects are almost three times as large, though the two are not statistically significant. This partially reflects the fact that there is a very high correlation between licensing exposure within industry groups and overall occupation group licensing exposure (0.852). Only when considering the overall effect regardless of industry (column 3) do we see a statistically significant relationship between group exposure and earnings.

Is this possible positive relationship robust or consistent across occupation groups? Following my cluster elimination procedure in Section 4.6, I re-estimate my spillover models while sequentially eliminating one of the 23 Census major occupation groups at a time. Figure E2 shows the result of this exercise. Essentially all of the perceived positive spillover effects of licensing on the focal occupation are driven by the occupations classified as "Construction and extraction occupations" by the Census.[34] When this group, which accounts for only 6% of total employment in my sample, is eliminated from the sample, the coefficients are nearly identical to the small and statistically insignificant within-industry exposure coefficients in Table E1. It is possible that complementarities in production occur exclusively in the construction sector, though interpreting the evidence this way would require a comprehensive explanation as to why construction is so unique among all sectors.

Given the above results, it is instructive to examine the total effect of licensing after taking into account spillovers coming through skill clusters and any coming through Census occupation groups. To do this, I perform a similar exercise to Section 4.5. The key difference in

[34]Further investigation shows it is the set of "Construction occupations" that drives this result, not extraction.

this exercise is that I estimate and predict weekly earnings for each worker while taking into account their own occupation licensing status, licensing exposure in their skill cluster, and licensing exposure in their Census occupation group. I then use the coefficients of this model to predict the distribution of earnings if licensing rates were zero for all three categories. If skill clusters (Census occupation groups) are capturing more consequential underlying characteristics of labor markets related to licensure, then the counterfactual distribution of earnings will be higher (lower) when licensing is set to zero relative to licensing set to its status quo levels.

Figure E3 shows the results of this exercise. Nearly the entire distribution is higher under the "no licensing" counterfactual regime than under the status quo even after accounting for within-group selection, agglomeration, or industry effects coincident with Census occupation groupings. Only the very top of the distribution appears to have lower predicted earnings under the "no licensing" regime. This implies, like in Figure 9, that licensing has an inequality-enhancing effect. Any net positive effects of licensing on earnings appear only at the top of the distribution of occupation earnings (above approximately \$1,300 per week).

These various comparisons suggest two main takeaways. First, occupation groups based on skill clusters appear to capture underlying choice dynamics in the labor market—especially related to occupational licensing—that are more consequential than when considering the occupation groupings used in the SOC and various Census systems. This would include connections between occupations that would represent some latent "outside option" broadly conceived. This may be unsurprising given the conceptual and historical context introduced above. Because Census occupation groups have more homogeneous licensing requirements, comparing someone that has already cleared a licensing hurdle into the industry (or is already on track to clear it) to someone else that has already cleared a licensing hurdle may mechanically blunt the possible effect of licensing on worker outcomes. Second, after accounting for latent features of Census occupation groups and their relationship to earnings and occupational licensing, cluster-based exposure appears to dominate, and the total effect of licensing as a feature of the labor market is negative for the vast majority of the distribution of workers.

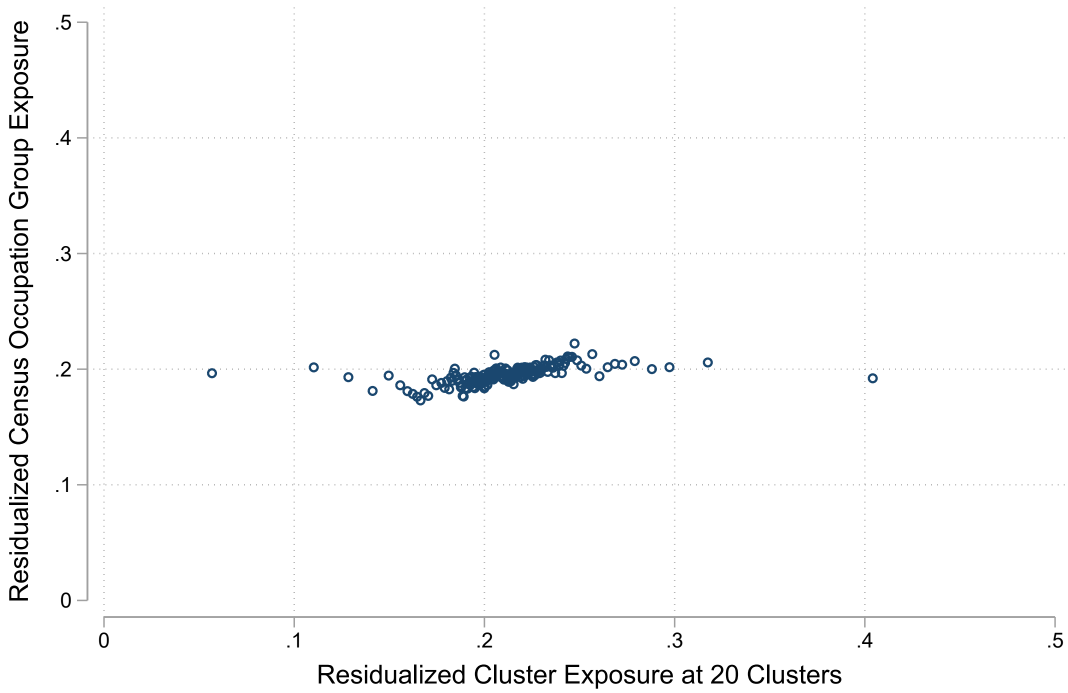
Finally, instead of relying on skill clusters or Census occupation groups, another sensible way of grouping occupations together is by major industry, of which there are 14 defined in the ACS. This differs from the complementarity exercise in Section 5.3 in that this grouping does not impose that workers have contact in the same narrow industry, but that the occupations in the same major group constitute viable outside options for a worker in that major industry group.

In Table E2, I show the results of estimating exposure in each occupation-industry group cell separate from Census occupation groups and skill clusters (column 1) and running a horse race with skill cluster groupings (column 2). The results show that using major industry groups to measure exposure also results in negative measured spillover effects on earnings. Like in Table 4, this effect persists once we add skill cluster exposure. In short, Census occupation groupings appear to be the outlier among the three approaches in resulting in null spillover effects, which is likely driven by the structures addressed above.

Figure E1: Licensing Exposure from Occupations within Group

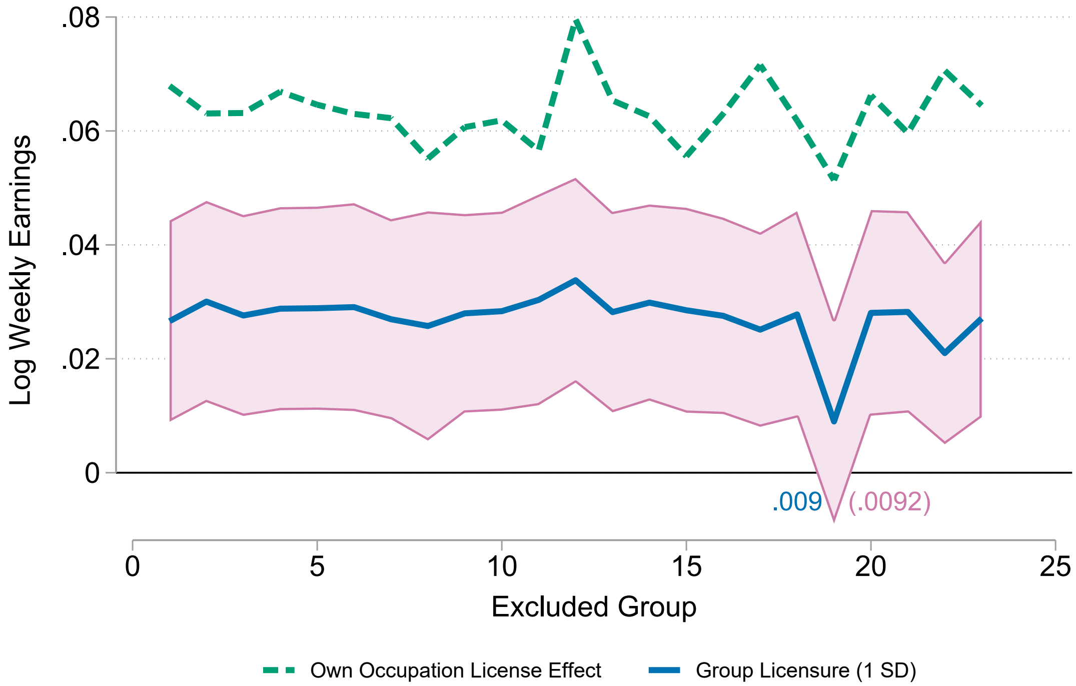
Panel A: Raw Exposure


Panel B: Residualized Exposure




Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Panel B values are residualized on occupation, state, and boundary fixed effects.

Figure E2: Exposure Effects Eliminating Census Occupation Groups




Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Coefficients are generated from the boundary discontinuity design detailed in Equation 1 using 23 Census occupation groups while eliminating one group at a time. The model includes fixed effects for occupation, state, and boundary. Group 19 corresponds to "Construction and extraction occupations."

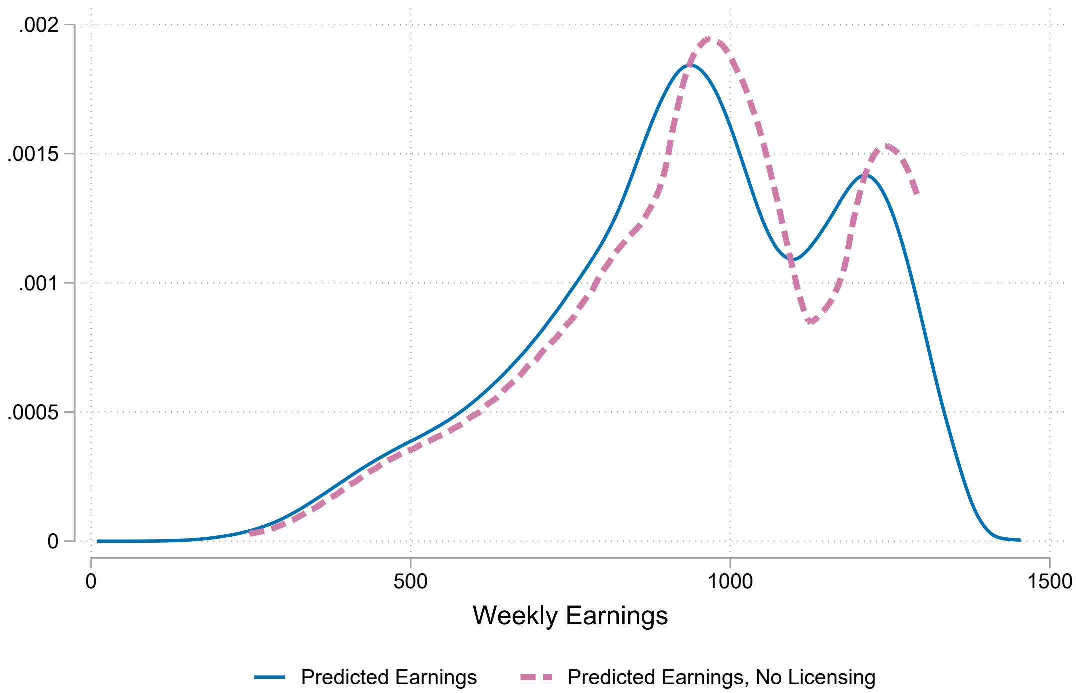
Table E1: Licensing Exposure Within Industry Groups in Census Occupation Groups

| VARIABLES | (1) Log(Weekly Earnings) | (2) Log(Weekly Earnings) | (3) Log(Weekly Earnings) |
|---|---|---|---|
| Own Occupation Effect | 0.0806*** | 0.0728*** | 0.0634*** |
| | (0.0232) | (0.0216) | (0.0195) |
| Group-Industry Exposure (1 SD) | 0.0117 | 0.00851 | |
| | (0.0157) | (0.0179) | |
| Group Exposure (1 SD) | | 0.0213 | 0.0273*** |
| | | (0.0169) | (0.00886) |
| Observations | 1,683,900 | 1,683,900 | 1,683,900 |
| R-squared | 0.459 | 0.459 | 0.459 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: Author's calculations of ACS, CPS, and O*NET data.
Notes: Estimates come from Equation 1 but include an additional measure of licensing exposure within the 14 major industry groups (NAICS) by 23 major occupation group cells. Standard errors are clustered at the occupation level.

Figure E3: Kernel Density of Predicted Earnings Accounting for Census Occupation Groups




Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Values are generated by predictions from the boundary discontinuity design detailed in Equation 1 using 20 skill clusters. The model also includes exposure to licensing from within each Census occupation group. The model includes fixed effects for occupation, state, and boundary. Predicted earnings are for the status quo and for setting licensing rates to zero for one's own occupation, skill cluster, and Census major occupation group.

Table E2: Licensing Exposure Within Industry Groups

| VARIABLES | (1)<br>Log(Weekly Earnings) | (2)<br>Log(Weekly Earnings) |
|---|---|---|
| Own Occupation Effect | 0.0918*** | 0.0955*** |
| | (0.0221) | (0.0215) |
| Major Industry Group Exposure (10 PPT) | -0.0112* | -0.0111* |
| | (0.00581) | (0.00581) |
| Skill Cluster Exposure (10 PPT) | | -0.0151** |
| | | (0.00700) |
| Observations | 1,688,506 | 1,688,506 |
| R-squared | 0.459 | 0.459 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: Author's calculations of ACS, CPS, and O*NET data.
Notes: Estimates come from Equation 1 but include an additional measure of licensing exposure within the 14 major industry groups (NAICS). Standard errors are clustered at the occupation level.

# Appendix F. Theoretical Appendix

## Appendix F.1. Simple Graphical Framework

A host of papers present models of a competitive labor market in which barriers to entry into specific occupations will result in fewer workers entering the occupation (Kleiner, 2000; Kleiner and Soltas, 2019; Blair and Chung, 2019). But one piece missing from the current literature is the set of choices made by those who exit or who are prevented from entering the occupation due to higher entry costs and the spillover effects of those choices on the structure of the labor market.

Consider the simple graphical frameworks in Figure F1 depicting possible responses to licensing restrictions in an unlicensed occupation closely related to a licensed occupation. In Panel A, which represents a labor supply spillover in an otherwise competitive market, workers prevented from entering the licensed occupation due to entry costs enter this similar occupation at higher rates. This shifts out the labor supply curve $S$ to $S'$, resulting in higher labor supply at $L'$ and lower wages at $W'$. The result is a combination of lower wages and higher employment. The size of the labor supply shift into this occupation depends on how closely the occupations are related in their skill dimensions, the ease of moving across occupations, and how prohibitive the licensing restrictions are for each prospective entrant.[35]

Those facing differential barriers with a licensing requirement or who are categorically ineligible to work in a licensed occupation will be more strongly affected in their occupation choices and therefore be the likely movers into unlicensed occupations. This might include women, who bear larger shares of home production responsibilities making occupational transitions more costly, foreign-born Hispanic workers most affected by citizenship, residency, or language requirements, or black workers, who are more likely than other racial groups to have a past experience with incarceration or experience labor market discrimination—statistical or "taste-based." This implies a composition shift among occupations.

As a brief example, consider the rising licensing requirements for being a physical therapist (PT) or occupational therapist (OT) cited in Cai and Kleiner (2016). Prior to the licensure of occupational therapy, some prospective entrants to PT might be deterred from PT and instead enter OT. As OT becomes more licensed, other prospective entrants may then be deterred from entering either occupation and instead enter something like athletic training, which requires a bachelor's degree in states where it is licensed, but in some states entirely lacks a governing body (Vargo et al., 2020). Even in the presence of a strong underlying skill endowment relevant to PT and OT, a larger share of workers enter the remaining, less-regulated occupation. This framework predicts higher employment and lower wages in athletic training.

In comparison to the competitive model, consider a model in which occupational licensing exacerbates monopsony power in the labor market. Such a model is discussed in Kleiner and Park (2010) in the context of dentists and dental hygienists in a single product market for dental services, but the monopsony context is worth exploring further. In Panel B of Figure F1, a person considering changing occupations into a licensed occupation but is deterred by the entry costs has fewer effective outside options. An entire branch of possible firms hiring in the licensed occupation becomes infeasible to such workers. This decreases the elasticity of labor supply to the firm, tilting the labor supply curve from $S$ to $S'$. A monopsonistic firm then employs workers at wage $W^M$ (below a worker's marginal revenue product) while employment falls to $L'$ because some workers may exit employment altogether if $W^M$ is below their reservation wage in the whole market. If the local pool of workers has reservation wages at

[35] In Appendix F.3, I discuss a model of skill transferability in a competitive labor market and how these parameters influence occupational choices when a licensing regulation is introduced.

or below $W^M$, employment may only fall marginally or not at all. The result is a combination of lower employment and lower wages, though reductions in employment are not a requirement for firms to pay workers below their marginal revenue product. In the prior example, raising licensing requirements in PT and OT may make entry from athletic training infeasible. A monopsonistic firm that employs athletic trainers may recognize this friction and thus has the ability to pay athletic trainers a lower wage because the threat of leaving the firm is less credible. The firm may also hire fewer new trainers.

## Appendix F.2. Monopsony Search Model

A monopsony search model can shed light on this dynamic. Black (1995) proposes a search model in which the presence of "prejudiced" firms that refuse to hire black workers may lead to higher search costs for black workers as their choices of "unprejudiced" firms are rarer, which lowers their reservation wages and therefore increases monopsony power of the "unprejudiced" firms over black workers. I adapt this model to my setting wherein a worker may search for a firm match both within and across occupations. An occupational licensing requirement raised in multiple outside occupations acts as an increase in the number of "prejudiced" firms that refuse to hire an unlicensed worker in a particular occupation because they legally cannot hire them.

Following Black (1995), suppose there is a $\theta$ share of firms who, due to their product markets, will hire licensed workers with skills in cluster $S$, and (1-$\theta$) share who will hire unlicensed workers in cluster $S$. Those with a license, $l$, and those without, $n$, face wage offers from "unprejudiced" firms, $u$, of $\omega_u^l$ and $\omega_u^n$, while only licensed workers receive wage offers from "prejudiced" firms at $\omega_p^l$. Parameter $\alpha$ is the utility value of job satisfaction in a firm-occupation match with a probability density function $f(\alpha)$. A worker searching for a job accepts a wage offer when $\alpha \geq u_r^l - \omega_j^l$, where $j = u, p$ and $u_r$ is reservation utility. Given $\kappa$ costs of the next search, a worker with a license in an occupation searches until the point she is indifferent, or when marginal search costs are equal to the marginal expected benefit of the next search:

$$\kappa = \theta \int_{\alpha_p^l}^{\infty} (\omega_p^l + \alpha - u_r^l) f(\alpha) d\alpha + (1-\theta) \int_{\alpha_u^l}^{\infty} (\omega_u^l + \alpha - u_r^l) f(\alpha) d\alpha \tag{F.1}$$

An increase in wages paid in firms and occupations in either the licensed or unlicensed statuses raises the reservation wage of a licensed worker. A rise in the share of firms that only hire licensed workers, which may occur with new licensing legislation, ambiguously changes licensed worker welfare depending on the change in wages between licensed and unlicensed occupations and firms.

For a worker without a license, the search will continue until:

$$\frac{\kappa}{(1-\theta)} = \int_{\alpha_u^n}^{\infty} (\omega_u^n + \alpha - u_r^n) f(\alpha) d\alpha \tag{F.2}$$

An increase in the share of firms only hiring licensed workers in the skill cluster strictly increases the search cost and therefore lowers the reservation wage of an unlicensed worker in the cluster. Because firms recognize this, they offer unlicensed workers lower wages, and any measured elasticity of labor supply to the firm with respect to offered wages becomes more inelastic. Some workers may find the remaining wage offers to be below their reservation wage and exit the market altogether or reduce their hours, leading to lower employment.

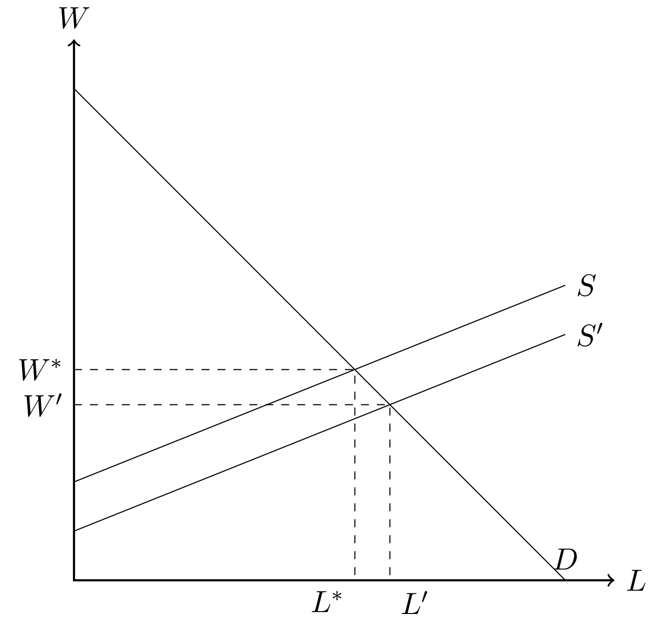
In the same framework, a worker that is part of a historically discriminated minority in the workforce (e.g. black workers, women) may find their outside options even more limited by occupational licensing. However, their individual returns to entering a licensed status then rise relative to the alternative, and they may take advantage of the signaling value of a license (Blair and Chung, 2018). In this case, the wage premium for obtaining a license will be higher for those in these demographic groups relative to others in the group, while the wage spillover penalty will be larger in the unlicensed status for these groups. In other words, the "own-license" effect will be more positive and the spillover effect will be more negative.

Many licenses contain requirements that may differentially increase $\theta$ depending on group characteristics. Requirements against any past felony conviction may differentially affect some black workers, while licenses whose exams are purely in English may negatively affect non-English speaking immigrants (half of which are Spanish speakers (Rumbaut and Massey, 2013)), and citizenship or residency requirements may disproportionately affect foreign-born workers. In that case as well, the spillover effect is expected to be larger.

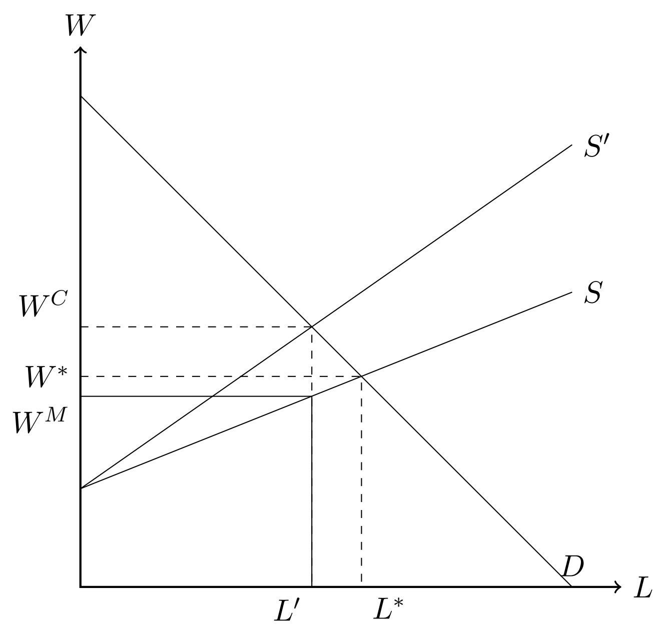
This framework predicts that as licensing increases within a cluster, equilibrium employment in the remaining occupations may also fall as monopsonistic firms hire fewer workers.

Figure F1: Competitive vs Monopsonistic Labor Market

Panel A: Competitive Market Labor Supply Shift




Panel B: Monopsony Model




Notes: An illustration of the possible spillover effects of occupational licensure onto other occupations in a labor supply (competitive) model vs a monopsony model.

## Appendix F.3. Model of Skill Transferability

The model in Shaw (1987) makes clear predictions about how skill transferability between occupations determines switching and investments into occupation-specific human capital. This model suggests conditions under which an individual in an occupation will change their occupation.

While my setting does not consider job changes per se, I conceptualize occupational choice as selecting an occupation that best matches with latent skills, either endowed or acquired through investment. Rather than past investment in the occupation's skillset, initial conditions are dependent on endowed skills when entering the labor market, either through family or public investments or innate ability. These can include any skills which make the individual suited for a set of occupations, like sociability, physical strength, cognitive ability, or leadership skills. The initial "occupation" represents the occupation for which the combination of an individual's endowed skills is best suited at baseline, or whose $I_0$ is largest.

Following Shaw (1987), I define the occupational human capital stock for a person in occupation $j$ at time $t$ ($I_t^j$) as:

$$I_t^j \equiv K_t^j + \gamma^{ij} K_{t_j-1}^i + ... + \gamma^{gj} K_{t_i-1}^g + I_0^j + \sum_{e=i,h,g} \gamma^{je} I_0^e \tag{F.3}$$

where an individual's human capital in occupation $j$ depends on time spent in the occupation since they entered the occupation ($t_j$) and on the human capital investments in all other occupations $i, h, ...g$ which were entered into at time $t_{i,h,...g}$. The final term is the sum of all initial endowments in skills related to each occupation. The endowment term gives a baseline for occupation choice structure. Essentially, all workers, as they enter the labor market, have a "default" occupation into which they would sort given their endowed comparative advantage. Further investment choices are afterward driven by comparison to this baseline. In short, this full equation represents the total investments through the current period in human capital for occupation $j$, including transferable skills in $i$ through $g$. Importantly, $\gamma^{ij}$ is the share of skills in occupation pair $i, j$ that is transferable between the two occupations.

Each $K^j$ is defined as the sum of all the earnings capacity invested in occupation $j$ in each year because time spent investing in human capital for an occupation is time not spent on production. Investment intensity, or the share of productive capacity used in developing human capital, is $k_t^j$, so realized earnings ($Y$) in the current period are some share of earnings capacity ($E$), where $Y_t = E_t(1 - k_t^j) + I_0$.

Simplifying a Mincer equation (Mincer, 1974) of earnings in which individual costs of investment $C^j$ directly translate into earnings through $K^j$ in the period after investment, income in the current period $t$ in occupation $j$ can be expressed:

$$Y_t^j = E^s + r^j (C_{t-1}^j + \gamma^{ij} C_{t_j-1}^i) - c_t^j + \gamma^{ij} I_0^i + I_0^j \tag{F.4}$$

Here, $E^s$ is earnings capacity or general human capital given formal schooling, and $r^j$ is a common rate of return to investments in $j$. The $C$ terms are at the individual level and represent the current stock of accumulated earnings capacity in $j$ until period $t-1$ as well as the earnings capacity due to skill transferability from occupation $i$ accumulated before the change to occupation $j$. The term $c_t^j$ captures current investment in $j$. In words, earnings capacity today is a function of schooling, returns to all accumulated investments in $j$, the share of investments in $i$ that are transferable to $j$, endowed capacity in $j$, and the share of endowed capacity in $i$ that is transferable to $j$ net of current investments in $j$.

In present value terms, given discount rate $r$, an individual will switch occupations from $i$ to $j$ when:

$$\{\gamma^{ij} r^i C_{t-1}^i - r^j C_{t-1}^j + (\gamma^{ij} I_0^i - I_0^j)\} \sum_{g=t}^{T} (1/(1+r)^g)) < 0 \tag{F.5}$$

and

$$\sum_{g=t}^{T}\sum_{h=t}^{g-1}\{(r^j c_h^j - c_g^j) - (r^i c_h^i - c_g^i)\}(1/(1+r)^g) \lesseqgtr 0 \tag{F.6}$$

Equation F.5 represents the loss of returns to past investments and endowments in occupation $i$. Because $\gamma^{ij} < 1$, there is a loss associated with switching occupations in which past investments into $j$ no longer reap rewards except through skill transferability. The present value of gains to investment in $j$ must be large enough to overcome the difference between 1 and the value of $\gamma^{ij}$.

Equation F.6 is the difference in the value of future investment in occupation $j$ vs occupation $i$. When the value of future investments in $j$ is larger, the worker will choose to absorb the costs of entering $j$ rather than $i$. There are two key predictions of this model: 1) the greater the skill transferability, $\gamma^{ij}$, the more probable a move between the two will be; 2) lower opportunities for investment in $i$ will increase the value of moving to $j$.

An occupational license in $i$ may affect the balance of these inequalities. A license that categorically blocks entry for some demographic groups such as non-residents, non-English speakers, or those who have been incarcerated sharply reduces opportunities for investment in $i$ and therefore increases the value of moving to $j$. The same holds if the costs of investment $c^i$ rise with additional education requirements, exams, or fees without offsetting returns through C. Alternatively, an occupational license may directly influence occupational skill substitutability by introducing requirements for an occupation that may be unrelated to the performance of the job.[36]

If the transferability of skills is highest in the $i, j$ combination over some set of other occupations, say, $i, h$, the first order choice is whether or not to move between $i$ and $j$. If $j$ is *also* licensed with large investment costs, the worker may move to the next comparison, $h$. In terms of my setup, this implies that occupational licenses will push individuals out of licensed occupations in their skill cluster and into the most related occupations in the same cluster, increasing labor supply in a competitive labor market, and reducing wages. If, however, licensing is widespread enough and adjustment costs are large, individuals may exit the cluster altogether.

[36]For example, Florida bill 851 required massage therapists, acupuncturists, dentists, pharmacists, and other health care professionals to be trained in spotting and reporting human trafficking violations and post signs regarding human trafficking in conspicuous places in their establishments as a condition of licensure. https://www.flsenate.gov/Committees/BillSummaries/2019/html/2089 (Accessed April 30, 2020).