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

Samuel Dodini
Norwegian School of Economics
samuel.dodini@nhh.no

June 2022

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 border match 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 consistent with a monopsony model where licensing increases search costs and reduces workers' outside options.

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

*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.

1 Introduction

Since the 1930s, economists have theorized about the possible consequences of imperfect labor markets (Robinson, 1933). Much of the recent empirical literature has focused on how imperfections in unregulated labor markets may negatively affect worker wages and employment. Sources of these imperfections include search frictions and switching costs (Webber, 2016; Ransom, 2021) and the concentration of labor demand (Azar et al., 2020; Dodini et al., 2020). Each of these may contribute to earnings and employment that are inefficiently low relative to a competitive equilibrium. There has been far less focus on the effects of possible market imperfections created by occupation-specific regulations.[1] This paper focuses on how occupational regulations affect workers by considering a growing source of strict oversight in the labor market: occupational licensing.

Occupational licensing is state-sanctioned permission to work in a particular occupation. These regulations on who can work in an occupation 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 affecting 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 market imperfections.

The current literature suggests that occupational licensing regulations in the US, most of which differ across states, have significant effects on the labor markets of the individual occupations being licensed.[2] There is also some evidence of wage spillovers for occupations that perform similar functions in the same narrow industries.[3] Finally, there is evidence that occupational licensing increases earnings inequality, primarily because the wage returns to having a license appear higher at the upper end of the education or income distribution (Kleiner and Krueger, 2013; Gittleman et al., 2018; Zhang and Gunderson, 2020).

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.[4] In particular,

[1]For example, the enforcement of non-compete or non-disclosure agreements (Starr et al., 2021; Lipsitz and Starr, 2022; Balasubramanian et al., 2020).

[2]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), and reduce interstate labor migration and occupational mobility (see Johnson and Kleiner (2017); Kugler and Sauer (2005) and Kleiner and Xu (2020)).

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

[4]Some studies estimate a licensing premium with a binary indicator for someone having a license without controlling for occupation (Kleiner and Krueger, 2013; Gittleman et al., 2018), which may entail some information about spillovers in the "licensed" versus "unlicensed" sectors. However, these studies do not con-

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. The answer also informs the literature in labor economics about how specific public policies contribute to labor market power.

This study addresses this question by testing for the presence of earnings and employment spillovers of occupational licensing on a set of counterfactual occupations. The lack of clear or systematic definitions of "counterfactual occupations" has been a key shortcoming 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. Spillovers may come through two competing mechanisms as well as sorting effects. First, licenses can have negative wage spillovers by raising costly barriers to entering one occupation and redirecting and increasing labor supply to unlicensed occupations. Alternatively, occupational licensing may increase monopsony power because licenses make outside options costlier to enter. In such a setting, employment in other occupations may fall rather than rising because firms with market power have the ability to hire fewer workers and pay lower wages, (Ashenfelter et al., 2010), particularly in smaller labor markets with fewer outside options. Licensing may also lead to sorting because of heterogeneous adjustment costs or differential impacts of licensing regulations across demographic groups. This is the first study to directly test for the presence of such labor market externalities across occupations.

I test for spillovers in three steps. First, in order to define a set of counterfactual occupations, I group occupations together based on their skill content. I use data from the O*NET database and non-parametric clustering techniques to group together occupations into clusters that require similar levels of key skills. The skills upon which I base these clusters come from Acemoglu and Autor (2011) and represent combinations of non-routine, routine, manual, cognitive, and interpersonal skills, an approach also taken in Dodini et al. (2020). This approach addresses a key need in the literature.[5] This is the first study to use this novel, skill-based approach to study the effects of labor market regulations in the United States. This provides a roadmap for future work to expand the set of applications for data on occupational skills.

Second, using recently available data from the Current Population Survey, I use the share of individual workers that indicate they are required to have a license as a proxy for the

sider spillovers within sectors and generally rely on strong assumptions about selection on observables, which complicates attributing information from this coefficient to spillovers.

[5]The application of this clustering approach to occupational skills evolved concurrently with Dodini et al. (2020). 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.

regulatory environment in each state (Kleiner and Soltas, 2019). This approach overcomes a core measurement challenge in the literature. 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. My empirical approach uses microdata from the American Community Survey in a state border match design to compare the earnings of workers in the same occupation in local labor markets on either side of a common state border. I control for state-level policies and economic variables that determine earnings across all labor markets in the state through state fixed effects. Conditional on these, workers across individual border pairs differ only in the share of the skill cluster outside their own occupation that is licensed based on the CPS individual licensing measures. I define the local labor market by ACS Public Use Microdata Areas (PUMAs), which are areas defined by the Census Bureau around counties (or groups of counties) and metropolitan areas that contain at least 100,000 people and collectively cover the entirety of the United States. Because the licensing environment is defined at the state level, state-level licensing shares are exogenous to local labor market factors.[6] After measuring the overall average effect of licensing spillovers, I test for differences in the effects of exposure to licensure across different subgroups, particularly across gender, race/ethnicity, nativity, and labor market size. This is the first study to examine both the existence of direct spillovers as well as how these spillovers differ across demographic groups. Using my estimates, I also calculate what the distribution of earnings within occupations would be if licensing were eliminated altogether. This exercise demonstrates the net effect that licensing has on earnings inequality when accounting for these spillovers.

Third, I estimate the effects of licensing exposure on employment and worker composition in each focal occupation. That is, for the same occupation, does having more licensure in a skill cluster reduce or increase employment in the focal occupation? Do the focal occupations differ in the types of workers they employ? The direction of this employment effect informs the underlying mechanism behind earnings spillovers. A positive employment spillover on other occupations is consistent with a labor supply mechanism, while a negative employment spillover is consistent with a monopsony mechanism and inconsistent with the labor supply explanation. My composition estimates shed light on the sorting effects of these regulations and would be predicted in both a labor supply and monopsony framework.

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 local labor markets on the other side of a state border. On the other hand, 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 1.5-2% *lower* for that worker. In other words, if every other occupation in

[6]My estimates are nearly identical when including fixed effects for local labor markets, which strengthens the case for this assumption.

one's skill cluster became fully licensed, earnings in one's own occupation would decline by approximately 15-20%. These negative effects are stronger for women, non-Hispanic black, and foreign-born Hispanic workers, which is consistent with these demographic groups being less able to absorb the costs associated with licensure or differential effects of the regulations themselves. Because these groups are in the lower portion of the income distribution, these effects imply that licensing and regulation externalities contribute to local income inequality. I present graphical evidence of this effect by showing the counterfactual distribution of predicted within-occupation earnings if licensing requirements did not exist in my sample. This 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%.

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. The negative employment effects are strongest in smaller labor markets, which aligns with the recent monopsony literature (Rinz, 2018; Dodini et al., 2020). 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 or have a Master's degree or PhD falls. In addition, the share of workers in the focal occupation that is Hispanic or foreign-born rises by over 8 percentage points as the cluster becomes fully licensed.

I augment my analysis with a placebo exercise in which I randomly assign occupations to clusters and recompute my estimates. These estimates show that licensing exposure in placebo clusters is uncorrelated with earnings, occupation composition, and employment. This exercise suggests that unobserved factors correlated with earnings, employment, and licensing rates at the state level do not explain the effects I find. For an unobserved economic factor or policy to be driving my results, such a factor would have to be correlated with licensing rates and with labor market outcomes in ways that *specifically* match the exact structure of each occupation's skill cluster assignment. In other words, any proposed omitted variable would have to be positively related to licensing rules and negatively related to earnings *only* through the very specific configuration of occupational groups generated by my clustering algorithm.

My findings are also robust to the inclusion of local labor market (PUMA) fixed effects. This specification limits identifying variation to only areas that share a border with multiple states. Unobserved differences in local labor markets correlated with licensing rules are, therefore, not driving my results. My findings also are robust to different choices about the optimal number of clusters and the sequential elimination of individual clusters from the analysis.

I also perform the same analysis using simple cross-state variation similar to Kleiner and Soltas (2019) with data from the 2015–2018 CPS Outgoing Rotation Group (ORG) and find similar results, implying that cross-border spillovers and the specific composition of workers in border PUMAs are not a pressing concern for my border design.[7] The pattern of results is also notably similar using a different measure of licensing intensity pulled from the text of licensing laws themselves in my border design (Redbird, 2016). Finally, I also find the same pattern of results using CPS ORG data back to 1983 when leveraging changes in licensing laws across states over time (Redbird, 2016). The direction and significance (statistical and economic) of spillover effects on wages and employment and the pattern of heterogeneous treatment effects in this exercise match and support those in my cross-sectional estimates. However, these time-varying estimates are likely subject to considerable attenuation bias (toward zero) due to measurement error, particularly when considering earlier years in the data.

This paper contributes to the growing empirical literature on the effects of labor market regulations on workers. In addition to the growing field of research on occupational licensing, recent studies have examined the effects of enforcing non-compete agreements and have found that enforcement of these agreements reduces worker wages (Starr, 2019; Starr et al., 2021; Lipsitz and Starr, 2022; Balasubramanian et al., 2020), including possible negative effects on those unconstrained by these agreements (Starr et al., 2019). The main mechanism through which these negative effects take place is through a decrease in the number of outside options available to a particular worker, an increase in search and switching costs, a decline in worker mobility, and an increase in firm monopsony power.

Particular to 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, 2020; Zhang and Gunderson, 2020). Synthetic control and other panel estimates of the effects of occupational licensing for specific occupations are approximately 7–10% and are similar to my estimates and the cross-sectional estimates found in other studies (Pizzola and Tabarrok, 2017; Thornton and Timmons, 2013; Carollo, 2020).

The main mechanism through which these wage effects in the prior literature appear is through reductions in labor supply to licensed occupations (approximately 20%) (Blair and Chung, 2019; Kleiner and Soltas, 2019), with some exceptions in occupations like nursing (DePasquale and Stange, 2016), coupled with licensed workers working more hours on the intensive margin (Bailey and Belfield, 2018; Kleiner and Soltas, 2019) and increases in prices in the product market (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, which is pertinent to overall labor supply choices, suggests that licenses decrease interstate migration by as much as 36% (Johnson and Kleiner, 2017). Finally, and importantly, occupational licensing reduces labor market fluidity as measured by job changes and can explain nearly 8% of the total change

[7] The main log wage regression coefficients are in Appendix Figure A13 and closely follow my main results for weekly earnings.

in occupational mobility over the last twenty years (Kleiner and Xu, 2020).

On the topic of spillovers, there are two strands of the literature related to my analysis. First, some studies estimate a wage premium for having a license using binary indicators for licensure as the treatment variable (sometimes at different quantiles of the earnings distribution), in part to compare to the effects of unionization (e.g. Kleiner and Krueger (2013); Gittleman et al. (2018); Zhang and Gunderson (2020)). These estimates may contain some information about possible "cross-sector" spillover effects of licensing, i.e. licensed on the unlicensed. However, licensing in one occupation may also spill over to affect earnings in *licensed* occupations as well, which makes any bias in attributing wage differentials to cross-sectors spillovers ambiguous. If the spillovers affect licensed (unlicensed) occupations more (less) intensely, then a coefficient on licensure will mean that the implied cross-sector effect will be a downward-(upward-) biased estimate of the total spillover effect. In addition, selection and composition effects and differences in econometric approaches elicit the need for caution when making direct comparisons between my models and the prior literature.

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, namely those which govern direct access to patients, have negative effects on the wages of occupational therapists because many services are substitutable between the two (Cai and Kleiner, 2016). When nurse practitioners, who act as a substitute for physicians in many medical services, are given broader scope for their practice, physicians' wages fall, while nurse practitioners' wages rise (Kleiner et al., 2016).[8] In the paper most related to my analysis, Kleiner and Park (2010) examine the effects of broadening the scope of practice for dental hygienists on the earnings and employment of both hygienists and dentists. They find that as regulations that allow hygienists to be self-employed are implemented, wages for hygienists rise by 10 percent, and employment among hygienists increases, 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 contributes to our understanding of the operation of regulated labor markets by identifying the broad effects of occupation-specific regulations on other occupations. In particular, this paper is the first to demonstrate that strict entry regulation comes at a broad cost: lower labor market earnings and employment for those in occupations that use similar skills. This study is also the first to show that these wage externalities are not consistent with a pure labor supply shift but that occupational entry restrictions increase labor market rigidity and thereby exacerbate 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

[8] Dillender et al. (2022) similarly find that earnings and lagged job postings both increased for nurse practitioners when their legal scope of practice expands.

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 workers in general, most of whom play no part in the legislative negotiations that ultimately determine the scope of these regulations.

2 Theoretical Frameworks for Spillovers

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 1 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.[9]

Those facing differential changes in barriers with a new 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

[9] In Appendix C, 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.

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 1, a person considering changing occupations into the 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 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 and may hire fewer new trainers.

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 θ share of firms who, due to their product markets, will hire licensed workers with skills in cluster S, and (1-θ) 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 ω_u^l and ω_u^n, while only licensed workers receive wage offers from "prejudiced" firms at ω_p^l. Parameter α 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 κ 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:

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

An increase in wages paid in firms and occupations in either the licensed or unlicensed sector 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:

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

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.[10]

From the product market perspective, as the cost of entry into competing product markets rises with licensing costs, product market power may increase. A simple example is the supply of massage therapists. Restricting the supply of independent operators reduces product market competition in addition to labor market competition. Recent research on the relationship between product market power and labor market concentration suggests the two are positively correlated (Marinescu et al., 2019; Qiu and Sojourner, 2019; Lipsius, 2018).

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 the licensed sector 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 sector for these groups.

Many licenses contain requirements that may differentially increase θ 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-

[10] One might ask how important occupational licensing would be in comparison to other costs of switching occupations in a monopsony framework. According to Cortes and Gallipoli (2018), only about 15% of total occupation switching costs is attributable to task-specific adjustment costs, so given general skill cluster matching, these skill-based adjustment costs are likely to be small. Licensing, however, imposes large time costs that may outpace task-specific costs.

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, 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. In addition, these negative wage effects will be larger in smaller labor markets due to fewer baseline search options. Labor market concentration and monopsony power have been shown to be higher in smaller labor markets where outside options are numerically limited by market size (Rinz, 2018; Dodini et al., 2020). The prediction of composition changes is the same as in a competitive model.

To summarize, the direction of any earnings and employment effects of occupational licensing regulations can inform us about the underlying mechanism. The key difference in the competitive context in relation to the monopsony context is the direction of employment changes: increases in employment in unlicensed occupations are suggestive of labor supply shifts in a competitive model and employment declines are suggestive of a monopsony effect, or at the very least, are inconsistent with the competitive labor supply explanation.

3 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. In my robustness tests, I also use data from the Current Population Survey Outgoing Rotation Group (ORG) dataset going back to 1983 and the Northwestern Licensing Database (NLD) (Redbird, 2016; 2017).

3.1 Current Population Survey

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. Redbird (2016; 2017) painstakingly organized a list of licensing requirements back to the 1970s based on the text of statutes in order to measure the effects of licensing on wages (the Northwestern Licensing Database). However, 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, "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 consistent with classical measurement error. 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 could lead to considerable attenuation bias—a bias exacerbated by the fixed effects models used in this literature and which is likely to increase if measurement quality deteriorates going back in time.

In 2015, the CPS began asking individual workers questions regarding licensing and certification, which helps address this measurement challenge. 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 endogenous. This exercise also allows me to incorporate differences in sub-occupational licensing status into broader occupational categories in the CPS. Returning to the anesthesiologist assistant example, given that the "physician assistant" occupation in the CPS would include "physician assistant" and "anesthesiologist assistant," and "family practice physician assistant," if these sub-categories are differently licensed across states, my aggregated measure will capture this variation across states within a single occupation code.[11]

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

[11] To give an example of the measurement challenge, using OES employment weights and the statutes in the Northwestern Licensing Database (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 would not be required to have a license under a binary (50% cutoff) licensing rule in the NLD, meaning licensing laws tend to underestimate licensing 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 preferred estimates, though both are attenuated toward zero, particularly in measuring an occupation's own wage premium. See Appendix Figure A20.

with licensing laws from other sources when the laws are well defined and understood. Though recent work (Carollo, 2020) suggests that in some cases there may be some undercounting of licensing rates in the CPS when compared to well-defined licensing laws for occupations licensed in all US states, this error is likely to be classical in nature due to classification errors in binary indicators, meaning my estimates may be slightly attenuated.[12]

On a fundamental level, because of these measurement challenges, current publicly-available datasets will entail a trade-off between cross-sectional accuracy in licensing in the CPS and time-varying but noisy measurements of licensing from legislative text. My analysis leverages the granular nature of the CPS in a border match design, though I show using variation over time in Section 5.5 evidence pointing to the same conclusions but with considerable attenuation bias.

3.2 American Community Survey

To construct my border match 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. Appendix Figure A1 shows maps of my border PUMAs in four Census Divisions. PUMAs map within states and across counties, are intentionally coincident with Metropolitan Statistical Areas in densely populated areas, and each contains at least 100,000 people. 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, nativity, and the size of the working-age population (18–64) in the PUMA, which serves as my measure of labor market size.

One limitation of the ACS generally is how it measures determinants of hourly wages: earnings and hours. Baum-Snow and Neal (2009) explain that part-time workers systematically under-report hours in the survey, leading to implausibly large estimates of their hourly wages. To avoid these measurement issues, 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).[13]

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. Recent work also suggests the CPS may undercount licensing rates in comparison to known laws in some of these universally licensed occupations

[12] Appendix E of Kleiner and Soltas (2019) details the econometric strength of the CPS measure, and the authors find that the effect of applying an empirical Bayes estimate of license shares is relatively small, particularly in state-occupation cells of sizes larger than 10. Because clusters are larger than occupations, finite-sample bias is less of a concern for these measures.

[13] While some of this earnings effect may be influenced by the intensive margin effect in which licensed workers work more hours than unlicensed workers, this equilibrium effect is important if spillovers also reduce the hours of other workers.

(Carollo, 2020), though my estimates are robust to their inclusion.[14] 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 244 PUMAs, 110 border match 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 may be primarily driven by the exclusion of parts of coastal California that have 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, which are defined at the state level.[15] There is, therefore, no significant composition difference in the distributions of occupational employment in these areas relative to the country as a whole.

3.3 O*NET

The Occupational Information Network (O*NET) database is the result of a survey fielded by the US Department of Labor. Incumbent workers and occupation experts are surveyed about over 400 attributes of each occupation. These include abilities required to perform the job, the type of tasks performed, and the skill level of the job. The survey also includes variables on knowledge, work style, interests, and work context variables. 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 based on the mean across responses, which I then standardize to be mean zero with standard deviation 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 type (Autor and Dorn, 2013; Autor, 2014).

These are:

1. Non-routine cognitive/analytical
2. Non-routine cognitive/interpersonal
3. Routine cognitive
4. Routine manual
5. Non-routine manual/physical
6. Non-routine interpersonal adaptability

[14] The inclusion of universally licensed occupations only makes the estimates less precise while having only a small effect on sample size.

[15] There is a similar difference in the border design in Blair and Chung (2019). However, my analysis using the full CPS sample (Appendix Figure A13) suggests the treatment effects are similar, meaning that the particular composition of the border sample is not a significant concern.

The components in the O*NET questionnaire used to define these skills are listed in Table 2. 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 that go beyond educational requirements alone, but characterize the abilities, either acquired or endowed, that are essential for someone to perform in 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 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) national figures 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.[16] In addition to these skill measures, 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.

4 Empirical Approach

4.1 Occupation Clustering

To classify occupations into similar groups based on their skill content and therefore 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.[17] 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. It 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.[18]

[16]When O*NET is characterized at the 8-digit SOC level, I first take the average of the skill measures at the six-digit level to match the 6-digit codes in the OES before collapsing to the Census occupation code level. There is little variation in the skill measures within six-digit SOC codes.

[17]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.

[18]For example, there are hundreds of occupations in the O*NET data which separately define the functions of workers who operate specific machinery in production or construction. The specificity of these occupational

Figure 2 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 distance between cluster members (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 line at some distance 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 (Acemoglu and Autor, 2011).[19] 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 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 B. 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 subset of the variables (Yeung and Ruzzo, 2001).

definitions without much difference in the skills necessary to operate these machines makes clusters that include these occupations very dense.

[19] Individual component outliers may influence this measure. 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.

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.

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 present the results of my validation exercises here. I use four validation measures common to clustering applications: 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 3 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. 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 Appendix A. 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.

As a test of the sensibility of the cluster assignments, I present the five most frequent occupations in each cluster at their most compact (20 clusters) in Appendix Table B1. The definitions appear sensible, 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 child care worker can personally attest to taking on multiple roles as a fitness/recreation worker, coach, and umpire—often simultaneously.

Because the goal of the algorithm is to characterize a set of alternatives that are viable, including median wages helps to minimize the possibility of matching occupations with similar skills but whose labor market returns vastly vary for idiosyncratic or industry-specific reasons. For example, though a professional athlete and a freight laborer may use similar physical and cognitive skills, the returns to these may differ dramatically. 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, in part because 43% of the variation in national median wages across occupations is unexplained by these skills. A worker's reservation wage would exclude a set of lower-wage occupations from consideration even if their skills were similar, while the infeasibility of entering many higher-wage occupations eliminates some from consideration even if their skills were similar. Excluding some measure of market returns from the clustering algorithm misclassifies cluster members and therefore the treatment variable (share of the cluster that is licensed), leading to measurement error by including data from irrelevant alternatives. Including national median wages bounds the distance between cluster members in wages conditional on being close on their skill components—essentially serving the function of trimming wage outliers from the group and classifying them in another.[20]

4.2 Border Match Design

In the experimental ideal, a researcher choosing to study the effects of occupational regulations like licensing on the labor market would allocate a random assignment of licenses and occupations to a treatment group taken from a pool of prospective workers and then observe the development of new market equilibrium outcomes in the treatment group in comparison to a control. To approximate this experimental ideal, 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

[20] See Appendix Table B2 for a demonstration of how even top focal occupations begin to include irrelevant or puzzling alternatives. For example, the most common occupations in cluster 5 are Police Officers, Computer Support Specialists, Sales Representatives, and Credit Counselors. This compares to cluster 6 when including wages, whose top occupations are Police Officers, Editors and News Analysis, Biological Scientists, Construction Inspectors, and Private Detectives. Appendix Figure B2 shows that there is a limited correlation between cluster licensure shares when including versus excluding wages from the algorithm. Overall, cluster licensure is significantly under-predicted in the upper half of the distribution when excluding median wages, mostly attributable to what is nearly a flat slope between the two in different ranges. If total licensure over all states in an occupation is correlated with wage returns to a package of skills, this suggests lower-wage occupations have irrelevant higher-wage occupations in their same cluster and vice-versa when excluding wages from the HAC exercise. See Section 5.5 for a separate approach that relies on skill-distance weighted exposure measures that exclude wages and leads to the same conclusions. Appendix B provides further justification for using these six skill measures and wages as inputs into the clustering algorithm rather than an alternative principal-component measure of skills.

the state level, and all workers in the state face the same licensing rules. Workers in different parts of each state live under different local economic conditions (near each state border) and have neighbors across the border that share in 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, therefore includes state, occupation, and border fixed effects:

$$y_{iocms} = \beta_0 + \beta_1 LicensedShare_{os} + \beta_2 LicensedShare_{cs}^{-o} + X_i'\beta_3 + \delta_o + \gamma_s + \tau_m + \varepsilon_{iocms} \quad (3)$$

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 border m.[21] Outcome y is log weekly earnings for my main specification. Coefficient β_1 captures the earnings effect of licensing individual i's entire *own* occupation category (o), whereas β_2 captures the effect of fully licensing all other workers in cluster c *outside* of occupation o (the focal occupation). Importantly, β_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" sectors, 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 occupation.[22]

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, nativity (native- vs foreign-born) in separate models. This allows the effects of licensure to vary across groups. I also calculate quartiles of the distribution of labor market size across PUMAs, which I define as the population of working adults with positive earnings in the PUMA. I then interact these indicators with my licensing variables. Labor market size is a particularly important measure, as theory and empirical evidence both suggest that smaller labor markets experience greater labor market concentration and monopsony power due to fewer outside options (Rinz, 2018; Dodini et al., 2020).

[21] For PUMAs that share borders with multiple states, I stack the sample and divide the sample weights by the number of borders.

[22] 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) noticeably 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.

To measure composition effects, I also use indicators for demographic groups as outcome variables. Instead of log weekly earnings, I replace y_{iocms} with binary indicators for race/ethnicity categories (i.e. indicators for being a female worker, Non-Hispanic White, Non-Hispanic Black, Hispanic, as well as foreign-born), broad education categories (i.e. Associate's, Bachelor's, Master's degrees and a PhD/professional degree) as well as a binary indicator for being age 18-25. For the composition regressions, I omit other individual controls. 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 β_2 tells usthe average change in the probability that a worker in occupation o 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 and run the same specification excluding individual characteristics:

$$EMP_{opcms} = \beta_0 + \beta_1 LicensedShare_{os} + \beta_2 LicensedShare_{cs}^{-o} + \delta_o + \gamma_s + \tau_m + \varepsilon_{opcms} \quad (4)$$

The own-occupation effect (β_1) measures the effects of licensure on employment in that occupation itself, while β_2 captures the employment spillovers. A pure labor supply explanation for the earnings effects I find would predict a negative β_1 and a positive β_2 coefficient as licensing pushes workers into other occupations using similar skills. A negative β_2 spillover coefficient on employment in the focal occupation is suggestive of monopsony power if earnings effects are also negative in my individual models in Equation 3.

4.2.1 Identifying Variation and Assumptions

The econometric challenge of identifying the causal effects of occupational licensing on earnings and employment using observational data are 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 labor demand or industry agglomeration; 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 border match design overcomes this obstacle by comparing workers in the same occupation on two sides of the same state border where the state line 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 across each specific border pair. Because licensure is determined at the state level, not the local level, differences in licensure across each specific border pair are then due to processes conditionally unrelated to residual determinants of worker earnings.

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 PUMAs in the state. These include regulatory conditions (e.g. minimum wage laws, education regulations, tax policy, overall state propensity to license, etc) as well as statewide shared economic conditions (e.g. industrial composition, historical comparative advantage, etc.). It is also key to remember that licensing intensity in my data is constructed at the state level as well.

The border 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 effect) 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.[23]

As a concrete example, consider PUMAs in the state of Virginia in Figure 4. In my regression, the occupation fixed effects ensure that comparisons come from cross-sectional 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 4, the states that border Virginia are written in different areas in the state. 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. The remaining variation in each outcome must then mechanically come at the sub-state level, either at the regional or individual level.

The border fixed effects take into account regional variation in economic conditions related to earnings and employment for carpenters. For example, PUMAs that border Maryland may have different economic conditions than PUMAs that border North Carolina, and the border fixed effects account for this. Finally, the border 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

[23]This approach is a methodological advance from much of the past literature on occupational licensing, which has generally considered individual licensing status as the treatment variable despite possibly endogenous selection to having a license and treats all other workers in the same occupation (or in some cases, across all occupations) as implicit control units. For example, Kleiner and Krueger (2013); Gittleman et al. (2018); Zhang and Gunderson (2020) take this approach to study the licensing wage premium and compare it to unionization. Many studies use a border match approach, including some in the occupational licensing literature (Blair and Chung, 2019; Black, 1999).

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 pair occur for 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 conditional on their differences in licensing exposure across each border pair: southern VA vs northern NC, southwestern VA vs eastern KY and northeastern TN, and northern VA vs eastern WV, western MD, and DC. The overall average treatment effect in my regression is, therefore, a weighted average of all the within-occupation comparisons across all occupations across all state border pairs across all PUMAs in my sample conditional on all systematic, shared characteristics within states.

These estimates of the average treatment effect will be unbiased measures of the causal effect as long as there are no factors related to licensure and earnings that systematically vary across all state borders. For example, for an unobserved factor to confound my estimates for carpenters in Virginia, that factor would have to consistently relate to licensing and earnings for carpenters in the same direction across every border pair. In other words, that factor or policy would have to differ from Virginia in a similar fashion as licensure rules in DC, Maryland, Tennessee, West Virginia, Kentucky, and North Carolina, which seems implausible in this setting.

In Section 5.5, I explore a large battery of robustness tests and specifications that vary the assumptions of the model in order to rule out a number of alternative explanations. These include a different approach to the border match design, a placebo exercise, the addition of PUMA fixed effects, using different data sources for licensing data and/or outcomes of interest, and leveraging variation in licensing rules over time. All my various tests support the results of my main approach.

5 Results

5.1 Earnings Premium and Spillovers

I first present the results for the overall earnings effects of widespread occupational licensure. Figure 5 plots the coefficients and confidence intervals for occupation spillovers for 4 to 20 clusters. For ease of visualization, coefficients are scaled to 100% licensure in one's own occupation and cluster. A 10 percentage-point increase in cluster licensure represents approximately a 1 standard deviation change (11 percentage points). The figure indicates that having 100% licensure for one's own occupation leads to an earnings premium of approximately 8%, a finding consistent with the findings in the prior literature (Kleiner and Krueger, 2013; Gittleman et al., 2018), including some that leverage cross-state policy variation over time in certain occupations (Carollo, 2020; Pizzola and Tabarrok, 2017). 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 1.5–2.5% on average. The confidence intervals rule out average effects smaller than -0.5–1% 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 1.5-2%.[24] To ease interpretation, I present the rest of my estimates using 20 skill clusters, the most conservative set of estimates. The results from varying the number of clusters are in Appendix A, and each follows a similar pattern to the overall estimates. Taken together, under strong assumptions about the equality of spillover effects across licensed and unlicensed occupations, an 8% wage premium within one's own occupation combined with a 15–20% penalty for full cluster licensure would be comparable to the overall "cross-sector" wage difference in the prior literature when estimated without occupational controls. Though caution in these comparisons is warranted, the implication is that a large share of the cross-sector wage differential may be attributable to negative spillovers.[25]

I find substantial heterogeneity in this effect across gender as well as race/ethnicity and nativity as detailed in Figure 6. Panel A shows the effects of licensure in one's own occupation, while Panel B shows the spillover effects of 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 17-18% 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 3%. The same coefficient is less than 1% for men. Non-Hispanic black workers and Hispanic workers experience larger earnings spillovers than their Non-Hispanic white counterparts. The point estimate for Hispanic workers is around -2% for a ten percentage point increase in cluster licensure compared to -1.5% for Non-Hispanic white workers, though the estimates are less precise at 20 clusters. Non-Hispanic black workers experience the largest penalty, with a point estimate of approximately 2.8%. 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, an idea explored in Blair and Chung (2018; 2019). As licenses that exclude those who have been convicted of a crime increase, 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 signal (or a signal of never having been convicted) may be higher in this case (Blair and Chung, 2018).

Most of the negative earnings spillover effect on Hispanic workers is driven by foreign-born Hispanic workers. The estimates indicate that there is essentially no earnings premium for foreign-born Hispanic workers in licensed occupations, perhaps because a smaller share of Hispanic immigrants can obtain a license when compared to other immigrant groups, be it for

[24] 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.

[25] Key estimates for the total average cross-sector wage differential are approximately 24% (Gittleman et al., 2018), approximately 20% for those above the median in Canada (Zhang and Gunderson, 2020), and 30% (Kleiner and Krueger, 2013).

education, language, or legal status reasons. Spillover effects for a ten percentage point increase in cluster licensure are nearly 3% compared to just over 1% for native-born Hispanic workers. 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. Both a direct labor supply effect into unlicensed occupation and a monopsony effect could explain this difference. In the monopsony case, the threat of leaving a firm to pursue another job or another occupation may be limited by concerns about legal work status.

Given the presence of possible statistical or taste-based discrimination against Non-Hispanic black workers as well as the additional imposition of citizenship or residency requirements for foreign-born workers, I expect spillover effects to be 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%.[26]

Finally, I estimate my model interacting my licensure measures with an indicator for quartiles of the size of the working 18–64 population in the PUMA to trace out heterogeneous treatment effects of spillovers over labor market size. Figure 7 shows that there is a clear relationship between the intensity of the earnings effects of licensure and labor market size. Panel A shows the interaction between own-occupation licensure and quartiles of labor market size, while Panel B shows the interaction between the spillover coefficients and size quartile. The largest labor markets (Quartile 4) exhibit a smaller earnings premium in licensed occupations (4-5%) and no spillover effect. For the other three quartiles, this relationship intensifies as market size declines. In the bottom three quartiles, the own-occupation effect is consistent at approximately 10% for full licensure, while the spillover effect is as large as -3.2% in Quartile 1 compared to -1.2% in Quartile 3 for a 10 percentage point increase in licensure.[27]

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 are less likely to be able to absorb the costs of licensing requirements. In addition, the spillovers are strongest in smaller labor markets that are likely to be less saturated with job openings and networks in which to search for a job.

[26]In contrast, due to the highly selective nature of immigration to the United States from European countries, there is no detectable earnings premium nor spillover effect for foreign-born, Non-Hispanic white workers. These results are available upon request.

[27]Estimates varying the number of clusters are in Appendix Figure A8.

5.2 Composition Effects

In addition to direct earnings effects, licensing spillovers may shift the distribution of workers within occupations in terms of educational attainment, sex, nativity, or race/ethnicity depending on differential ability to absorb the costs or returns to obtaining a license. To test this, I estimate linear probability models on binary indicators for sex, education categories, race/ethnicity groups, nativity, and an indicator for being age 18–25 using the same specification as my earnings model, except I exclude other individual controls. I again plot the coefficients and confidence intervals for within-cluster spillovers set at 20 clusters.

Figure 8 shows that as other occupations in the cluster become more licensed, the likelihood that a worker in the focal occupation is a woman or holds an advanced degree falls. The probability of having a Master's degree falls by 0.75 percentage points with a 10 percentage point increase in cluster-wide licensure outside the focal occupation. Relatedly, workers in the focal occupation are more likely to be Hispanic or foreign-born. Increasing cluster licensure by 10 percentage points leads to an increase in the likelihood that workers in the focal occupation are Hispanic or born outside the US of 0.8 and 1 percentage points respectively. These are the largest spillover effects I find across all outcomes.

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 into the remaining unlicensed occupations, although the most advanced degrees do decline marginally. It does not appear that shifting human capital, per se, is responsible for the decline in earnings. Rather, there is a shift in the gender and race/ethnicity composition of the focal occupation, as well as a marginally significant increase in the share of young workers age 18–25. Widespread licensing appears to push some men (women) out of (into) licensed occupations and into (out of) unlicensed occupations, as evidenced by the fact that the share of women in the focal occupations shifts downward. Hispanic workers and foreign-born workers filter out of licensed occupations in the skill cluster and into the remaining unlicensed occupations. Overall, the composition effects, while statistically significant, are small relative to the observed earnings effects.

5.3 Employment

To understand the other mechanisms underlying the earnings effects I observe, I estimate a border match model of employment within each occupation-PUMA cell.

Figure 9 indicates that overall employment in each occupation falls by approximately 5 workers when the occupation is fully licensed. However, as licensure increases across the cluster, overall employment in the focal occupation *falls* by 20–30 workers when the rest of the cluster is licensed. Given that the ACS is a 1% sample, this implies overall negative labor supply effects of approximately 500-3000 workers in the typical occupation-PUMA cell. If occupational licensing increases labor market power, monopsonistic firms may employ workers at rates lower than they otherwise would in a competitive market, leading to effects consistent with the observed declines in employment I find.

Like the earnings effects discussed previously, the employment effects differ widely across labor market sizes. Figure 10 shows that the negative coefficients on employment in the focal occupation are particularly pronounced in smaller labor markets. The effect is approximately 45 fewer workers in the focal occupation in the smallest labor markets, compared to zero for the largest.[28]

Taken together, these results indicate that widespread licensing in a skill cluster lead to negative employment effects. This is particularly true in smaller labor markets. The strong negative employment and earnings effects of licensing spillovers appear suggestive of monopsony power, which I discuss in Section 6.

5.4 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. To contextualize these effects in the distribution of incomes, I present graphical evidence of the counterfactual kernel density distribution within occupations 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 3 and predict individual earnings based on this model. Then, setting licensure 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 license in their own occupation and no licensure in their cluster?"

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 explains the relatively uneven densities, which we would not expect in an unconditional distribution. Notably, this counterfactual exercise 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.

The results are in Figure 11. Panel A shows that after eliminating licensing in the sample, predicted weekly earnings shift rightward across most of the conditional distribution. Panel B shows the differences in the densities and suggests that 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.

[28] Estimates varying the number of clusters are in Appendix Figure A9.

To formalize the comparison between the two distributions of predicted weekly earnings, I generate distributional statistics for each: the ratio of the 90/10, 90/50, and 10/50 percentile ratios and the Gini coefficient. Table 3 shows the comparison of these statistics across the distributions. There are significant changes in within-group inequality as a result of eliminating occupational licensing in my sample. The ratio of the 90th to 10th percentile of weekly earnings would fall by nearly 4%, and the ratio 90th to 50th percentiles would fall by 2.5%. 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%.

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. Because many workers in "universally" licensed professions earn particularly high incomes (e.g. physicians, attorneys, pilots), the results also imply that eliminating licenses for which there is not a national consensus for their usefulness (or where states differ in their licensing rules) would reduce earnings inequality in the unconditional distribution of earnings as well 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.

5.5 Robustness to Alternative Explanations

As an alternative approach to the border match design, I re-estimate my models changing the operating fixed effects. I omit state fixed effects and instead add interacted fixed effects at the occupation-by-border-pair level. This narrows identifying variation in the model in a slightly different manner than my combination of state, occupation, and border fixed effects. Nevertheless, the spillover results from this approach are nearly identical to my main models, which lends further credibility to my approach to the border match design (see Appendix Figure A2).

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 Appendix Figure A3). 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 relative skill distance. The results generally confirm the rest of my analysis, but with relatively wider standard errors (that make heterogeneous effects noisy) and a different interpretation: the coefficient on 100% licensure implies that every occupation in the entire state is fully licensed. 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.

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

As such, in order to be driving my estimates, any state-level policy difference across borders must affect certain border PUMAs differentially, and that differential policy effect must be systematically pointing to the same direction. For example, state industry policies must affect PUMAs at the Virginia-Maryland border differently than they do the Virginia-North Carolina border, and that differential effect must be correlated with the difference in licensing across the VA-MD and VA-NC borders, and so on for all borders.

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 because we know they 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 then 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 with the same specification as Equation 3 in the ACS. If the licensing environment is correlated with state variables that are also correlated with the distribution of earnings, then the relationship between licensing exposure and earnings and employment should not significantly change.[29]

The results of this exercise are in Figures 12 and 13. Panel A of Figure 12 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

[29]Another implicit indication that overall licensure is not a significant driver of my results 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.

general 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 6. Similarly, Panel B shows that there is no relationship between occupational composition and licensing exposure in placebo clusters, indicating that the propensity to license does not reflect occupational composition differences across localities. With regards to employment, Figure 13 shows that there is no general relationship between overall employment in each occupation and licensing exposure in placebo clusters. When considering labor market size, there is a slightly positive relationship between employment and the tendency to license in the largest local labor markets, though the relationship is not large nor statistically significant for the other three quartiles.

This exercise greatly strengthens the case for a causal interpretation of the spillover effects I have identified. By showing that licensing exposure within placebo clusters does not result in significant estimates, I show 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 skill cluster structure from my clustering algorithm.

To ensure that other unobserved characteristics of the local labor market in the PUMA are not biasing my results (e.g. cross-border differences in labor demand specific to each PUMA), I estimate my border match sample including PUMA fixed effects in Section 5.5 and include "universally" licensed occupations in the analysis, 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 DC and souther-western Virginia PUMAs that border Kentucky and Tennessee. Results from this exercise show that my border design is robust to unobserved characteristics of the hyper-local labor market. This is because the average treatment effect now comes from comparing the earnings of workers in a single PUMA to those on the other side of two state borders. Going back to the Virginia example, this specification compares a carpenter in Fairfax County (in northern Virginia) to a carpenter in Washington, DC and also to a carpenter in Bethesda, Maryland, and the average treatment effect for carpenters in Fairfax County is the weighted average of these two pairwise differences conditional on their differences in licensing rates after accounting for unobservable local economic conditions and policies in the county. Appendix Figures A10, A11, and A12 show these results 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 are not biasing or driving my baseline model.

One concern about a border match design of this nature is the possibility of spatial spillovers and cross-state commuting. If individuals move across the state border to avoid occupational regulations or if workers in one jurisdiction commute to and work in another jurisdiction, this should bias my estimates of the labor market effects of licensure towards zero, meaning my estimates would be a lower bound. This is because licensed workers may live in one jurisdiction and contribute to the licensing rate in that state while actually working with their license in another. However, because licensing rates are calculated at the state level and not the PUMA level, this bias is likely to be small. That my results are almost identical with PUMA fixed effects strengthens the argument. Similarly, workers less exposed to licensure in their own state may nevertheless experience effects from licensing in the bordering state. With the inclusion of PUMA fixed effects, identifying variation comes from a smaller set of local labor markets, but the results are similar.

To further 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 pair fixed effects. The results in Appendix Figure A13 show a similar wage premium to Kleiner and Soltas (2019) and spillover estimates that are very similar to my border match design. It is, therefore, unlikely that sample selection in my border areas or peculiarities in cross-border commuting and/or spatial spillovers are driving my results.

As an additional check, I re-estimate my earnings regressions at 20 clusters while sequentially eliminating a cluster at a time. This allows me to pinpoint if my results are driven by any particular cluster, large or small. Figure A14 indicates that the overall earnings estimates are not sensitive to any particular cluster. For two of the clusters, my estimates fall from -0.15 to -0.1 log points, though the difference is not statistically significant. For these tests by gender and race/ethnicity, see Appendix Figures A15 and A16. I also show the employment effects regressions in this same format in Figure A17. The effects on total employment, while visually sensitive to the exclusion of Cluster 1, are not statistically significantly different upon excluding it. There is still a sizable negative employment effect. I show a similar graph for my composition regressions as well in Appendix Figures A18 and A19.

Abstracting away from the measurement issues discussed in Section 3, I use the Northwestern Licensing Database (Redbird, 2016) to estimate the earnings spillover effects of licensure in my border match design and present those estimates in Appendix Figure A20. The result is an attenuated own-occupation earnings premium and a slightly smaller spillover effect compared to my base model. That the attenuation is more pronounced in the own-occupation effect is notable because individual occupations are smaller and the consequences of measurement error are more pronounced relative to larger clusters that are aggregations of several occupations. However, 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—to drive my results, such a factor must: 1) differentially affect earnings and employment in specific PUMAs without spilling over into PUMAs on the other side of the shared state border (from the simple cross-state model and the PUMA fixed effects model); 2) hold across the types of occupations being considered (from the models eliminating clusters); 3) would have to hold even when controlling for conditions of the local labor market (from the PUMA fixed effects model); 4) be positively correlated with cluster-specific propensities to license in a way that is specifically correlated with skill cluster structure from O*NET (from the placebo model); and 5) be correlated with licensing from two different databases (from the NLD model). Tax policy, minimum wages, industrial relations and unionization policies, industry composition, local labor demand, 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.

5.6 Extension: Time-Varying Measures of Licensure

Finally, 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.[30] Here, identification of the spillover effects and the within-occupation effect of licensing comes from variation in licensing laws across states over time within occupations and within skill clusters.

To construct this 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 3 but I include occupation, state, and year fixed effects:

$$y_{iocst} = \beta_0 + \beta_1 LicensedShare_{ost} + \beta_2 LicensedShare_{cst}^{-o} + X_i'\beta_3 + \delta_o + \gamma_s + \tau_t + \varepsilon_{iocst} \quad (5)$$

As in my earlier analysis, I also interact these time-varying measures of licensing with my different demographic groups to examine heterogeneous treatment effects. To demonstrate the effect of time-varying intensity of measurement error, 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

[30] Carollo (2020) 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 cross-sectional estimates. That database was created as part of a set of recent working papers but is, however, still under construction (according to the author) and is not available for public release.

and estimate a similar repeated cross-sectional model to the earnings equation above, omitting individual controls.

The results for the spillover coefficients are in Table 4.[31] 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 standard deviation of 0.162 compared to a standard deviation of 0.11 in the border match sample, I can rescale these estimates to match in terms of standard deviations. 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 border match estimates. 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).

Looking at heterogeneous treatment effects, I find similar wage patterns of penalty differentials as I find in my cross-sectional 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 6 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 Figure 4, a 10 percentage point increase in cluster licensure outside the focal occupation reduces employment in the focal occupation in the state by approximately 13–34 workers depending on the sample dates. At an average of approximately 20 PUMAs per state, my border match estimates imply an average statewide reduction of 55–60 workers in each focal occupation with the same change in cluster licensure. The estimates in Panel B of Table 4 for my time-varying model are, therefore, smaller than those in my border match design. However, the differences are similar to how the wage estimates in Panel A differ from the earnings estimates in Panel B of Figure 6 (a ratio of the standardized estimates of approximately 0.6).

In summary, although considerable measurement error may bias these estimates toward zero, I find suggestive evidence of significant negative spillovers in hourly wages on similarly

[31]The own-occupation effects (β_1) are small and imprecisely estimated for all repeated cross-sectional estimates, so I omit these from the table. Because occupation cells are smaller than cluster cells, measurement error in licensure may be more consequential in terms of attenuation bias. I see the same phenomenon with the NLD in Figure A20, where the own-occupation effect in my border match design changes far more than the spillover effect.

skilled occupations in the CPS using variation in licensing laws over time, including similar patterns in the relative size of the heterogeneous spillover effects across race/ethnicity and gender when compared to my border match estimates. I also find significant negative spillover effects of cluster licensure on employment in the focal occupation, just as in my main estimates. Overall, this exercise supports the totality of my border match results regarding spillovers.

6 Discussion

The pattern of lower earnings and lower employment in the focal occupation as a result of cluster-wide licensure is consistent with an increase in monopsony power in the local labor market.

Two key implications of monopsony theory are: 1) that even firms in what are ostensibly competitive labor markets can exhibit monopsony power if there are substantial costs to the worker for a job change; and 2) firms with monopsony power may employ fewer workers and pay lower wages than otherwise equivalent firms in competitive local labor markets (Ashenfelter et al., 2010).

I argue that 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. This view is supported by recent work that explores the use of more comprehensive definitions of a "local labor market" for workers in the measurement of the effects of labor market concentration and concludes that incorporating outside options is an important component (Schubert et al., 2019; Dodini et al., 2020). It is clear from the past literature that licensing increases labor market rigidity across occupations. Kleiner and Xu (2020) find that workers that are licensed are 24% less likely than unlicensed workers to have recently switched into their occupation. That transitions between occupations fall logically fits into a monopsony framework in which incumbent workers cannot credibly threaten to leave a low-paying firm. As licensing increases the cost of leaving a firm to pursue outside options within and across skill clusters as well as across state lines, the set of available options that are feasible for them to enter shrinks, which may exacerbate the low 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.

There is also 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.[32]

The literature also suggests that immigrants supply labor to the firm much less elastically than their native-born counterparts, which Hirsch and Jahn (2015) predict leads to a pre-

[32] This also is related to the fact that women generally perform more non-routine, cognitive work than men on average, and these tasks as performed by those with more education are more exposed to monopsonistic behavior by firms (Bachmann et al., 2019; Dodini et al., 2020).

dicted 7% wage penalty. 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). These two points together may partially explain why native-born and foreign-born Non-Hispanic black workers and women face the largest earnings penalty.

Finally, the case for a monopsony explanation is bolstered by the observation that both the negative labor supply spillovers and earnings penalties are stronger in smaller labor markets. The literature on labor market concentration suggests that smaller labor markets experience higher levels of concentration and also exhibit a stronger negative relationship between concentration and wages (Rinz, 2018; Dodini et al., 2020). Switching costs may be lower in large labor markets because of the physical proximity of available jobs and a wide set of available choices from which a worker may select. Relatedly, smaller labor markets may also imply a smaller product market for services performed by licensed workers. Limiting the entry of product market competitors in a smaller market leads to larger relative changes in product market power, which is positively correlated with labor market concentration (Marinescu et al., 2019; Lipsius, 2018; Qiu and Sojourner, 2019). For example, a massage therapist in a small town may be one of only a few producers of that service, whereas, in a larger labor market, that is unlikely to be the case. The imposition of costly licensing requirements leads to relatively larger market share changes in both the product and labor markets for that service in smaller areas.

Taken together, a reduction in market power by employers would increase incomes across the distribution, leading to a reduction in inequality both within and across occupations.

7 Conclusion

This analysis has presented the first evidence of substantial direct labor market spillovers from occupational licensing in the United States using a border match 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 increase in licensure. 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. 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. That total employment in related occupations falls is consistent with a monopsony model in which licensure increases search and adjustment costs, reduces a worker's outside options, and reduces a monopsonistic firm's incentives to hire new workers. 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, and significantly reduce pre-tax earnings inequality within occupational groups.

While the analysis presented here shows significantly lower employment as well as worker composition shifts as a result of licensing spillovers, 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. Future work in this area may attempt to directly measure the effects of occupational licensing on labor market power in particular industries or occupations using administrative or other data.

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, 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. This analysis suggests that 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 that 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—even within occupations.

The employment and earnings effects of licensing and other labor market regulations, if broadened to include more occupations, may lead to 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.

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.

Azar, José, Ioana Marinescu, and Marshall Steinbaum. 2020. "Labor Market Concentration." *Journal of Human Resources* 1218–9914R1.

Bachmann, Ronald, Gökay Demir, and Hanna Frings. 2019. "Labour Market Polarisation and Monopsonistic Competition."Technical report, Working Paper.

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

Balasubramanian, Natarajan, Jin Woo Chang, Mariko Sakakibara, Jagadeesh Sivadasan, and Evan Starr. 2020. "Locked In? The Enforceability of Covenants Not to Compete and the Careers of High-Tech Workers." *Journal of Human Resources* 1218–9931R1.

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.

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.

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.

Carollo, Nicholas A. 2020. "The Impact of Occupational Licensing on Earnings and Employment: Evidence from State-Level Policy Changes."Technical report, Working paper, available at: https://www. ncarollo.net.

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

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

Cortes, Guido Matias, and Giovanni Gallipoli. 2018. "The Costs of Occupational Mobility: An Aggregate Analysis." *Journal of the European Economic Association*, 16(2): 275–315.

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."Technical report, National Bureau of Economic Research.

Dillender, Marcus, Anthony T Lo Sasso, Brian J Phelan, and Michael R Richards. 2022. "Occupational Licensing and the Healthcare Labor Market."Technical report, National Bureau of Economic Research.

Dodini, Samuel, Michael Lovenheim, Kjell Salvanes, and Alexander Willén. 2020. "Monopsony, Skills, and Labor Market Concentration."(DP15412): , URL: https://t.co/LUg2Z4qB6x?amp=1.

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."Technical report.

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.

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.

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*, DOI: http://dx.doi.org/10.1257/jep.14.4.189.

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(2): , DOI: http://dx.doi.org/10.1086/669060.

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."Technical report, National Bureau of Economic Research.

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*, DOI: http://dx.doi.org/10.1007/s11149-017-9333-y.

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*, Forthcoming, URL: https://doi.org/10.1287/mnsc.2020.3918, DOI: http://dx.doi.org/10.1287/mnsc.2020.3918.

Lipsius, Ben. 2018. "Labor Market Concentration Does Not Explain the Falling Labor Share." *Available at SSRN 3279007*.

Marinescu, Ioana Elena, Ivan Ouss, and Louis-Daniel Pape. 2019. "Wages, Hires, and Labor Market Concentration." *Available at SSRN 3453855*.

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

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.

Qiu, Yue, and Aaron Sojourner. 2019. "Labor Market Concentration and Labor Compensation." *Available at SSRN 3312197*.

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.

Ransom, Tyler. 2021. "Labor Market Frictions and Moving Costs of the Employed and Unemployed." *Journal of Human Resources* 0219–10013R2.

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*, DOI: http://dx.doi.org/10.1177/0003122417706463.

Rinz, Kevin. 2018. "Labor Market Concentration, Earnings Inequality, and Earnings Mobility." *Center for Administrative Records Research and Applications Working Paper*, 10.

Robinson, Joan. 1933. *The Economics of Imperfect Competition.*: Macmillan.

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. 2019. "Mitigating Monopsony: Occupational Mobility and Outside Options."

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

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.

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.

Webber, Douglas A. 2016. "Firm-Level Monopsony and the Gender Pay Gap." *Industrial Relations: A Journal of Economy and Society*, 55(2): 323–345.

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.

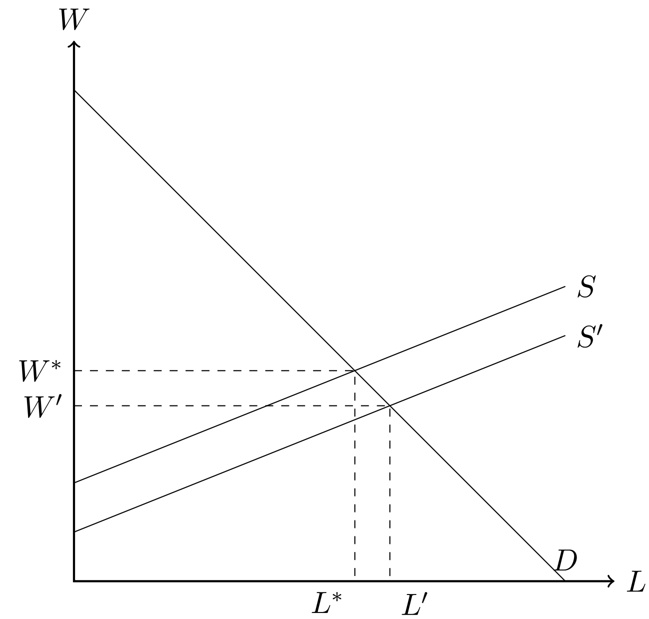
Yeung, Ka Yee, and Walter L. Ruzzo. 2001. "Principal Component Analysis for Clustering Gene Expression Data." *Bioinformatics*, 17(9): 763–774.

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.

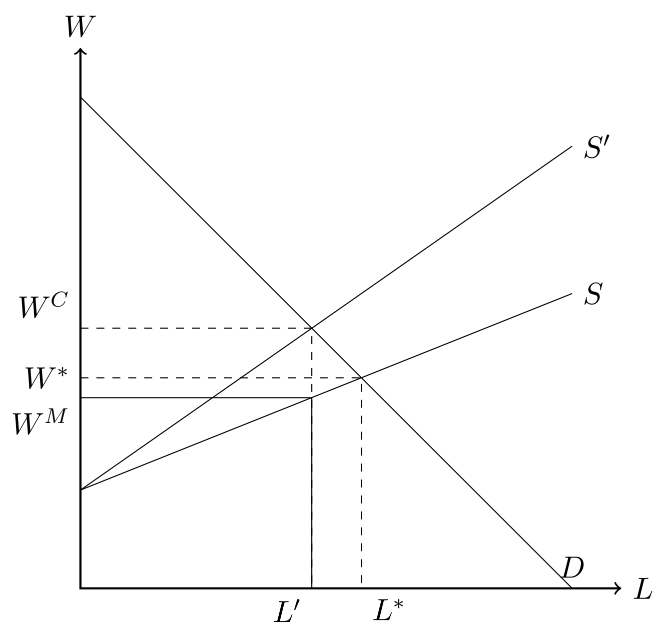
Figures

Figure 1: Competitive vs Monopsonistic Labor Market

Panel A: Competitive Market Labor Supply Shift

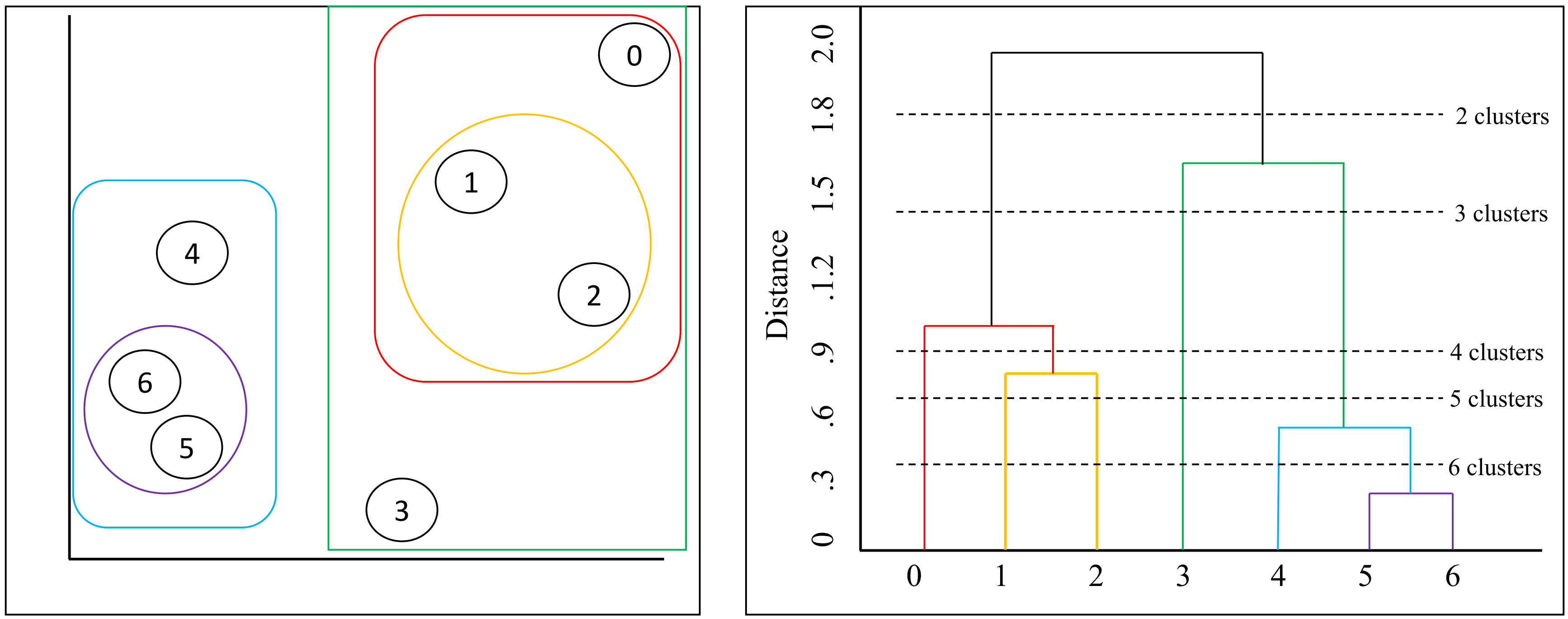


Panel B: Monopsony Model



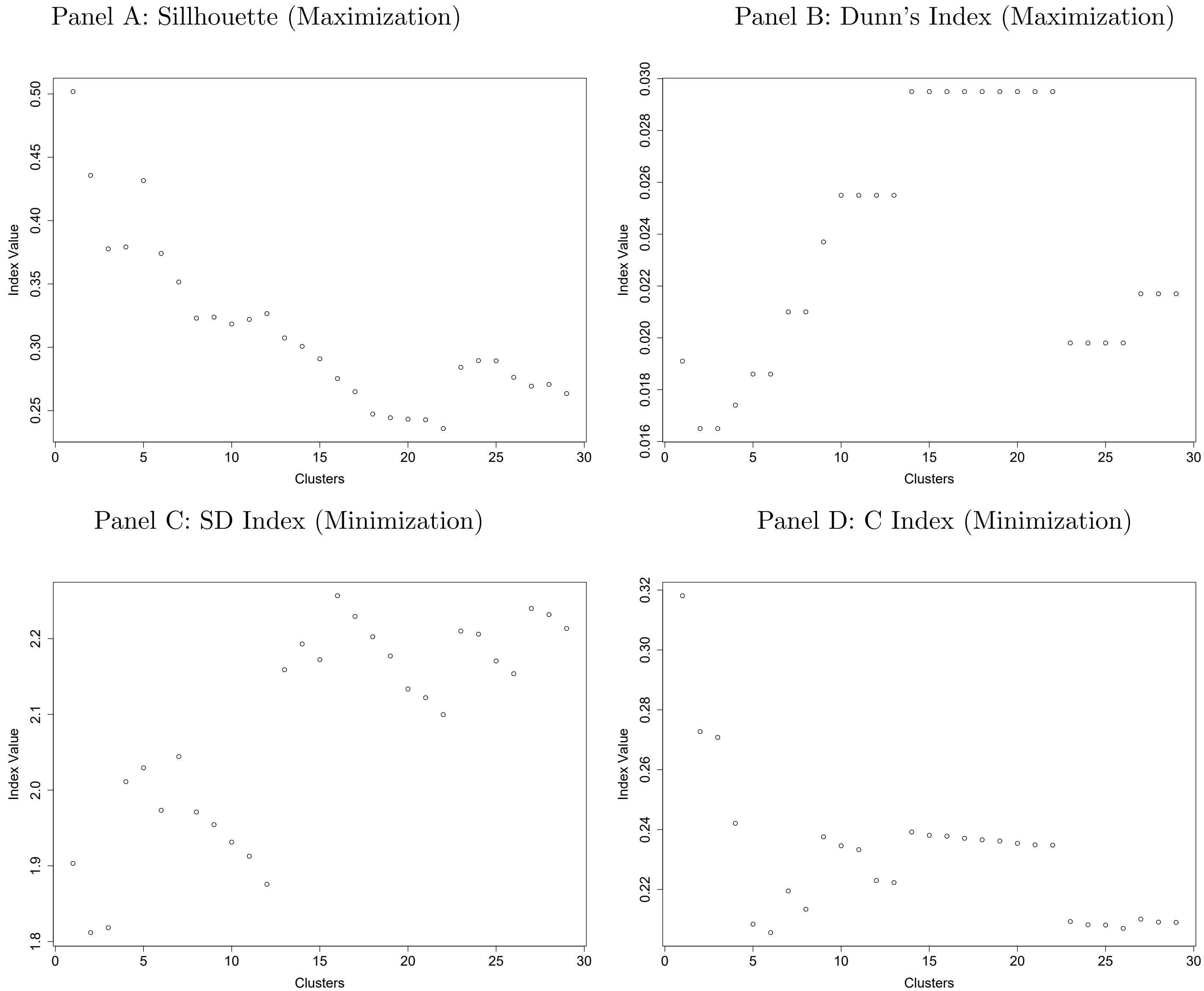
Notes: An illustration of the possible spillover effects of occupational licensure onto other occupations in a labor supply (competitive) model vs a monopsony model.

Figure 2: Hierarchical Agglomerative Clustering Example



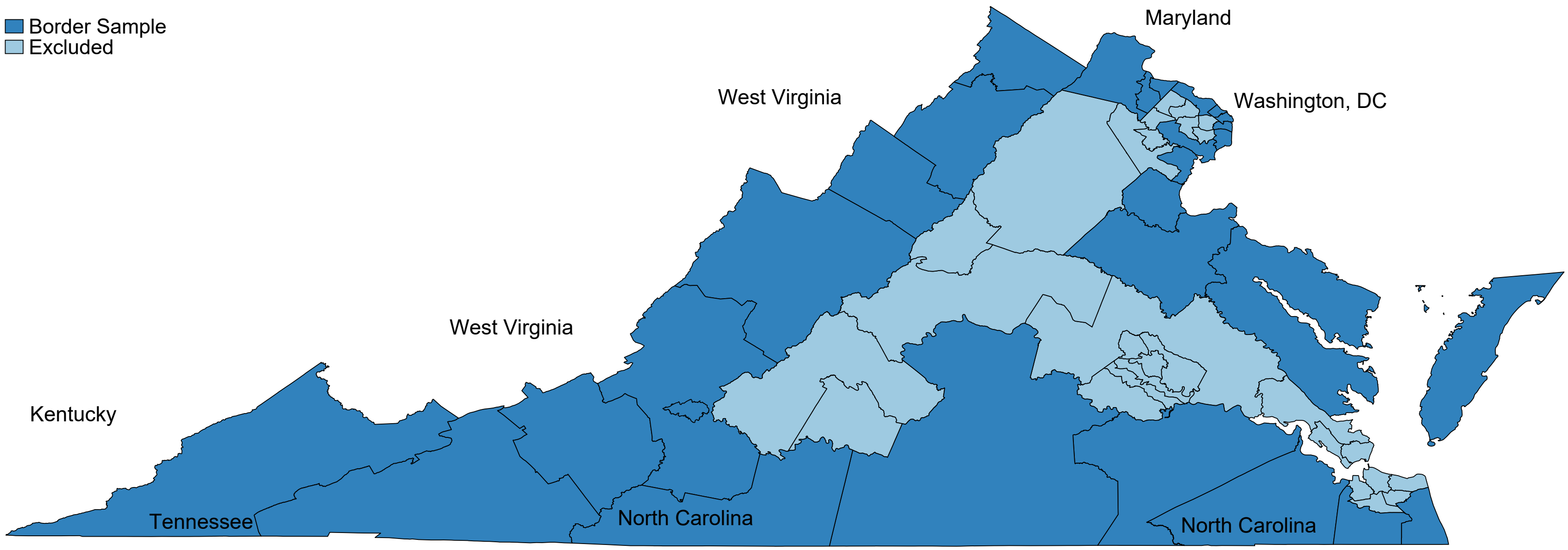
Notes: Moving up the dendrogram, clusters are sequentially merged. The researcher can then choose cut points at a certain number of clusters or at a maximum distance value within clusters. First, groups 5 and 6 merge to form the purple cluster. Next, this purple cluster merges with group 4 to form the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

Figure 3: Cluster Validation Exercises



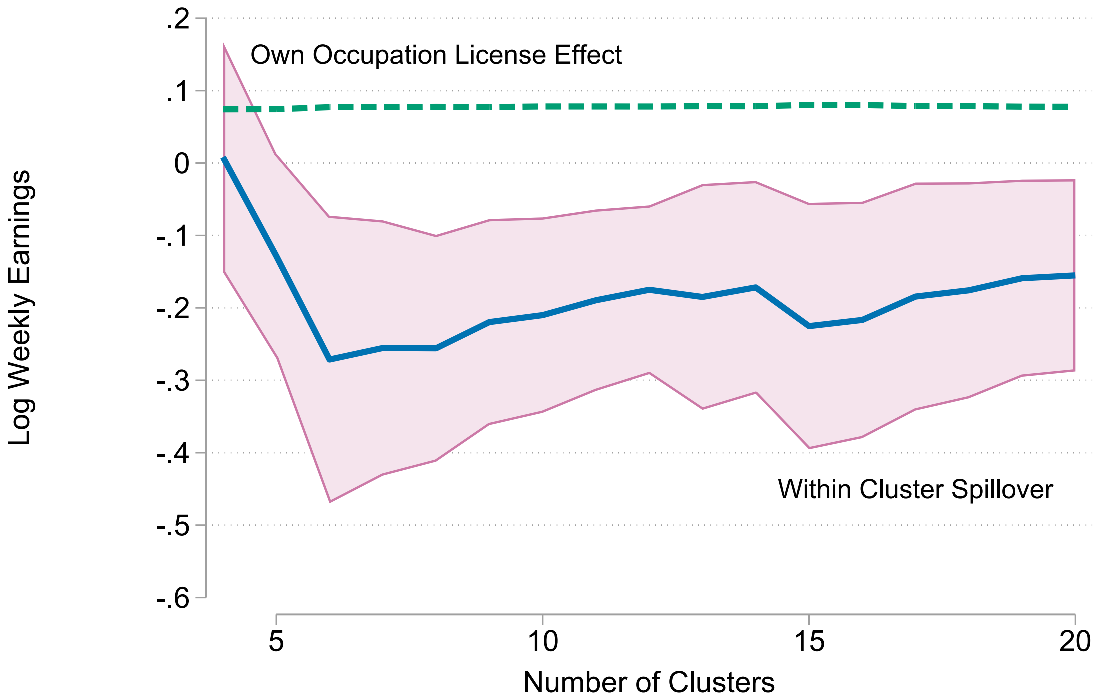
Source: Author's calculations of O*NET skills data following six skills in Acemoglu and Autor (2011) and median log wage.
Notes: Clusters are generated using the HAC approach detailed in section 4.1.

Figure 4: Border Match 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 3 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.

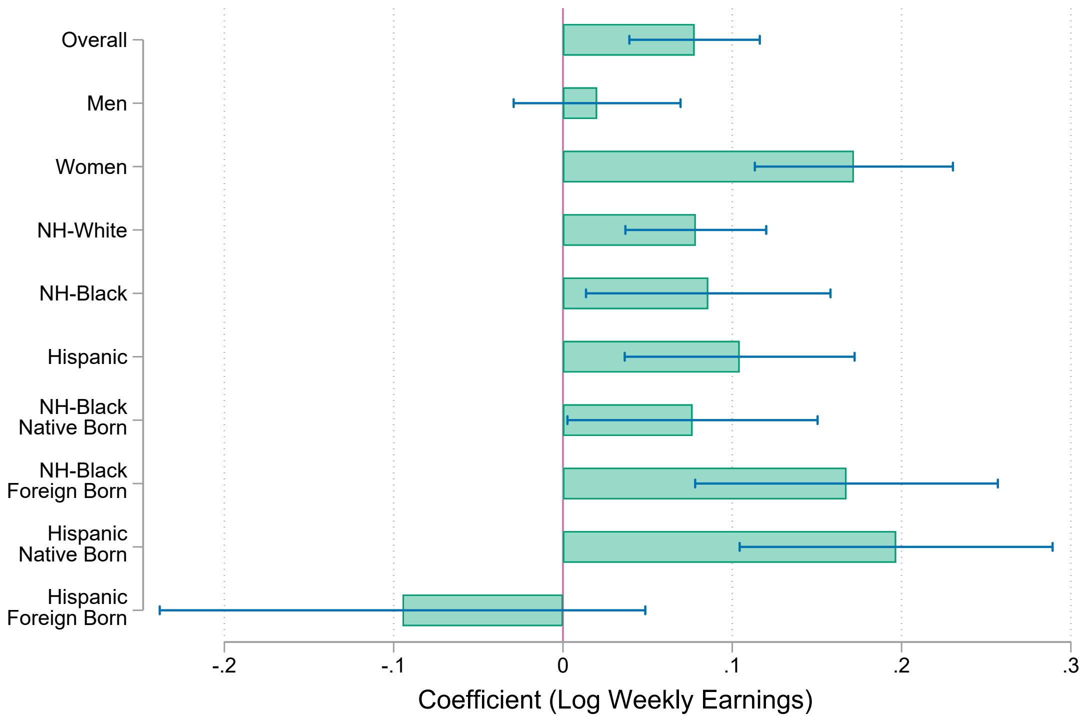
Figure 5: Coefficients of Log Weekly Earnings by Number of Clusters



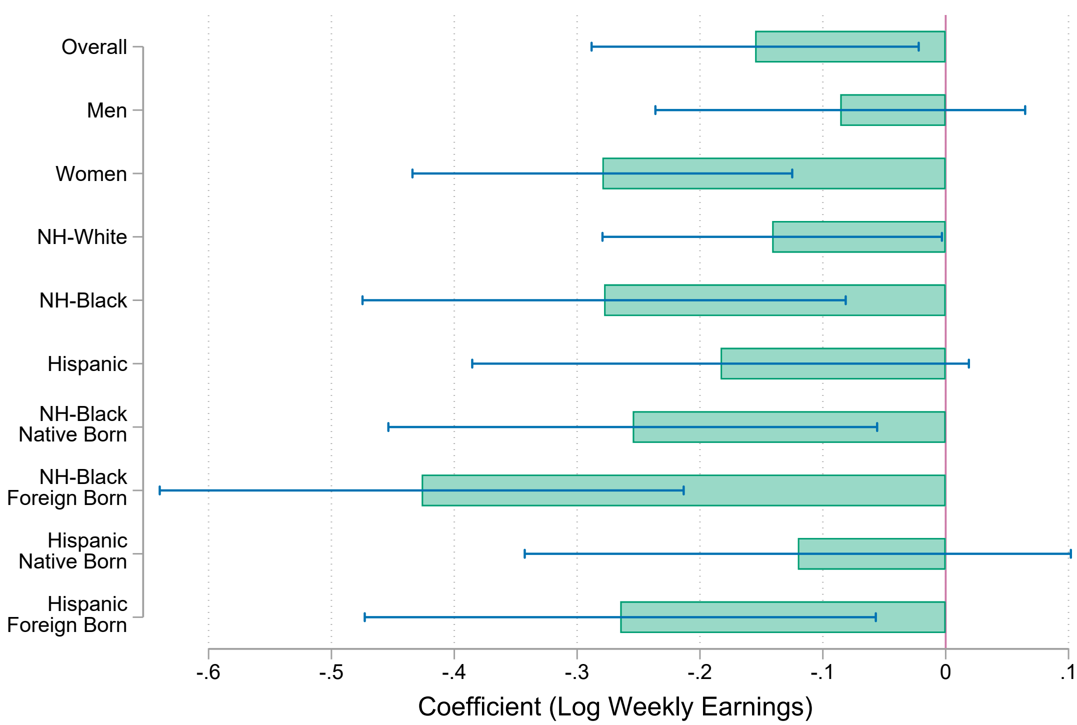
Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Coefficients are generated from the border match design detailed in Equation 3. Standard errors are clustered at the occupation level. 95% confidence intervals in red. For ease of visualization, spillover coefficients are based on 100% licensure of an occupation's cluster outside their own occupation. Models include occupation, state, and border pair fixed effects and controls for race/ethnicity, sex, age, and age squared.

Figure 6: Coefficients of Log Weekly Earnings by Subgroup at 20 Clusters

Panel A: Own Earnings Effects



Panel B: Within-Cluster Spillover Effects

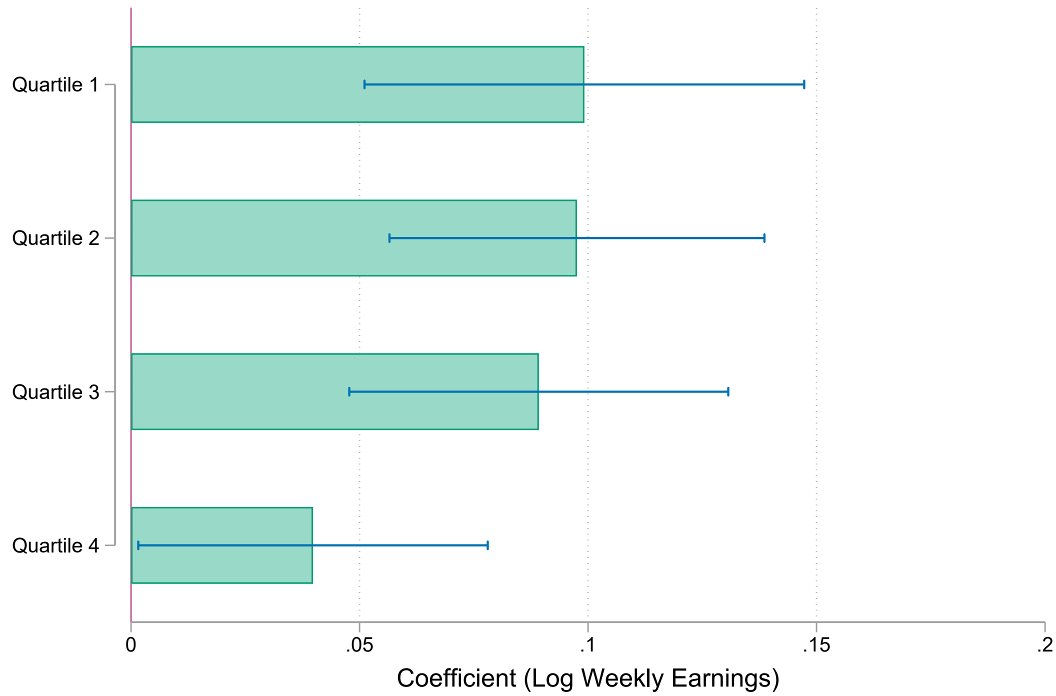


Source: Author's calculations of ACS, O*NET, and CPS licensing data.

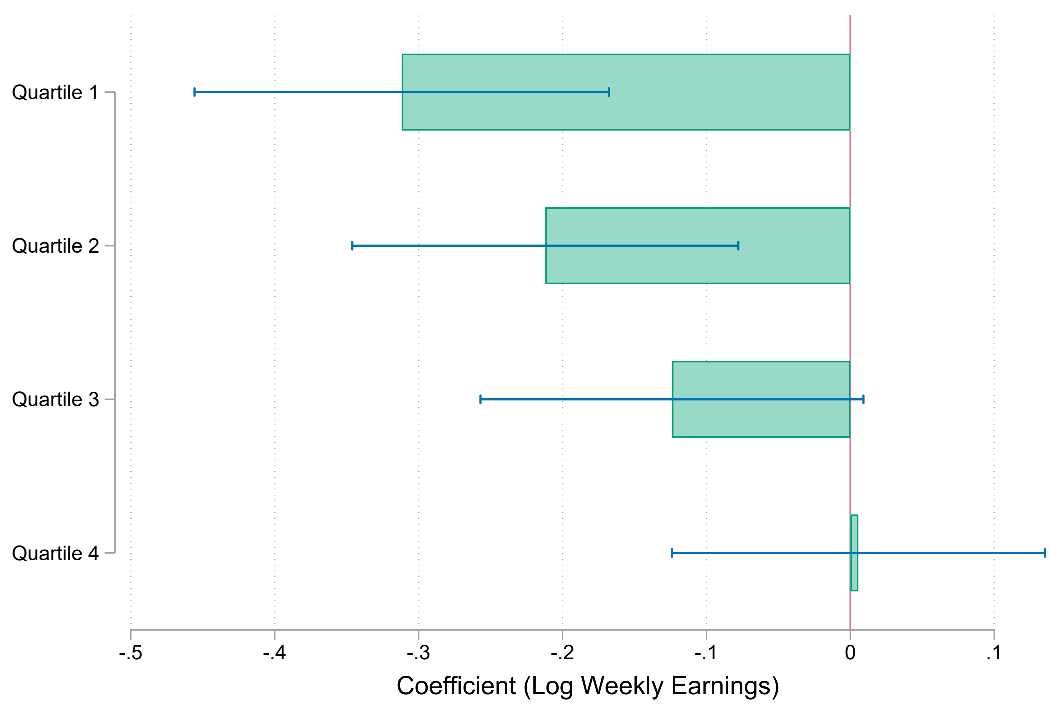
Notes: Coefficients are generated from the border match design detailed in Equation 3 using 20 skill clusters. Bars represent point estimates, and the bands represent the 95% confidence interval clustered at the occupation level. For ease of visualization, spillover coefficients are based on 100% licensure of an occupation's cluster outside their own occupation. Models include occupation, state, and border pair fixed effects and controls for race/ethnicity, sex, age, and age squared.

Figure 7: Coefficients of Log Weekly Earnings by Labor Market Size at 20 Clusters

Panel A: Own Earnings Effects

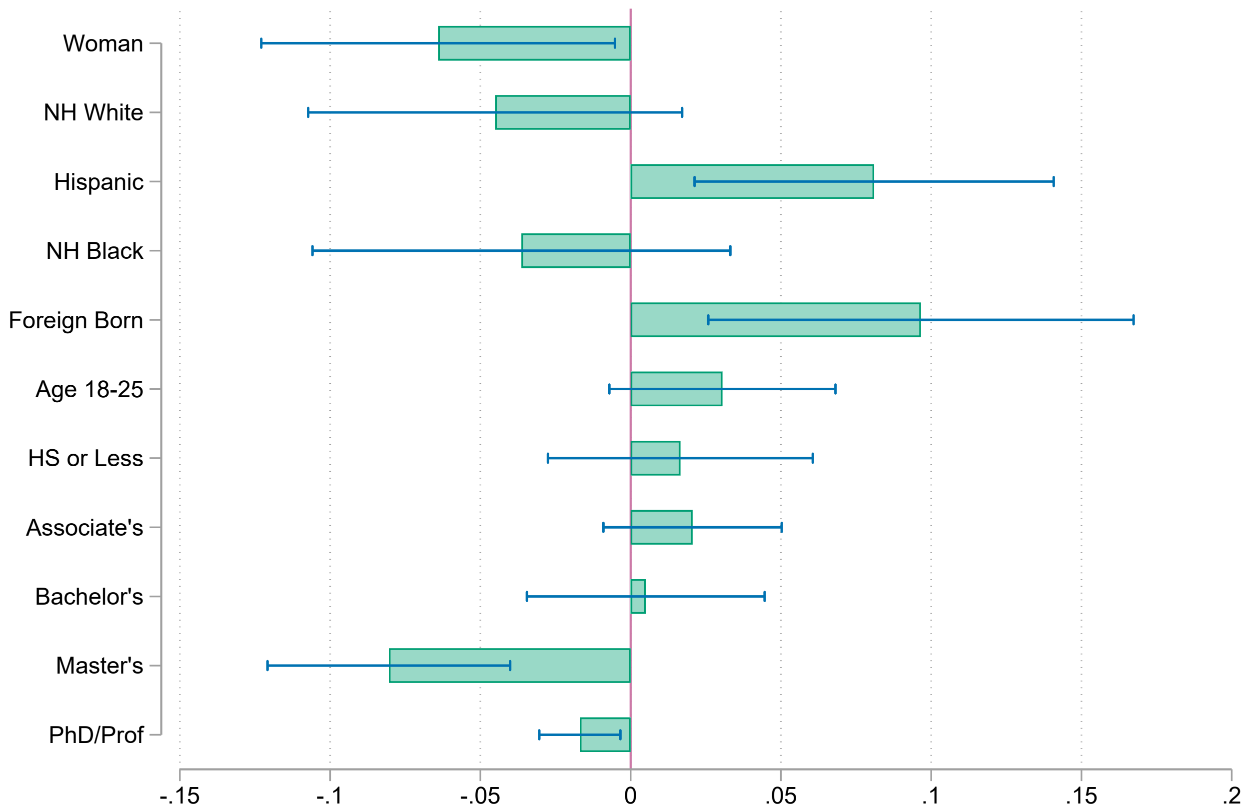


Panel B: Within-Cluster Spillover Effects



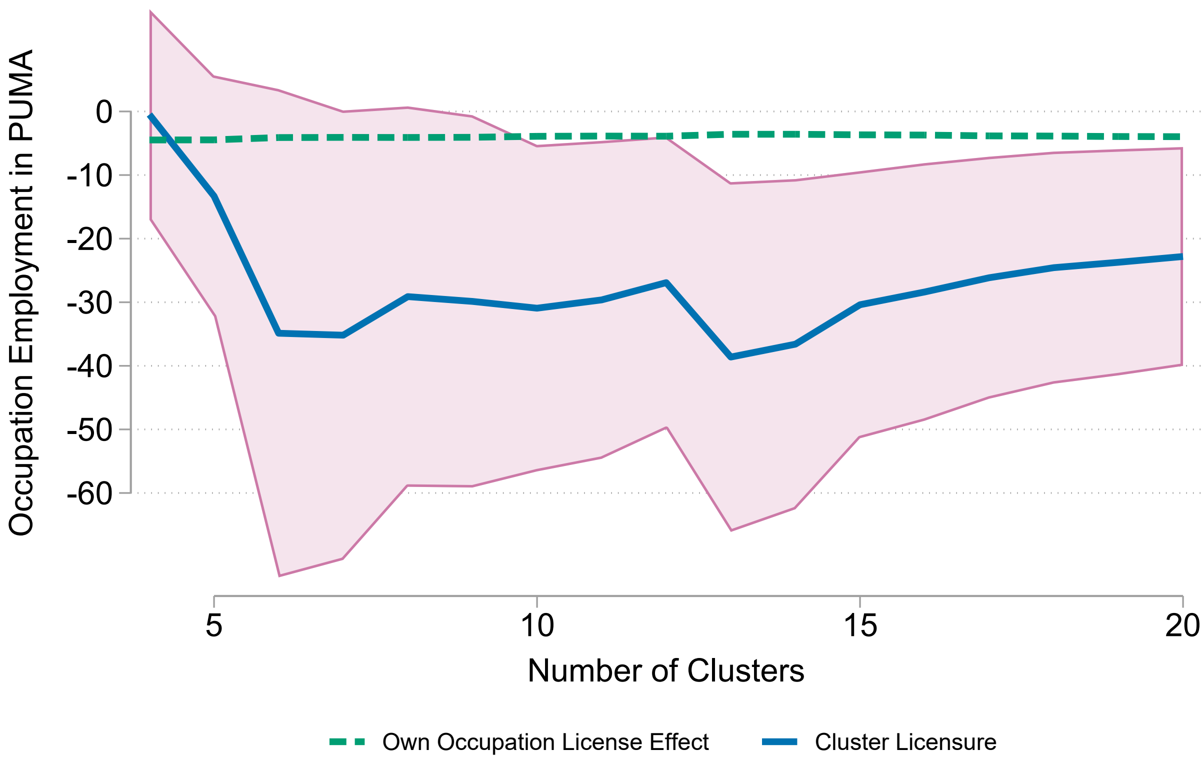
Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Coefficients are generated from the border match design detailed in Equation 3 using 20 skill clusters. Bars represent point estimates, and the bands represent the 95% confidence interval clustered at the occupation level. For ease of visualization, spillover coefficients are based on 100% licensure of an occupation's cluster outside their own occupation. Quartiles are defined by the size of the working population age 18–64 within each PUMA. Models include occupation, state, and border pair fixed effects and controls for race/ethnicity, sex, age, and age squared.

Figure 8: Composition Effects of Licensing Spillovers, 20 Clusters



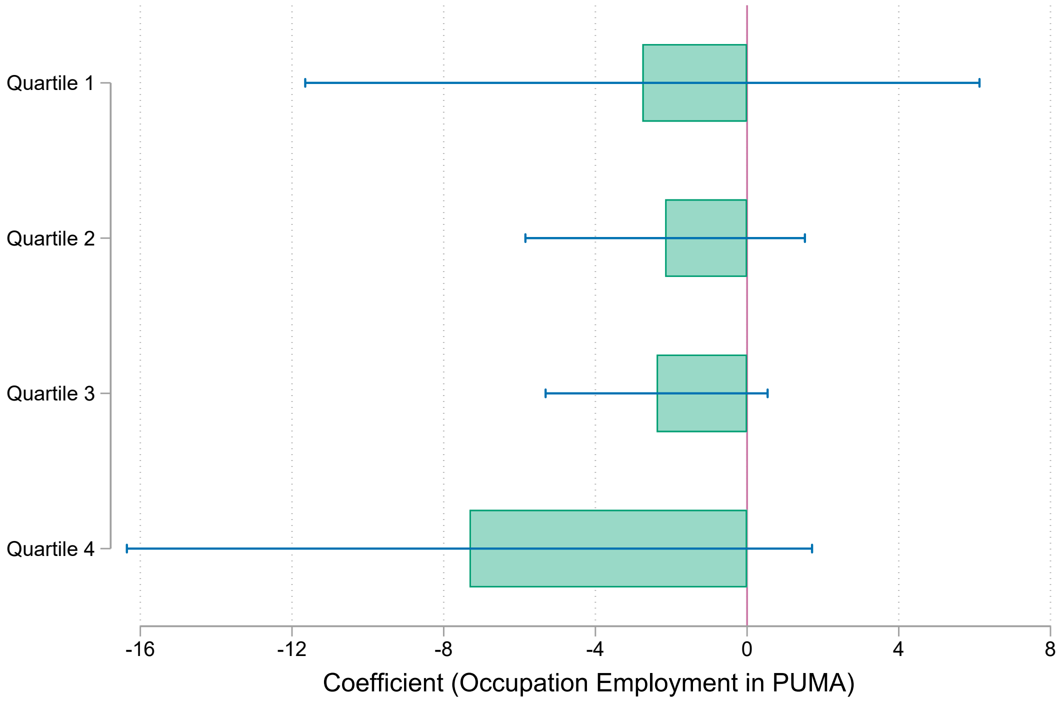
Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Coefficients are generated from the border match design detailed in Equation 3 for linear probability models on binary outcomes. Bars represent point estimates, and the bands represent the 95% confidence interval clustered at the occupation level. For ease of visualization, spillover coefficients are based on 100% licensure of an occupation's cluster outside their own occupation for 20 total defined clusters. Models include occupation, state, and border pair fixed effects.

Figure 9: Employment Effects of Licensing Spillovers by Number of Clusters

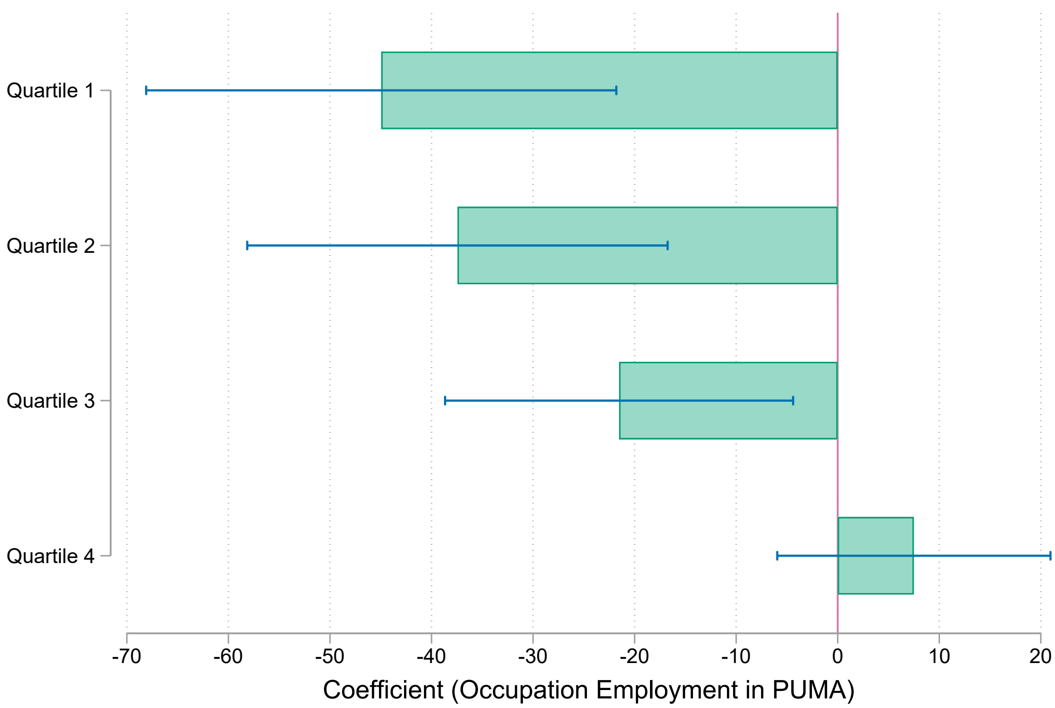


Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Coefficients are generated from the border match design detailed in Equation 3 collapsed into PUMA-occupation cells. Standard errors are clustered at the occupation level. 95% confidence intervals in red. For ease of visualization, spillover coefficients are based on 100% licensure of an occupation's cluster outside their own occupation. Models include occupation, state, and border pair fixed effects.

Figure 10: Coefficients of Employment by Labor Market Size at 20 Clusters
Panel A: Own Employment Effects

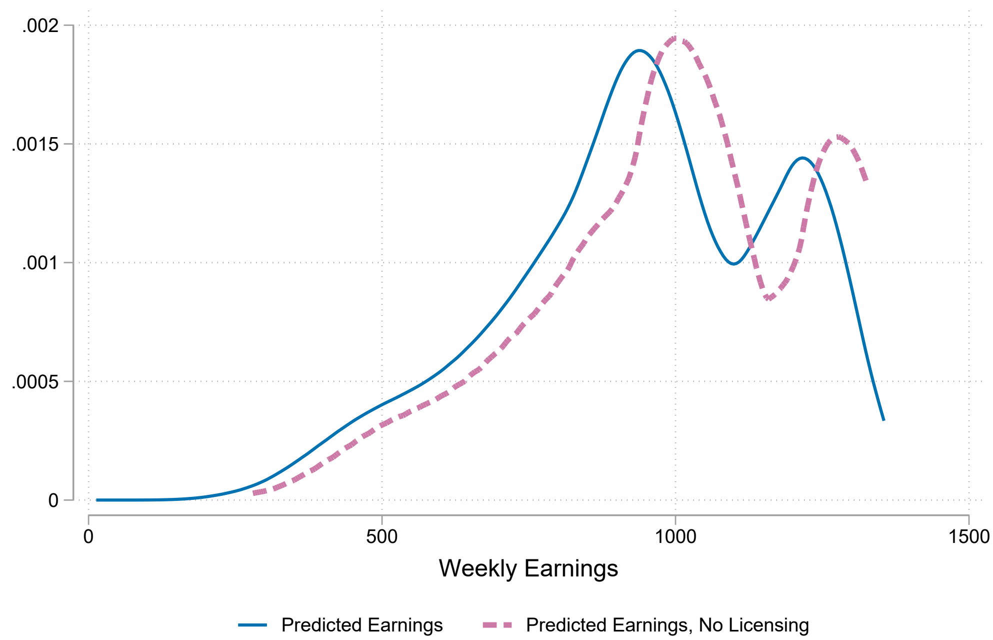


Panel B: Within-Cluster Spillover Effects



Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Coefficients are generated from the border match design detailed in Equation 3 using 20 skill clusters. The data are collapsed into PUMA-occupation cells. Bars represent point estimates, and the bands represent the 95% confidence interval clustered at the occupation level. For ease of visualization, spillover coefficients are based on 100% licensure of an occupation's cluster outside their own occupation. Quartiles are defined by the size of the working population age 18–64 within each PUMA. Models include occupation, state, and border pair fixed effects.

Figure 11: Kernel Density of the Counterfactual Distribution of Weekly Earnings
Panel A: Kernel Density of Predicted Earnings

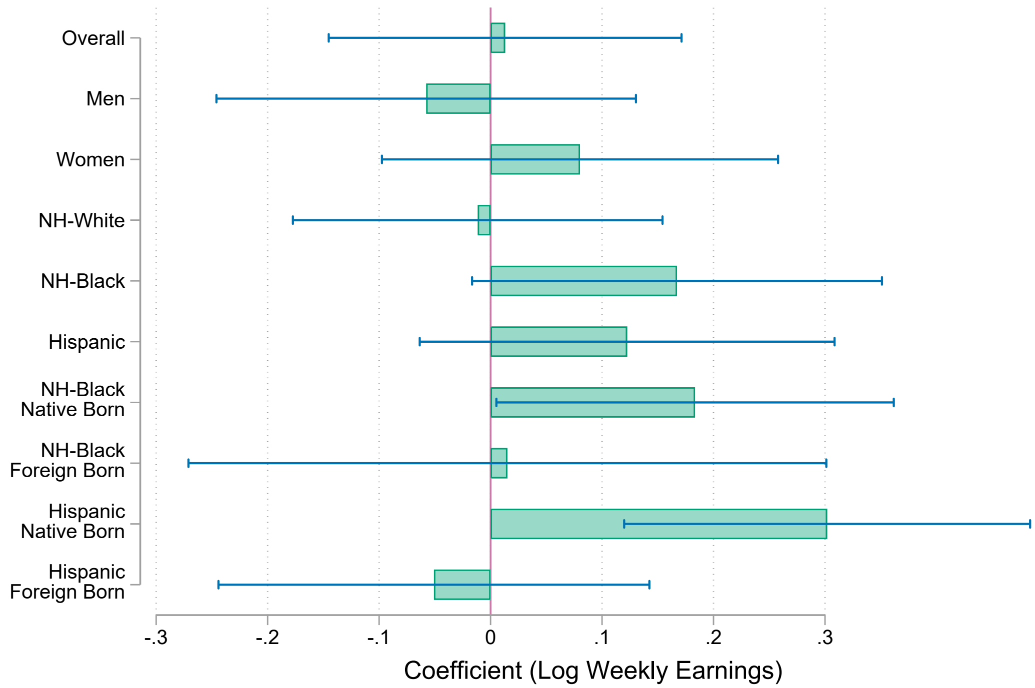


Panel B: Difference in Densities

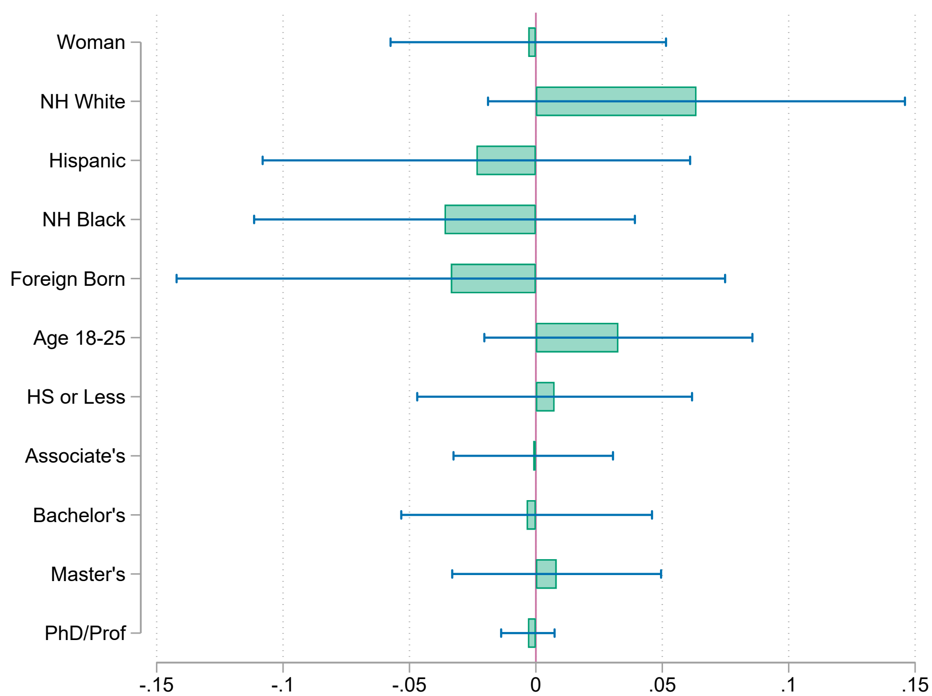
Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Coefficients are generated from the border match design detailed in Equation 3 using 20 skill clusters. The model includes fixed effects for occupation, border pair, and state. Predicted earnings are for the status quo and for setting licensing rates to zero for one's own occupation and skill cluster.

Figure 12: Spillover Coefficients of Log Weekly Earnings, 20 Placebo Clusters

Panel A: Earnings Effects

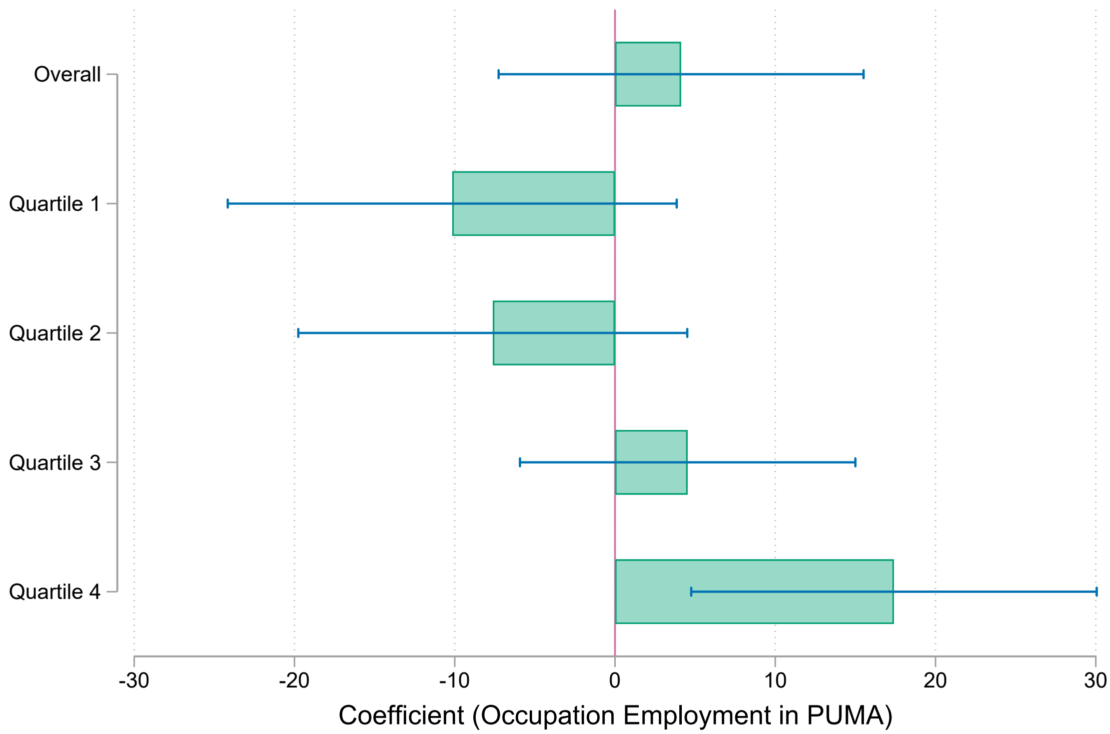


Panel B: Composition Effects



Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Coefficients are generated from β_2 in the border match design detailed in Equation 3 using 20 placebo skill clusters. Bars represent point estimates, and the bands represent the 95% confidence interval clustered at the occupation level. For ease of visualization, spillover coefficients are based on 100% licensure of an occupation's cluster outside their own occupation. Models include occupation, state, and border pair fixed effects and controls for race/ethnicity, sex, age, and age squared.

Figure 13: Coefficients of Employment by Labor Market Size at 20 Placebo Clusters



Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Coefficients are generated from β_2 in the border match design detailed in Equation 3 using 20 placebo skill clusters. Bars represent point estimates, and the bands represent the 95% confidence interval clustered at the occupation level. For ease of visualization, spillover coefficients are based on 100% licensure of an occupation's cluster outside their own occupation. Quartiles are defined by the size of the working population age 18–64 within each PUMA. Models include occupation, state, and border pair fixed effects.

Tables

Table 1: Summary Statistics by Sample

	Border Sample		Full Sample	
	Mean	SD	Mean	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
N	1,337,103		4,578,382	
PUMAs	244		982	
Occupations	410		410	
Border Pairs	110		N/A	

Source: Author's calculations of ACS, CPS, and O*NET data.
Notes: Clusters are based on description in Section 4.1. ACS samples are from 2014-2017 corresponding with CPS individual licensing data from 2015-2018.

Table 2: Components of Latent Skill Measurements

Occupational Skill Area	O*NET Variables
Non-Routine, Cognitive, Analytical	“Analyzing data/information” “Thinking creatively” “Interpreting information for others”
Non-Routine, Cognitive, Interpersonal	“Establishing and maintaining personal relationships” “Guiding, directing and motivating subordinates” “Coaching/developing others”
Non-Routine, Manual, Physical Adaptability	“Operating vehicles, mechanized devices, or equipment” “Spend time using hands to handle, control or feel objects, tools or controls” “Manual dexterity” “Spatial orientation”
Routine, Cognitive	“Importance of repeating the same tasks” “Importance of being exact or accurate” “Structured v. Unstructured work (reverse)”
Routine, Manual	“Pace determined by speed of equipment” “Controlling machines and processes” “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 3: Distributional Statistics of Predicted Weekly Earnings with vs without Licensing within Occupational Groups

	Status Quo	Without Licensing	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%

Source: Author's calculations of ACS, CPS, and O*NET data.
Notes: Distributional statistics are based on the predictions from Equation 3 as described in Section 5.4.

Table 4: CPS Repeated Cross-Section Estimates from the NLD by Sample Start Year

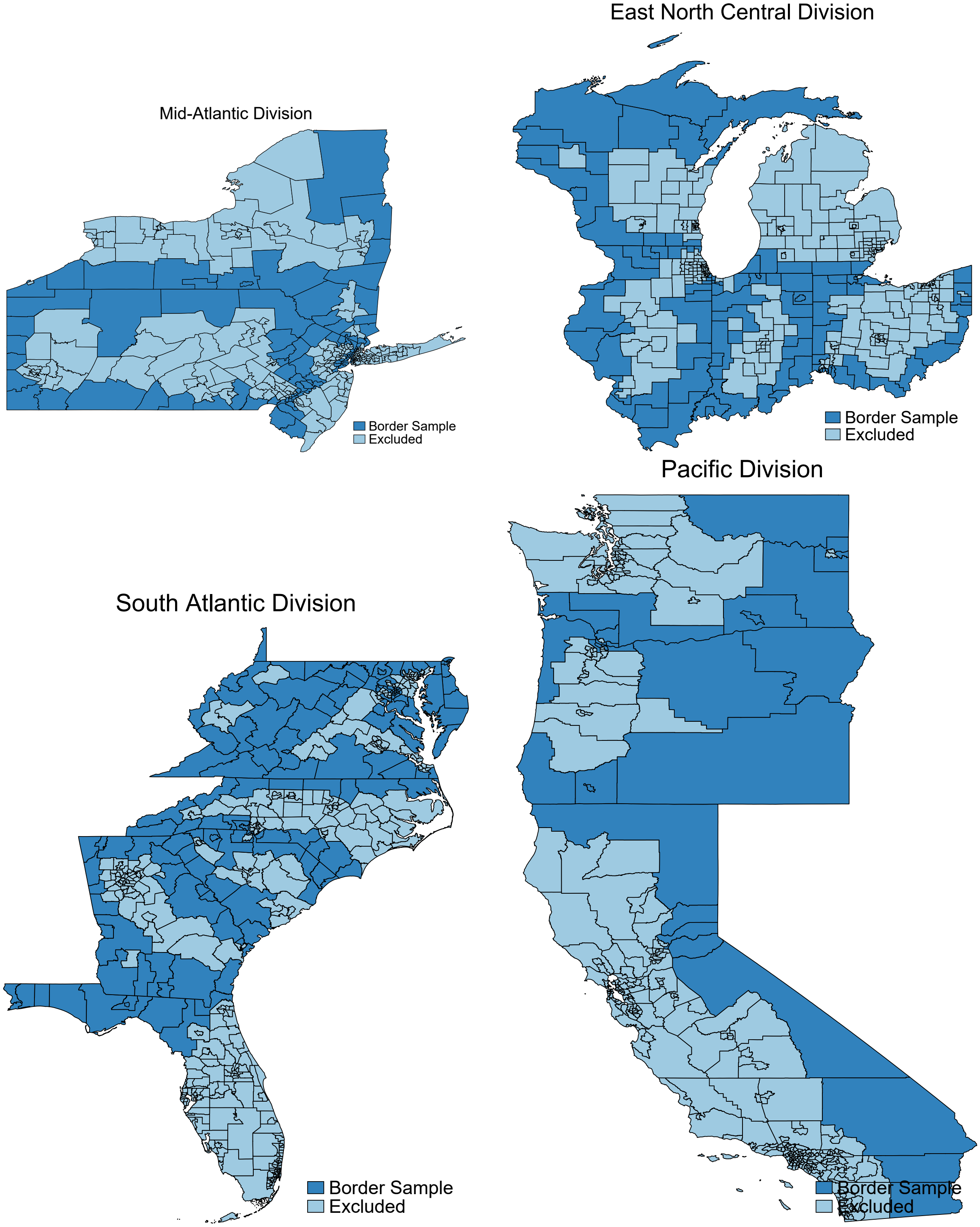
VARIABLES	(1) 1983	(2) 1994	(3) 2001	(4) 2010
Panel A: Log Wage Effects				
Avg Cluster Spillover Effect (100%)	-0.0398**	-0.0452***	-0.0699***	-0.0715***
	(0.0167)	(0.0131)	(0.0135)	(0.0159)
Spillover Effect: NH White	-0.0338*	-0.0361**	-0.0640***	-0.0661***
	(0.0190)	(0.0149)	(0.0168)	(0.0193)
Spillover Effect: NH Black	-0.106***	-0.128***	-0.146***	-0.139***
	(0.0340)	(0.0360)	(0.0342)	(0.0349)
Spillover Effect: Hispanic	-0.0655**	-0.0761**	-0.0944***	-0.105***
	(0.0259)	(0.0301)	(0.0301)	(0.0325)
Spillover Effect: Women	-0.0265	-0.0498	-0.0770**	-0.0825**
	(0.0387)	(0.0340)	(0.0310)	(0.0327)
Spillover Effect: Men	-0.0447**	-0.0427**	-0.0665***	-0.0665***
	(0.0198)	(0.0178)	(0.0161)	(0.0169)
Observations	5,364,304	3,694,393	2,664,510	1,208,935
Panel B: Employment Effects				
Avg Cluster Spillover Effect (100%)	-146.0	-134.0	-245.5**	-335.4***
	(92.46)	(103.9)	(112.2)	(129.2)
Observations	500,523	361,388	269,778	127,447
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 100% 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.

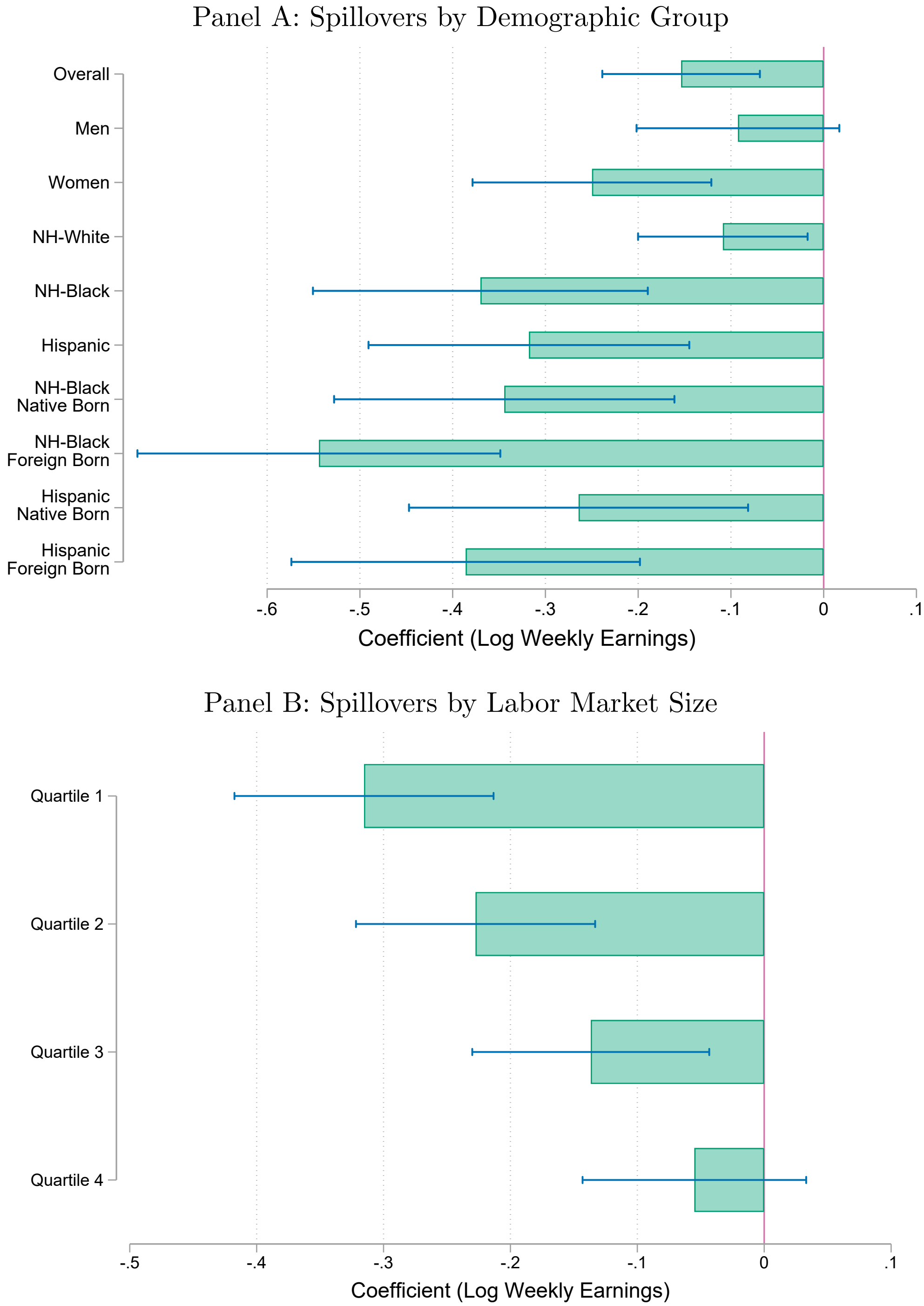
A Figures and Tables Appendix

Figure A1: Border Sample PUMAs in Select Census Divisions



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

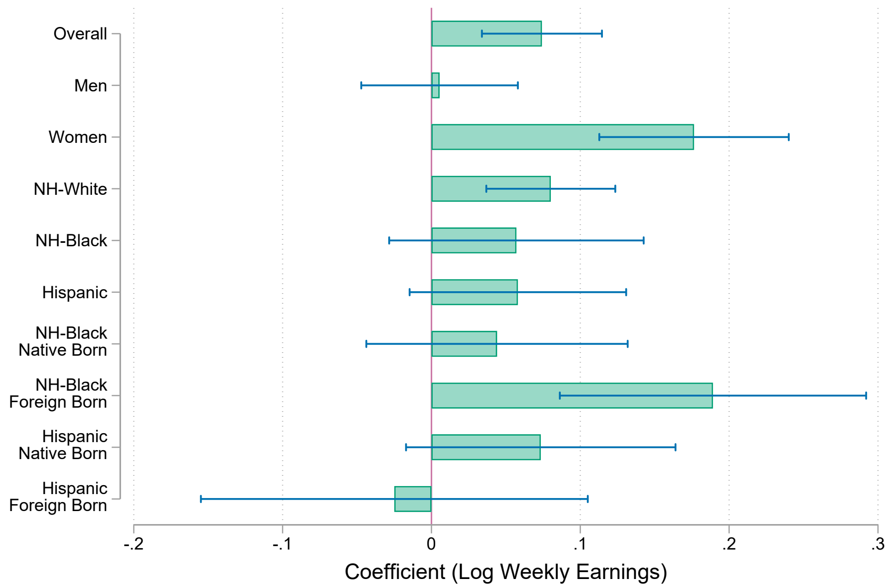
Figure A2: Coefficients of Spillover Effects by Subgroup at 20 Clusters, Occupation by Border Fixed Effects



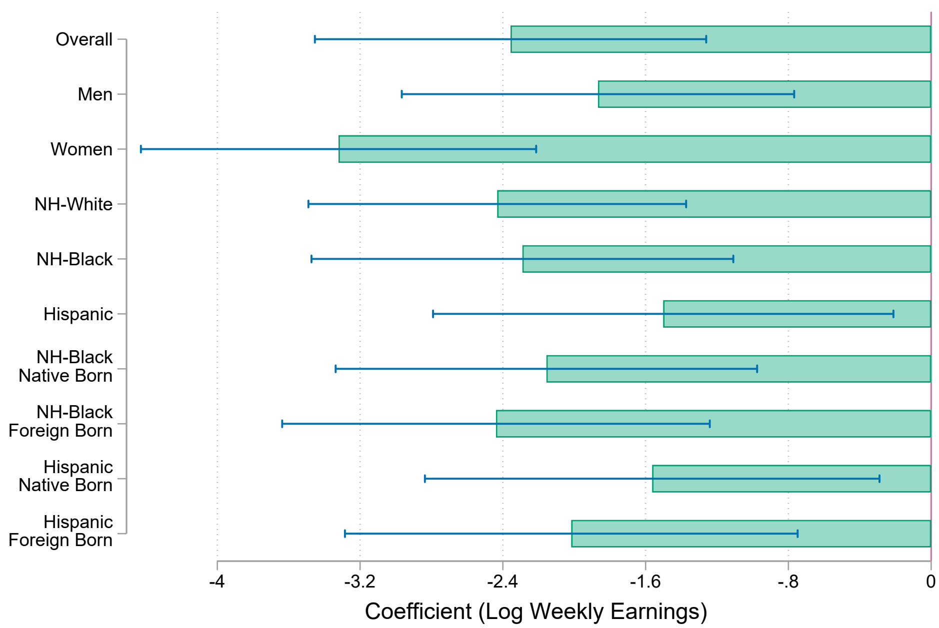
Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Coefficients are generated from the border match design with occupation-by-border-pair interacted fixed effects using 20 skill clusters. Bars represent point estimates, and the bands represent the 95% confidence interval clustered at the occupation level. For ease of visualization, spillover coefficients are based on 100% 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: Coefficients of Log Weekly Earnings by Skill-Distance Weighted Exposure

Panel A: Own Earnings Effects

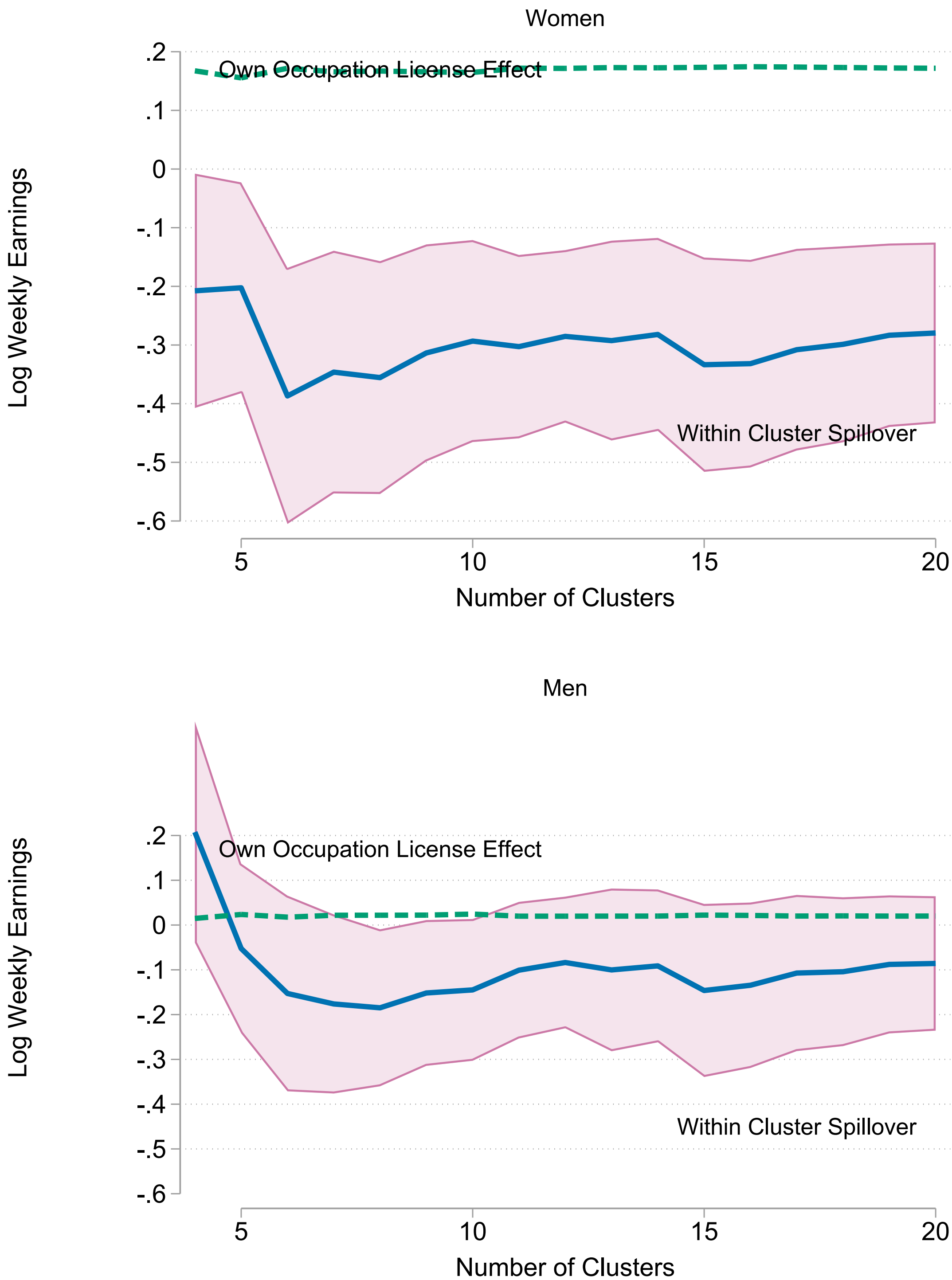


Panel B: Skill-Distance Weighted Spillover Effects



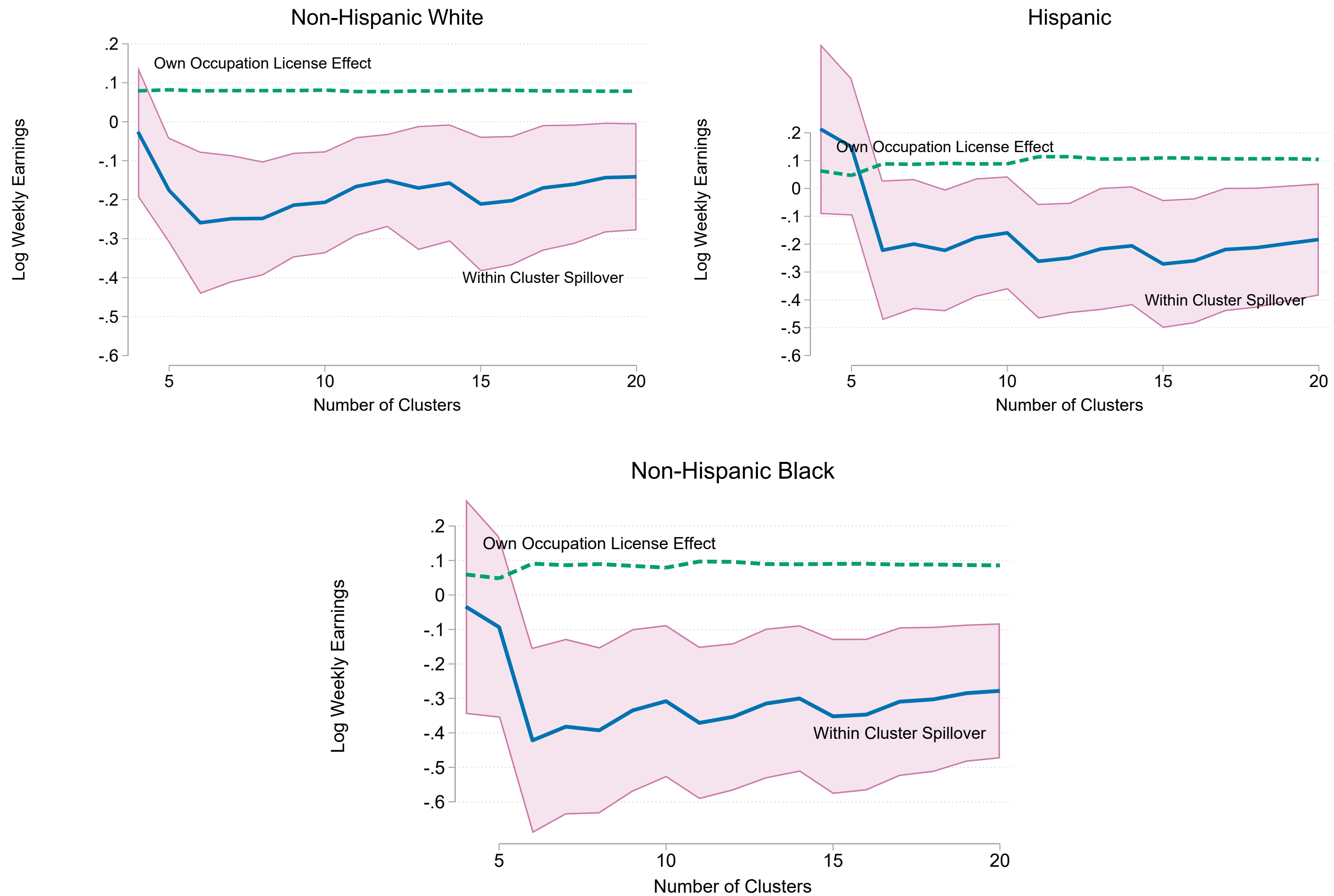
Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Coefficients are generated from the border match design detailed in Equation 3, except replacing cluster licensing exposure with skill-weighted exposure. Exposure to licensure is defined as licensing rates in every occupation in each state weighted by the skill similarity of each occupation to each other. Bars represent point estimates, and the bands represent the 95% confidence interval clustered at the occupation level. Spillover coefficients are based on 100% licensing exposure for all occupations in the state. Models include occupation, state, and border pair fixed effects and controls for race/ethnicity, sex, age, and age squared.
Scaling by a standard deviation (0.0228) implies effects larger than the size of my estimates based on skill clusters for a standard deviation change in exposure (-5.4% vs -1.6%). However, interpretation in this model is that *all* occupations in a state have become fully licensed rather than just the skill cluster.

Figure A4: Coefficients of Log Weekly Earnings by Number of Clusters, by Gender



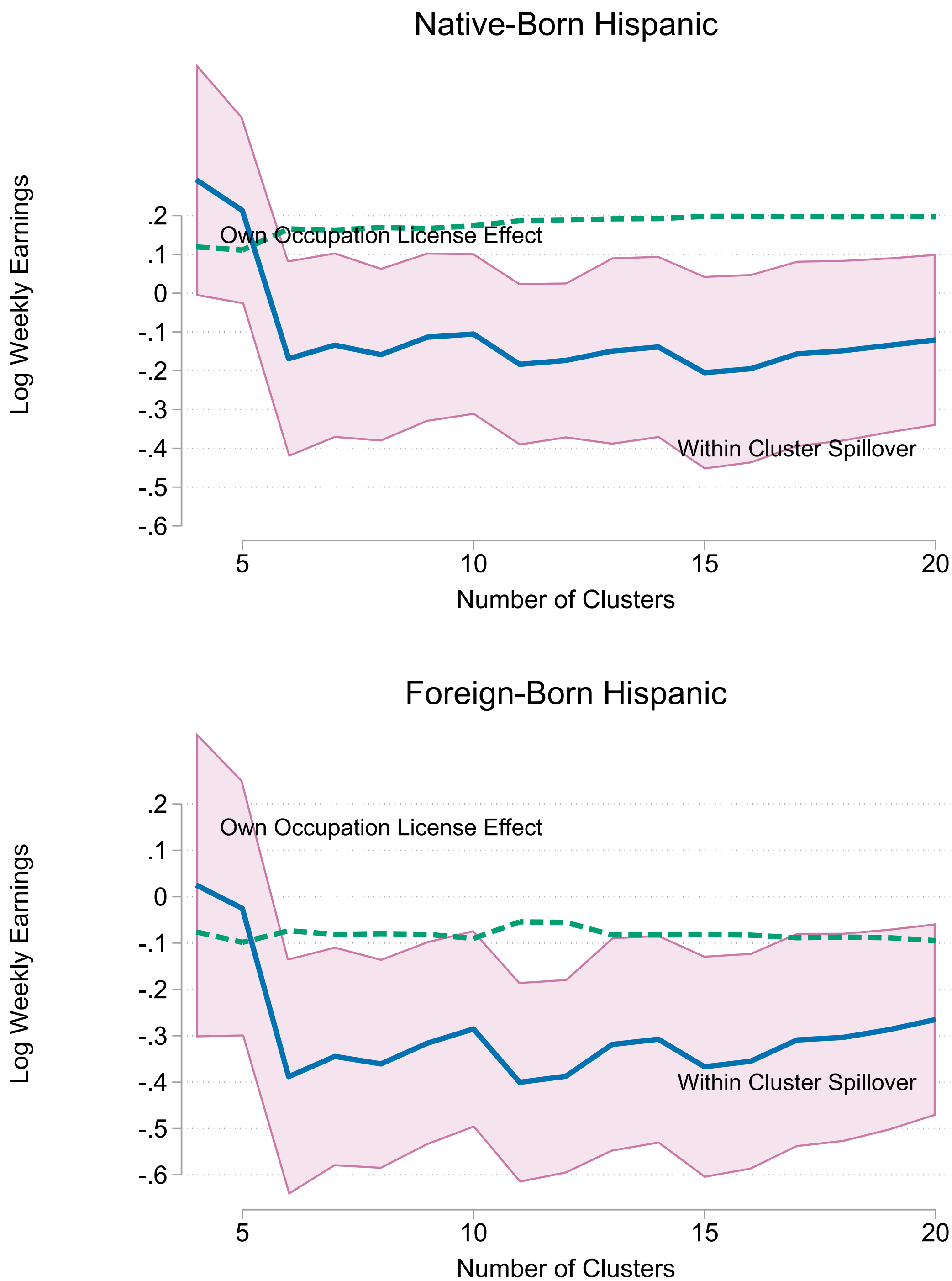
Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Coefficients are generated from the border match design detailed in Equation 3. Standard errors are clustered at the occupation level. 95% confidence intervals in red. For ease of visualization, spillover coefficients are based on 100% licensure of an occupation's cluster outside their own occupation.

Figure A5: Coefficients of Log Weekly Earnings by Number of Clusters, by Race/Ethnicity



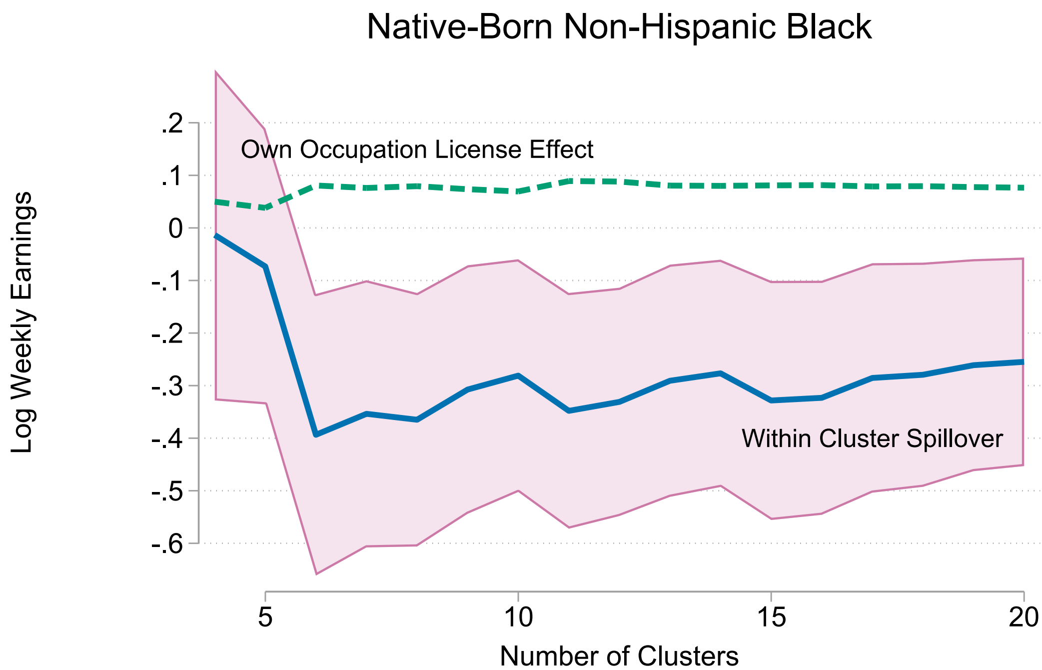
Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Coefficients are generated from the border match design detailed in Equation 3. Standard errors are clustered at the occupation level. 95% confidence intervals in red. For ease of visualization, spillover coefficients are based on 100% licensure of an occupation's cluster outside their own occupation.

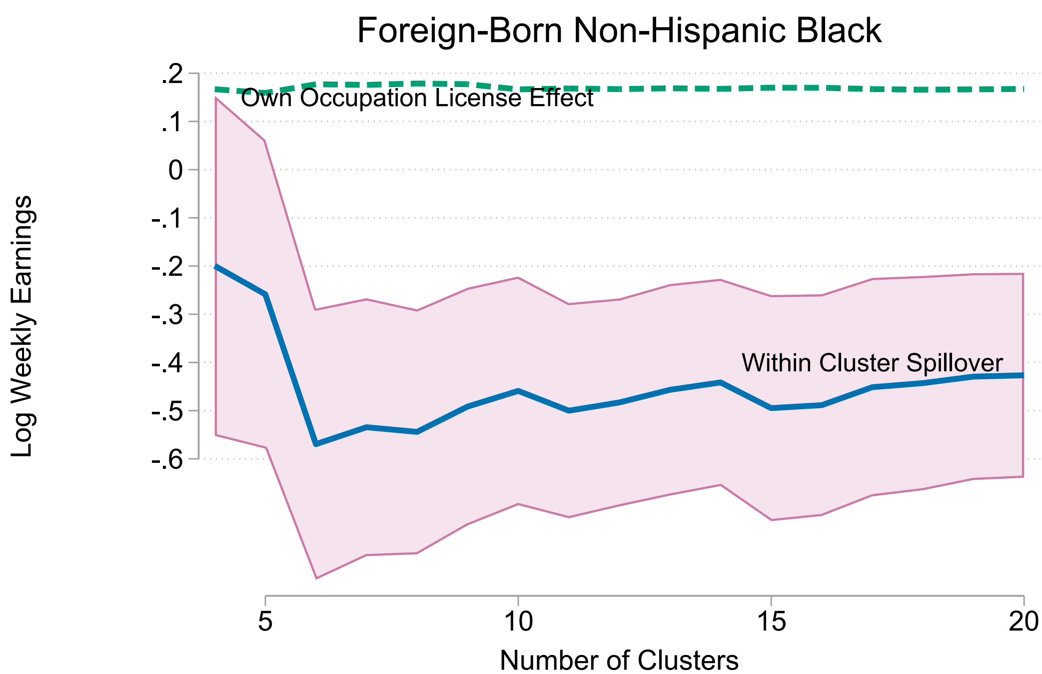
Figure A6: Coefficients of Log Weekly Earnings by Number of Clusters,
Hispanic Workers by Nativity



Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Coefficients are generated from the border match design detailed in Equation 3. Standard errors are clustered at the occupation level. 95% confidence intervals in red. For ease of visualization, spillover coefficients are based on 100% licensure of an occupation's cluster outside their own occupation.

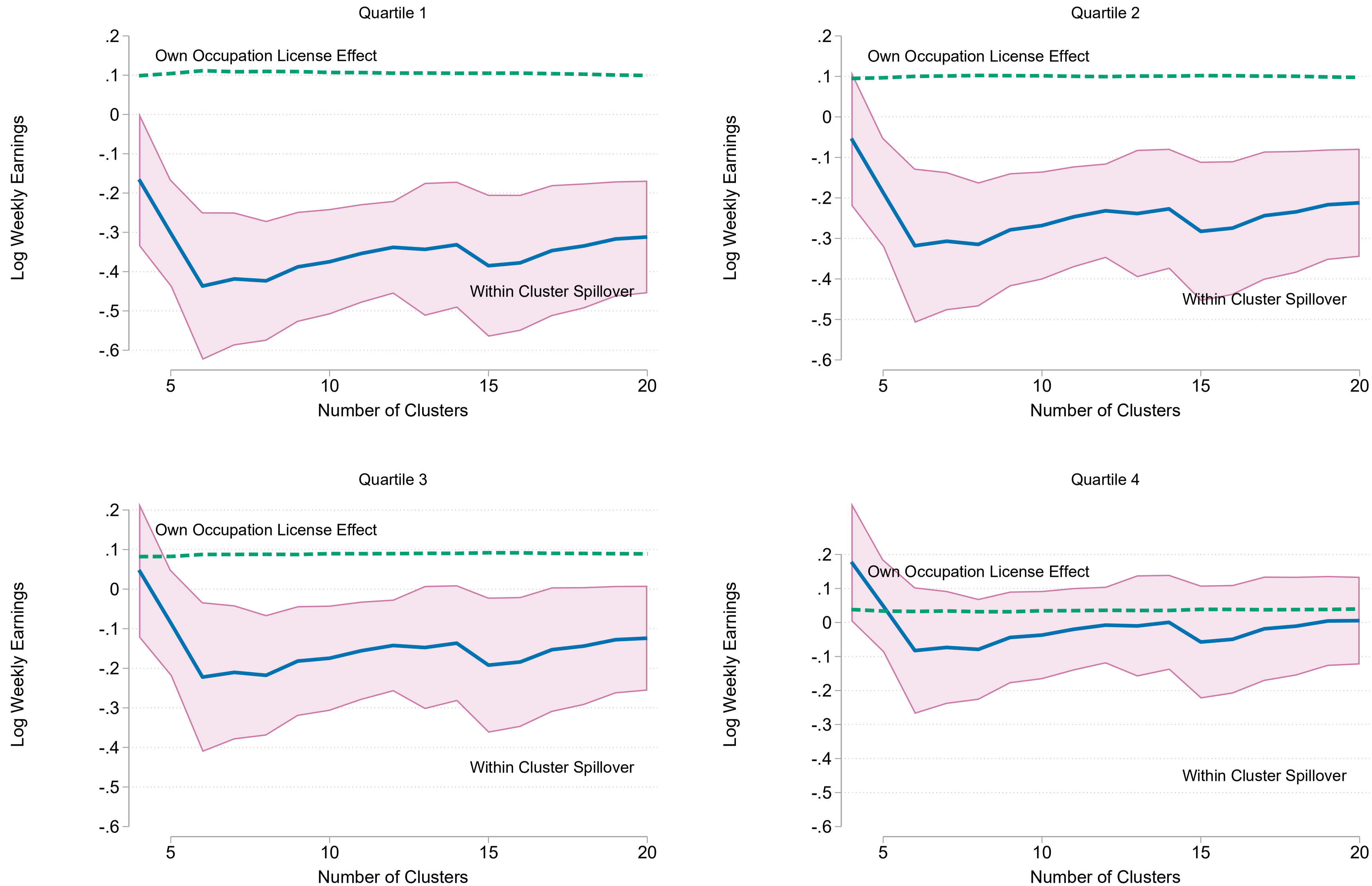
Figure A7: Coefficients of Log Weekly Earnings by Number of Clusters,
Non-Hispanic Black Workers by Nativity





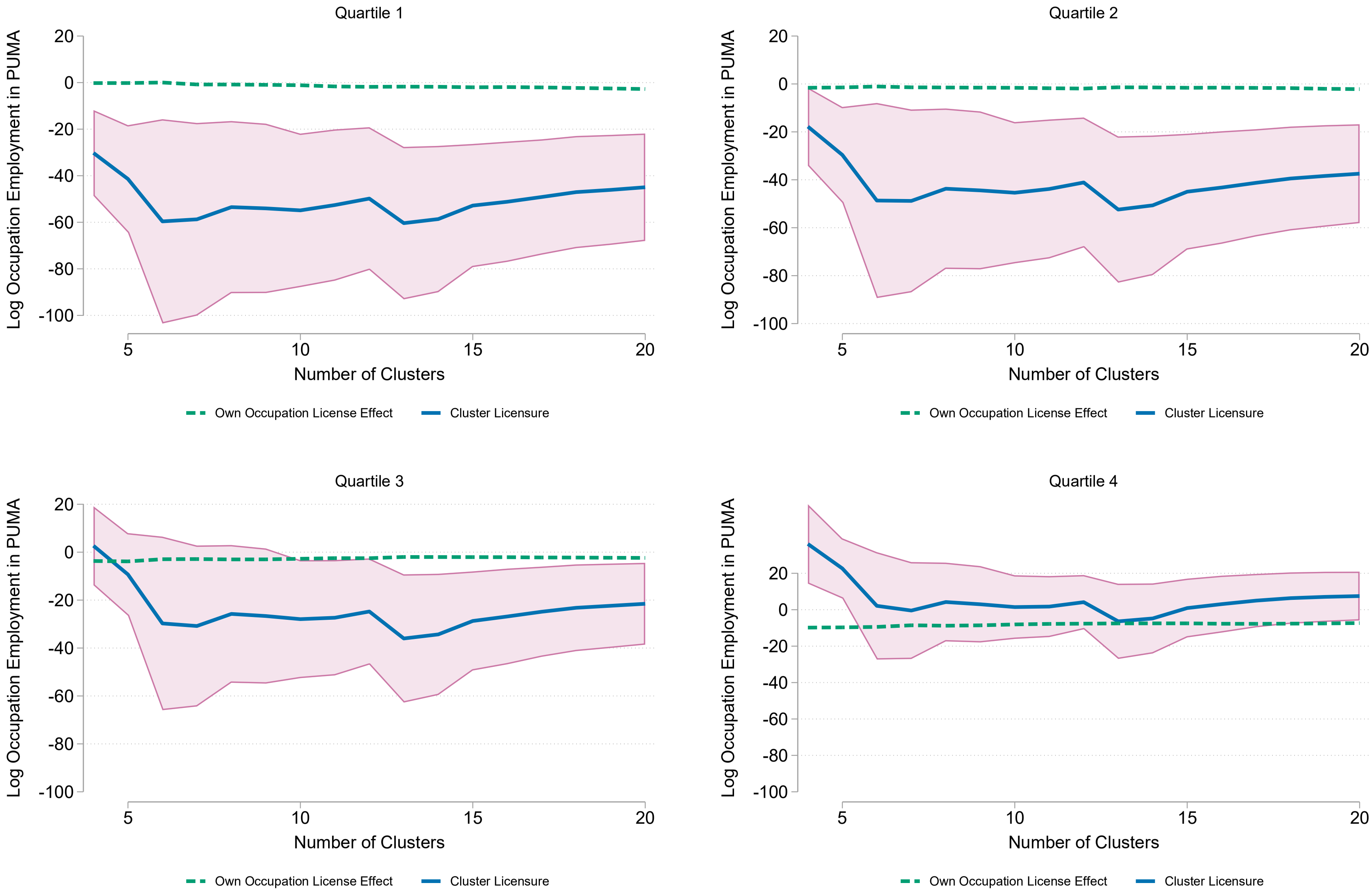
Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Coefficients are generated from the border match design detailed in Equation 3. Standard errors are clustered at the occupation level. 95% confidence intervals in red. For ease of visualization, spillover coefficients are based on 100% licensure of an occupation's cluster outside their own occupation.

Figure A8: Coefficients of Log Weekly Earnings by Number of Clusters,
By PUMA Size Quartile



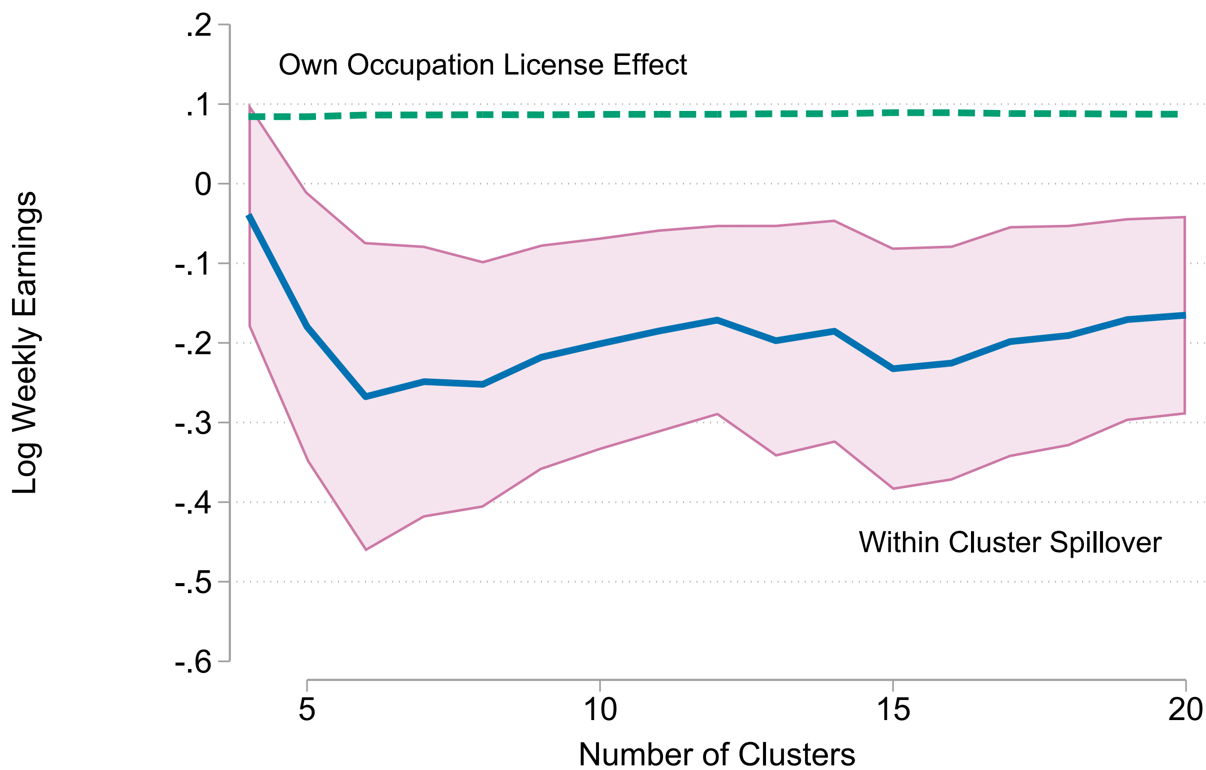
Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Coefficients are generated from the border match design detailed in Equation 3. Standard errors are clustered at the occupation level. 95% confidence intervals in red. For ease of visualization, spillover coefficients are based on 100% licensure of an occupation's cluster outside their own occupation.

Figure A9: Coefficients of Employment by Number of Clusters,
By PUMA Size Quartile



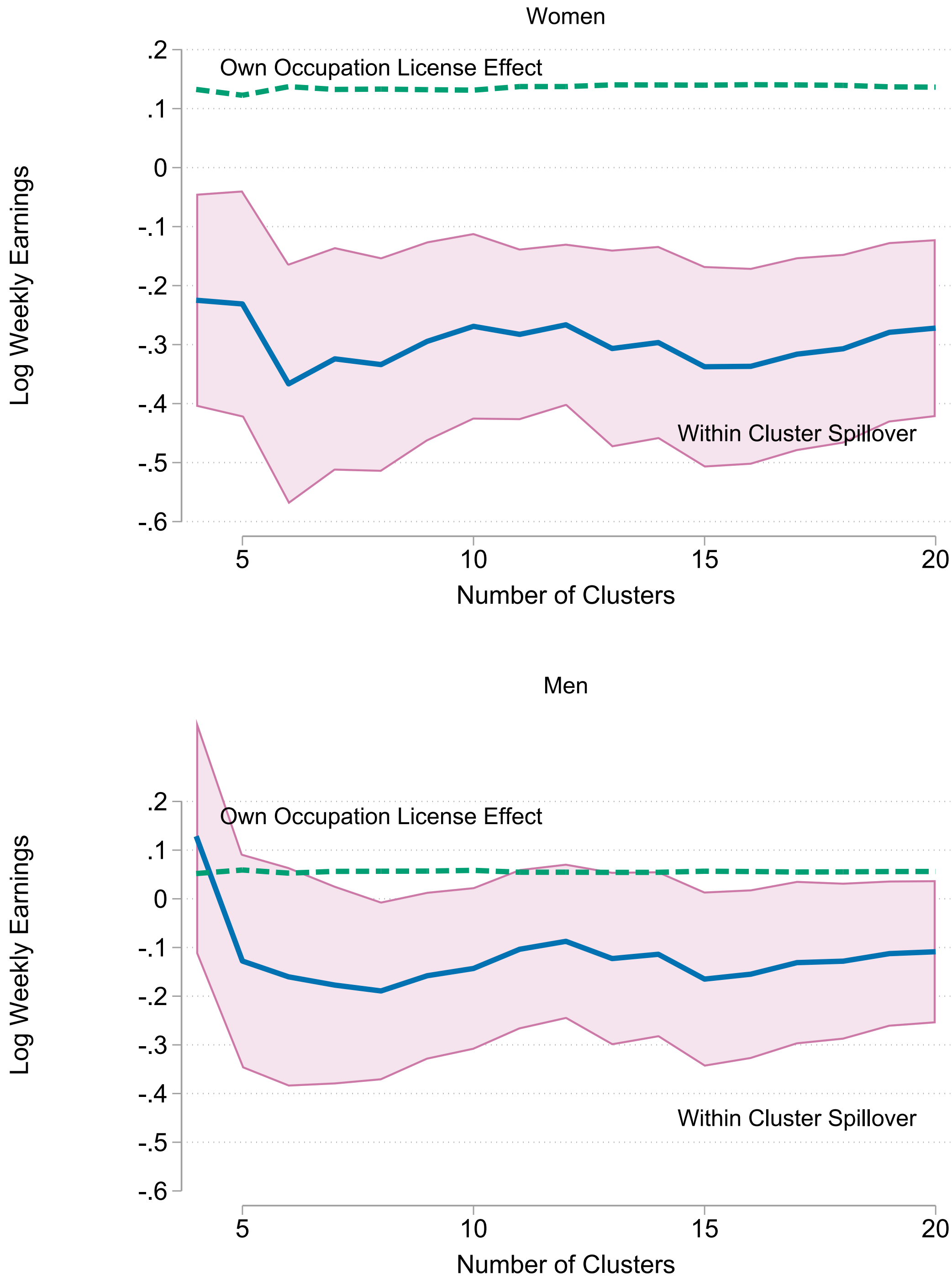
Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Coefficients are generated from the border match design detailed in Equation 3. Standard errors are clustered at the occupation level. 95% confidence intervals in red. For ease of visualization, spillover coefficients are based on 100% licensure of an occupation's cluster outside their own occupation.

Figure A10: Coefficients of Log Weekly Earnings by Number of Clusters
All Occupations, Adding PUMA FE



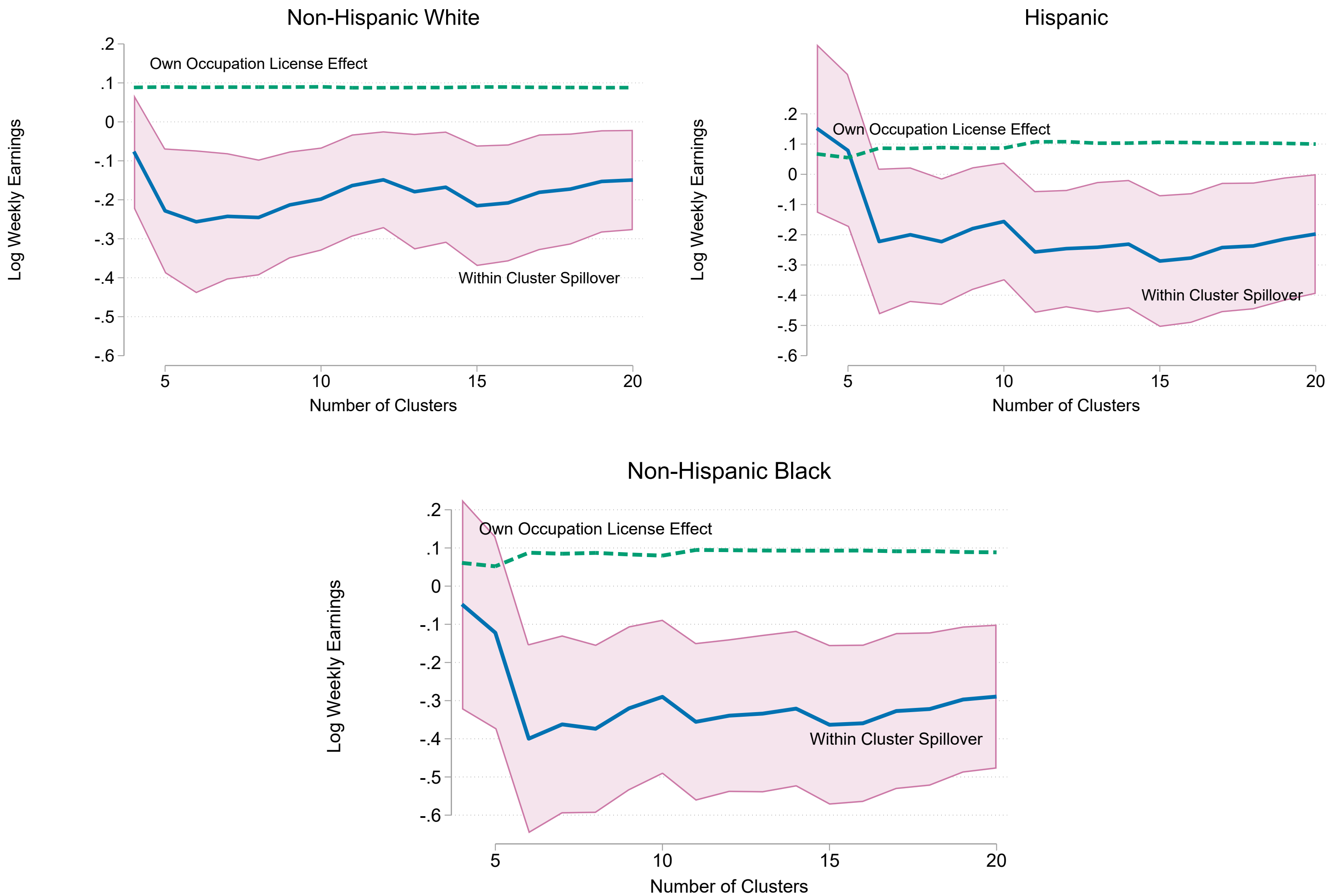
Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Coefficients are generated from the border match design detailed in Equation 3. Standard errors are clustered at the occupation level. 95% confidence intervals in red. For ease of visualization, spillover coefficients are based on 100% licensure of an occupation's cluster outside their own occupation.

Figure A11: Coefficients of Log Weekly Earnings by Number of Clusters, by Gender
All Occupations, Adding PUMA FE



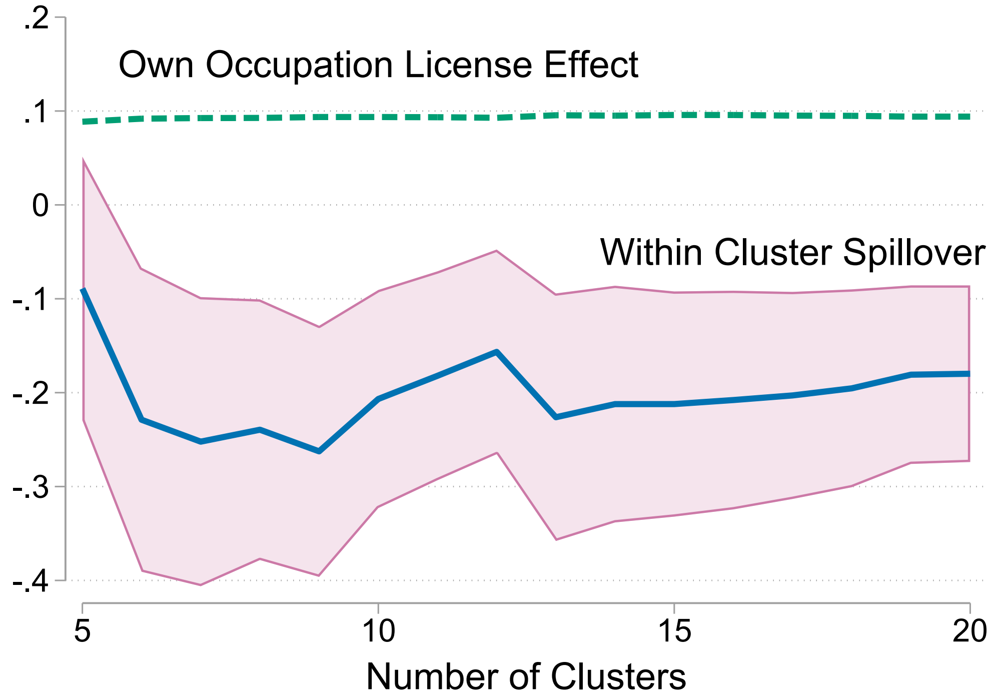
Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Coefficients are generated from the border match design detailed in Equation 3. Standard errors are clustered at the occupation level. 95% confidence intervals in red. For ease of visualization, spillover coefficients are based on 100% licensure of an occupation's cluster outside their own occupation.

Figure A12: Coefficients of Log Weekly Earnings by Number of Clusters, by Race/Ethnicity
All Occupations, Adding PUMA FE



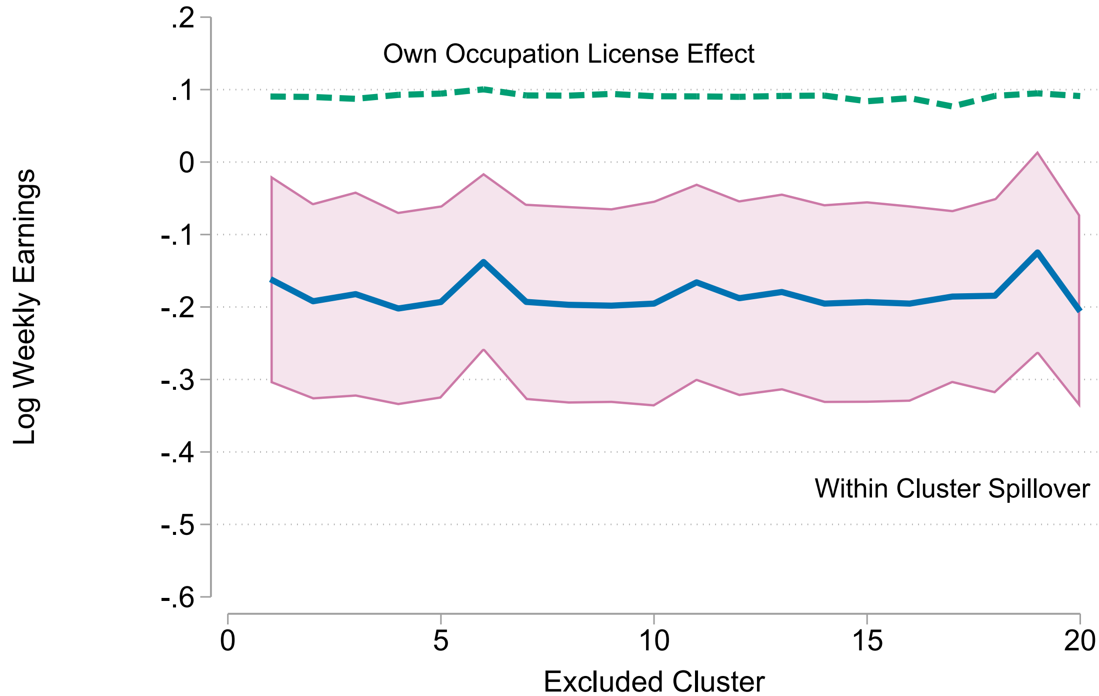
Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Coefficients are generated from the border match design detailed in Equation 3. Standard errors are clustered at the occupation level. 95% confidence intervals in red. For ease of visualization, spillover coefficients are based on 100% licensure of an occupation's cluster outside their own occupation.

Figure A13: Log Wage Effects with CPS 2015-2018
State and Occupation Fixed Effects



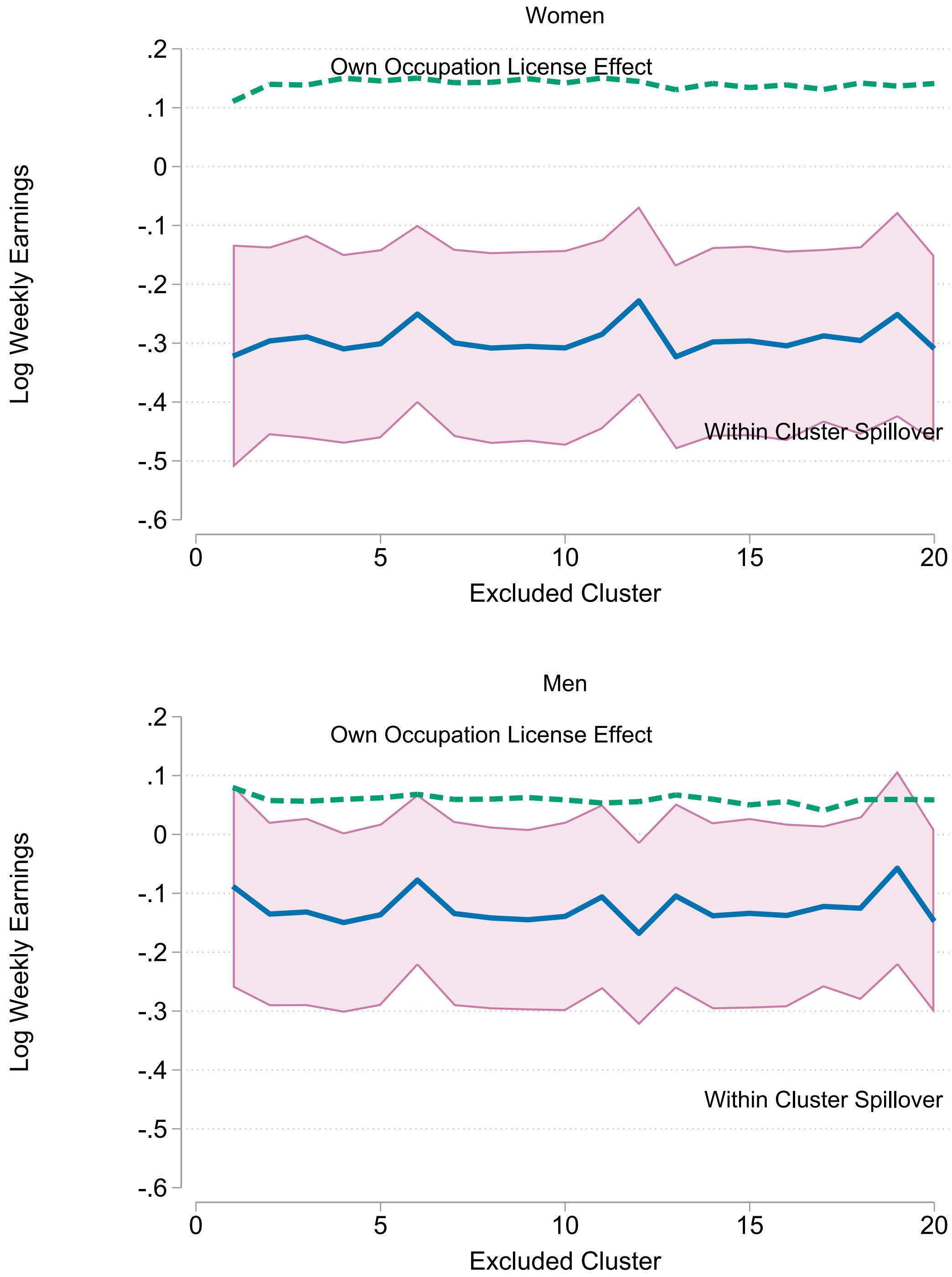
Source: Author's calculations of O*NET, and 2015-2018 CPS.
Notes: Coefficients are generated from estimates of log hourly wage in the CPS on individual sex, race/ethnicity, age, age squared, and state and occupation fixed effects. Standard errors are clustered at the occupation level. 95% confidence intervals in red. For ease of visualization, spillover coefficients are based on 100% licensure of an occupation's cluster outside their own occupation.

Figure A14: Earnings Effects, Sequentially Removing Clusters
20 Clusters



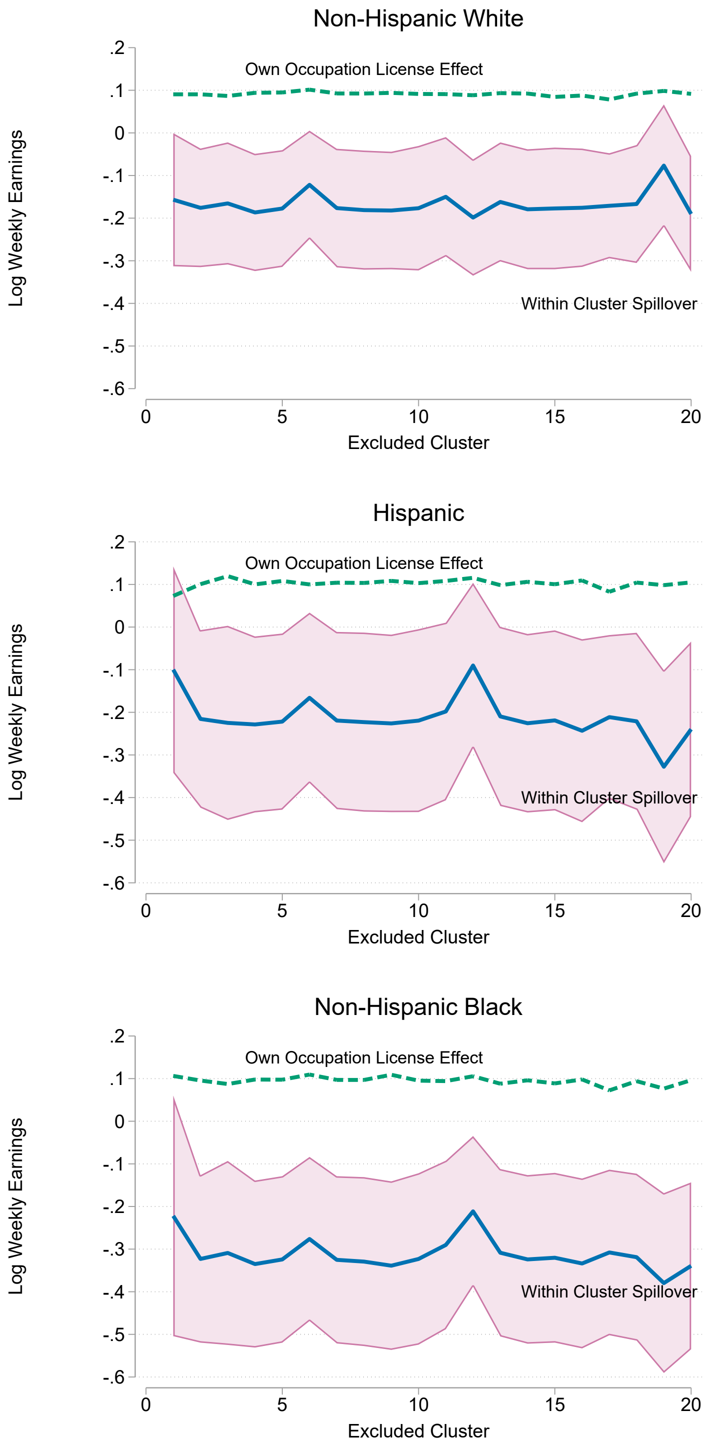
Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Coefficients are generated from the border match design detailed in Equation 3. Standard errors are clustered at the occupation level. 95% confidence intervals in red. Spillover coefficients are based on 100% licensure of an occupation's cluster outside their own occupation for 20 total defined clusters, eliminating one cluster at a time.

Figure A15: Earnings Effects, Sequentially Removing Clusters
By Gender, 20 Clusters



Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Coefficients are generated from the border match design detailed in Equation 3. Standard errors are clustered at the occupation level. 95% confidence intervals in red. Spillover coefficients are based on 100% licensure of an occupation's cluster outside their own occupation for 20 total defined clusters, eliminating one cluster at a time.

Figure A16: Earnings Effects, Sequentially Removing Clusters
By Race/Ethnicity, 20 Clusters

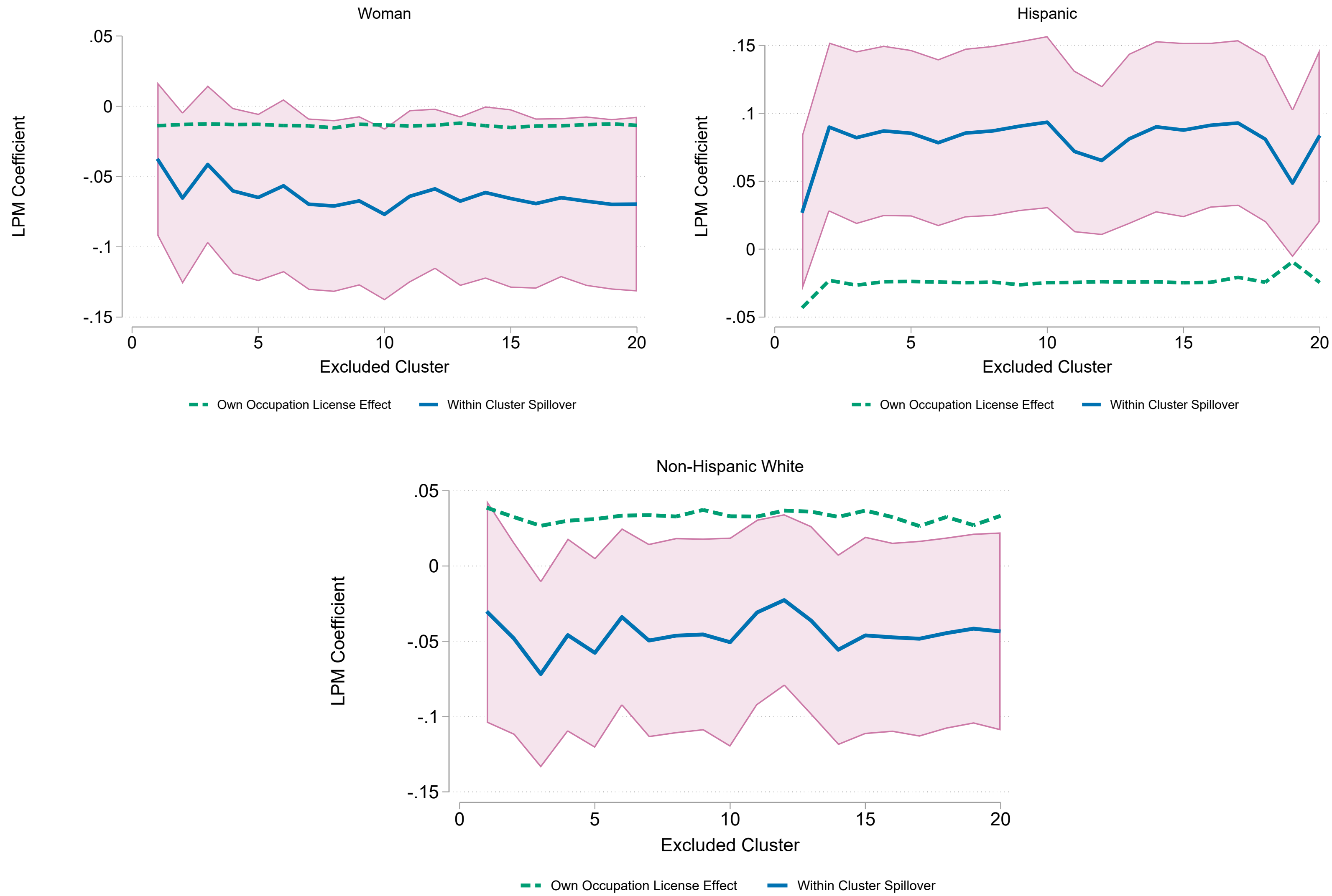


Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Coefficients are generated from the border match design detailed in Equation 3. Standard errors are clustered at the occupation level. 95% confidence intervals in red. Spillover coefficients are based on 100% licensure of an occupation's cluster outside their own occupation for 20 total defined clusters, eliminating one cluster at a time.

Figure A17: Employment Effects, Sequentially Removing Clusters
20 Clusters

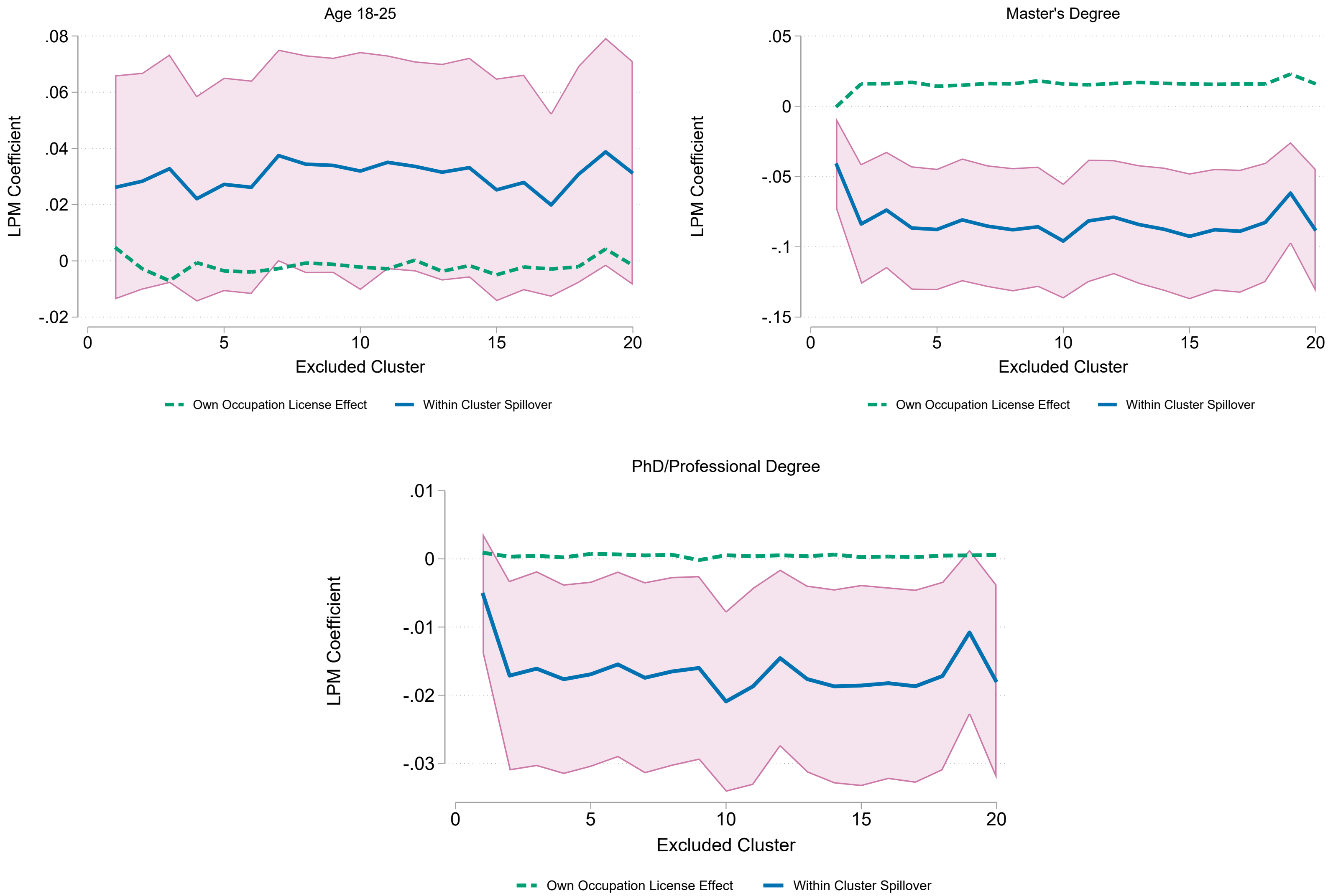
Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Coefficients are generated from the border match design detailed in Equation 3. Standard errors are clustered at the occupation level. 95% confidence intervals in red. Spillover coefficients are based on 100% licensure of an occupation's cluster outside their own occupation for 20 total defined clusters, eliminating one cluster at a time.

Figure A18: Composition Effects, Sequentially Removing Clusters, Sex and Race/Ethnicity
20 Clusters



Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Coefficients are generated from the border match design detailed in Equation 3. Standard errors are clustered at the occupation level. 95% confidence intervals in red. Spillover coefficients are based on 100% licensure of an occupation's cluster outside their own occupation for 20 total defined clusters, eliminating one cluster at a time.

Figure A19: Composition Effects, Sequentially Removing Clusters, Age and Education
20 Clusters



Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Coefficients are generated from the border match design detailed in Equation 3. Standard errors are clustered at the occupation level. 95% confidence intervals in red. Spillover coefficients are based on 100% licensure of an occupation's cluster outside their own occupation for 20 total defined clusters, eliminating one cluster at a time.

Figure A20: Log Earnings Effects using Northwestern Licensing Database

Source: Author's calculations of ACS, O*NET, and Northwestern Licensing Database (NLD) data.
Notes: Coefficients are generated from the border match design detailed in Equation 3 using the Northwestern Licensing Database (NLD). Standard errors are clustered at the occupation level. 95% confidence intervals in red. Spillover coefficients are based on 100% licensure of an occupation's cluster outside their own occupation in the NLD.

B Clustering Appendix

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 B1, 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.

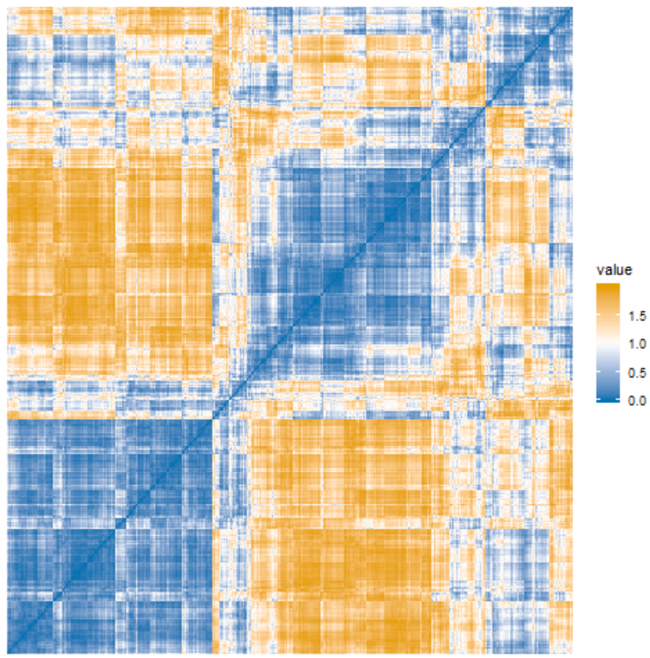
The figure is a colored representation of the dissimilarity matrix. Each occupation is represented on both axes, and the diagonal of the matrix is the distance between each occupation with itself (zero). Darker blue regions represent small differences between occupations along the dimensions considered. 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 B3 below compares the "height" of the various dendrogram connections between occupations along the three measures considered in Figure B1. 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.

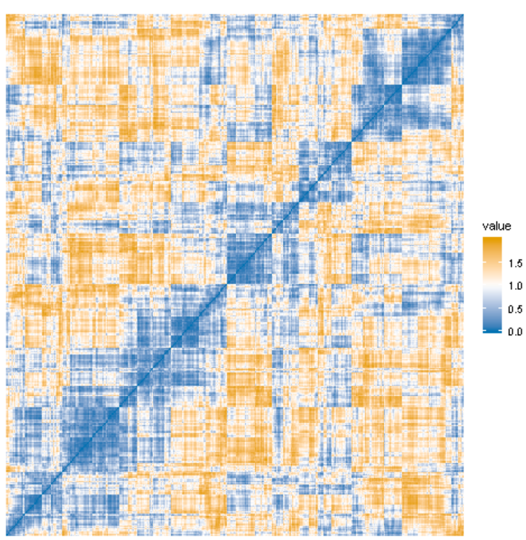
Overall, the measures in Acemoglu and Autor (2011) generate more compact clusters and greater separation between clusters than when clustering over the principal components of the O*NET data. The computer science literature bears this out, stating that in many cases, the principal components of the data, 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.

Figure B1: Correlative Distance Values Between Occupations

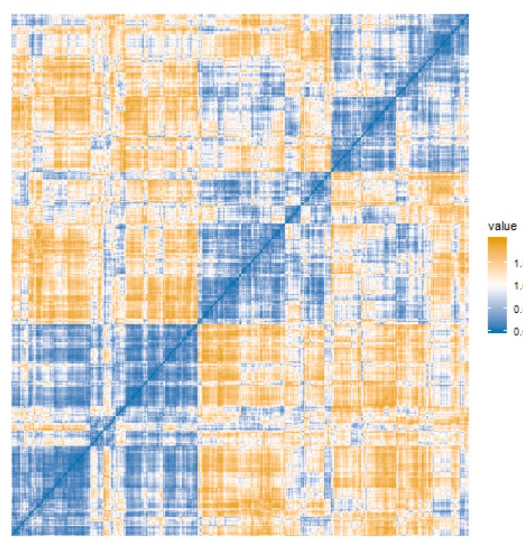
Panel A: Skills in Acemoglu and Autor (2011)



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 B2: Correlation Between Cluster Licensing Shares, Including and Excluding National Median Wages by Occupation from HAC

Source: Author's calculations of ACS, O*NET, and CPS licensing data.
Notes: Shares are generated from the border match design sample at 20 skill clusters where clusters are defined including and excluding national median wages by occupation from the HAC algorithm.

Table B1: 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 4.1. ACS samples are from 2014-2017.

Table B2: Top 5 Focal Occupations by Cluster, Excluding National Median Wages by Occupation from HAC

Occupation	Cluster	Freq	Rank
Managers, Nec (Including Postmasters)	1	161944	1
Elementary And Middle School Teachers	1	161405	2
First-Line Supervisors Of Sales Workers	1	156541	3
Retail Salespersons	1	111932	4
Customer Service Representatives	1	98746	5
First-Line Supervisors Of Construction Trades And Extraction Workers	2	30855	1
Farmers, Ranchers, And Other Agricultural Managers	2	15456	2
First-Line Supervisors Of Landscaping, Lawn Service, And Groundskeeping Workers	2	4659	3
Photographers	2	4030	4
First-Line Supervisors Of Farming, Fishing, And Forestry Workers	2	2630	5
Accountants And Auditors	3	75768	1
Postsecondary Teachers	3	59665	2
Computer Scientists And Systems Analysts/Network Systems Analysts/Web Developers	3	59412	3
Management Analysts	3	27784	4
Office And Administrative Support Workers, Nec	3	23195	5
Nursing, Psychiatric, And Home Health Aides	4	71725	1
Waiters And Waitresses	4	70596	2
Office Clerks, General	4	51245	3
Personal Care Aides	4	44406	4
Food Service And Lodging Managers	4	36221	5
Police Officers And Detectives	5	35700	1
Computer Support Specialists	5	25173	2
Sales Representatives, Services, All Other	5	23563	3
Credit Counselors And Loan Officers	5	12527	4
Compliance Officers, Except Agriculture	5	9999	5
Secretaries And Administrative Assistants	6	136243	1
Software Developers, Applications And Systems Software	6	47609	2
Receptionists And Information Clerks	6	41797	3
Security Guards And Gaming Surveillance Officers	6	33186	4
Health Diagnosing And Treating Practitioner Support Technicians	6	24673	5
Life, Physical, And Social Science Technicians, Nec	7	8709	1
Construction And Building Inspectors	7	3318	2
Appraisers And Assessors Of Real Estate	7	2872	3
Engineering Technicians, Except Drafters	8	16298	1
Dental Assistants	8	11273	2
Database Administrators	8	4678	3
Artists And Related Workers	8	4657	4
Computer Operators	8	3559	5
Production, Planning, And Expediting Clerks	9	13599	1
Agricultural And Food Science Technicians	9	1763	2
Surveyors, Cartographers, And Photogrammetrists	9	1447	3

Prepress Technicians And Workers	9	992	4
Bookkeeping, Accounting, And Auditing Clerks	10	47800	1
Inspectors, Testers, Sorters, Samplers, And Weighers	10	32325	2
Food Preparation Workers	10	31460	3
Billing And Posting Clerks	10	19213	4
Diagnostic Related Technologists And Technicians	10	15015	5
Janitors And Building Cleaners	11	87855	1
Laborers And Freight, Stock, And Material Movers, Hand	11	84622	2
Stock Clerks And Order Fillers	11	59355	3
Construction Laborers	11	53641	4
Other Production Workers Including Semiconductor Processors And Cooling And Freezing Equipment Operators	11	50467	5
Aircraft Mechanics And Service Technicians	12	7138	1
Stationary Engineers And Boiler Operators	12	3551	2
Tool And Die Makers	12	2481	3
Precision Instrument And Equipment Repairers	12	2166	4
Television, Video, And Motion Picture Camera Operators And Editors	12	1958	5
Massage Therapists	13	4535	1
Models, Demonstrators, And Product Promoters	13	1473	2
Cashiers	14	106546	1
Combined Food Preparation And Serving Workers, Including Fast Food	14	12551	2
Host And Hostesses, Restaurant, Lounge, And Coffee Shop	14	8114	3
Entertainment Attendants And Related Workers, Nec	14	6591	4
Counter And Rental Clerks	14	2840	5
Carpenters	15	34059	1
Painters, Construction And Maintenance	15	14140	2
Heating, Air Conditioning, And Refrigeration Mechanics And Installers	15	13725	3
Bus And Truck Mechanics And Diesel Engine Specialists	15	12739	4
Firefighters	15	12405	5
Cluster 16 is entirely universally licensed occupations			1
Security And Fire Alarm Systems Installers	17	2037	1
Animal Control	17	302	2
Agricultural Workers, Nec	18	34934	1
Bus And Ambulance Drivers And Attendants	18	20930	2
Crossing Guards	18	1671	3
Motor Vehicle Operators, All Other	18	1133	4
First-Line Supervisors Of Production And Operating Workers	19	39633	1
First-Line Supervisors Of Food Preparation And Serving Workers	19	20068	2
First-Line Supervisors Of Housekeeping And Janitorial Workers	19	7236	3
Counter Attendant, Cafeteria, Food Concession, And Coffee Shop	19	5017	4
Computer, Automated Teller, And Office Machine Repairers	20	6750	1

Radio And Telecommunications Equipment Installers And Repairers	20	6138	2
Electrical And Electronics Repairers, Transportation Equipment, And Industrial And Utility	20	700	3

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

Notes: Clusters are based on description in Section 4.1 but excluding occupations' national log median wages in the clustering algorithm. ACS samples are from 2014-2017.

Table B3: 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.

C 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:

$$I_t^j \equiv K_t^j + \gamma^{ij} K_{t_j-1}^i + ... + \gamma^{gj} K_{t_i-1}^g + I_0^j + \sum_{e=i,h,g} \gamma^{je} I_0^e \tag{6}$$

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, γ^{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:

$$Y_t^j = E^s + r^j(C_{t-1}^j + \gamma^{ij} C_{t_j-1}^i) - c_t^j + \gamma^{ij} I_0^i + I_0^j \tag{7}$$

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:

$$\{\gamma^{ij} r^i C^i_{t-1} - r^j C^j_{t-1} + (\gamma^{ij} I^i_0 - I^j_0)\} \sum_{g=t}^{T} (1/(1+r)^g)) < 0 \tag{8}$$

and

$$\sum_{g=t}^{T} \sum_{h=t}^{g-1} \{(r^j c^j_h - c^j_g) - (r^i c^i_h - c^i_g)\}(1/(1+r)^g) \lesseqgtr 0 \tag{9}$$

Equation 8 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 γ^{ij}.

Equation 9 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, γ^{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.[33]

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.

[33]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).